1968

1768


Encyclopædia Britannica;
OR, A
DICTIONARY
OF
ARTS and SCIENCES,
COMPILED UPON A NEW PLAN.
IN WHICH
The different Sciences and Arts are digested into distinct Treatises or Systems;
AND
The various Technical Terms, &c. are explained as they occur in the order of the Alphabet.
ILLUSTRATED WITH ONE HUNDRED AND SIXTY COPPERPLATES.
By a Society of GENTLEMEN in Scotland.
IN THREE VOLUMES.
VOL. I.
EDINBURGH:
MDCCLXXI.
Plate LVI.
CONIC SECTIONS.
Plate CXX.
Treble Cliff
Tenor Cliff
Bass Cliff
Encyclopædia Britannica;
OR, A
DICTIONARY
OF
ARTS and SCIENCES,
COMPILED UPON A NEW PLAN.
IN WHICH
The different Sciences and Arts are digested into distinct Treatises or Systems;
AND
The various Technical Terms, &c. are explained as they occur in the order of the Alphabet.
ILLUSTRATED WITH ONE HUNDRED AND SIXTY COPPERPLATES.
By a Society of GENTLEMEN in Scotland.
IN THREE VOLUMES.
VOL. I.
EDINBURGH:

Plate X.
CONIC SECTIONS.
Plate XVIII.
Encyclopædia Britannica;
DICTIONARY
RTS and SCIENCES,
COMPILED UPON A NEW PLAN.
IN WHICH
AND
VOL. I.
The
CORINTHIAN
ORDER

ENCYCLOPÆDIA BRITANNICA

THE UNIVERSITY OF CHICAGO

The Encyclopædia Britannica
is published with the editorial advice of the faculties
of The University of Chicago and of a
committee of members of the faculties of Oxford, Cambridge
and London universities and of a committee
at The University of Toronto

*

"LET KNOWLEDGE GROW FROM MORE TO MORE
AND THUS BE HUMAN LIFE ENRICHED."

ENCYCLOPÆDIA BRITANNICA

VOLUME
20

First Published in 1768

by A Society of Gentlemen in Scotland

ENCYCLOPÆDIA BRITANNICA, INC.

William Benton, Publisher

CHICAGO · LONDON · TORONTO · GENEVA · SYDNEY · TOKYO · MANILA

PRINTED IN THE U.S.A.

Library of Congress Catalog Card Number: 68-10064

FOUNDED A.D. 1768

ENCYCLOPÆDIA BRITANNICA

Volume 20

SCHÜTZ *TO* SPEKE

SCHÜTZ, HEINRICH (1585–1672), the greatest German composer before J. S. Bach, was born on Oct. 8, 1585, at Köstritz. In 1599 he became a chorister at Kassel where the landgrave of Hesse-Kassel provided him with a wide general education. In 1608 Schütz entered the University of Marburg where he began to study law, but in 1609 he went to Venice where for three years he studied music at the landgrave's expense; his chief teacher there was Giovanni Gabrieli. In Venice Schütz wrote his first known works, a set of Italian madrigals for five voices (1611). In 1613 he returned to Germany and went to Leipzig to resume his legal studies; shortly afterward, however, the landgrave offered him the post of second organist at the court in Kassel. In 1614 he went to Dresden to supervise the music for the christening of the son of the Elector of Saxony, and eventually the landgrave was persuaded to allow him to take a permanent post in the electoral chapel. In 1619 he married Magdalene Wildeck, who bore him two daughters. In 1628 Schütz again visited Venice where Claudio Monteverdi was now the chief musical figure; it is possible that Schütz studied with him. Three years after his return to Dresden, Schütz left the elector's court, which was being seriously affected by plague and by the turmoils of the Thirty Years' War, and from 1633 to 1635 he was *Kapellmeister* to the royal court of Copenhagen. From 1635, apart from further visits to the Danish court, he remained, in spite of his frequent pleas for dismissal, in the elector's service at Dresden until his death there on Nov. 6, 1672.

After the early set of madrigals, almost all Schütz's known works are vocal settings of sacred texts, with or without instruments. Of his known secular works, *Daphne* (1627), the first German opera, and compositions for the marriage of Johann Georg II of Saxony in 1638, were lost. Schütz's special achievement was to introduce into German music the new style of the Italian monodists (as typified in Monteverdi's work) without creating an unsatisfactory hybrid: his music remains extremely individual and German in feeling. After the Latin of *Symphoniae sacrae I* (1629) he used the vernacular. The first German requiem was his *Musikalische Exequien* (1636), for soloists and choir, in which the writing for solo voice or duet is often florid in the Italian manner, while the choral sections are firmly based on German chorale tradition. The final section is for double choir, a form which recalls Schütz's studies with the earlier Venetian composers. Other principal works from the middle of his life are two sets of *Kleine geistliche Konzerte* (1636, 1639) for solo voice and continuo, *Geistliche Chormusik* (1648), and *Symphoniae sacrae II* and *III* (1647, 1650) for various combinations of voices and instruments. In all these Schütz's strong dramatic sense is noticeable. The Christmas oratorio of 1664 for soloists, choir, and instruments foreshadows his austere last works. These are *a cappella* Passions, settings of the text of the Gospels according to Matthew, Luke, and John; in these works even the sparing vocal figuration of the Christmas oratorio is absent. The plain scriptural text is delivered by the appropriate soloist in a kind of recitative, generally syllabic, while the words of the Jews, high priests, etc., are set as brief polyphonic choruses.

A complete edition of Schütz's works was begun in Germany in 1955. (C. P. Co.)

SCHWÄBISCH GMÜND, a town in the *Land* (state) of Baden-Württemberg, Federal Republic of Germany, lies in the valley of the Rems, 33½ mi. (54 km.) E of Stuttgart. Pop. (1961) 41,050. It is a centre of the jewelry industry and after World War II craftsmen from Gablonz (Jablonec), Czech., introduced the manufacture of glassware. Besides the Romanesque Johanneskirche and the Gothic Heiligkreuzmünster the town has half-timbered, Baroque, and Rococo houses, and remains of the old surrounding walls. It was chartered in 1162.

SCHWANN, THEODOR (1810–1882), German physiologist known as the co-originator, with M. J. Schleiden, of the cell theory, and the first to use this term, was born at Neuss in Rhenish Prussia on Dec. 7, 1810. After studying at Cologne, Bonn and Würzburg, he graduated in medicine at Berlin in 1834. There he assisted Johannes Peter Müller (*q.v.*) with his experimental work in physiology. Schwann in 1838 was called to the chair of anatomy at the University of Louvain, and in 1847 went as professor to Liège, where he continued until his death, at Cologne, on Jan. 11, 1882. While working with Müller, Schwann's attention was directed to the physicochemical basis of life. Between 1825 and 1837 several workers demonstrated that yeasts of beer and wine are cells that multiply by budding.

In 1839 Schwann was the first to observe internal spores of yeasts, and demonstrated the necessity to digestion of the presence

of a ferment which he called pepsin. In 1837 he began to investigate the laws of muscular contraction, discovering the striated muscle in the upper part of the esophagus, and, later, the myelin sheath, covering the peripheral axons, which now bears his name (*see* REGENERATION: *Nerve Regeneration*). Schwann's *Microscopic Investigations on the Accordance in the Structure and Growth of Plants and Animals* (1839; Eng. trans., 1847) set forth his hypothesis that both animal and vegetable tissues are to be traced back to cells and that the cells of each are identical in character. His cell theory was taken up by Rudolf Virchow (*q.v.*), who disproved Schwann's hypothesis of the discontinuity of cell formation.

SCHWARZBURG, a Thuringian dynasty which gave its name to two sovereign states in Germany before 1918. The ancestors of the dynasty were counts of Käfernburg, a castle near Arnstadt on the Gera River south of Erfurt, in the 11th century. By 1109 they were also counts of Schwarzburg, another castle, on the Schwarza River west of Saalfeld. The two sons of Count Günther III (d. 1196) divided their lands between Käfernburg and Schwarzburg lines. The Käfernburg line died out in 1385, and its lands (including Saalfeld) passed to the rival dynasty of Wettin (*q.v.;* and *see* also SAXONY; THURINGIA). Meanwhile the Schwarzburg line was from 1275 split between Schwarzburg-Schwarzburg and Schwarzburg-Blankenburg branches (the modern Bad Blankenburg stands northeast of the old Schwarzburg). Schwarzburg-Schwarzburg between 1305 and 1334 got possession of Rudolstadt, on the Saale River just below the Schwarza confluence; and about 1330 it acquired Leutenberg, on the Sormitz tributary of the Saale above Saalfeld, so that it was subsequently known as the Schwarzburg-Leutenberg line. When it died out in 1564, its lands reverted to Schwarzburg-Blankenburg. The latter line had provided a German king in 1349 (*see* GÜNTHER); but of more historical consequence was its acquisition of Frankenhausen in the 1340s and of Sondershausen in 1356, since both these lands, beyond the Unstrut River, lay considerably to the north of the main Schwarzburg block. In Thuringia, the Schwarzburgs and the separate dynasty of Reuss (*q.v.*) were ultimately the only ones to conserve independence from the more powerful House of Wettin.

Though the Blankenburg lands had been subdivided, Günther the Rich (d. 1552) reunited them all. He also adopted Lutheranism. After the Leutenberg inheritance, his sons effected another partition, from which the lines of Schwarzburg-Sondershausen and Schwarzburg-Rudolstadt emerged (1584). Both Sondershausen and Rudolstadt, however, had exclaves in one another's vicinity.

Despite Wettin objections, the counts of Schwarzburg-Sondershausen were recognized as princes of the Holy Roman Empire in 1697, and those of Schwarzburg-Rudolstadt in 1710. Both principalities, as sovereign states, entered the Confederation of the Rhine in 1807 and the German Confederation in 1815 (*see* GERMANY: *History*). Schwarzburg-Rudolstadt received a constitution in 1816, and another in 1854, modified in 1870; Schwarzburg-Sondershausen received one in 1841 and another in 1849, superseded in 1857 (with many subsequent modifications). Both states declared themselves for Prussia in the Seven Weeks' War and entered the North German Federation (1866) and the German Empire (1871). On the death, childless, of Charles Günther of Sondershausen in 1909, both states came under the rule of Günther Victor of Rudolstadt, in personal union. On the German Revolution of 1918 the prince abdicated, and both states became republics before merging themselves, in 1920, in the new state Thuringia (*q.v.*).

SCHWARZENBERG, FELIX, PRINCE ZU (1800–1852), Austrian statesman who restored the Habsburg monarchy to prestige as a great power after the revolutions of 1848–49, was born at Krumau in Bohemia on Oct. 2, 1800, the second son of Joseph, prince zu Schwarzenberg. He entered the army in 1818, but was transferred to the diplomatic service (without resigning from the army) in 1824 and worked in the embassies in Lisbon, St. Petersburg, Paris, and London (where his amour with Jane, Lady Ellenborough, wife of the lord privy seal, led to her being divorced). A protégé of Metternich, he was appointed Austrian minister to Sardinia-Piedmont in 1838 and to the Kingdom of the Two Sicilies in 1844.

During the Neapolitan revolution of 1848 Schwarzenberg returned to Vienna. Having meanwhile been promoted to the rank of general, he then joined Joseph Radetzky's army in northern Italy (*see* ITALIAN INDEPENDENCE, WARS OF). Wounded at Goito, he was sent back to Vienna to dissuade the government from surrendering Lombardy and to obtain reinforcements for Radetzky. Successful in this mission, he was appointed military governor of Milan, but soon returned to Vienna to have his wound treated.

When revolution broke out in Vienna on Oct. 6, 1848, Schwarzenberg tried to induce the military commander to make a stand and remained there till Oct. 13—four days after being summoned to join the imperial court, then on its way to Olmütz. On the advice of his brother-in-law, Alfred, prince von Windischgrätz (the field marshal on whom the court depended), Schwarzenberg was bidden to form a government on Oct. 19. On Nov. 21 he was declared prime minister and foreign minister.

Schwarzenberg procured the abdication of the imbecile emperor Ferdinand (Dec. 2, 1848) and the elevation to the throne of the 18-year-old Francis Joseph (*q.v.*). From the first he showed regard for the young emperor's self-respect: he always presented his advice in such a way that Francis Joseph could believe himself to have made the decisions.

Schwarzenberg at first seemed ready to cooperate with the Kremsier Parliament (*see* AUSTRIA, EMPIRE OF). On March 4, 1849, however, he imposed on the monarchy a totally centralistic constitution; and on March 7 he dissolved the Kremsier Parliament. His policy of centralization displeased Windischgrätz, who thought that order could be restored only if the various lands of the empire preserved their historic autonomy. Dissension grew worse when Windischgrätz demanded that Austria accept Russia's offer of military help to suppress the Hungarian rebellion. Schwarzenberg, seeing how humiliating this would be for Austria, engineered the recall of Windischgrätz from the command in Hungary (April 13); but a few days later, with Hungary almost completely overrun by the rebels, Schwarzenberg had to accept Russia's offer.

The defeat of the revolution convinced Schwarzenberg that the various nations of the Habsburg monarchy had forfeited their historic rights. In his centralizing policy he was upheld by his two successive ministers of the interior—namely Franz, Graf von Stadion, and Alexander, later Freiherr von Bach—and, not least, by the emperor himself. Schwarzenberg wanted to use the army and the civil service as the basis for the reconstruction of the state, in which the Germans were to have the leading role. He consequently sought to enhance Austria's position within the German Confederation, into which he wanted to bring the whole Habsburg monarchy, so as to realize the economic "empire of 70,000,000" envisaged by his minister of commerce, Karl Ludwig von Bruck.

When the National Assembly at Frankfurt abortively offered the German imperial crown to Frederick William IV of Prussia (*see* GERMANY: *History: The German Confederation, 1815–66*), Schwarzenberg recalled the Austrian delegates from the Assembly. As he could not conciliate Prussia, he began trying to revive the Confederation in its old form. In May 1850 a federal diet met again in Frankfurt, comprising representatives of Austria and ten German states; and Schwarzenberg, who could rely on a well-organized army, then prepared for war against Prussia. Diplomatic intervention by Russia checked his plans at the last moment, but even so, at Olmütz in November 1850, he forced the Prussians to renounce their scheme for a united Germany without Austria. At the Dresden conference of the German states, however, his plan for a reform of the federal constitution and for the admission of all the Habsburg lands into the Confederation was rejected. He did not even manage to introduce Austria into the German customs union.

By the so-called *Silvesterpatent* (New Year's Eve decree) of Dec. 31, 1851, Schwarzenberg abolished the constitution granted in 1849. This abolition not only damaged the prestige of the young emperor, whom Schwarzenberg thus induced to break his word to his subjects, but also inaugurated an era of neo-absolutism.

Schwarzenberg died in Vienna on April 5, 1852. His death was an irreparable loss for the emperor Francis Joseph. Handsome and pleasure-loving, he had been a coldly realistic aristocrat. He had led Austria back to its old eminence as a great power after the most serious crisis of its history. His political concepts were clear, and he was prompt to perceive the right solution for the problems of international affairs. Yet in domestic affairs the return to German-based centralism became meaningless when Austria could not win hegemony in Germany; and the repeal of the constitution left only a shaky foundation for the great power which Schwarzenberg had resurrected.

Bibliography.—E. Heller, *Fürst Felix zu Schwarzenberg: Mitteleuropas Vorkämpfer* (1933); A. Schwarzenberg, *Prince Felix zu Schwarzenberg* (1946); R. Kiszling, *Fürst Felix . . .* (1952); E. Crankshaw, *The Fall of the House of Habsburg* (1963). (K. O. v. A.)

SCHWARZENBERG, KARL PHILIPP, Prince zu (1771–1820), Austrian field marshal and diplomat, one of the foremost personalities of the Napoleonic Wars, was born in Vienna on April 15, 1771, the son of Prince Johann Nepomuk zu Schwarzenberg. Joining the army in 1787, he served against the Turks in 1788–89. Promoted major as early as 1790, he proved his talents as a cavalry leader in the War of the First Coalition against France (*see* French Revolutionary Wars). He was promoted general in 1796 and lieutenant field marshal in 1800. In the War of the Second Coalition (*see* Napoleonic Wars) he covered the withdrawal of the right wing of the archduke John's army after the defeat at Hohenlinden (1800). In the War of the Third Coalition he won a victory at Jungingen and managed to reach Bohemia with his division after the defeat at Ulm (1805). Vice-president of the Supreme War Council from 1805 to 1808, he had a share in the army reform which made possible Austria's initial successes in the war of 1809. Meanwhile, he had undertaken diplomatic missions to Russia (1801), to Bavaria (1805), and to Russia again (1808–09); on this last mission, he persuaded the Russian emperor Alexander I to delay his support of the French in the war of 1809. In the Battle of Wagram (1809) Schwarzenberg distinguished himself in command of a cavalry corps.

After the Peace of Vienna (Schönbrunn), Schwarzenberg went as ambassador to France. He thus helped to negotiate the marriage between Napoleon and Marie Louise (1810) and the agreement for the war of 1812 against Russia. Napoleon appreciated Schwarzenberg's qualities and wanted him to command the Austrian forces against Russia, since the archduke Charles refused to do so. Schwarzenberg accepted this post, but conducted himself in accordance with Metternich's policy of holding back; and his retreat into Austrian territory in the winter of 1812–13 greatly facilitated the junction of the Russians with the Prussians. Thereafter he was a spokesman for bringing Austria into the war against Napoleon. When this policy prevailed (August 1813), he was promoted field marshal and was made supreme commander of the principal force of the Allies. Following the plans of Joseph Radetzky (*q.v.*), he united the Allied forces for the Battle of Leipzig (*q.v.*) in October. He likewise directed the operations of the main army in France in 1814.

When the Congress of Vienna (*q.v.*) was reorganizing Europe, Schwarzenberg stressed the military objections both to a revival of the Habsburgs' imperial authority in Germany and to a restoration of the former Habsburg possessions on the Upper Rhine, since such measures would have overtaxed Austria's strength. He also opposed Prussia's demand for the whole of Saxony, since this would have meant the encirclement of Bohemia. Prematurely aged, he suffered a paralytic stroke in 1817 and died at Leipzig on Oct. 15, 1820. His letters to his wife, Maria Anna Theresia, Gräfin von Hohenfeld, whom he had married in 1799, are edited by J. F. Novak (1913).

Bibliography.—A. Prokesch-Osten, *Denkwürdigkeiten aus dem Leben des Feldmarschalls Fürsten Schwarzenberg*, 2nd ed. (1861); H. Kerchnawe and A. Veltzé, *Feldmarschall Karl Fürst zu Schwarzenberg* (1913); H. Rössler, *Österreichs Kampf um Deutschlands Befreiung*, 2nd ed. (1945). (Fr. Hu.)

SCHWARZSCHILD, KARL (1873–1916), German astronomer, whose researches were of great range and versatility, was born on Oct. 9, 1873, at Frankfurt am Main, Ger. His exceptional ability in science was evident in a paper on the theory of celestial orbits he wrote at the age of 16. As a student, Schwarzschild came under the inspiration of the astronomer Hugo von Seeliger at Munich. In 1901 he became professor and director of the observatory at the University of Göttingen, leaving in 1909 to become director of the Astrophysical observatory at Potsdam. He served actively in World War I, during which he contracted an illness of which he died on May 11, 1916, at Potsdam.

Schwarzschild's contributions to physical science are significant in several fields. In observational astronomy, he was the first to introduce and apply precise methods in photographic photometry; the first to employ a coarse grafting in front of the telescope objective for photometric and astrometric work; and he developed certain basic methods of analysis of solar spectra taken during eclipses. In theoretical astronomy Schwarzschild introduced the concept of radiative equilibrium in astrophysics and was the first clearly to recognize the role of radiative processes in the transport of heat in the atmospheres of the stars. He also initiated some basic mathematical methods for the treatment of these problems.

Schwarzschild laid the foundations of modern statistical methods in astronomy, and also made the most important contributions, after Carl Friedrich Gauss, to geometrical optics and the theory underlying the design of optical instruments.

Although generally considered an astronomer, Schwarzschild also made fundamental contributions to theoretical physics and to relativity. In the former, he was one of the great pioneers in developing Niels Bohr's theory of atomic spectra. In a great paper, the proofs of which he read on his deathbed, Schwarzschild developed (independently of Arnold Sommerfeld) the so-called general "rules of quantization" and gave the complete theory of the Stark effect and initiated the quantum theory of molecular spectra. Similarly, in developing the general theory of relativity, Schwarzschild gave the first exact solution of Einstein's general gravitational equations leading to the discovery of what is now called Schwarzschild's line element describing the geometry of space in the neighbourhood of a mass point. (Su. C.)

SCHWEINFURTH, GEORG AUGUST (1836–1925), German botanist and traveler who made his name as the explorer of the Bahr el Ghazal region of the west Nile Basin, was born in Riga on Dec. 29, 1836, the son of a German wine merchant who had settled in Riga. He studied botany at Heidelberg and Berlin and early became interested in the plants of Africa. In 1863 he traveled down the Red Sea to Suakin and overland to Khartoum, botanizing as he went. On his return to Berlin in 1866 he interested the Royal Academy of Science in the botanical exploration of the hitherto little-known region watered by the Bahr el Ghazal, for which he received a grant from the Humboldt Institute. Arriving at Suakin in Sept. 1868 and at Khartoum in Jan. 1869, he ascended the White Nile and traveled westward by way of Lake No with a party of ivory traders.

Primarily a botanist, Schweinfurth had also a good knowledge of zoology and ethnology; he spent nearly three years in the region, visiting the Niam-Niam, the Bongo, and the Mangbetu peoples; and his is the first authoritative account of the Congo pygmies. Crossing the Nile-Congo watershed he reached, in March 1870, the westward-flowing Uele, later discovered to be an affluent of the Congo. He received the Royal Geographical Society's Founder's Medal for 1874, his discovery of the Uele having helped toward a final estimate of the extent of the Nile system.

From 1875 to 1888 Schweinfurth lived in Cairo, traveling each winter, notably to the Lebanon in 1880, Socotra in 1881, and the Yemen in 1888, interesting himself in botany, geology, paleontology, and archaeology. In 1876 he attended the African conference summoned to Brussels by King Leopold, and in 1888 moved to Berlin where his herbarium was housed. Until 1914 he went to Africa frequently. He died in Berlin on Sept. 19, 1925. He was a keen observer and an accurate and humorous recorder; *The Heart of Africa* (1873) is excellent reading and a reminder of the part played by German explorers in African discovery.

See S. Passarge, *Mitteilungen der Geographischen Gesellschaft in Hamburg*, vol. 37 (1926); and an unsigned obituary in *Geogr. J.* vol. 67 (1926). (D. Mn.)

SCHWEITZER, ALBERT (1875–1965), Alsatian philosopher, theologian, musician, mission doctor, and winner of the Nobel Peace Prize in 1952, was born on Jan. 14, 1875, at Kaysersberg, Upper Alsace, the eldest son of a Lutheran pastor. The family soon moved to Günsbach, in the Münster Valley, which remained Schweitzer's European home. At the *Gymnasium* at Mühlhausen, where he went when he was 9, he showed interest in history and natural science. At 18 he entered Strasbourg University to study theology and philosophy. His mind soon began to turn to the problems of the Synoptic Gospels, and during his military service (1894) he pursued the researches into the life of Jesus which, despite many other labours, he continued throughout his life. After studying in Paris, he took his doctorate in philosophy at Strasbourg with a treatise on *Die Religionsphilosophie Kants* (1899). He then studied at Berlin, returning to Strasbourg as lecturer in philosophy and preacher at St. Nicholas' Church, and taking his doctorate in theology in 1900. In 1903 he became principal of a theological college attached to the university. He had already published, in 1901, *Das Abendmahlsproblem auf Grund der wissenschaftlichen Forschung des 19. Jahrhunderts und der historischen Berichte* and *Das Messianitäts- und Leidensgeheimnis Jesu* (Eng. trans., *The Mystery of the Kingdom of God*, 1914), preparing the way for *Von Reimarus zu Wrede: eine Geschichte der Leben-Jesu-Forschung* (1906; *The Quest of the Historical Jesus: a Critical Study of Its Progress from Reimarus to Wrede*, 1910), which established him as a world figure in theological studies. In these, he made a radical and exhaustive demonstration of the eschatological view of the life of Jesus—the view that Jesus' ministry was dominated by his knowledge of his messiahship and his expectation of the imminent end of the world. In two books on St. Paul, notably *Geschichte der Paulinischen Forschung* (1911; *Paul and His Interpreters*, 1912), and in later and briefer writings, he continued to expound the view that the New Testament is pervaded by eschatological expectation.

© KARSH OF OTTAWA FROM RAPHO GUILLUMETTE

SCHWEITZER

Meanwhile, Schweitzer was also studying music. He had begun his career as organist at Strasbourg in 1893, when E. Münch, the initiator and conductor of Bach concerts there, asked him to play the organ accompaniment for performances of Bach's cantatas and Passions. In Paris, he studied the piano with a former pupil of Liszt. Becoming interested (1896) in organ-building and restoration, he published in 1906 a booklet on the art of organ-building and -playing in France and Germany (Eng. trans. 1927). In 1902 C. M. Widor, his organ teacher in Paris, recognizing that, as well as being a gifted organist, Schweitzer was a Bach interpreter of rare and original perception, asked him to write a study of Bach's life and art. Although already working on *The Quest of the Historical Jesus*, Schweitzer agreed, and the intended essay grew into a substantial book, *Jean Sébastien Bach: le musicien-poète* (1905), written in French while he was lecturing and preaching in German at Strasbourg. Its publication brought him acclaim in France and Germany. Called upon for a German translation, Schweitzer rewrote the book in German at twice the length, and it was from this edition (1908) that the English translation (by Ernest Newman, with additions and corrections by Schweitzer) was made (two volumes, 1911). Following his usual practice, he gave the reader a thorough historical and analytical background of the subject, of the origins of the chorale, of the cantata, and of the Passion music; he reviewed the forms of music and art within the history of thought and presented Bach as a deeply religious mystic whose music was impersonal and unselfconscious, as cosmic as the forces of nature. In a style matching his theme, he described the composer as a musician-poet and the supreme pictorial artist in sound; and he gave detailed, copiously illustrated (and sometimes controversial) instructions for playing Bach's works. His *J. S. Bach* remains a classic study for the depth and breadth of its interpretation and for its rich spiritual content.

In 1905, a scholar and an organist of wide repute with an unbounded academic future before him, Schweitzer announced to his friends his intention to realize a decision made in 1896. He began to study medicine, to qualify himself as a mission doctor to the people of equatorial Africa, and in 1906 he resigned his university appointments. Later, in *Zwischen Wasser und Urwald* (1921; *On the Edge of the Primeval Forest*, 1922), he explained that he had found his simple motive in the parable of Dives and Lazarus (Luke 16:19–31), identifying Dives with the white man, endowed with all the benefits of culture and science, and Lazarus with the Negro, exploited and oppressed and lacking even medical treatment for his disease and pain. With characteristic thoroughness he took the six years' course in medicine and surgery, continuing meanwhile some of his major literary and musical work. In 1912 he married Hélène Bresslau, daughter of a well-known Strasbourg historian and herself an accomplished scholar, who trained as a nurse to share her husband's renunciation and adventure in discipleship until her death in 1957. In 1913 they went to the Gabon province of French Equatorial Africa (after 1960, the Republic of Gabon). The site for Schweitzer's hospital at Lambaréné, on the forested banks of the Ogowe (Ogooué) River, was provided by the Paris Missionary Society, which had declined his services because of his unorthodox theological views; but in all else his was an independent enterprise. He equipped and maintained his hospital from the proceeds of organ recitals and lectures on his visits to Europe and from royalties on his books (and later also from gifts and grants from individuals and foundations in many countries).

Building his hospital with his own hands and African help, operating far into the night under the most primitive conditions, and entering his long fight with leprosy, with sleeping sickness, and with a host of tropical diseases, Schweitzer began to turn his active intellect to the problem of world civilization. Opportunity for the development of his thought was given in 1914 by the brief internment at Lambaréné of himself and his wife as German subjects, and he continued to work more intensively on his thesis when in 1917 they were taken back to Europe and eventually (1918) to a prison camp in Provence. The outcome was the publication in 1923 of the first two volumes of his *Kulturphilosophie*. The first, *Verfall und Wiederaufbau der Kultur* (*The Decay and the Restoration of Civilization*, 1923), is a brief introduction, while the second, *Kultur und Ethik* (*Civilization and Ethics*, 1923), is a brilliant review of the history of ethical thought leading up to his own original and positive contribution of "reverence for life" as the true and effective basis for a civilized world.

Early in 1924 Schweitzer returned to Africa to rebuild his derelict hospital and to renew its work. Famine, pestilence, floods, and lack of adequate help made this fresh start even more formidable than the first beginning, and he decided to move to a better site about 2 mi. (3 km.) farther up the Ogowe. There his practical powers were devoted to the creation of a larger and more efficient hospital. Over the years the hospital village grew, and with the discovery of new drugs for the treatment of leprosy a large leper colony came into being near by. In April 1963, when he received an address of felicitation on the jubilee of Lambaréné from his supporters in 28 countries, there were in the hospital about 350 patients with their relatives, and in the leper colony 150 patients, all served by a staff of 30 white and 30 nonwhite doctors, nurses, and helpers.

Besides this principal preoccupation, Schweitzer published *Die Mystik des Apostels Paulus* (1930; *The Mysticism of Paul the Apostle*, 1931), more mature than his previous pieces of original New Testament scholarship; and *Die Weltanschauung der indischen Denker: Mystik und Ethik* (1935; *Indian Thought and Its Development*, 1936), a book that grew from one chapter of the draft of the continuation of his *Kulturphilosophie*. He returned for brief visits to his home at Günsbach (where he had used the proceeds of the Goethe Prize he received in 1928 to build a house for himself and his colleagues, and where he received many visi-

tors), and to give lectures and organ recitals in Germany, Great Britain, the Netherlands, Scandinavia, and France; and he accepted an invitation to take a leading part in the Goethe bicentenary celebrations at Aspen, Colo., in 1949. He also made recordings and resumed his editing of Bach's music, begun with Widor in 1911 (*Bachs Orgelwerke*, 1912–14). Intermittently he returned to the third volume of the *Kulturphilosophie*—though not with the intention that this should be published during his lifetime—and to a restatement of his theological thought. His address on receiving in 1953 the 1952 Nobel Peace Prize, *Das Problem des Friedens in der heutigen Welt* (1954; *The Problem of Peace in the World of Today*, 1954), had a worldwide circulation. He was received by Queen Elizabeth II in 1955 when he was appointed an honorary member of the British Order of Merit. In 1958 he broadcast from Oslo three appeals to the world, published as *Friede oder Atomkrieg* (*Peace or Atomic War?*, 1958). Schweitzer died at Lambaréné on Sept. 4, 1965, and was buried there.

Schweitzer's astonishing and lasting capacity for arduous physical and mental labour was supported by an exceptionally strong physique. Tall and broad, of relentless energy and acute concentration, his face at once forceful and compassionate, he had immediate and magnetic charm for all who met him. Shrewdness was evident in his business arrangements and in the administration of his hospital settlement, together with an almost patriarchal feeling in his retention of control over every detail of its work and management. The practical thoroughness underlying his powers of creation and interpretation was reflected also in his passion for the building and repair of organs and in his exhaustive mastery of the work of his predecessors in theology and philosophy.

Bibliography.—*See* his *Aus meiner Kindheit und Jugendzeit* (1924, *Memoirs of Childhood and Youth*, 1925), which has become a classic of its kind, *Aus meinem Leben und Denken* (1931; *My Life and Thought*, 1933), and *Afrikanische Geschichten. . . .* (1938; *From My African Note-book*, 1938); also *Music in the Life of A. Schweitzer*, with selections from his writings, trans. and ed. by C. R. Joy (1951); *A. Schweitzer, An Anthology*, ed. by C. R. Joy (1952); G. Seaver, *A. Schweitzer: the Man and His Mind* (1947; 5th ed. 1955); O. Kraus, *A. Schweitzer: His Work and His Philosophy* (trans. by E. G. McCalman, 1944); E. N. Mozley, *The Theology of A. Schweitzer for Christian Inquirers. . .* (1950); N. Cousins, *Dr. Schweitzer of Lambaréné* (1960); G. McKnight, *Verdict on Schweitzer* (1964); H. Clark, *The Philosophy of A. Schweitzer* (1964); E. Anderson, *The Schweitzer Album* (1965). (J. D. N.)

SCHWENCKFELD, KASPAR, von Ossig (1489–1561), German reformer and preacher, lay leader of the Protestant Reformation in Silesia, founder of a widespread movement known as the Reformation by the Middle Way, and spiritual ancestor of the Schwenkfelder Church, a small Protestant denomination extant in the United States. A prolific author, he laboured throughout his life for the principles of religious freedom, a labour which forced him into exile and hiding for the greater part of his career.

Schwenckfeld was born of the nobility and grew up on the family estate in Lower Silesia. He studied at the universities of Cologne and Frankfurt and served as counselor to various courts from 1511 until 1523. His spiritual awakening came in 1518. In 1525 he visited Martin Luther in Wittenberg to submit his views on the Eucharist. This conference closed on a note of complete disagreement, and Schwenckfeld returned to Silesia to further develop his own theology and reformatory plan. The Middle Way, as his approach was thereafter designated, endeavoured to strike a spiritual course between the Catholics and Lutherans, both of which, he claimed, were swerving away from Christ and dwelling on external symbols. He subsequently published strong anti-Catholic and anti-Lutheran views which resulted in his final dismissal by the duke of Liegnitz. In 1529 he voluntarily left Silesia, taking up residence at Strasbourg where he became acquainted with Sebastian Franck, Melchior Hoffman, Michael Servetus and Paracelsus. There, too, he encountered Huldreich Zwingli for the first time, although Zwingli had earlier published Schwenckfeld's book on the sacraments in Switzerland. They were unable to reconcile their differences, and Schwenckfeld was not invited to participate in the Marburg colloquy of 1529, where his views were subsequently dismissed. In 1533 Schwenckfeld defended his doctrines and the principles of religious liberty against Martin Bucer in the synod of Strasbourg. However, the leaders of the orthodox groups maintained their control over the synod and Schwenckfeld left Strasbourg, residing variously in Augsburg, Tübingen and Ulm. The publication (1539) of a book in proof of his most characteristic doctrine—the deification of the humanity of Christ—led to active persecution by the Lutherans and his expulsion from Ulm. The next year (1540) he published a refutation of the attacks upon his doctrine with a more elaborate exposition of it, under the title *Grosse Confession*. The book emphasized the differences between the Lutherans and Zwinglians on the doctrine of the Eucharist at a moment when efforts were being made to reconcile them. An anathema was accordingly issued from Schmalkald against Schwenckfeld (together with Sebastian Franck); his books were placed on the Protestant "index"; and he himself was made a religious outlaw. He went into hiding.

His followers then broke completely from orthodox church circles and formed small societies and brotherhoods. Schwenckfeld lived the life of a fugitive from 1540 until his death, frequently changing his abode, writing under a variety of pseudonyms and answering his critics with a constant flow of pamphlets and books. He died at Ulm on Dec. 10, 1561.

Schwenkfelder brotherhoods and conventicles remained active in southern Germany after his death, particularly in Strasbourg, Speyer and Ulm, but their vitality was virtually destroyed during the Thirty Years' War. The last important nucleus of adherents lived in the principality of Liegnitz, Lower Silesia. In 1719 a Jesuit mission was sent from Vienna to Silesia to convert them by force. In the wake of this effort, and its subsequent persecutions, more than 500 Schwenkfelders fled into Saxony and 200 of these ultimately found homes in the new world. When Frederick II (the Great) of Prussia captured Silesia in 1740, he extended his protection to those who remained in the province and issued an edict to the exiles in Pennsylvania, offering to return to them their expropriated homes if they would journey back to Silesia.

After 50 years of pioneer life in southeastern Pennsylvania, the exiles took their first steps toward formal organization, forming in 1782 the Society of Schwenkfelders. This society was the forerunner of the present-day Schwenkfelder Church, incorporated in 1909, consisting of five congregations, all near Philadelphia, with a combined membership of approximately 2,600 members. Each church is incorporated and self-sustaining, all joining in a general conference that meets twice yearly. Worship is nonliturgical and theology Christocentric.

Bibliography.—*Corpus Schwenckfeldianorum*, vol. i–xix (1907–61); S. G. Schultz, *Caspar Schwenckfeld von Ossig* (1946); Rufus Jones, *Spiritual Reformers in the 16th and 17th Centuries* (1914); Oswald Kadelbach, *Ausführliche Geschichte Kaspar v. Schwenkfelds und der Schwenkfelder* (1865); H. W. Kriebel, *The Schwenkfelders in Pennsylvania* (1904); S. K. Brecht (ed.), *The Genealogical Record of the Schwenkfelder Families* (1923). (A. S. Be.)

SCHWERIN, a town in the northwest of the German Democratic Republic and the capital of the *Bezirk* (district) of Schwerin, lies on the southwest shore of the Schweriner See (lake), 111 mi. (179 km.) NW of Berlin by road. Pop. (1964) 91,210. Schwerin, which until 1952 was the capital of the *Land* (state) of Mecklenburg, was a Wendish settlement in 1018; the German town and its charter date from 1160. The urban district is dotted with lakes, and a 19th-century castle on the site of an earlier one stands on an island between the Schweriner See and Burg See. The cathedral, mainly Gothic (13th–15th century), was completed in 1892 with a tower 384 ft. (117 m.) high. The town's industries include the manufacture of machinery (ship fittings, gears, etc.), plastics, furniture, cables, clothing, and the production of food items, sterilized milk, and cigarettes. Schwerin is a railway junction and a focal point of roads.

Schwerin Bezirk extends northward from the Elbe to within about 12 mi. (19 km.) of the Baltic Sea, and is bounded westward by the Federal Republic of Germany. Area 3,348 sq.mi. (8,671 sq.km.). Pop. (1964) 593,722. The *Bezirk*, constituted in 1952 from parts of the *Länder* of Mecklenburg and Brandenburg, is divided into the urban district of Schwerin and ten rural districts. The other principal towns are Güstrow (*q.v.*), Wittenberge, and Parchim. The district is mainly agricultural (rye, sugar beet,

potatoes) with some forestry, and there is extensive peat-cutting on the outskirts of Schwerin town.

SCHWINGER, JULIAN SEYMOUR (1918–), U.S. physicist, shared the 1965 Nobel Prize for Physics with R. P. Feynman and S. Tomonaga (*qq.v.*) for their contributions to quantum electrodynamics. His independent work in the 1940s substantially contributed to methods for mathematically treating the behaviour of interacting elementary particles such as those encountered in the decay of radioactive materials (*see* PARTICLES, ELEMENTARY: *Renormalization: Dispersion Theory*). He was born in New York City on Feb. 12, 1918, and entered the City College of New York at the age of 14; he received an A.B. (1936) and Ph.D. (1939) from Columbia University. From 1939 to 1945 he served first as National Research Council fellow, then doing research and teaching (in order) at the University of California (Berkeley, where he worked with J. Robert Oppenheimer), Purdue University, and Massachusetts Institute of Technology. In 1945 he joined the faculty at Harvard University, where he served as professor of physics. Earlier (1951) he was the co-recipient of the first Albert Einstein Award.

Schwinger's publications include *Quantum Electrodynamics* (1958), a group of selected papers assembled under his editorship.

SCHWYZ, a town and canton in central Switzerland. The town, capital of the canton, lies at the foot of the Grosser Mythen (6,230 ft. [1,899 m.]), 16 mi. (26 km.) E of Lucerne and 3 mi. from Brunnen, its port on Lake Lucerne. Pop. (1960) 11,007, including the hamlets of Ibach, Seewen, and Rickenbach. The main buildings of interest include several 16th- and 17th-century patrician houses, the Baroque church of St. Martin (1769–74), the Kerchel chapel, the ancient St. Peter's nunnery (rebuilt 17th century), the historical museum, and the 17th-century town hall. The Archives (1935) houses the charter of the confederation and has a collection of old federal charters and banners; in its garden is the statue of "National Defense." Schwyz is a popular and picturesque summer resort, with a mild climate. It is the starting point for extensive mountain tours, and the well-known skiing regions of Ibergeregg, Holzegg, and Stoos are nearby.

SCHWYZ CANTON is one of the ancient Forest cantons of central Switzerland. Its total area is 351 sq.mi. (908 sq.km.), of which roughly three-quarters is reckoned as productive (forests covering about 75 sq.mi. and vineyards about 30 ac. [12 ha.]); of the rest, 20 sq.mi. is occupied by the larger lakes (chiefly parts of Zürich and of Lucerne, a small area of Zug and the whole of Lowerz). Pop. (1960) 78,048, mostly German-speaking Roman Catholics. Its loftiest point is the Böser Faulen (9,199 ft. [2,804 m.]), while two of the highest summits of the Rigi (the Kulm, 5,909 ft., and the Scheidegg, 5,453 ft.) are within its borders; but, on the whole, the land is hilly rather than mountainous. It has two main valleys, the Muota, receiving the waters of Lake Lowerz and draining into Lucerne, and the Sihl, which receives the Alp River on which Einsiedeln stands; the reinforced stream, though formed near the head of the Lake of Zürich, flows for a long distance roughly parallel to it and enters the Limmat River below the lake.

Schwyz is in the diocese of Coire. The largest town, after Schwyz, is Einsiedeln (pop. [1960] 8,792), a great pilgrimage centre noted for its Black Virgin and its Benedictine monastery. The canton is essentially pastoral, its local breed of brown cattle being much esteemed, particularly in north Italy, but there is some industrial activity (textiles) near the Lake of Zürich and Schwyz. The 30 cantonal communes are grouped into six administrative districts. The legislature (*Kantonsrat*) is composed of 105 members and the elections, since 1907, are carried out by proportional representation. The executive (*Regierungsrat*) of seven members is elected by popular vote and, like the larger body, holds office for four years. The two members of the federal *Ständerat* and the three of the federal *Nationalrat* are also chosen by popular vote.

The canton has few main railways, the principal one being a portion of the main Saint-Gotthard line between Sisikon and Küssnacht. Arth-Goldau (memorable for the great landslide of 1806) is a railway junction, with a line to Zug and another past Biberbrücke (junction for Einsiedeln) toward Wädenswil. A mountain railway also terminates at Arth-Goldau for the ascent of the Rigi Kulm. Of other mountain lines in the canton the most important are the electric cogwheel railway from Brunnen to Axenstein and the funicular from Schwyz to Stoos (4,265 ft.), the steepest in Switzerland.

History.—The valley of Schwyz is first mentioned in 972 as "Suittes." Later, a community of freemen settled at the foot of the Mythen, subject only to the count of the Zürichgau, as representing the German king. In 1240 the community obtained from the emperor, Frederick II, the privilege of being subject immediately to the empire. Its territory then included only the district around the village of Schwyz and the valley of the Muota. But in 1269 it bought Steinen and Rothenthurm. Schwyz took the lead in making the famous Everlasting League of Aug. 1, 1291, with the neighbouring districts of Uri and of Unterwalden (Obwalden-Nidwalden), its position and political independence specially fitting it for this prominence. An attack by Schwyz on Einsiedeln was the excuse for the Austrian invasion that was gloriously beaten back in the Battle of Morgarten (Nov. 15, 1315). In the history of the league Schwyz was always to the fore, so that its name in a dialectal form, Schweiz (Switzerland), applied to the three Forest cantons as early as 1320 and extended to the whole confederation by 1352, though it formed the official name only from 1803 onward. After the victory of Sempach (1386) Schwyz greatly extended its borders. An alliance with Einsiedeln in 1397 ended in 1434; between 1386 and 1436 the March (the region near the upper Lake of Zürich) was acquired; in 1402 Küssnacht was bought, and in 1440 the Höfe, the parishes of Wollerau, Feusisberg, and Freienbach, situated on the main lake of Zürich. All these districts were governed by Schwyz as subject lands, the supreme power resting with the *Landsgemeinde* (or assembly of all male citizens of full age). Schwyz joined the other Forest cantons in opposing the Reformation, and took part in the Battle of Kappel (1531), in which Huldreich Zwingli fell. In 1798 Schwyz, including Gersau (free from 1390), formed part of the République Telliane (or Tellgau), set up by the French, which a week later gave way to the Helvetic Republic. In 1803 the separate canton of Schwyz was again set up. Schwyz joined, in 1832, the League of Sarnen, and in 1845 the Sonderbund. In 1832 the outer districts (Einsiedeln, the March, Küssnacht, and Pfäffikon) formed themselves into a separate canton, which led to a federal occupation of the old canton in 1833 and to the dissolution of the new. In 1838 strife broke out in the older portion of Schwyz between the richer and poorer peasants over the use of the common pastures. The cantonal constitution of 1848 put an end to the ancient *Landsgemeinde*; it was revised in 1876 and in 1898. *See also* SWITZERLAND: *History; Administration and Social Conditions.* (P. Sc.)

SCIACCA, a coastal town and episcopal see in the province of Agrigento in southern Sicily. Pop. (1961) 30,596. Though the town has a modern appearance it retains some of its ancient fortifications: the town walls (*c.* 1550), the ruins of the castle of Luna (a Spanish family of the 15th and 16th centuries who were rival lords of the town with the Perollo family), and the Porta San Salvatore with 15th-century bas-reliefs. The 14th-century Steripinto palace and the 15th-century church of Sta. Margherita are of architectural interest. Sciacca is the birthplace of the historian Tommaso Fazello (1498–1570) and the painter Mariano Rossi (1731–1807). The town, occupying the site of the Roman Thermae Selinuntiae, is a thermal resort; nearby are the hot springs of Mt. San Calogero (ancient Mons Cronius).

(M. T. A. N.)

SCIATICA is pain along the course of the sciatic nerve, down the back of the leg. It often develops either immediately, or after an interval of several hours to a few days, following an unusual movement or exertion that places a strain on the lumbar portion of the spine where the nerve has its roots. The pain may be made worse by coughing and sneezing, and by flexion of the neck. It is relieved by positions that minimize traction on the sciatic nerve; for this reason a person with sciatica tends to walk with the affected leg bent at the knee and externally rotated at the hip. Sciatica is commonly associated with rupture of a portion of an intervertebral disc into the spinal canal. Most cases,

including those in which the nerve is displaced by protruding disc substance, can be treated effectively by bed rest on a firm mattress for several days to a few weeks. In some cases, however, surgical decompression of the nerve is necessary to prevent repeated attacks of disabling pain or to relieve serious, progressive disturbances in function of the sciatic nerve, which take the form of weakness and sensory loss in the leg.

Sciatica usually is a benign, if temporarily disabling, reflection of human structural imperfection. However, it may also develop as a symptom of some local compressive lesion other than a protruded intervertebral disc; or it may indicate a more generalized disorder involving the peripheral nerves (*see* NEURALGIA; NEURITIS). (S. SN.)

SCIENCE, HISTORY OF. This article treats the history of science under the following headings:

I. INTRODUCTION

Is it possible to define what is meant by science? Perhaps not, for science is one of the major activities of the human mind, in this resembling art, religion, or philosophy, which can be understood only in relation to their past history. Science may perhaps be regarded as a mood in which the world is considered. No man is always in the same mood, and no man of science remains permanently in the scientific mood. In evaluating the judgments of men of science outside their own department it is well to remember this, for these judgments are sometimes to be rated very low.

It is very difficult to discuss science as a whole. The Latin word *scientia* meant "knowledge," but the modern usage covers only certain kinds of knowledge whose area, however, is now so vast that no man can grasp more than a fraction of them. Moreover, "scientific" knowledge is extremely diverse, extending from subatomic reactions to mental processes; from mathematical laws of thermodynamics to the economics of race relations; from the births and deaths of stars to the migration of birds; from the study of ultramicroscopic viruses to that of extragalactic nebulas; from the rise and dissolution of cultures and of crystals to the rise and dissolution of atoms and universes. It includes both knowledge of the workings of living bodies and knowledge of the laws of thought, together with the nature of their disturbances. Can this great diversity of topics be brought under any one formula?

All involve systematic and unbiased observations. The examination of records of observations by trained minds leads to classification; from classifications general rules or "laws" are deduced; these laws may be applied to further observations; failures in correspondence between new observations and accepted laws may result in alterations of the laws; and these alterations lead to yet further observations; and so on. This is usually held to constitute the "method" of science. (*See* SCIENTIFIC METHOD.)

A short cut may, however, be taken by a mental process of which little is known; but even then the final appeal is to observation, often in the form of experiment. Further, science is always developing and is not a static body of knowledge. To summarize, science is a never-ending search for judgments to which the universal assent of experts may be obtained.

In English, as in other languages with vocabularies even more largely Latin-derived, the complex adjectival form of the word "science," namely "scientific" (*i.e.,* "knowledge-making"), has been steadily displacing simpler, shorter, and more natural formations, such as "sciential," "scientic," and their variants, since the beginning of the 17th century. The acceptance of the form "scientific" follows closely the growing prestige of what has come to be called "science" (earlier "natural philosophy"). This correlation of concept and adjective becomes intelligible in the light of the growing awareness that science is the making of knowledge and not knowledge as such, so that it has become more nearly equated with "research." When the process of making knowledge ceases, science subsides into static or recessive tradition. Science, nevertheless, does not exclude tradition, but necessarily involves developing tradition. It is, of course, true that atomic physics, for example, may be advanced without exact knowledge of the careers, achievements, and mental processes of Max Planck and Albert Einstein; and there are episodes of research wherein the researcher is wise to confine himself temporarily to his own thoughts. But the man of science inherits an age-old way of thinking; and only by extending knowledge previously won can he build new knowledge.

Despite their philosophical importance, controversies on the nature of knowledge, on the question of whether our knowledge is real or whether there is another and deeper reality to which we cannot reach, and the even more pressing debate as to how or whether science can give certitude are of no aid in defining, delimiting, or understanding the nature of science (*see* KNOWLEDGE, THEORY OF). Science can treat the outer world solely on the level of phenomena ("things that appear," "appearances"), which can appear only to the senses. The senses may yield results that are ultimately contradictory or in which, at least, the mind can find no harmony. But the quieting of the mind on such things is ultimately a task of philosophy or religion or both.

Yet by technological devices the sphericity of the earth, the discontinuity of matter and of forces, the movements of atoms, the bending of light by the sun's mass, and even the mutual convertibility of mass and energy may be demonstrated to the senses. Although it is true that science frequently leaves phenomena altogether, to mount into an atmosphere of abstract symbols, usually mathematical, it then descends again to the prediction or demonstration of phenomena. Phenomena must ultimately be sensed, and mathematical considerations, however recondite, and scientific instruments, however intricate, are but delicate, remote, and specialized ways of extending sense-experience, though sense-experience may for some sciences (and perhaps eventually for all) be ultimately reducible to scale readings.

II. THE AGE OF ANONYMOUS SCIENCE

Perhaps 400,000 years ago the weapons or tools of primitive men began to assume a symmetrical form involving some mental image of the object before it was wrought. This implied an adaptation of means to ends, based primarily on trial and error, a form of experiment used even by lower animals, but for men, a great step on the road to science. About 30,000 years ago men succeeded in portraying animals in positions of movement. A further passage to exact observation and the recording of nature occurred when they became food growers about 13,000 years ago. Sowing and reaping times were chosen according to the lunar and solar cycles (*see* CALENDAR: *Primitive Calendar Systems*). Numbering of days became necessary with city life. More settled agriculture needed specialized tools, and technology began to develop.

The age of stone passed into that of metals, which involved treatment of ores, extraction, and working. The development of rights in land demanded surveying: tradition has it that the an-

nual Nile flood rendered necessary an annual remeasurement of the fields of Egypt. Thus geometry (literally "earth measurement") was born. The cutting up of animals for food and the examination of their entrails for divination yielded, especially in Mesopotamia, some knowledge of bodily structure. These activities are among the sources of metallurgy, mathematics, and anatomy.

As society became yet more complex, commerce increased. The stewards and priests of palaces and temples needed records, so that systems of numerical notation were invented; and ultimately writing developed from pictographs.

Investigation of the ancient empires, notably the Babylonian, has revealed a far more extensive and systematic accumulation of astronomical and mathematical information than was formerly suspected. Science can therefore no longer be called a Greek product. Cumulative records began in the 4th millennium B.C. in the river valleys of both Egypt and Mesopotamia: but those that have survived are so discontinuous that the history of science in these ancient empires still cannot be written. *See,* however, CALENDAR: *Egyptian Calendar; Babylonian and Assyrian Calendars.*

(C. SI.)

III. INDIAN SCIENCE

1. General.—The scientific movement that originated in India spread widely. Its writings consisted of numerous treatises, commentaries, manuals, and technical dictionaries, which, in translation, gave rise to new movements in other parts of the East. In the West, it influenced and was influenced by the traditions of western Asia and Greece. It was based on rational principles in the sense that it sought natural explanations for observed phenomena and on a theoretic logic against which ideas were tested. The oldest Indian texts, older than the known scientific treatises, are already preoccupied with man and the external world and are almost devoid of magic, so much in vogue in other contexts.

Indian science concentrated chiefly on astronomy, physiology, and psychology; it neglected physics, geography, geology, and zoology. It made great progress in plant knowledge; but less in order to understand plants than to list their properties for food, medicine, or technical use. It composed (somewhat late) a chemistry which is mostly alchemy or materia medica.

Although Indian science speculates rationally, seeks detailed knowledge, observes, and forms explanatory hypotheses, rarely does it experiment, except in psychology, where it made its greatest progress thanks to its techniques of body and mind control. Elsewhere its theories were founded on insufficient facts, owing to its exclusively practical purpose in observation and lack of systematic and exact research. The theories gave mental satisfaction but were not fitted to a superficially regarded reality. This intellectual satisfaction, allied to a withdrawal within its ancient heritage while India was under foreign domination, eventually arrested the progress of science after a long period of brilliant development. But its tradition of study and logic was not destroyed, and modern India has adapted this intellectual inheritance to current international research.

2. Beginnings of Cosmogony and Astronomy.—The ancient Brahmanic and Vedic texts frequently refer to a general concept of natural law, Good Order, or Norm: the so-called *rta.* To it is opposed the *anrta,* the "Not-Good-Order." (The Old Persian *arta,* Avestic *asha,* shows this idea in Indo-Iranian prehistory.) This Good Order, later to be called *dharma,* the right ordering of all things, is universal, being cosmic, social, and moral. The universe shows forth the *rta* by the regular progress of the stars and the cycle of the seasons, especially the rains, upon which all life in India depends. Ritual ceremonies must be regulated by the Good Order of nature, so as to maintain or increase the prosperity of king and people. Hence the seasons and the positions of the heavenly bodies were early recorded on an approximately exact, luni-solar calendar (*see* CALENDAR: *Hindu Calendar*).

The religious texts correlate heavenly bodies and deities, but do not forecast human destiny by the stars. They explain certain astronomical phenomena by mythical stories eventually seen to disagree with observation; *e.g.,* that of King Moon, going into a "decline" on account of the complaints of his wives, the constellations, who take exception to his devotion to one of them at the expense of the rest—while the calendar records the moon's uninterrupted passage through each constellation in turn.

The Brahmana texts state that the world was created by the Word of the Veda, and assimilate the number of Vedic stanzas to that of the "moments" (*muhurta*) of the year, *i.e.,* 10,800 (30 moments a day $\times$ 360 days), the number of syllables in the stanzas amounting to 432,000 (this being also the number of years of the cycle of the universe called *kali yuga*). These numbers, or their submultiples or decimal multiples, were interpreted as the key numbers of the heavenly revolutions and of the laws of the universe and were therefore preserved in classical astronomy. Indeed, $108 = 27$ (basic number of *nakshatras*) $\times$ 4 (number of moon's phases); and $432 = 27 \times 16$ (number of portions of the lunar half-month). (*See* CHRONOLOGY: *Hindu.*)

The Vedic texts explain human and animal life by the circulation of breath (*prana*) in the body (see *Medicine,* below). These texts call doctors unclean, which shows that they originated in different regions from those where the *Ayurveda,* "science of longevity," was elaborated. (*See* MEDICINE AND SURGERY, HISTORY OF: *Traditional Medicine and Surgery in the Orient: India.*)

3. Early Mathematics.—The beginnings of mathematics are unknown; but the sacrificial formulas of the *Yajurveda* (*see* SANSKRIT LITERATURE: *Vedas*) bear witness to the use of high numbers among special names (up to *arbuda* = 100,000,000). The oldest mathematical texts are aids to Vedic religious ritual, and deal chiefly with the building of the *vedi* (sacrificial platforms) and the *citi* (heaps, tumuli, altars). The *Yajurveda* lists the kinds of altars, but it does not follow that the manuals are contemporary with it, for the geometrical problems solved in them apply to modifications (possibly later) of the buildings. Conversion theorems for figures concerning fixed proportions are given; and, in particular, the theorem of the square of the hypotenuse of a right-angled triangle and the sum of the squares constructed on the other sides is expounded. The texts belong in general to the high period of Vedic ritual (1st century B.C.).

4. Classical Mathematics and Astronomy.—Mathematics other than ritual geometry was mostly linked with astronomy; and from the first centuries A.D. astronomical treatises generally included mathematics. (However, the manuals of Mahāvīrācarya, 9th century, and Mallanna, 11th century, were for calculation only. The Bakhshali manuscript from Kashmir [*c.* 12th century] is a collection of arithmetical problems and solutions.)

The astronomical treatises contain astrology. From about the 4th century B.C., especially, the notion of sidereal influence on mankind is evidenced by proper names; *e.g.,* Candragupta, "the Moon's protégé," Brhaspatimitra, "Who has Jupiter for friend." But other names, such as Agnimitra, "Who has Fire for friend," imply that the planetary divinities were no more than protectors like the rest. Astrology properly so called appears in the first Christian centuries, with phonetically transcribed Greek terms in the Sanskrit texts. Its vast literature utilized the findings of scientific astronomy, on which there survive classical treatises and summaries of and allusions to lost treatises.

Classical tradition is represented, first of all, by the (lost) *siddhanta* "solutions" (Arabic *Sindhind*) of Paitamaha (Brahman), Vasistha, Paulisa (Paulus Alexandrinus), the Romaka (*i.e.,* Roman), and Surya (the Sun), the most important being the last named. They are known from a description and critique of their main data in the *Pañcasiddhantika* of the astronomer-astrologer Varaha-mihira (6th century A.D.). The chief works by individual authors are the *Aryabhatiya* ("[the work] of Aryabhata," whose author was 23 in A.D. 499 and wrote about the beginning of the 6th century); Brahmagupta's *Brahma-sphuta-siddhanta* (A.D. 628; *see* BRAHMAGUPTA); and Bhaskara's *Siddhantasiromani* (1150; *see* BHASKARA). These are not in a uniform tradition. Aryabhata stands apart especially in accepting the rotation of the earth on its axis. All consider time to be an indefinite cycle of "great yugas;" *i.e.,* periods at the end of which all the planets will return to the same position after a whole number of complete simultaneous revolutions. The 12-sign zodiac (*q.v.*) also appears now. (*See* CHRONOLOGY: *Hindu.*)

Astronomical tables were condensed into verse; and this led to the invention of several methods of expressing figures in words with a numerical value or in letters and combinations of letters (*see* ALGEBRA, HISTORY OF). The shifting of figures in position to signify their change of value and the use of zero in the form of a special sign or a space were known to Babylonia (in the sexagesimal system) before India; but India came first with decimal numeration and the notation of numbers with nine figures and zero (or gap). From India this system spread to the Arabs and the West (*see* MATHEMATICS, HISTORY OF). It is implied in Aryabhata's work by his method of extracting square roots. Trigonometry appears with the substitution (from the *Suryasiddhanta* and in Aryabhata) of the half-chord of the double arc for the chord of an arc (*see* TRIGONOMETRY: *History*). Aryabhata gives rules like: the product of two factors equals half the difference between the square of their sum and the sum of their square. Brahmagupta develops algebra as a general method for finding whole solutions to indeterminate equations of the second degree. The highest development is reached with Bhaskara, who devotes the first two parts (*Lilavati* and *Bijaganita*) of his work on astronomy to arithmetic and algebra, and who breaks down complex movements into uniform, instantaneous, combined movements. In the 18th century, a Sanskrit version of an Arab translation of Euclid was made for Maharaja Sawai Jai Singh II of Jaipur (1699–1743). Jai Singh also investigated European astronomy and built in the north of India several great observatories with huge instruments made of masonry (Jaipur, Delhi, Mathura [Muttra], Benares [Varanasi], Ujjain). The gnomon and the clepsydra had been, on the whole, the essential tools of ancient astronomy. Classical astronomy also used the armillary sphere and various circles, half-circles and quarter-circles. Bhaskara invented an astrolabe (*phalakayantra*).

5. Medicine.—Classical Indian medical knowledge is called *Ayurveda*, "knowledge of long life," the two chief traditions being those "of Atreya" (6th century B.C.) and "of Dhanvantari." The most important recension of the former is the *Agniveśatantra*, held to have been prepared by Agniveśa, one of Atreya's pupils, and edited by Caraka (possibly the Charaka who was doctor to the Indo-Scythian king Kanishka, *c.* 2nd century A.D.). This edition, the *Caraka Samhita*, includes some later additions. The Dhanvantari tradition, revised and completed by a certain Nagarjuna believed to be the Buddhist philosopher of this name (2nd–3rd century A.D.), is represented by the classical *Suśruta samhita* (named for Dhanvantari's disciple Suśruta), its present form dating from the first Christian centuries. The *Suśruta samhita* stresses surgical teaching and the *Caraka Samhita* diagnosis and prognosis. These works are the basis of a Sanskrit medical literature, some of which was translated into Tibetan and some from Tibetan into Mongolian. The *Rgyud bzi*, or "Four Books," is still popular in Tibet.

Such literatures exist in other Indian languages, especially Tamil, whose chief medical works, in the classical tradition, are attributed to Agastya and Teraiyar. Two other fashionable medical traditions of South India are that of the Sittar or "Perfect Ones," containing an adumbration of chemistry; and *yunani* or "Ionian" (*i.e.*, Greek) medicine, transmitted through the Arabs and used by the Muslims.

Theoretically, classical medicine consists of: (1) general surgery; (2) eye, ear, nose, and throat treatment; (3) general therapeutics; (4) science of disease-causing demons; (5) child care; (6) antidotes; (7) strengthening or restoration of youth; and (8) aphrodisiacs. But the great treatises rarely follow these divisions. First they comprise a general section consisting of discussion of a doctor's qualifications and of the various methods of healing. Then they deal with the occasions of falling ill; the body (*i.e.*, anatomy and embryology); and healing procedures. Most have supplementary treatises, such as Suśruta's *Uttaratantra* on ophthalmology. Caraka has, for example, the *Indriyasthana* on the sensory and motor faculties. The great medical texts are unsystematic collections of teachings, as is shown by their title of *samhita* ("collections").

Their doctrines, however, are coherent. The body is made up of a combination of the five great elements of nature in order of decreasing density: earth, water, fire, wind, and empty space. Earth and space, represented by the "full" organs (*e.g.*, the liver) and "hollow" organs (*e.g.*, the stomach), are inert but animated by wind under the form of breath; by fire under the form of bile (to "cook" food by digestion); and by water under the form of phlegm. These three elements (*dhatu*) cause illness when they are overexcited or inhibited. The body, circulated throughout by them, is composed of seven other elements also called *dhatu*, which result from combinations of different proportions of the world's raw materials. These are: (1) chyle; (2) blood, which is chyle dyed red by fire (*i.e.*, bile); (3) flesh; (4) fat; (5) bone; (6) marrow; and (7) sperm. All possess the vital juice of energy, centred in the heart.

Wind, or breath, has five forms: "in front," to assure respiration and swallowing; "going upward," to produce speech; "concentrated," to blow up the digestive fire; "diffused," to move the limbs; and "going downward," to assure excretion and childbirth. Bile and phlegm, too, each take five forms, heating and illuminating for the bile, lubricating and binding for the phlegm.

Anatomy shows acquaintance with bones, joints, and muscles but imperfect knowledge of the internal organs, although rudiments of dissection can be seen in the examination of corpses steeped in running water. Great importance is attached to the vital organs and large agglomerations of blood vessels and nerves. The heart is the seat of the spirit (*manas*), toward which the sensory impressions converge under the impulse of the breath. Classical physiology is thus essentially pneumatist.

Temperament and bodily functioning are determined by the predominance of one of the three elements and by climate, the seasons, behaviour, and diet. Sickness (apart from wounds, drownings, or diabolical possession, especially in children) is due to bodily changes, chiefly in the three organic elements, caused by stress. The medical texts hardly envisage supernatural or determinist causes of illness. For example, according to Caraka, smallpox is caused by a perturbation of bile and phlegm and not by a malignant goddess, as popular belief would have it.

Pathology classes maladies comprehensively according to the symptoms. There are fevers, skin affections, urinary troubles, sugar urine (*i.e.*, diabetes), etc. Within these broad classes, forms of illness are distinguished according to the signs of perturbation of this or that element. Diagnosis, therefore, is more of the underlying disturbances provoking the illness than of the illness itself.

Healing seeks simultaneously to cure symptoms, to abolish causes, and to reestablish, by exciting or calming medicines, the normal relations of breath, bile, and phlegm. For this, hygiene, correct diet, and drugs (mostly vegetable) are very important. Pharmaceutical preparations are powders, pastes, steepings, decoctions, infusions, etc.; doses are taken in water, oil, or butter. Decongestants, leaching, blistering (by caustics), fomentation, blasting with vapours, and bringing on of sweats are separately classified. Fumigation and inhalation apparatus, surgical instruments, and bandages are fully described. Surgery is practised mostly on limbs, but includes daring operations such as trepanning, removal of cataract and hernia, grafting, and even sutures of the bowels, these last being done by bringing together the edges of the wound and putting on ants to bite them, pulling off the ants' bodies, and leaving their heads behind.

Caraka studies in great detail the conditions affecting validity of the doctor's observations, diagnoses and prognoses, and procedures. Rational though his writing is, however, even to having the air of a treatise on logic, he admits methods of forecasting the course of an illness which are pure divination; *e.g.*, interpretation of dreams and of omens encountered by the physician on his way to the patient. (This would seem to be due to Persian influence in the west of India previous to the 4th century B.C.) On the other hand, he accounts for premonitory dreams of death, for example, by perturbation of the three elements.

Veterinary medicine for horses and elephants is based on the same principles.

New methods of diagnosis and treatment developed. The examination of the pulse, for example, became a complicated method of interpreting disturbances of health. But medicine owed most of its enrichment to chemistry and psychosomatic

techniques, which developed outside the classic tradition.

6. Chemistry and Alchemy.—Some chemical preparations are prescribed by Suśruta and Caraka. The remarkable early development of metallurgy is attested by the renown of Indian iron, an example being the pillar of almost pure iron at Delhi with an inscription of the 4th century A.D.

Caraka sketches a theory of chemical reactions in connection with the tastes (acid, salty, etc.), which distinguish substances, and can combine or neutralize one another.

Chemistry texts properly so called begin with the *Rasaratnakara* attributed to the Buddhist sage Nagarjuna, whom several traditions describe as an alchemist. There are numerous medieval and modern texts. Some treat of the preparation of medicines, others of definitely alchemical matters such as the elixir of life, transmutation of metals, etc. Mercury and its salts play a key role in all these processes. There are "great essences" (chiefly cinnabar; *i.e.*, red mercuric sulfide), which are bodies in their natural state; and "lesser essences"; *i.e.*, metals and salts. Chemical procedures are chiefly extractions, purifications, and calcination. The efficacy of medicines is increased by repeated intense heating in a closed vessel. Percolated and organic substances (*e.g.*, vinegar, urine) are often used as solutions.

There were other chemical traditions in southern India (not exclusively related to the Sanskrit ones), which survive in Tamil texts.

Physics and zoology were neglected, physics being chiefly represented by atomist philosophical theories of the universe (*see* INDIAN PHILOSOPHY: *Six Systems*).

7. Psychology and Psychosomatic Techniques.—The most original and advanced Indian science was psychology, which made use of psychical and physiological techniques, *i.e.*, yoga, for the achievement of mastery over mind and body.

The analysis of the workings of feeling and thought begun in the last centuries B.C. (Upanishads and old Buddhist texts) resulted both in an idea of the importance of subconscious life attested by memory and in a realization of the link between bodily and psychic functions. Every conscious experience leaves a trace in the soul. These traces are not lifeless imprints but come together in "psychic constructions," which in turn form an ethereal body within the physical one. This ethereal body governs the tendencies and reactions of the soul, and can be reincarnated after death in a new physical body corresponding to the tendencies animating it. The direction of its development could be controlled by the yoga-regulated selection of its experiences. Yoga techniques consist in discipline of the attention, neutralizing external impressions; fixity of meditation; the placing of the conscious mind in a chosen attitude; and finally the setting up of corresponding "constructions" in the unconscious mind. To achieve this the yogin calls on physiological techniques which facilitate fixing of the attention and neutralization of external impressions; makes use of postures, attitudes, exercises, and, above all, breath control. This last is of special importance, partly because of the pneumatist theory of physiology, and partly through experience of the effect of control on heart rhythm and psychosomatic functioning. In addition, some forms of yoga borrow ideas about the vital points (*marman*) from the *Ayurveda*. These points are referred to a theoretical anatomical form related to the movements of breath in the body during the exercises and to symbolical correspondences between microcosm and macrocosm. (*See* also YOGA.)

8. Spread of Indian Sciences.—The Indian sciences very probably had links with Greek science in Persian Achaemenid times (559–330 B.C.). The *Hippocratic Collection* mentions some Indian medicines: it also contains a treatise "On the Winds," which expresses classical Indian ideas. In astronomy, calendar durations decided upon by Heraclitus and the Babylonian Berossus correspond to Indian values. Indian science mostly spread, however, to Tibet, China, Japan, Indochina, and Indonesia. The astronomical system of *nakshatras* (constellations of the ecliptic entered by the sun and moon) is balanced by the 12-year zodiacal cycle of Jupiter used by the Chinese (*see* CALENDAR: *Chinese Calendar; Hindu Calendar*). Ill-defined relations existed between the Chinese and Indian alchemical systems. The medical sciences were not related; but some of the Taoist techniques for breath regulation and control of body and thought seem to have been influenced, if not inspired, by yoga brought to China in parallel with Buddhism. (J. L. A. F.)

IV. CHINESE SCIENCE

1. General.—The beginnings of Chinese science were contemporaneous with and parallel to those of Greek science, but soon developed the characteristic conceptions of correlation, association, and intuitive interconnection of the elements of the cosmos.

According to the Neo-Confucianist principle, the *Li* ("law," "reason") (*see* CHINESE PHILOSOPHY), all phenomena and all existing objects are like the contents of a basketmaker's shop, in which individual elements are without interest and only the finished product matters. The harmony of the universe results from the correct interrelationship of its parts. Thus, whoever knows the world (the macrocosm) knows man (the microcosm) by analogy.

The idea of the hierarchical order of the elements resulted in a construction which was completely arbitrary, but of such ethical, aesthetic, and cultural value that scientific progress was almost inconceivable.

Chinese science did not spread as did Indian, Greek, and Arab-Persian science. It has even seemed to be isolated; but in fact it was constantly enriched by foreign contributions and in turn exported its techniques (though not so much its speculations) to the West, from the 3rd century B.C. (Han dynasty). There was direct contact with Europeans from the 16th century. But the acquisitions of Chinese science from elsewhere were not equal to those of Arab science from Greek.

2. Astronomy.—Astronomy was not clearly distinguishable from astrology and was much more important for social and political life than as a science. The sovereign, at the centre of space and time, binding heaven and earth together, ruled over the cosmic order by imposing his calendar upon his subjects, just as his Western counterparts imposed their coinage.

The calendar was luni-solar. The shell oracles of the Yin dynasty (1st millennium B.C.) show that the solstices (recorded on gnomons) defined the limits of the year, just as the moon defined the month. To make lunar and solar periods agree, a fictitious intercalary month was inserted every third civil year. Thus the Metonic cycle, a 19-year period containing seven intercalary months, was discovered. (*See* CALENDAR: *Chinese Calendar.*)

The perfecting of the calendar, especially for forecasts of eclipses and other spectacular phenomena, was the chief aim of Chinese astronomy.

The chief instruments were the gnomon and the clepsydra or water clock. There were also armillary spheres and celestial panoramas moved hydraulically to reproduce the movement of the heavenly bodies. The toothed-wheel mechanism (for astronomical clocks) was discovered in the 11th century A.D. by Su Sung (1020–1101). Chinese astronomers distinguished the equator from the ecliptic; determined the length of the tropical year as distinct from the sidereal year; learned the precession of the equinoxes; observed sunspots and eclipses; and made star maps and catalogues.

The astronomer Kuo Shou-Ching (1231–1316) may perhaps have invented spherical trigonometry.

3. Mathematics.—Spills and the abacus were used for calculation. A group of spills, arranged vertically or horizontally on a flat surface, represented each figure. Errors in calculation could result from accidental displacement, however. The abacus (*q.v.*), being steady, was better (forms of it are still used in Asia and eastern Europe, great speed in calculation being attainable). Multiplication tables were used at least from the 6th century B.C. The calculating book of Sun Tzu (*Sun-tzu suan-ching*) is a manual of logistics believed to be of this period, containing a multiplication table from nine times nine to one times one. A manuscript in the Bibliothèque Nationale in Paris, brought to France from Tun-huang by Paul Pelliot (1908), gives a fragment of table; another, also from there, contains a Tibetan translation of a frag-

ment of Chinese table. Such tables were necessary on account of the extreme accuracy required for the use of both spills and abacus, for numbers had to appear in succession and workings-out could not be recorded.

Chinese mathematical works were compilations for the solution of the practical problems of engineering, business, and administration. Rare were the true scholars to whom mathematics was an end in itself; as can be seen from Chinese treatment of the Pythagorean theorem and the calculation of π.

The first part of the *Calculating Book of the Gnomon* (*Chou-pei-suan-ching*), 5th–3rd century B.C., applies Pythagoras' method to the right-angled triangle with sides 3, 4, and 5; and the method seems to have been thought of as a practical procedure rather than as a theorem.

Another portion of the *Calculating Book*, composed about the 2nd century B.C., uses the number 3 for the value of the relation of circumference to diameter; this value is again discussed in the *Nine Chapters on the Mathematical Art* (*Chiu-chang suan-shu*), with commentary by Liu Hui (*c.* mid-3rd century A.D.). Liu Hui uses the method of inscribed polygons, and obtains the number 3.14 with a polygon of 192 sides. A 5th-century astronomer and mathematician, Tsu Chung-Chih, would seem to have calculated π with six decimals and to have obtained an excellent approximation (π between 3.1415926 and 3.1415927); but the work believed to be his was lost about the 10th century.

The Chinese algebrists of the second half of the 13th century were perhaps the world's greatest. Chu Shih-Chieh (*c.* 1299) outdid all his predecessors, anticipating Pascal, Leibniz, P. Ruffini, and W. G. Horner; only to be forgotten.

Modern Western scholars find that Chinese mathematics, despite its strongly archaic character, can enrich modern mathematics with forgotten perspectives and new attitudes.

4. Physics.—There was much stress on acoustics, because of the political and social importance of music in ancient China. The pitch standard ("yellow bell," *huang-chung*) was a bamboo flute. This standard served also for the sound wavelength. With it, Prince Chu Tsai-Yü discovered the equally tempered scale about 1595 (*i.e.*, before M. Mersenne [1636], J. S. Bach, and J. P. Rameau). He also left a treatise on time and a perpetual calendar (1584).

5. Chemistry.—Chemistry never succeeded in freeing itself from alchemy, which began in the 4th century B.C., culminated during the 2nd and 3rd centuries A.D., and decayed from the 14th to the 17th centuries. Its relations with the other alchemical systems of Europe and Asia are ill known. Ko Hung (*c.* 281–340) was head of a scientifico-mystical school which sought potions of longevity and immortality (not for the philosopher's stone). Their adumbration of natural philosophy, had it been better exploited, could have helped the development of science and technology, as witness *The Exploitation of the Works of Nature* (*T'ien-kung k'ai-wu*, 1637). (*See* also ALCHEMY; TAOISM.)

6. Natural Sciences.—In geology, knowledge of natural gases, oil, mineral deposits, and fossils ("dragons' bones") is very ancient. Chang Heng built (A.D. 132) the first known seismograph, and earthquakes were recorded very early.

Botany was highly developed; China's vegetation was rich and many useful foreign plants were imported. The first botanical gardens appeared during the Han dynasty. Li K'an published a *Treatise on Bamboo* (*Ch'u-p'u hsiang-lu*) *c.* A.D. 1299. The study and drawing of vegetables is older and is a topic of numerous treatises. Gardening flourished from T'ang times (7th–10th centuries A.D.), when new varieties of flowers were obtained by selection and grafting.

There were treatises on zoology too; and animals studied included the unicorn, phoenix, and dragon. The Chinese had little gift for rearing mammals (except for Pekingese dogs), but showed remarkable skill with insects, mollusks, fish, and birds. They also raised bees, silkworms, and crickets (for fighting). From the 10th century A.D. they bred fish, creating new varieties from goldfish. In the 13th century they tried to obtain artificial pearls. Lastly, they used birds as messengers (pigeons), fishers (cormorants), hunters (falcons), for fighting (quails, cocks), and as pets.

7. Medicine.—Medicine was based on the two principles of the elements and the vital breaths. There were 12 paired, symmetrical vessels (canals); linking organs and viscera, and continuously circulating the vital breath, the blood, and the *yin* (female principle) and the *yang* (male principle). The Jesuit translators equated these last with Galen's moisture and warmth.

Illness was diagnosed by inspection, examination of the tongue, and, above all, study of the pulse. Measles, smallpox, beriberi, rickets, scurvy, and goitre were thus identified; their causes were traced and they were treated by diet. Certain skin diseases also were known (scabies, leprosy, syphilis). Dietetics and hygiene were highly developed, although for centuries they served the illusory purpose of helping to secure immortality.

In sex, the "bedroom treatises," from the Han dynasty onward, distinguished between ordinary intercourse with the aim of impregnation and *coitus reservatus*, in which the man controls his genital reflexes so that he does not ejaculate.

The extremely rich body of medical materials is expounded in the *Pen-Ts'ao*, exhaustive studies of mineral, vegetable, and animal products, used for medicine, pharmacology, biology, dietetics, chemistry, industry, folklore, history, geography, and even philology. The most famous is the *Pen-Ts'ao Kang Mu* of Li Shih-Chen (1518–93), the only important 16th-century work outside the system of Galileo and Vesalius. In modern China and Japan, the *Pen-Ts'ao* are the subject of a flourishing science (pentsaology).

Ancient Chinese treatments included moxibustion (cauterization by wormwood leaves), acupuncture (*q.v.;* needle puncture acting on the 12 canals), movement therapy, breathing and relaxation techniques; all of which are of renewed interest to modern medicine. *See* also CHINA: *History*. (P. HU.; M. WO.)

V. GREEK SCIENCE

1. Beginnings and the Classical Period (*c.* 600–*c.* 300 B.C.).—The figure traditionally associated with the beginning of science among the Greeks is Thales of Miletus in Asia Minor, who flourished in the first half of the 6th century B.C. He made certain geometrical discoveries, though the elements of geometry, according to Greek tradition, came from Egypt (and those of astronomy from Mesopotamia). The Egyptians, however, had not generally reached beyond an empirical use of certain special relations of such figures as triangles and rectangles, pyramids and spheres. The question as to how far they generalized mathematical conceptions is still under discussion. On the other hand, there is evidence from their art that they knew and constantly practised the "golden section." Thales or his Greek contemporaries succeeded in generalizing such cases and made other discoveries in elementary mathematics.

By the 6th century B.C. Greek-speaking peoples had founded colonies in southern Italy and Sicily, whose intellectual activity, especially that of Pythagoras and his followers, was significant for science (*see* PYTHAGORAS AND PYTHAGOREANISM). The use of letters to express numbers by Phoenicians, Hebrews, and Greeks encouraged the Pythagorean belief in the independent existence of numbers and their mystical and magical application. It was the Pythagoreans who gave to the word "mathematics"—which first meant simply "learning"—its special relationship to number. Their conception of numbers as the elements of all things and of the heavens as a numerical and musical scale gave to science the important conclusion that there is a correspondence between the working of the human mind and the working of nature; and their doctrine of the relation of the four elements of matter to the four "humours" composing the human body was the first attempt to trace the rules of the external world to the working of man's body. By the middle of the 5th century B.C. the Athenian school of thought was commencing the specialization thenceforth characteristic of science, in the form of recognition of the independence of mathematics and medicine. Exemplars of this tendency were the geometer Hippocrates of Chios (*c.* 460 B.C.) and Hippocrates of Cos (fl. 400 B.C.), known as the "father of medicine." (*See* also MEDICINE AND SURGERY, HISTORY OF: *Early Greek and Roman Medicine.*)

The intellectual history of the 4th century B.C. is filled by the

gigantic figures of Plato and Aristotle, who are considered here only in their relation to science.

Many of Plato's thoughts assume a mathematical guise, and he thought that other studies should conform to the certitude of mathematics. He highly esteemed astronomy, regarding the motions of the heavenly bodies as perfect geometric forms; and for his followers astronomy became a field for the exemplification of mathematics rather than (as today) for its application.

Plato regarded the mathematical form of the universe as evidence of the rational mind of its Creator. "God," he is said to have said, "ever geometrizes." To deny the existence of mind as a separate entity was, he held, to assume the universe to be the result of accident, which was a denial of the validity of philosophy. It is not inconsistent with this view that Plato respected Hippocrates the physician, who "was the first who separated science from philosophy." But the trend of Platonism in general and of ancient Platonism in particular was usually away from observational activity, although there have been many evident exceptions, and Platonism has often been helpful to science both in stressing its quantitive aspect and in opposing an entrenched and static Aristotelianism. It was from Pythagorean teachers that Plato derived the so-called "Platonic bodies," the five regular polyhedra which have equal sides and equal angles. Many centuries later mathematicians proved that the possible number of regular bodies is only five, and it was from a consideration of these bodies that Kepler developed the first unitary scheme of the universe (A.D. 1596).

Aristotle devoted his incomparable genius to systematizing and organizing the whole area of knowledge. His earliest and, from the modern scientific point of view, his best efforts were on biological topics. The whole of his science and indeed the whole cast of his mind was deeply influenced by his firsthand observations of living things. In his *Parts of Animals*, he sets forth his view of the relation between biology and "physics," the latter being for him a general description of the universe. He says:

> Of the things constituted by nature some are ungenerated, imperishable, eternal; others subject to generation and decay. The former are excellent beyond compare and divine, but less accessible to knowledge. The evidence that might throw light on them and on the problems which we long to solve respecting them is furnished but scantily by our senses. On the other hand, we know much of the perishable plants and animals among which we dwell. We may collect information concerning all their various kinds, if we but take the pains. (Somewhat paraphrased.)

Living things are for Aristotle the type of existence, and existence as a whole presents, for him, evidence of design. He attempted to analyze the nature of generation, of heredity, and of sex; and treated many other topics. There is a profundity in his biological thought which gives it a permanent value. He was a first-class observing naturalist in the modern sense. (*See* also BIOLOGY: *History*.)

Aristotle, like Plato, had Pythagorean tendencies, which he exhibits in his physical scheme. He emphasized the "perfection" of the circle and of the sphere, on which therefore the world is modeled. For him the heavens are a series of concentric, crystalline, mechanized spheres arranged round the earth as a central body.

The mechanical scheme of the universe set out by Aristotle and his successors suggests a series of geared wheels and may have been suggested to the Greeks through some such complex apparatus. It was the basis of mankind's theory of the universe for 2,000 years. It may thus be summarized: (1) Matter is continuous; (2) All mundane things are made up of four "elements," which in their turn manifest the four "qualities"; (3) Stars and planets move with uniform circular velocity, embedded in crystalline spheres, centred round the earth. Each sphere is subject to the influence of those beyond; (4) Circular, changeless, eternal movement is perfect order. It contrasts with the rectilinear movement which prevails on our changing and imperfect earth; and (5) The universe is limited in space and within an outmost sphere. It is unlimited in time, being subject as a whole neither to creation nor to destruction.

2. Alexandrian Science: First Period (300–30 B.C.).—For the 500 years after Aristotle the scientific centre of the world was Alexandria. There mathematics assumed a prominent position. Among its first exponents was Euclid (fl. 300 B.C.), whose *Elements of Geometry* determined instruction in the subject for the next 22 centuries. (*See* GEOMETRY.)

Another Alexandrian teacher, Aristarchus of Samos (fl. *c.* 270 B.C.), made the first scientific attempt to measure the distances of the sun and moon from the earth and their relative sizes. He estimated the sun as 18 times more distant than the moon (instead of more than 346 times). That he thought the sun far larger than the earth may have suggested to him that it was improbable that a large body would revolve round a relatively minute one. Anyhow, he held that the earth rotates, and revolves round the sun.

Hipparchus (fl. 146–127 B.C.), the greatest astronomer of antiquity, erected at Rhodes the first recorded observatory. He developed trigonometry, by which numerical calculations can be applied to figures drawn on either plane or spherical surfaces. He made numerous observations and collated the records of Babylonian and earlier Greek astronomers to see if astronomical changes had taken place in the course of the ages. These comparisons led him to his two brilliant conceptions of the precession of the equinoxes and of planetary movements. (*See* HIPPARCHUS.)

Anatomy and physiology became recognized disciplines at Alexandria. The first medical teachers of the school, contemporaries of Euclid, began the practice of dissecting the human body publicly. They compared its structure with that of animals and opposed both Plato and Aristotle by regarding the brain as the sole seat of the intelligence. They observed that arteries, unlike veins, pulsate, but did not ascribe this to the heart's action, thinking that it was an activity of the arteries themselves. They thought of the nerves as hollow and as conveyors of the hypothetical "nervous fluid." They distinguished between the main brain, or cerebrum, and the lesser, or cerebellum; observed the cerebral convolutions of both man and animals; and associated their greater complexity in man with his intelligence. Their experiments on animals led them to distinguish between the posterior nerve roots of the spinal cord, which convey sensations, and the anterior, which convey the motor impulses.

From about 400 B.C. the doctrine of atoms (rejected by Aristotle) was current in the Greek world. Later, though important philosophically, it had on the whole a retarding effect on Greek science despite some influence on Alexandrian anatomy and physiology. Some Alexandrian anatomists supposed that nonatomic "air," equivalent to pneuma (*see* below), is taken in by the lungs and passes to the heart, where it enters the blood and is changed into a peculiar kind of pneuma—the "vital spirit"—which is sent to the various parts of the body by the arteries. It is thus carried to the brain, where it is further altered into another pneuma, the "animal spirit." This, they believed, reaches different parts of the body through the hollow nerves.

One Alexandrian school considered the pneuma that circulates in the body to be ultimately drawn from the air, or world-pneuma. This assumption gave a physiological basis to the philosophical conception of the spirit of man as part of the world-spirit frequent in later writings of the Stoic school (*see* STOICS).

The biological work of the early Alexandrian Theophrastus (*c.* 372–*c.* 287 B.C.), a pupil of Aristotle, should be mentioned. His botanical writings are the best-arranged ancient biological treatises. Among his many acute and accurate observations is his clear distinction between monocotyledons and dicotyledons. Interesting too is his attempted distinction of sex in plants, successful only for palms. His splendid botanical works had a definite influence on modern biology.

The greatest of the Alexandrian school was Archimedes (*c.* 287–212 B.C.) of Syracuse, constantly recalled in the mathematical construction of the Archimedean spiral (*see* CURVES, SPECIAL: *Section 38*); in the mechanical construction of the screw of Archimedes (*see* ARCHIMEDES, SCREW OF) for raising water; and for his exposition of the doctrine of levers.

Perhaps the earliest existing work of Archimedes is *On Plane Equilibria*. It sets forth certain fundamental principles of me-

chanics as rigorous geometric propositions. The work opens with his famous "postulate": "Equal weights at equal distances are in equilibrium; equal weights at unequal distances are not in equilibrium but incline toward the weight at the greater distance." This postulate developed into the principle of the steelyard (*see* WEIGHING MACHINES), and led Archimedes in the end to the discovery of the centre of gravity in a variety of geometric figures.

Among the achievements of Archimedes is his method of measuring the areas of curved figures and surfaces, most simply expressed in the effort to "square the circle." His predecessors had broached the important idea of "limits"; Archimedes, however, employed limits systematically.

The principle is that a square can be inscribed within a circle, the sum of its sides being evidently less than the circumference and its area less than that of the circle. The number of sides can be doubled to make an eight-sided figure within the circle. The proposition remains true, however many the sides; but as their number is increased the discrepancy decreases. "In the limit," when its sides become no more than points, the polygon may be conceived as becoming the circle. Archimedes realized that this limit can be approached as nearly as we wish (*see* LIMIT).

Archimedes proved that the area of a circle is equal to that of a triangle of base equal to the circumference of the circle and of height equal to its radius, by finding the ratio between circumference and diameter. To do this he sought the limit approached by the sides of regular polygons both inscribed in and circumscribed on the circle. The limits for their ratio to that of the diameter he found to lie between $3\frac{10}{71}$ and $3\frac{10}{70}$. The latter has since been generally accepted as a good approximate value of the quantity known as π.

He applied a comparable process to other curves. In his treatise *The Quadrature of the Parabola* he brings both an inscribed and a circumscribed rectilinear figure into relation with a curve. The two rectilinear figures, as it were, compress the curve, one from within and the other from without, until they coincide with it. This procedure, together with the use of mechanics for solutions subsequently demonstrated by geometry, leads to the consideration of his extremely important treatise *The Method*.

For the most part, Archimedes, like other Greek men of science, gives his proofs but does not tell how he reached them. In *The Method*, however, addressing Eratosthenes (*see* below) he recalls the discoveries which he had sent to his friend on a former occasion and then says that he is now explaining how he made them.

The Method applies two principles: (1) that a plane figure may be regarded as an aggregate of an infinite number of parallel lines; and (2) that consideration of the respective weights of two plane figures, one curved and one rectilinear, may reveal the approximate area of the curved figure. The same process may be applied to demonstrate relationship between the volumes of solid curved figures considered as aggregates of an infinite number of parallel plane figures. It amounts to a practical solution of problems of the relation between areas or volumes by analysis, mechanical or other, after which the investigator returns to a synthetical mathematical process. He thus gains by experience some estimate of the solution before he seeks its demonstration.

An Alexandrian successor of Archimedes, Apollonius of Perga (fl. 220 B.C.), developed the knowledge of conic sections. The familiar terms parabola, ellipse, and hyperbola derive from him. The idea of conics was for more than 1,500 years a purely intellectual exercise, generations of mathematicians discussing the properties of these curves, which were not known in nature. Then in 1618 Kepler discovered that planets move in elliptical orbits. Surely here the human mind showed itself attuned to nature, whose ways it had unwittingly explored for all these ages.

Among the Alexandrian luminaries was Eratosthenes (*c.* 276–*c.* 194 B.C.), the greatest scholar of his age. His most important achievement was the measurement of the globe of the earth by an operation of beautiful simplicity (*see* GEODESY: *Early History*).

Having measured the earth, Eratosthenes considered the known parts of it. Here, like all his predecessors, he had self-imposed limitations, regarding the habitable world as wholly within the Northern Hemisphere and forming only about one-third of that. He considered the land surface of the world to be longer from east to west than from north to south and estimated the distance from the Atlantic to the Eastern ocean as 78,000 stadia (about 7,800 geographical miles) and the distance from the parallel of the Cinnamon land (Ceylon) to that of Thule as 38,000 stadia. As he estimated the circumference or equator at 250,000 stadia, Eratosthenes could also estimate the circumference at the parallel of the "Pillars of Hercules" (Strait of Gibraltar), which he knew was also that of Rhodes. This fundamental parallel (latitude 36°) passed through other important points—the westernmost point of Spain, for example, and the southern points of Italy and of Greece—and along the Taurus Mountains. Eratosthenes estimated the total circumference of the world at this parallel to be 200,000 stadia. All but 78,000 stadia of this was sea.

At right angles to the parallel of Rhodes, Eratosthenes determined a north-south line between Alexandria and Syene. This, produced northward, he regarded as passing through Byzantium and thence to the Dnieper River. Southward, he considered that it passed to Meroe and then along the Nile.

These fundamental lines, together with those on other parallels of latitude and lines of longitude, are sufficiently accurate for constructing an outline world map in which the Mediterranean area is recognizable.

3. Alexandrian Science Under the Roman Empire (30 B.C.–A.D. 200).—Despite the dominance of Rome, science remained Greek in spirit and language. It was finally synthesized by two pupils of the Alexandrian School, Ptolemy (fl. *c.* A.D. 140) for cosmology and geography and Galen (*c.* A.D. 130–*c.* 200) for anatomy and biology.

Ptolemy's *Almagest* was of the highest significance for later astronomical development. Its basic cosmic conceptions come from his predecessors but are expounded with the utmost skill. Ptolemy invoked epicycles to explain the movements and behaviour of the planets, employing them to resolve some errors and inconsistencies of Hipparchus. However, he retained eccentrics to explain certain elements in these movements.

Ptolemy possessed a series of astronomical instruments, the existence of which proves the high degree of mechanical skill evolved during generations in exact metalwork. Among his greatest achievements is the determination of the distance of the moon by parallax: the method is in principle that still in use. Ptolemy estimated the moon's distance to be 59 earth radii, which is not very far from the truth. Working on an eclipse method of Hipparchus, he estimated the sun, however, to be only 1,210 earth radii distant, which is one-twentieth of the truth. He had no way of estimating the distances of the lesser planets but followed tradition in accepting apparent speed as the main test of nearness. His scheme passed to the Middle Ages. (*See* PTOLEMAIC SYSTEM.)

Ptolemy's other great work, his *Guide to Geography*, was a product of knowledge brought to him by Roman imperial expansion. He developed his own manner of representing on a plane surface the curved surface of the earth. In his "projection" the parallels of latitude are arcs of concentric circles. Chief among the parallels are the equator and the circles passing respectively through Thule, through Rhodes, and through Meroe. The meridians of longitude are converging straight lines.

The great biological synthesis of antiquity was made by Galen, a native of Pergamum in Asia Minor. Galen studied at Alexandria but spent most of his active life in Rome. He elaborated a complete physiological scheme, generally accepted until modern times. Three kinds of pneuma or spirit are involved in addition to the world-pneuma or air. The basic principle of life was drawn from the air by breathing. Entering the body through the windpipe, it passes, Galen thought, to the lung and thence to the left ventricle of the heart, where it encounters the blood. His view of the changes that then take place in the blood was most ingenious and was based on experiments, but the errors that it involved remained current till the 17th century and beyond.

Galen's scientific works are among the most influential of all time. Nevertheless, he established no school and had neither disciples nor direct followers. On his death anatomical and physio-

logical inquiry ceased abruptly. Mathematical and mechanical research continued a little longer than biological, but they pass into a silver period in the 3rd century. For science in general the Middle Ages begin effectively about A.D. 200.

VI. MEDIEVAL SCIENCE

There followed a period well-nigh barren in positive scientific results until versions of the Greek scientific works appeared in Arabic in the 10th and 11th centuries and in Latin, mostly from Arabic, in the 13th and 14th. Though these later "Middle Ages" added little to the growth of scientific knowledge, they did something for its coherence and presentation. The Eastern and the Western share are here treated separately.

1. Islamic Science (*c.* 850–*c.* 1200).—In the West, science disappeared with the Latin Roman Empire. In the Greek-speaking Eastern Roman Empire, a disintegration began that was never arrested. Much valuable Greek scientific literature was, however, translated into Syriac, the language of the heretical Nestorian Christians. Between 750 and 850 these versions were revised and many others added. From 850 to 950 Nestorian translators at Baghdad rendered the Syriac versions into Arabic. This presentation of Greek science in Arabic is the primary source of "Arabian science," which left a deep impress on the Latin world.

The Arabic language and culture spread afar: to Portugal in the west, as far as China in the east, and over many degrees of latitude. The most characteristic Arabic scientific developments were in alchemy, in mathematics, in astronomy, and in medicine.

Alchemical is a term used to describe very diverse kinds of literature, most of it highly mystical. From Egypt, from Byzantium, and perhaps from China, however, Arabic-speaking "alchemists" derived many recipes and methods for industrial processes, and their experiments demanded special apparatus. The greatest Arabic-writing alchemist, Rhazes (d. *c.* 930), makes the earliest known suggestions for furnishing a chemical laboratory. Industrial technology and alchemical apparatus of Arabic origin had an important role in the rise of modern chemistry.

In mathematics, only the Greeks attained so high a standard as the Hindus. By the 9th century, the Islamic peoples were using the Indian system of numerical notation, the so-called "Arabic numerals." The most influential work propagating this was by the Persian Mohammed ibn Musa al-Khwarizmi (*c.* 825), from whose name the medieval word for arithmetic, *algorism*, was formed. The title of the Latin version of his *Algebra* is the first Western use of the word in the mathematical sense: it means "rectification," that is, transposition of negative terms of an equation to the opposite side. (*See* ALGEBRA, HISTORY OF.) The Arabs also considerably developed the geometrical and optical works of the Greeks.

Astronomy and astrology were constant preoccupations of the Islamic world (*see* ASTROLOGY; ASTRONOMY: *History of Astronomy*). They were specially developed in Spain, groups of experts in them being formed at Córdoba and at Toledo. The Toledan tables of positions of stars were drawn up in 1080, their authors seeking to replace the Ptolemaic by a strictly concentric system.

In medicine, as in alchemy, the first and greatest original Muslim writer was Rhazes, whose erudition was all-embracing. He made the first distinction between measles and smallpox. It is impossible to mention Arabic medicine without recalling Avicenna (980–1037; *see* MEDICINE AND SURGERY, HISTORY OF: *Arabian Teaching and the School of Salerno*), whose work was for centuries standard in Latin as it still is in Arabic. He was scientifically much inferior to Rhazes. The physiology and anatomy typified by him was a grossly deteriorated version of the Greek. The great Arabic contribution to medicine is the introduction of new vegetable drugs, many of which are still in use.

2. Science in the Latin West (1100–1450).—Until the 11th century no Western scientific movement needs consideration, since Spain in those earlier centuries was within the orbit of Eastern culture. In the 11th century the West began to come into relation with the East through Latin translations of Arabic works. These, during the following three centuries, came mostly from Spain but also from Sicily, from Provence, and from Syria.

Medieval Latin had at first no technical scientific vocabulary. The translators, therefore, often transliterated Arabic words, and thus many Arabic names of stars, of chemical substances, of apparatus, of plants, and even of anatomical parts passed into Latin. Some have survived in modern vernaculars. Moreover, Aristotle's views of the structure of the universe as conveyed in these Arabic-Latin translations from his works and from those of Ptolemy provided the medieval world-picture. The recovery of the works of Aristotle was a major factor in the 13th-century revival of intellectual coherence and gave to scholastic science its essential character. This "medieval Aristotle" was later modified by a slow infiltration of direct translation from the Greek.

Great emphasis has often been placed on medieval "forerunners" of science, notably on Roger Bacon (*c.* 1220–*c.* 1292); his scientific contribution was, however, small, as was that of several other medieval writers who sought to outline a philosophy of science. There were experiments with the compass (perhaps under Arabian influence); the problem of the path of light within a spherical lens was partially solved on a mathematical basis; a parabolic burning mirror was constructed or at least attempted; and a solitary genius made a workable astronomical clock of great complexity. But these achievements, scattered over two or three centuries, are minute compared with those of the best Greek or Arabic centuries. They hardly justify a revision of the standard view that modern science arose in the 15th and 16th centuries, primarily as a recovery of Greek science. But the better empirical technology of the new science was the product of the medieval centuries; and its method of exposition was a contribution of scholastic thought.

Astronomy—which cannot at this stage be distinguished from astrology—was certainly the main scientific interest of the scholastic age, but its practical results are meagre. Western knowledge of astronomy was largely based on the activity of King Alfonso X the Wise of Castile, who collected at Toledo a body of scholars, mostly Jews, who had access to Arabic sources and calculated a set of astronomical tables (1252; *see* ALPHONSINE TABLES).

In pure mathematics the original achievement of the scholastic age was small, though a borrowed element, the so-called "Arabic" notation, was the source of the modern method of notation, which, however, was adopted only very slowly. The *Liber abaci* of Leonardo of Pisa (*c.* 1170–1230), which advocates this system with great skill, appeared in 1202, and is the first book by a Latin Christian to employ it. Other works of Leonardo were much more original, but being before their time had less influence.

VII. THE RENAISSANCE AND EARLY MODERN SCIENCE

1. The Humanistic Century (1450–1550).—The Western recovery of the Greek classics from surviving Greek traditions and texts was in full action by 1450, and the scientific heritage had been completely recovered by 1550. This century saw also the introduction and spread of printing and illustration.

The first Latin exponent of the value of experiment is the German cardinal Nicholas of Cusa (1401–64), whose recorded careful experiment on a growing plant, proving that it absorbs some weight from the air, is the first modern formal biological experiment and the first experimental proof that air has weight. Carried to its logical conclusion it would have ended the old doctrine of the pneuma. Nicholas wrote a book on the use of the balance and showed that he could apply the experimental method in detail. His views led him to believe that the earth moves, though he reached no formal astronomic theory.

Copernicus (1473–1543), despite the vastness of the scientific change involved in his theory of the earth's revolution around the sun (*see* ASTRONOMY: *History of Astronomy;* COPERNICUS, NICOLAUS), was a scholar more than an observer, and conservative. Although he was aware of error in the Ptolemaic system and although he knew that some Greek philosophers (including, it is believed, Aristarchus) had held that the earth moved, his heliocentric theory retained the Ptolemaic ideas of the sphericity of the universe and its limitation by the sphere of the fixed stars. He retained also the conception of uniform circular motions of the celestial bodies, as well as some of the Ptolemaic epicycles (the

smaller of the circles in which the sun and planets moved).

Important during the humanistic century is the impact of the main technical instrument of Renaissance art, namely perspective. This new device, though it rapidly became a convention, is truly scientific because, given certain conditions, its truth can be demonstrated experimentally and its accuracy estimated. Leonardo da Vinci (1452–1519) was its greatest scientific exponent, but it was the product of many minds. It produced sounder and simpler plans of engines of all kinds and was of great aid in many technical arts. All developments involving geometry in three dimensions became more easily intelligible. Above all it made possible the adequate representations of living things and their parts. Thus Leonardo could both illustrate many devices and make clear on paper, for the first time, many details of the structure of human and animal bodies.

This same movement produced in the first half of the 16th century many exact studies of plant form. Its most characteristic product was the *De humani corporis fabrica* (1543) of Andreas Vesalius (1514–64), the first great scientific monograph in the modern manner, which appeared within a few weeks of the work of Copernicus. Copernicus is essentially a medieval figure, Vesalius a modern who employed all the resources of the art of his age to illustrate personal observations. But it must be remembered that the researches of Vesalius were solidly and consciously based on knowledge of the 1,300-year-old findings of Galen.

2. The New Status of Greek Science.—Medieval thought on the material world was essentially based on that of the Greeks and especially on Aristotle. By 1500 the whole works of Plato and almost the whole of Aristotle were available in reliable Latin translations direct from Greek. The Renaissance saw a revival of Platonic thought. Moreover, from about 1550 an accumulation of biological works based on Aristotle tended to confirm him as "the master of those that know" in that field. At about the same time, however, in the fields of mechanics and of cosmology, cracks were beginning to appear in Aristotle's system. These were brought out clearly by Simon Stevin (1548–1620) in the last years of the 16th century. He investigated the Aristotelian theory of the fall of bodies. His name is associated with the method of resolution of forces, with the distinction of stable and unstable equilibrium, and, above all, with the law of equilibrium on an inclined plane in a famous demonstration of the impossibility of perpetual motion. He perceived that science could not well advance until its mathematical framework had been improved. This was facilitated by his introduction of decimal fractions. Several other contemporary devices helped in the same direction: for example, the employment of letters to represent quantities (1591), the conception of conic sections as stages in a series from the line pair to the circle (1604), and the invention of logarithms (1614). With such equipment and with the improvement in the technique of instrument making, notably of lenses, the great exponents of scientific method of the earlier part of the 17th century —Galileo (1564–1642), Kepler (1571–1630), Descartes (1596–1650)—could proceed with their work. By the middle of that century Aristotle's physics and cosmology had fallen, while his biology was on its way to being rebuilt on his own lines; mathematics had become a recognized instrument of research in the physical sciences; the contributions of Archimedes were recognized, even when applied in the biological field; alchemy had effectively been replaced by a chemistry built mainly on the technological traditions of craftsmen but aided by the apparatus of the alchemists; the exploration of the structure and function of the human body had become a profession. Science itself had emerged as a vocation supported by its specialized philosophers, by its specialized exponents in the arts of discovery in various departments, and by its skilled instrument makers. Science had thus become consciously mature and was prepared for its formal organization in the period which followed. All this was exemplified in the brilliant personality of Galileo.

3. The Beginning of the Modern Scientific Outlook.—The outlook of the first half of the 17th century thus involved on the negative side a rapid waning of the scholastic reliance on ratiocination, and on the positive a new reliance on experience and especially on experiment as a way to acquire knowledge. A man of science of the 20th century could converse freely with one of 1650, with a very few of 1600, with a handful of 1550, but hardly with one of 1500. What had happened in this first and critical half-century of modern science, from 1600 to 1650?

In cosmology, the explanation of planetary movements by Kepler and Galileo's discoveries with the telescope had destroyed geocentric thinking. They had done away with the moral connotation of physical events expressed in the idea of the perfection of the sphere. Incidentally they had destroyed astrology.

There was a new view of the nature of matter. The systematization of mineralogical knowledge and the increased literacy of industrial technicians had given birth to chemistry and destroyed alchemy, removing there too a moral significance from material phenomena.

To express his findings in mathematical terms became the ideal of the physical experimenter. Broadly, this was acceptance of a relation between number and form, for which Descartes is still commemorated in the term "Cartesian co-ordinates."

The idea of acceleration is a development of that of mathematical expression but one of such significance that it demands special mention. Its introduction by Galileo was a great seminal idea for all aspects of physics and notably for celestial physics.

Physiology set physics as its ideal. The discovery of the circulation of the blood by William Harvey (1578–1657) and the chemical investigation of the bodily processes by J. B. van Helmont (1577–1644) made possible the first comprehensive treatment of the subject by Descartes.

Aristotelian methods were extended to the classification of animals and of plants. "Natural history" perhaps first achieved this status at the hands of Joachim Jung (1587–1657).

From this point on it is harder to consider science as a whole. Specialization began. Having dichotomized into mathematics and medicine at its dawn about 400 B.C., science had now developed physico-chemical and biological branches. Soon there would be much more complex division and recombination.

VIII. DETERMINISM AND POSTDETERMINISM

From the second half of the 17th century to the end of the 19th men lived in a world of which the material framework had been outlined by Isaac Newton (1642–1727). The attitude of mind called "scientific determinism," sometimes rashly equated with "scientific materialism," is not quite as old as that. Newton's great work, *Philosophiae Naturalis Principia Mathematica*, was published in 1687. It had little immediate effect: it could at first be read only with difficulty even by mathematicians, and its demonstrations were intelligible to a mere handful of Newton's contemporaries, none of whom, moreover, was interested in introducing them to less scientific readers. Its wider influence began to be felt after Newton's death.

Newton left no doubt as to what he regarded as the principles of science and his hopes for it. All its difficulties, he held, "seem to consist in this—from the phenomena of motions to investigate the forces of nature, and then, from our knowledge of these, to demonstrate other phenomena." And he goes on: "I wish I could derive all phenomena of nature by some kind of reasoning from mechanical principles; for I have many reasons to suspect that they all depend upon certain forces by which the particles of bodies are either mutually attracted and cohere in regular figures or are repelled and recede from each other." Thus he hoped to fit all material events into a framework of relatively simple and mathematically expressible rules. That is what he regarded as the task of his "philosophy" and what for two centuries was regarded as the task of "science."

Newton's marvelous genius ran at its ease only along the line of demonstration to the senses, and it was as a means for such demonstration that his superb mathematical apparatus was designed. He was also much interested in theology and spent much time on it; but for philosophy (in the modern sense) he had a definite distaste. The influence of his science on philosophy (and doubtless through it on religion) has mostly been unwitting, though no less important on that account. But this scientific giant pre-

sents a different figure when he works in a field ill-suited to his trained talents, and with tools the use of which he had not learned in the best tradition: Newton's incursions into chemistry, which consisted of experiments on alchemy, make a depressing story of waste of most precious effort.

In spite of all this, Newton formulated, more clearly than any of his predecessors, those primary statements on which "scientific determinism" came to be based. The term was introduced about the middle of the 19th century, but the passages just quoted from Newton's *Principia* contain the essential elements in the determinist faith. Some of its exponents have, however, extended the conception of "phenomena" to include events within the mind itself, treating them also as determinate.

The conception that mental events are determinate has been much more revolutionary for religion than anything in the Newtonian system proper. That man's body works on ascertainable mechanical principles had indeed seemed obvious to Descartes before Newton and as long ago as 1627. The followers of Descartes directed the thought of the age for about a century. The Cartesians, however, recognized with their master that men are more than mechanical systems. Only if man's entire nature, including his thinking, could be fitted into a long chain of causal development could his actions and his mind be treated as truly determinate. It was in this very direction that the conceptions of Descartes and Newton were extended in the middle and later years of the 19th century, especially after 1859, in which year appeared both Darwin's *Origin of Species* and Marx's *Zur Kritik der politischen Ökonomie*. The former contains the germ of biological determinism, the latter that of psychological determinism. Both are sources of the doctrine of social determinism.

Admittedly in the 18th century the French encyclopaedists and some of the first English utilitarians had been determinist; as, in some sense, had been certain Renaissance and medieval heretics; and a number of thinkers of antiquity. But it was Newton who first demonstrated a law of physical movement which seemed wholly unrelated to any spiritual order. It was this that gave to the conception of determinance an immediacy as well as a practical workaday aspect that it had never previously worn.

Although Newton, like many thinkers, sought always to cast his scientific conclusions into a mathematical form, in fact much science is not, or has not yet become, mathematical.

So far as the 19th century is concerned, it was the geologists and the evolutionary biologists who raised most acutely the issues primarily responsible for "the conflict between religion and science," by demonstrating that man's place in nature is very different from that which had been taught. Evolutionary doctrine quickly affected historical, social, linguistic, and even literary studies. No line of scientific thought has so rapidly seized so many minds as that associated with Charles Darwin (1809–82). Now, however, the thunders of the evolutionary conflict rumble only from the further frontiers of philosophical discussion.

But in considering this and other great seminal ideas it must always be borne in mind that science never considers and cannot consider the world as a whole. It is essential to any science that it proceed by abstracting a part of the universe, to be considered by and for itself. This fragmentation carries with it the secret of the scientific triumph and the secret of a common philosophic fallacy. While one science may combine with another to form a new science with its own technique (for example, entomology and psychology have together produced insect psychology), and while one science can use the results of another, yet each science can describe only its own little bit of the universe in its own terms. These terms are derived by a comparison of yet smaller bits of the universe with other smaller bits. Such terms have little or no application outside the science for which they were devised.

Gross errors have been made, even by scientific men, through transference of scientific terms from their original field of reference. This may be seen from, for example, the mass of fruitless disputation that has arisen around such words as "evolution," "instinct," "element"; or the employment of the words "race," "tradition," "heredity," "nationality," "value" to give a scientific appearance to nonscientific judgments; or, again, the endless confusion between the scientific, the philosophic, and the theological usages of such terms as "substance," "theory," "idea," "cause," "purpose," "function." The language of the sciences is peculiarly difficult to master, for language is a product of complex and ever-changing social conditions.

The definition of terms cannot be discussed in this article; it must suffice to indicate that there cannot be a "science" of the whole universe: for it is impossible to attain this by adding the sciences together, and there are vast regions of experience, such as art, literature, and philosophy, that are refractory to scientific treatment. In any event, there is no such thing as a unitary scientific treatment of the material universe.

Nevertheless, a science must not be mistaken for a mere fragmentation of knowledge, despite the necessary minuteness of scientific analysis, which may give to the untrained an impression of triviality. The objective of scientific analysis is to reach a series of points—a natural frontier—from which the mind can return along its tracks to arrange its findings into a theory or generalization. This frequently involves a reunion of the findings of workers in different departments. Often these findings become the starting places for new analytic explorations and the bases of yet further generalizations and thus of new sciences.

The great scientific generalizations are, however, essentially abstractive; and by verbal though not real paradox, the astronomer's "expanding universe" is as much an artificially separated fragment of the universe as the minutest entomological research. The "expanding universe" of the relativists is, after all, not the universe of ordinary thought, but simply a technical term referring to something the existence and growing size of which must be assumed in order to fit together the mathematical deductions from certain physical measurements.

During the 20th century it became increasingly apparent that a necessary condition for scientific observation is the postulation of certain basic data, things "given" or taken for granted (*i.e.*, metaphysical foundations); as is shown in the famous saying of Archimedes: "*Give* me a place on which to stand, and I can move the world." This necessity is strikingly illustrated by Einstein's theory of relativity, the essence of which may be expressed by saying that the very act of observation affects that (*i.e.*, the behaviour of light from the stars) which is to be observed. Similarly, in connection with intra-atomic physics, it was found that the more accurately the position of a particle could be specified, the less accurately could its velocity be predicted, and vice versa. The principles of "indeterminancy" or "uncertainty" (*see* UNCERTAINTY PRINCIPLE) and of relativity run athwart the whole view of scientific determinism and are of great importance for the consideration of the ultimate validity of scientific demonstration. It cannot, however, be assumed that, because minutely exact prediction may be impossible, physical action is ultimately indeterminate. The element of indeterminancy may be in the researcher rather than in nature; and in any case whether it is or is not possible to ascertain all the determinate elements in nature is an entirely different question. Moreover, as neither physical relativity nor physical indeterminancy is involved in the individual observational sciences, a scientific theory can generally be safely launched regardless of them. It should, however, be remembered that scientific conclusions are essentially partial or temporary; that even the great generalizations are not goals but starting points for further exploration; and that the sciences can at best describe only a part of the universe.

To the question "What does science say about religion?" the strictly true answer is "Nothing"; if by religion is meant the view that the universe has a meaning, an idea of what its meaning is, and an idea of man's place in it. Admittedly, science can analyze the effects of a particular religious system on the outward conduct of a society or of an individual; but it cannot pass judgment upon its validity nor touch the essential nature of existence, of which man is aware through "consciousness."

Consciousness is the ultimate datum, the thing taken for granted; the judge, as it were, before whom science must recite its narrative of experiences of phenomena. Their recital, and that alone, is the role of science.

See also Arabic Philosophy; Indian Philosophy; Logic, History of; Philosophy. *See* also the *History* sections of the articles on particular sciences; also the references under "Science, History of" in the Index. (C. Si.)

Bibliography.—G. Sarton, *Introduction to the History of Science*, 3 vol. (1927–48), *The Appreciation of Ancient and Medieval Science During the Renaissance, 1450–1600* (1955); A. Wolf, *A History of Science, Technology, and Philosophy in the 16th and 17th Centuries*, 2nd ed. rev. by D. McKie (1950), *A History of Science, Technology, and Philosophy in the 18th Century*, 2nd ed. rev. by D. McKie (1952); C. Singer *et al.* (eds.), *A History of Technology*, 5 vol. (1954–58); M. Schrimpf, *La collection Souan-king che-chou* (1963); Li Yen, *Histoire des sciences mathématiques en Chine* (1963); Yuan Han-ching, *History of Chemistry in China* (1956); P. Huard and Ming Wong, *La Médecin chinoise* (1964); V. B. Datta and A. N. Singh, *History of Hindu Mathematics* (1935–38); G. R. Kaye, *Hindu Astronomy* (1924); Sir P. C. Raya, *A History of Hindu Chemistry* (1907–25); J. Filliozat, *The Classical Doctrine of Indian Medicine*, Eng. trans. (1964).

SCIENCE, TEACHING OF. In most countries, science instruction begins in the elementary school. The teacher at this level cannot be a specialist in science teaching, for he must also teach the other subjects of the curriculum, but there is a growing tendency to employ specially trained science consultants who do demonstration teaching in science and who assist the regular teachers in their work.

The teaching of science as a separate subject generally begins with the middle or junior high school (typically, grades six to eight or seven to nine; *i.e.*, second to fourth or third to fifth form). Instruction at this level in Great Britain, the United States and many other countries offers, under the heading of general science, all major aspects of the subject. Many countries, however, follow the French system of offering separate courses in physics, chemistry and biology, starting with simple and descriptive accounts in the sixth or seventh grade (second or third form). The instruction in these fields is typically cyclical, and the instruction continues throughout the secondary school. The Anglo-U.S. pattern of general science is followed by courses in physics, chemistry and biology in the later secondary-school years. Less rigorous courses such as physical science are sometimes offered for less capable students.

Postsecondary-school science courses are many and diverse. Universities offer sequences in each major field of science, from introductory and general courses to highly specialized courses which provide intensive instruction in restricted fields leading to research or to technical vocations. Technical schools provide specialized training for industrial work; science instruction in such schools is focused more on the practical than on the theoretical aspects of science.

Goals of Science Teaching.—The modern world requires a very large number of skilled technicians and practitioners as well as of persons who can add to the store of fundamental knowledge through their own research. A primary responsibility of the science teacher, therefore, is to foster the development of the scientist. But there is a second goal, no less important, which is that of educating each person, to the degree of his capabilities, in scientific understanding to enable him to cope intelligently with his environment and to contribute both to his own well-being and to that of society. Another way of looking at goals is to examine the nature of science itself. Science is, first, a body of tested facts and concepts that satisfactorily interpret natural phenomena and disclose causal relationships, and, second, a means for discovering such facts and principles and for applying them in the solution of problems. There is no single "scientific method," but there are general procedures which distinguish the scientific methodology as a discipline from the methodologies of other fields such as art or philosophy. This aspect of science, heavily emphasized in graduate programs of universities, formerly had little attention at lower levels, particularly in European schools, but emphasis on scientific attitudes and methods as an instructional goal is increasing.

Methods of Science Teaching.—Methods are a reflection of goals. For example, the development of knowledge sufficient to pass the rigorous governmental examinations has been a chief instructional goal in the French *lycées*. As these examinations emphasize detailed memory of facts and principles (and, at least formerly, largely ignored reflective analysis of data, applications of principles, and so forth), instruction has placed a premium on memorization. The formal lecture method is widely employed, and students are expected to take notes and to remember what the teacher has taught. Laboratory work is seldom used, and even demonstrations are generally neglected in favour of a systematic and highly verbal treatment of a logically organized body of scientific facts. Practice in the U.S., as another example, has tended in the opposite direction. The secondary schools have placed a much stronger emphasis on individual laboratory work, demonstrations, problem solving and informal lecturing and discussion. About the middle of the 20th century, however, an American trend toward greater rigour and systematic instruction was paralleled by a European trend toward greater flexibility with less emphasis on strictly memory work. (R. W. Bu.)

SCIENCE FICTION deals with the human drama, the conflicts and adventures, arising out of scientific discovery in the future. There are two basic types. Science fiction proper, an almost step-by-step development of possibilities from known scientific or social data, is exemplified by Arthur C. Clarke's *The Sands of Mars* (1952), which works out the conditions encountered by the first explorers on that planet; by Hal Clement's *Mission of Gravity* (1954), which deals similarly with an entirely alien world and its beings in another solar system; and by George Orwell's *1984* (1949), which presents an uncannily perceptive extrapolation from contemporary totalitarian social patterns. Science fantasy, on the other hand, can leap directly to whatever farfetched assumptions may be necessary to the story. Obeying the sole requirement of dramatic plausibility, it permits the imagination not only to go beyond the known and proved but to contradict it when that is necessary, as exemplified by Alfred Bester's *The Demolished Man* (1953), Ray Bradbury's *The Martian Chronicles* (1950) and Edward E. Smith's *Lensman* series (various dates). Needless to say, the two basic types are frequently combined.

The major themes of science fiction are: (1) space travel to and from other planets, solar systems and galaxies, including the exploration, settlement and exploitation of other worlds as well as encounters with, or between, extraterrestrial life forms; (2) time travel to the future or the past and, similarly, travel to "alternate universes" which are usually different versions of our own; (3) psychological and biological changes in man, brought about by nature or science, and similar changes in other species; (4) "supernormal" powers and talents, achieved either through technology or the advancement of such "fringe sciences" as parapsychology; (5) science applied, directly or indirectly, to human relations for either constructive or destructive purposes. These themes are clearly illustrated in such short-story anthologies as Groff Conklin, *The Best of Science Fiction* (1946) and *A Treasury of Science Fiction* (1948); August Derleth, *Beyond Time & Space* (1950); and Raymond J. Healy and J. Francis MacComas, *Famous Science Fiction Stories, Adventures in Time and Space* (1957).

History.—Even in ancient times certain science-fiction themes were used as vehicles for social satire and criticism and for tales of wonder. These treatments dealt with what was then admittedly impossible. True science fiction, however, assumes a "this *can* happen" attitude which could not exist until the scientific method was defined and until the mechanization of everyday life made great numbers of people aware of science. It was not until the latter part of the 19th century, especially in the works of such writers as Jules Verne and H. G. Wells (*qq.v.*), that science fiction began to acquire its 20th-century forms and its functions not only of entertainment but of speculation and prophecy. In this last it has been more accurate technologically than socially.

As a separate and self-aware literary form, science fiction may be said to date in the United States from Hugo Gernsback's founding in 1928 of the magazine *Amazing Stories*. In 1937 it entered a new epoch when John W. Campbell, Jr., took over the editorship of *Astounding Science Fiction* and established a story pattern emphasizing not only the element of scientific speculation but those of character and human drama as well. Advances in

nuclear power and rocketry during and after World War II increased its popularity tremendously. In the United States the number of specialized magazines rose from 8 in 1945 to more than 30 a few years later, and the number of novels and anthologies increased even more rapidly. Serious journals of comment and opinion took note of science fiction. The general magazines, motion pictures, radio and television began to use it much more often. A specialized critical literature came into being.

This phenomenon was especially marked in the United States and the rest of the English-speaking world; however, in varying degrees it took place in most civilized nations—for science fiction had accomplished something unique in the history of literature. The central technological achievements of the period—radar, television, automation, electronic brains, atomic power for peace and war, rockets and earth satellites—had been science-fiction commonplaces for a generation before their practical realization. With a remarkable degree of general accuracy science fiction had heralded and defined the electronic age, the atomic age and the age of space.

Influence.—Through its readers science fiction had considerable sociological influence. Surveys showed that many of the readers of the better science-fiction magazines were young scientists and technologists or students destined for these fields. Similarly, many science-fiction writers, such as Arthur C. Clarke, Isaac Asimov, Robert Heinlein, John Taine and Edward E. Smith, were themselves scientists or engineers. Science fiction influenced many young people to choose the sciences as a career; it stimulated the imaginations of many scientists in the forefront of discovery; it helped to orient many intelligent laymen to the scientific world of the present and the future.

Nevertheless, science fiction did not break with literary tradition. The same artistic rules apply to it. Its characters must be as fully developed as any others; its plot structures must be as coherent; its emotional impacts must be just as strong. That is why, with the infinite variation of its themes, it holds at its best so much for readers and writers alike.

BIBLIOGRAPHY.—J. O. Bailey, *Pilgrims Through Space and Time: Trends and Patterns in Scientific and Utopian Fiction* (1947); Everett F. Bleiler, *The Checklist of Fantastic Literature* (1948); Reginald Bretnor (ed.), *Modern Science Fiction: Its Meaning and Its Future* (1953); J. J. Bridenne, *La Littérature française d'imagination scientifique* (1950); B. Davenport, *Inquiry Into Science Fiction* (1955); B. Davenport *et al.*, *The Science Fiction Novel* (1959); D. B. Day, *Index to the Science-Fiction Magazines, 1926–1950* (1952); L. Sprague de Camp, *Science-Fiction Handbook: the Writing of Imaginative Fiction* (1953); L. A. Eshbach (ed.), *Of Worlds Beyond: the Science of Science Fiction Writing* (1947); Damon Knight, *In Search of Wonder* (1956); Patrick Moore, *Science and Fiction* (1957). (RD. BR.)

SCIENTIFIC MANAGEMENT: *see* MANAGEMENT SCIENCES.

SCIENTIFIC METHOD is a description now reserved for the procedure by which we gain knowledge in empirical studies such as physics, chemistry, and physiology. At one time the word "science" was applied to all systematic studies or organized bodies of knowledge, including mathematics and theology. There was a tradition derived from Plato and Aristotle that the system to be expected in a science should be such as we have in mathematics; that is, that all the asserted propositions which are not themselves self-evident should be derived from others that are self-evident. In accordance with this usage the word "scientific," first introduced in its Latin form as a rendering of *epistemonikos* in Aristotle's *Posterior Analytics*, i, 2, was applied only to reasoning which produced knowledge like that which we have in mathematics. For many centuries it remained a synonym for "demonstrative." Thus John Locke says in his *Essay Concerning Human Understanding*, iv, 3, § 26, "How far soever human industry may advance useful and experimental philosophy in physical things, *scientifical* will still be out of our reach"; and he explains his meaning later in the paragraph by the remark "*Certainty* and *demonstration* are things we must not in these matters pretend to."

The modern use of "scientific" noticed above appears to have been established in England about the middle of the 19th century. At the beginning of that century physicists and chemists were still called philosophers, and the tools of their trade philosophical instruments. The word "scientist" was invented by William Whewell in 1840, and the earliest evidence in the *Oxford English Dictionary* for an acknowledged particularization of the word "science" in its special modern sense comes from 1867.

Clearly there may be many different methods by which scientists gain knowledge in different fields of research. Thus Archimedes found a way of determining specific gravities, and A. H. L. Fizeau found a way of measuring the speed of light. In modern times the special methods of research used in the various sciences have become more numerous and more complicated, so that a large part of the training of a future scientist is the imparting of techniques. Knowledge of these is important not only because it is necessary for one who is to make new contributions as an experimenter but also because it is an indispensable link between observation and theory.

A physiologist may use a cathode ray oscillograph without being an expert in electronics, but it will not help him to get evidence for or against a hypothesis about the working of nerves unless he understands its own working at least in outline. Since such a special method involves application of scientific results already attained, it cannot be expounded without an exposition of science. But when we speak of scientific method in the singular, we ordinarily mean some pattern of reasoning common to all empirical sciences, and that is the subject of this article.

Traditional Views: From Aristotle to J. S. Mill.—If there is any common method of the sciences it must obviously be related to what is common in the aims of the various kinds of scientists. Aristotle gives two accounts of the matter which had great influence on subsequent thought.

In the first place Aristotle introduces the notion of induction (*q.v.*) as a way of learning distinct from demonstration and logically prior because it provides the premisses needed for demonstration. But in his *Prior Analytics*, ii, 23, he talks of induction as a kind of syllogism in which we reach a universal conclusion from an exhaustive survey of the cases it covers; and in his *Posterior Analytics*, i, 1 and 18, he talks of induction as the establishment of a universal truth by consideration of an instance or instances which reveal to thought the necessity of the connection asserted. In modern times these two distinct procedures have been called summative and intuitive induction respectively. Neither can be identified with the procedure of ampliative induction, by which universal propositions are in fact established in the empirical sciences. For scientists do not claim to have proved that copper is a good conductor of electricity by examining every specimen of copper, nor do they claim to have discovered the necessity of the connection by careful reflection on some single experiment. Secondly, Aristotle says in various places (*e.g.*, in *Posterior Analytics*, i, 2) that science is knowledge of causes (*see* CAUSALITY). Perhaps he meant by causes in this context much the same as we mean by "explanations," but his detailed account of the various kinds of causes that he recognized would not be accepted by modern scientists as a statement of their objectives, and the doctrine is not in any case connected with either of the accounts of induction noticed above.

For more than 2,000 years, however, these two pronouncements of Aristotle determined the course of discussion about scientific method even among those philosophers who professed hostility to his doctrines. Thus Francis Bacon, David Hume and J. S. Mill (*qq.v.*) all assumed that the business of the empirical scientist was to establish universal propositions about causal connection, though they differed from Aristotle in the accounts that they gave of causes. Some part of scientific activity may indeed be described as a search for causes of kinds of phenomena, but it now seems clear that this is not a satisfactory description of all that scientists try to do. We need consider here only the general outline of the doctrine which was introduced by Francis Bacon in his *Novum Organum* of 1620 and accepted for a long time after as a correct account of the method of the new science.

According to Bacon it is the business of the scientist to discover the "forms" of phenomena, but in his usage this word "form" (derived presumably from Aristotle's theory of formal causes) means what would now be called a necessary and sufficient condition.

Thus F is the form of P if and only if P never occurs without F but always occurs with F. If either is capable of varying in degree, it is required also that the other should vary in similar fashion. For the discovery of forms in this sense the proper procedure is to digest our information about the phenomenon in three tables showing presence, absence and degrees: the first includes positive instances of the phenomenon, the second negative instances (*i.e.*, cases which are in most respects like known positive instances but do not in fact exemplify the phenomenon under investigation) and the third instances which vary in degree. Together they should enable us to refute all false suggestions put forward in reply to the question "What is the form?" and so to reach the true answer by elimination. We can say, for example, that anything which does not vary in degree when the phenomenon under investigation varies cannot be its form. It is assumed, of course, that every phenomenon must have a form, or necessary and sufficient condition.

J. S. Mill, whose *System of Logic* (1843) had great influence in the 19th century, wrote of causes rather than forms, but defined a cause as an invariable antecedent and modeled his account of induction on Bacon's. He spoke of it, however, as employing a number of different methods: for example, a method of agreement, in which the cause of a phenomenon is revealed by the consideration that it is the only circumstance (other than the phenomenon itself) in which positive instances agree; and a method of difference, in which the cause is revealed by the consideration that it is the only circumstance (other than the phenomenon itself) in which a positive and a negative instance differ.

This doctrine of scientific method could not have obtained the popularity that it once enjoyed if it corresponded to nothing at all in scientific inquiry. Mill's method of agreement is indeed used in the purely observational sciences and his method of difference in the experimental sciences. But the doctrine does not cover the whole field of scientific activity and moreover is misleading in its suggestion that the business of discovery can be reduced to rules. Even when the work of a scientist can be described without distortion as a search for necessary and sufficient conditions, it is not correct to assume that he can find these by application of a standard procedure.

Bacon and Mill both take for granted that there is no serious difficulty in drawing up an exhaustive list of circumstances which deserve consideration in a search for the necessary and sufficient condition of a phenomenon; yet there is in fact nothing to indicate what should be considered in this connection as a single circumstance. When we single out for attention half-a-dozen distinct features of events, say A, B, C, D, E, and F, why should we not also take account of their conjunctions and disjunctions, that is to say, of complex features such as might be represented by "A and B," "C or D" or even more complicated formulas like "[A or (B and C)] and [(D and E) or F]"? And if we allow for all this complication what assurance can there be that the half-dozen features with which we started are the only elementary features that deserve consideration? Mill admits in a somewhat half-hearted way that there may be what he calls plurality of causes, but he does not recognize that his methods are defective in general for lack of any systematic enumeration of possibilities.

Varieties of Scientific Activity.—It is very difficult indeed and perhaps not very profitable to find a formula which will adequately characterize all scientific activity. For while some scientists are concerned for all or most of their time with the making of generalizations from experience, others are concerned more with the elaboration of deductive systems in which such generalizations may be derived from hypotheses that cannot themselves be tested directly because they deal with unobservable entities such as electromagnetic waves. And apart from these two groups, whose methods may be called primary and secondary induction respectively, there are others whose interest in nature is more like that of a historian. Thus the palaeontologist who tries to trace the biological evolution of man wants to discover what happened in certain places at certain times, and the geologist and the cosmologist are equally interested in the course of past events. All use generalizations, but none of them is primarily interested in these for their own sake. If such inquiries are to be called scientific, it is not even correct to say that all empirical science is inductive, unless indeed the meaning of the word "inductive" is stretched to cover not only the secondary induction noticed above but also any reasoning concerned in any way with generalizations from experience. Perhaps the best we can say is that the pursuit of science is the search for knowledge and understanding through formulation of the laws of nature. This excludes the collection of disconnected scraps of information and draws attention to the special importance of natural laws in our conception of science. Whatever can be said usefully about scientific method in general must be related to these.

The beginning of science is classification (*q.v.*). Whatever we recognize must be of some kind, and language even at its lowest level is full of general words. But the same things may be classified in many different ways; *e.g.*, according to size, to shape, or to colour. Each classification is based on real characters of the things classified, and each may be useful for some human purpose. There are, however, some that seem more important than others, those, namely, for which the classifying terms of our most primitive language are nouns as distinct from adjectives. Thus "grass," "tree," "stone," "cow," "horse" and "dog" are said sometimes to mark natural kinds, whereas "red," "square" and "hard" serve only for artificial grouping. The point of the distinction is not that we can make anything red or square or hard by calling it so but rather that recognition of a thing as a piece of grass or as a tree or as a stone is the gaining of an important clue to what may be expected of it. If the thing before us is a blade of grass, then it will probably grow bigger, but not to a height of more than a few feet; when touched, it will bend; after a few months it will turn brown and gradually decay; and so on. It is not necessary, however, that I should satisfy myself of the thing's behaviour in all these respects before venturing to call it grass: fortunately I can recognize grass by its appearance and then make a number of prophecies about it with fair confidence. There is very little, on the other hand, that I can safely assume about the future behaviour of a thing merely because it is red.

Here, then, in the classifications called natural, we have the beginning of our intellectual mapping of nature. Long before there was any conscious pursuit of science men took it for granted that there were bounds to natural possibility; *i.e.*, that not all the conceivable combinations of characters were realized in nature. We cannot discard all their assumptions without abandoning also the language inherited from them; but we can and do improve on their classifications. Many of the terms for metals, which our children learn with ease because of the importance of metals in modern life, have been added to the vocabulary of men within historic times.

Primary induction is the deliberate attempt to find more laws about the behaviour of the things that we can observe and so to draw the bounds of natural possibility more narrowly. In so far as it is successful, it enables us to make more inferences from the observed to the unobserved and, in particular, more predictions by which we may guide our conduct. But this is not its only value. Already at this stage there is intellectual satisfaction to be had in the discovery of laws: we have a beginning of explanation when we can say that a piece of wire has melted in a candle flame because it is lead and lead always melts at the heat of a candle flame.

Sometimes scientists or writers on science who are interested chiefly in the more advanced stages of science speak of primary induction as though it scarcely deserved to be called scientific. And it would indeed be a great mistake to suppose that science is never anything more than the making of generalizations in the fashion of 17th-century naturalists. But generalization from experience is an indispensable feature of science. In the more advanced sciences such as physics and chemistry, observations are directed by an interest in theory (*q.v.*); but what the scientist deduces from his theory and then tries to test by observation or, if possible, by experiment (*i.e.*, observation in circumstances over which he has control) is always some universal proposition. Here primary induction comes late in the temporal order of a scientist's activities, but it is still primary in the sense that it is basic in his final argument and logically independent of the superstructure of theory. For a generalization which has been confirmed by experi-

ence only after it has been deduced from a theory is not invalidated by the discovery of a flaw in the theory, though it may perhaps lose something in reliability when it is no longer linked with other generalizations through the theory.

With primary induction we may conveniently include arguments of the kind sometimes called proportional generalizations. If we have observed many *A* things and found that the frequency of *B* things among them has remained fairly constant in the neighbourhood of a fraction *f*, we are inclined to say in general terms that the proportion of *A* things which are *B* is about *f*. Unless the class of *A* things is known to be finite, we cannot take this remark as an ordinary statement of proportion; but in practice we may use it as a guide to expectation, when we wish to know whether something which is *A* is likely to be also *B*. Often it is said to be a statement of the probability (*q.v.*) of an *A* thing's being *B*.

Secondary induction is the attempt to incorporate the results of primary induction in an explanatory theory covering a large field of inquiry. At this level in science it may be impossible to make progress without the introduction of concepts far removed from those used by primitive men in their classification of perceptible objects. In retrospect we can see that physical theory has contained such concepts for a long time (*e.g.*, the concept of energy); but until the 20th century the fact could be ignored because it was still possible to provide in imagination some sort of model for the interpretation of the statements of physical theory. When it was supposed that electromagnetic waves were transmitted through an ethereal jelly, the model was already inappropriate in some features but not in so obvious a way that it could not still be taken seriously by great scientists. Since then the difficulty of bridging the gap between theory and common sense has become more obvious. Modern views on this subject are discussed in EXPLANATION.

The Role of Hypothesis.—From the nature of the scientific enterprise it is clear that the reasoning by which it proceeds cannot be reduced to validating schemata like those of deduction; that is, to patterns of inference which guarantee the truth of the conclusion provided only that the premisses are true. There may of course be deductive arguments in natural science, but the laws and theories which are said to be established by scientific research are not deduced from observations. On the contrary, each is put forward first as a hypothesis (*q.v.*) which together with certain initial data entails the facts of observation. Our effort is to construct a deductive system "from the wrong end up" but not a trivial system in which the hypothesis itself could be deduced from the facts of observation. To be worth consideration a hypothesis of law must restrict the field of possibility more than that field is already known or assumed to be restricted; and similarly a suggested theory must be more restrictive than the set of laws which it is supposed to coordinate and explain; in other words, it must have some consequences other than those we already accept.

For these reasons it has been said that the conclusions of ampliative induction are never more than probable; but this pronouncement is unfortunate in some ways. The word "probable" is ordinarily used to recommend propositions when we are not completely satisfied about them; but we do in fact claim to know some results of induction (for example, that arsenic is poisonous), and adoption of the suggestion that we should apply the word "probable" to all results of induction would therefore lead to a blurring of a useful distinction. On the other hand, it is important to appreciate that the cases in which we say that we know some result of induction differ only in degree from cases in which we are not prepared to commit ourselves by use of the word "know"; and this is at least part of what philosophers have tried to convey by saying somewhat paradoxically that the results of induction are never more than probable. In mathematics, a general proposition such as C. Goldbach's conjecture of 1742 that every positive even integer is the sum of two primes (*see* NUMBERS, THEORY OF) may perhaps be regarded as highly probable because it has so far survived all attempts to find a counter-example. But no accumulation of evidence of this kind would ever justify a claim to knowledge of the truth of the generalization: for that, it would be necessary to produce an argument of a totally different kind, and the possibility of such an argument is at least conceivable. In natural science, however, there can be nothing corresponding to mathematical proof, and the customary requirement for the use of the word "knowledge" is therefore different.

The theory of statistics (*q.v.*) supplies principles for the marshaling of complex evidence in many different inductive studies, but if we take "scientific method" to mean a rule or set of rules by obedience to which any fairly intelligent person can make discoveries in natural science whenever he wishes, we must admit that there is no scientific method. For the all-important act in scientific discovery is the finding of a hypothesis that will survive testing, and this cannot be reduced to a routine. When we are looking for a necessary and sufficient condition of a phenomenon, or, less exactingly, for a sufficient condition only, we naturally consider only circumstances which are to be found accompanying some instance; and attempts were made by Bacon and by Mill to formulate rules of primary induction on this basis.

But the notion of accompanying is not very precise, and it is in fact impossible to set out any exhaustive list of circumstances that deserve consideration, because an intellectual search of this kind admits of no device like the systematic quartering of the ground by which we can make sure of success in a search for some material object that has been lost. An argument from analogy (*q.v.*) may perhaps suggest an interesting hypothesis here or in secondary induction, but again there is no infallible procedure for discovering fruitful analogies. The finding of the right hypothesis is always a problem, rather than a task. But in this it resembles the finding of proofs in most parts of mathematics, and it must not be thought that success in either case is just a matter of luck.

If we take "scientific method" in the looser sense of a program or policy, then undoubtedly there is a scientific method, namely the policy of making hypotheses for the purposes of primary and secondary induction as described above. Once formulated, this seems to be obviously rational, and some philosophers of modern times are inclined to talk of such scientific activity as though it were nothing but the commonest of common sense. But the deliberate pursuit of knowledge by this method is a comparatively recent enterprise of mankind: at the beginning of the 17th century, when Bacon wrote his *Novum Organum,* it was a novelty not very well understood; and a thinker as great as Descartes could maintain that there was no essential difference between physics and geometry. A hundred years later, after the founding of the Royal Society and the great triumphs of Sir Isaac Newton, the method was still something that could arouse debate among intelligent men, as may be seen from the curious remarks of Newton in the *General Scholium* that he added at the end of his *Principia.* In 1739 Hume, who thought Newton "the greatest and rarest genius that ever rose for the ornament and instruction of the species," found himself driven nevertheless to the conclusion that induction could be nothing but customary association of ideas (which Locke had called "a sort of madness") because it was clearly not to be justified as a form of deduction. If the method or policy now seems obviously wise, that is because we understand our situation better than our predecessors did and are therefore no longer inclined to think we can get what we want in any other way.

Presuppositions and Aims.—It has sometimes been suggested that scientific method involves use of a postulate called the Principle of the Uniformity of Nature. According to those who proclaim it, this principle is needed as a supplementary premiss by addition of which inductions can be converted into valid deductions. But no one has ever succeeded in formulating the proposed principle in any way which makes it seem at all plausible, and we have seen reason to believe that there is a fundamental mistake in the program of those who advocate it. Science is not advanced by any method which can be represented as inference from the data of sense by application of universally valid rules, and it has no presuppositions of the sort discussed in the older manuals of induction.

Scientists are always concerned with the elaboration and testing of hypotheses which may conceivably be refuted by further experience, and their attitude is scientific only in so far as they are prepared to admit this. If at any time they put forward theories

for which they refuse to admit the conceivability of falsification, they have abandoned scientific method, even though they may produce masses of evidence which is supposed to be confirmatory. It has been alleged in particular that dialectical materialists and psychoanalysts, in spite of their pretensions to be called scientists, are guilty of such obscurantism. On the other hand we can reasonably say that the scientific enterprise is a search not only for special uniformities (*i.e.*, regularities expressible in empirical generalizations) but also for uniformity in the sense of theoretical simplicity. James Clerk Maxwell, for example, is accounted a very great scientist not because of any work which he did in Baconian fashion but because he showed how the separate studies of heat, light, magnetism, and electricity could be united in one elegant theory. It is impossible, of course, to say in advance what success will be achieved in the search for conceptual simplicity, but it is interesting to notice two different ways in which scientists have tried to get what they want.

The first method is by concentration of attention on those features of the world which seem to admit of mathematical treatment; *i.e.*, on the so-called primary qualities of things (number, shape, size, and motion or rest), in distinction from the secondary qualities (warmth, colour, taste, smell, etc.) which are correlative to our special senses and were described by Locke as the powers of bodies to produce sensations in us by the action of their minute parts on our sense organs. This concentration of attention began already in antiquity when Democritus put forward his theory of atoms, and it is the principle of the demarcation of physics as that science has been studied since the 17th century. Though the hypotheses of secondary induction favoured by 20th-century physicists differ greatly from those formulated 300 years earlier, they are still in terms of primary qualities.

Theoretical physics is essentially mathematical physics, and it was even maintained by Einstein that the ideal of physical explanation should be the geometrization of the basic laws of nature. Einstein meant by this, not that physics could be developed *a priori* as a branch of pure mathematics (which is perhaps what Descartes hoped), but rather that it might be possible to choose from among the abstract geometries discussed by pure mathematicians one which in application to nature would provide a unified field theory of all forces. He did in fact geometrize the law of gravitation within his general theory of relativity, but the prospect of the geometrization of all the fundamental laws of physics seems much more remote now than when he first conceived his ideal. Obviously the unification and simplification of science which can be achieved by the method characteristic of physics is not available for dealing with those events which have been deliberately excluded from the realm of physics, namely experiences.

The second general method of simplifying our scientific picture is that of presenting less fundamental regularities in nature as logical consequences from the conjunction of more fundamental regularities with initial conditions of a historical or geographical kind. Evolutionary explanations are of this pattern. They were tried first in humane studies during the 18th century, and from there passed in the 19th to biology. In the 20th century they have appeared even within physical science. It is now customary, for example, to think of atoms, planets, stars, and galaxies as relatively stable products of a process of cosmic evolution. Within this grand scheme it is natural to think of the occurrence of sense experiences as connected with the appearance of organisms of a certain complexity; but we have as yet no explanation of the novelty, and if, in the phrase favoured by some biologists and philosophers, we describe it as a case of emergence (*q.v.*), we do no more than label the problem.

The Scope of Scientific Method.—The prestige acquired by physics, chemistry and physiology from the middle of the 19th century became so great that there have been many attempts to apply the methods of these sciences in other fields. These efforts have in turn provoked attempts to erect barriers beyond which scientists should not presume to go. In particular there has been much talk of demarcating a boundary between the proper spheres of science and of religion. It is desirable therefore to consider the scope of scientific method; *i.e.*, whether there are any limits to its application.

Because the various special sciences have all been separated successively from the matrix of philosophy (*q.v.*), those who advocate the application of scientific method to new fields sometimes think of themselves as liberators and champions of enlightenment against obscurantism. At times the attitude of professional philosophers toward science may perhaps have given some justification for this view; but in general it is a mistake to think of philosophy as opposed to science. For in the past the word "philosophy" was used, as we have seen, to cover all that is now called science, and in its more restricted modern usage it signifies the search for clarity by the study of the relation of concepts. So long as scientific advance does not involve the asking of any questions about the relations of concepts of high generality science has little, if any, contact with philosophy in the modern sense, and those who are actively engaged in the advance may come to think of philosophical inquiries as fruitless and even perverse. This was the situation in chemistry, for example, at the beginning of the 20th century. But when the development of scientific theory leads to suggestions which are at variance with common sense or at least so far removed from the realm of common sense that they cannot be illuminated by familiar analogies, scientists themselves say that their work has a philosophical aspect; *i.e.*, that it involves clarification of concepts. This happened in physics with the appearance of the relativity and quantum theories.

Conversely, philosophers of the 20th century who are especially interested in the progress of science have sometimes described their work as scientific. Bertrand Russell, for example, published in 1914 a book entitled *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*, and thereafter the phrase "scientific philosophy" occurred often in philosophical literature. This way of speaking seems, however, to involve an unjustified extension of the use of the word "scientific." Those who adopt it wish no doubt to convey the impression that the philosophy which they favour is a permanent intellectual acquisition and not merely the product of a fashion or the expression of personal attitude. But there is usually no difficulty in distinguishing the conceptual inquiries of the philosopher from the empirical researches of the natural scientist, and there is nothing to be gained from a blurring of the distinction.

When we say that psychology and sociology are empirical sciences, rather than departments of philosophy in the modern sense of that word, what we have chiefly in mind is the possibility of using primary induction in these studies. But those who issued the first declarations of independence on behalf of these studies sometimes revealed in their claims an extravagance like that which is too often found in political nationalism. Thus Émile Durkheim maintained in his *Règles de la méthode sociologique* (1894) that sociological laws were not reducible to those of any other science; and in psychology also it has sometimes been held that explanation can be achieved only by the introduction of hypotheses which are themselves purely psychological. It is clear that for research to be practicable there must be a division of academic labour and that there can be no specialization without the distinction of one science from another. But this does not entail the denial of all possibility of co-operation or require that the distinction between primary and secondary induction, which is of the greatest importance in physics, should reappear separately in all other branches of science. There may of course be principles of explanation other than those of physics, but the question whether the laws of one department of science can be explained by reference to those of another is to be settled in each case by scientific inquiry, not by *a priori* pronouncements from separatists or universalists.

Since psychology and sociology have been recognized as sciences it is sometimes asked whether scientific method can be applied in the study of human history. This is a confusing question. Those who ask it may perhaps mean by "scientific method" no more than the use of modern techniques such as the radio-carbon method of dating a wooden object. Or again they may mean simply a dispassionate and unprejudiced approach to controversial issues. In either case the answer is "Yes." If, however, the questioners wish

to know whether history can become a branch of science, the answer seems to be that history cannot be a science like physics, chemistry or even sociology, since it deals with individual persons and events, but that it may be scientific in the same sense as astronomy and geography, which also apply generalizations to individual objects; and that in a way already noticed it may provide evolutionary explanations of certain generalizations about human behaviour (*e.g.*, in anthropology and philology).

In all ages history has in fact exhibited the patterns of explanation favoured by its authors. When men believed in magic, their history was full of magical causation. When it was the ideal of explanation to relate everything to the divine purpose, history was conceived as theodicy or the tracing of God's plan. When in the second half of the 19th century the principle of the survival of the fittest was thought to be the supreme principle of explanation in all biology, there were soon historians ready to apply it in their accounts of the vicissitudes of men and societies.

In all this there is nothing to indicate any need for drawing a boundary to the application of scientific method. On the contrary, it seems absurd even to suggest that there might be such a boundary within the realm of fact as opposed to value. For whenever we make an assertion that cannot be checked either by hard thinking alone or by observation alone, we commit ourselves to judgment by scientific method. This is not to say that we cannot reasonably support any of our statements by appeal to authority, but rather that in this field all rationality, including the rationality of relying on any authority, depends in the last resort on the rationality of scientific method as described above. For even those who cite miracles as evidence of the reliability of their authority assume premisses derived from induction, as for example that events of certain sorts do not happen without divine intervention.

Valuations as opposed to statements of fact are not established by scientific method. This is indeed a truism, and if we wish to speak of a limit to the applicability of scientific method we may perhaps say that we have found it here. But we must be cautious in the way in which we draw the boundary. For discoveries made by scientific method are undoubtedly relevant to some of our valuations. Although chemical analysis of an artist's materials does not contribute to an aesthetic judgment of his work, scientific knowledge about the consequences of a man's act may be of the greatest importance for a moral judgment of its rightness. What we must guard against here is the suggestion that moral distinctions are meaningless unless they can be identified completely with distinctions of the sort that scientists can study.

See also references under "Scientific Method" in the Index.

BIBLIOGRAPHY.—R. B. Braithwaite, *Scientific Explanation* (1953); W. C. Kneale, *Probability and Induction* (1949); G. H. von Wright, *A Treatise on Induction and Probability* (1951); K. R. Popper, *The Logic of Scientific Discovery* (1959). (W. C. K.)

SCILLY ISLES, a group of about 140 small islands and islets, lying off Cornwall, Eng., 21–31 mi. (34–50 km.) W by S of Land's End. The origin of their name has never been authoritatively settled (possibly from the Cornish *Silya* "conger eel"); they are thought to be the Cassiterides (Tin Islands) of the Phoenicians, Greeks, and Romans, and the Silures of the Romans. The true islands number 48 with an area 6.3 sq.mi. (16.3 sq.km.). Apart from the 5 inhabited islands, 18 are capable of bearing grass, and 25 are barren. The larger islands vary in height from a little below to a little above 100 ft., the highest point, 166 ft., being found on St. Mary's. The islands are wild and picturesque, with sheer cliffs and many large caves hollowed out by the Atlantic. Because of the reefs and shoals by which these shores are surrounded, navigation becomes perilous in rough weather and on an outlying rock to the southwest is Bishop Light, first built, with infinite difficulty, in 1858; other lighthouses are on Round Island and Penninis Head (St. Mary's). The islands are part of an almost submerged granite mass that has been broken up by the sea. This mass was probably joined to the mainland between the Armorican folding and the Upper Cretaceous periods, but during Pliocene times it was submerged, since when there has been an elevation from time to time as can be seen in the raised beaches and the terracing of the rocks. According to legend, the Scilly Isles are the only visible remains of the land of Lyonnesse (*q.v.*). It is interesting to note that Solinus, the Latin grammarian and compiler, in A.D. 240 and Sulpicius Severus, the Christian writer, in 400 spoke of the Scillies in the singular, "Siluram Insulam." Some of the islands, *e.g.*, St. Agnes and Gugh, are joined together by bars of shingle and blown sand, and St. Mary's really consists of three islands joined by sand bars. Olaf I Trygvesson (*q.v.*) in his saga tells of his visit to the Scillies in the 10th century and his conversion there to Christianity, which he then introduced to Scandinavia.

The climate of the islands is unusually mild, the range of temperature being from 8° to 14° C (46° to 58° F). Snow is rarely seen. Fuchsias, geraniums, and myrtles attain an immense size, and aloes, cactus, and prickly pear flourish in the open. The gardens of the governor on Tresco Island are subtropical in character. The raising of early asparagus, spring vegetables, and flowers is the principal industry. From the start of flower growing in 1881 and a recorded figure of 65 tons of cut flowers in 1885, this trade had expanded by the mid-1960s to more than 1,200 tons, representing 60,000,000 blooms. The acreage under flowers in the mid-1960s was 582. Grown in the open air in winter, protected from the Atlantic gales by hedges, the Scilly flowers are more popular than those forced under glass. The islands are of great biological interest. They lie farther south than the rest of Great Britain and with their milder climate have a flora and fauna akin to those of the European continent. Three species of wild plant are found in the isles and nowhere else in Britain. The Scilly shrew mouse *Crocidura cassiteridum*, although not related to any species in Britain, is closely related to a species found in France. Among the immense variety of sea birds, the roseate tern is the rarest British breeding tern and the Manx shearwater has its only British breeding site in the islands. In 1957 the Nature Conservancy, the Duchy of Cornwall, and Lt.-Commander T. M. Dorrien-Smith, the lessee, agreed to treat many of the uninhabited islands as a "Site of Special Scientific Interest."

The inhabited islands are St. Mary's, Tresco, St. Martin's, St. Agnes, and Bryher. The total population in 1961 was 2,288, of which 1,736 lived on St. Mary's. Hugh Town, St. Mary's, is the capital, occupying a sandy peninsula. The town possesses a harbour and a roadstead where large vessels can lie at anchor. Governed by a unique type of county council, the islands are part of the St. Ives parliamentary division of Cornwall. They are served by steamers from Penzance and by air from St. Just, and there is radio and cable communication with the mainland.

On Tresco there are ruins of an abbey and of two fortifications called Oliver Cromwell's Tower and King Charles's Tower. Rude pillars and circles of stones and barrows are common, the most remarkable being a barrow on the Isle of Samson, 58 ft. in girth and containing a very perfect kistvaen, or sepulchral chamber of stone. There is no written evidence concerning the Scillies until the reign of Henry I. The king gave the isles to the abbot and Church of Tavistock (Devon), and secular priests were temporarily substituted for regulars by the abbot in 1345. Sharing the dignity of lords of Scilly with the abbot was the family of Blanchminster (de Albo Monasterio) in the reign of Edward I. Ralph de Blanchminster held of the earldom of Cornwall lands in Scilly at a yearly service of 300 puffins in 1345. In 1484 value of the lands was estimated at 40*s*. in time of peace and nothing in time of war. The Blanchminsters resisted and imprisoned the coroner of Cornwall and in 1319 were granted a coroner of their own. In 1547 Silvester Danvers, a co-heir of the Blanchminsters, sold his moiety of Scilly to Sir Thomas Seymour, by whose attainder in 1549 it fell to the crown, which already possessed the church land and revenues. In 1571 Elizabeth I leased the islands for £10 a year to Francis Godolphin, who built Star Castle above Hugh Town in 1593. During the Civil War, Hugh Town gave shelter to Prince Charles until his escape to Jersey in 1646. The islands were occupied by Roundheads for two years, but in 1648 declared for the king with Sir John Grenville as governor. After Grenville's surrender to Adm. Robert Blake in 1651, Cromwell's Tower was built on Tresco. Early a haunt of pirates, the islands were afterward notorious for smuggling. In 1834 Augustus Smith succeeded the

Godolphins as lessee and introduced compulsory education and other measures to improve the condition of the inhabitants. His nephew, T. A. Smith-Dorrien-Smith, succeeded him in 1872, and in 1933 A. A. Dorrien-Smith, the latter's son, surrendered to the Duchy of Cornwall the four islands of St. Mary's, St. Martin's, St. Agnes, and Bryher, retaining the others on a 99-year lease.

Bibliography.—W. Borlase, *Observations on the Islands of Scilly* (1756); H. J. Whitfield, *Scilly and Its Legends* (1852); J. G. Uren, *Scilly and the Scillonians* (1907); E. L. Bowley, *The Fortunate Islands* (1945); Geoffrey Grigson, *The Scilly Isles* (1950); C. C. Vyvyan, *The Scilly Isles* (1953); G. Forrester Matthews, *The Isle of Scilly* (1960). (C. Ki.)

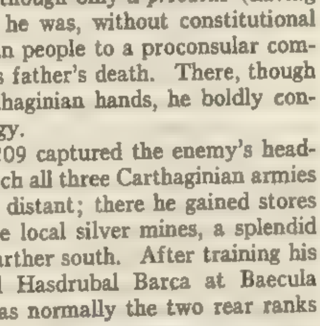
SCIPIO (originally meaning "staff"), the name of a patrician branch of the *gens Cornelia* in ancient Rome. The most important Scipios are given below. Others include Lucius Cornelius Scipio Barbatus (consul 298 B.C.) and his son Lucius Cornelius Scipio (consul 259, censor 258) who attacked Corsica in 259; both were buried in the "Tomb of the Scipios" outside Rome on the Appian Way.

Publius Cornelius Scipio (d. 211 B.C.), consul in 218, sailed with an army to southern Gaul to prevent Hannibal's advance on Italy (*see* Hannibal). Having arrived too late, he himself returned to Italy but boldly sent his army on to Spain under his elder brother Gnaeus to check the Carthaginian forces still there. In northern Italy he hoped to fight delaying actions against Hannibal along the tributaries of the Po. Being repulsed and wounded at the Ticinus, he retired to the Trebia where he was joined by his colleague Tiberius Sempronius Longus, who insisted on fighting and was defeated (December 218; *see* Trebbia).

In 217 Scipio was sent as proconsul to Spain; in a battle near the Ebro he and his brother smashed Hasdrubal's attempt to break through to Italy (215), and by 212 they had captured Saguntum (modern Sagunto). From this base they could move farther south; advancing separately, however, they both met disaster and death, Publius on the upper Baetis (Guadalquivir), Gnaeus in the hinterland of Carthago Nova (Cartagena) (211). In spite of this final defeat, they had for seven years denied Hannibal the resources of Spain.

Publius Cornelius Scipio, surnamed Africanus, the elder (236–184/183 B.C.), son of the Publius above, is said to have saved his father's life at the engagement on the Ticinus and, as military tribune, to have strengthened Roman resistance to Hannibal after the Carthaginian victory at the Battle of Cannae (216). After his curule aedileship (213), though only a *privatus* (having been neither praetor nor consul), he was, without constitutional precedent, appointed by the Roman people to a proconsular command in Spain in the year after his father's death. There, though all south of the Ebro was in Carthaginian hands, he boldly continued his father's offensive strategy.

By a brilliant swoop Scipio in 209 captured the enemy's headquarters, Carthago Nova, from which all three Carthaginian armies in Spain happened to be ten days' distant; there he gained stores and supplies, Spanish hostages, the local silver mines, a splendid harbour, and a base for advance farther south. After training his army in new tactics he defeated Hasdrubal Barca at Baecula (Bailén) in Baetica (208); whereas normally the two rear ranks of a Roman army closely supported the front line, Scipio in this battle, under a screen of light troops, divided his main forces, which fell on the enemy's flanks. When Hasdrubal broke away, ultimately to join his brother Hannibal in Italy, Scipio wisely declined the impossible task of trying to stop him and decided rather to accomplish his mission in Spain, the defeat of the other two Carthaginian armies still there. This he brilliantly achieved in 206 at the Battle of Ilipa (*q.v.*; Alcalá del Río, near Seville), where he held the enemy's main forces while the wings outflanked them. He then secured Gades (Cádiz), thus making Roman control of Spain complete.



BY COURTESY OF THE ROYAL COLLECTION OF COINS AND MEDALS, COPENHAGEN

PUBLIUS CORNELIUS SCIPIO AFRICANUS THE ELDER, PORTRAIT ON A SILVER COIN FROM CARTHAGO NOVA

Elected consul for 205, Scipio boldly determined to disregard Hannibal in Italy and to strike at Africa. Having beaten down political opposition in the Senate, he crossed to Sicily with an army consisting partly of volunteers. In 204 he landed with perhaps 35,000 men in Africa, where he besieged Utica. Early in 203 he burned the camps of Hasdrubal Gisco and his Numidian ally Syphax. Then, sweeping down on the forces which the enemy were trying to muster at the Great Plains on the upper Bagradas (modern Souk el Khemis, on the Medjerda, in Tunisia), he smashed their army by a double outflanking movement.

After his capture of Tunis the Carthaginians sought peace terms, but Hannibal's subsequent return to Africa led to their renewing the war in 202. Scipio advanced southwestward to join the Numidian prince Masinissa (*q.v.*), who was bringing his invaluable cavalry to his support. Then he turned eastward to face Hannibal at the Battle of Zama (*q.v.*); his outflanking tactics failed against the master from whom he had learned them, but the issue was decided when the Roman and Numidian cavalry, having broken off their pursuit of the Punic horsemen, fell on the rear of Hannibal's army. Victory was complete and the long war ended; Scipio granted comparatively lenient terms to Carthage. He was named in honour Africanus. (*See* also Punic Wars.)

In 199 Scipio was censor and became *princeps senatus*. Though he vigorously supported a philhellenic policy, he argued during his second consulship (194) against a complete Roman evacuation of Greece after the ejection of Philip V of Macedonia, fearing that Antiochus III of Syria would invade it; his fear was premature, but not unfounded. In 193 he served on an embassy to Africa and perhaps also to the east. After Antiochus had advanced into Greece and had been ejected by a Roman army, Scipio's brother Lucius was given the command against him, Publius serving as his legate (190); together the brothers crossed to Asia, but Publius was too ill to take a personal part in Lucius' victory over Antiochus at Magnesia (for which Lucius took the name Asiagenus). (*See* also Antiochus: *Antiochus III the Great.*)

Meantime in Rome Scipio's political opponents, led by the elder Cato, launched a series of attacks on the Scipios and their friends. Lucius' command was not prolonged; the generous peace terms which Africanus proposed for Antiochus were harshly modified; and the "trials of the Scipios" followed. On the trials the ancient evidence is confusing: in 187 an attack on Lucius for refusing to account for 500 talents received from Antiochus (as war indemnity or personal booty?) was parried; and Africanus himself may have been accused, but not condemned, in 184. Anyhow, his influence was shaken and he withdrew from Rome to Liternum in Campania, where, embittered and ill, he died (184/183).

Scipio married Aemilia, daughter of the consul L. Aemilius Paullus (Paulus) who fell at Cannae. He had two sons (Publius, debarred by ill health from a public career, adoptive father of Scipio Aemilianus; and Lucius, praetor in 174) and two daughters, one of whom became the mother of the Gracchi. A man of wide sympathies, cultured and magnanimous, Scipio easily won the friendship of such men as Philip of Macedonia and native princes in Spain and Africa, while he secured the devotion of his own troops. Though essentially a man of action, he was also something of a mystic in whom contemporary legend saw the favoured of Juppiter Capitolinus as well as a spiritual descendant of Alexander the Great. One of the greatest soldiers of the ancient world, by his tactical reforms and strategic insight he created an army which defeated even Hannibal and asserted Rome's supremacy in Spain, Africa, and the Hellenistic east. His Greek sympathies led him to champion Rome's mission in the world as protector of Greek culture; he preferred to establish Roman protection rather than direct conquest and annexation. For ten years (210–201) he had commanded a devoted army at the people's wish; but, though convinced of his own powers, he offered no challenge to the dominance of the Roman nobility ensconced in the Senate, except by normal political methods (in which he showed no outstanding ability). Reaction against his generous foreign policy and against his en-

couragement of Greek culture in Roman life led to his downfall amid personal and political rivalries, but his career had shown that Rome's destiny was to be a Mediterranean, not merely an Italian, power.

PUBLIUS CORNELIUS SCIPIO AEMILIANUS, surnamed AFRICANUS (the younger) and NUMANTINUS (185/184–129 B.C.), was the younger son of the general L. Aemilius Paulus (*q.v.*), under whom he served at the Battle of Pydna (*q.v.*) in 168 B.C. He was adopted by the elder son of Scipio Africanus, Publius, whose name he assumed. His early years are described by Polybius, who gained his friendship. In 151, when the Romans had met with defeats in Spain, he volunteered for service there as military tribune; his integrity, personal bravery and honest diplomacy, reminiscent of the qualities of his adoptive grandfather, were appreciated by Spanish tribes long sickened by Roman treachery.

In 150, when on a military mission to Masinissa, he was asked by the Carthaginians to mediate in their war against Masinissa, but he achieved no result. In the following year, after Rome had declared war on Carthage, he served in Africa as military tribune with conspicuous ability, and in 148 Masinissa, at the point of death, asked him to arrange the future of Numidia; Scipio divided it among the king's three sons. Returning to Rome to stand for the aedileship, he was by a special vote of the people elected to the consulship of 147 and was granted the command against Carthage. Scipio returned to Africa with his friend Gaius Laelius and possibly also Panaetius. He blockaded Carthage and in 146 destroyed the city, enslaving the survivors and establishing the province of Africa. He celebrated a triumph and received the name of Africanus (Minor).

His censorship in 142 was marked by traditional sternness. Tiberius Claudius Asellio, whom he demoted from the equites, prosecuted him in 140. Accompanied by Panaetius, he served on an embassy to Egypt, Syria, Pergamum, and Greece (140–139). On his return he prosecuted L. Aurelius Cotta for extortion (138) and supported a measure to introduce secret ballot in judicial assemblies of the people (137). Military disasters and inefficiency in Spain required the dispatch of a capable soldier; to this end, since reelection to the consulship had been forbidden in 151, by special dispensation he was elected consul for 134. After reestablishing discipline in the Roman army he brought the war in Spain to an end by the blockade and destruction of Numantia in 133 (traces of his camps and earthworks there still survive) and received the additional cognomen Numantinus.

He returned to Rome after the death of his cousin and brother-in-law Tiberius Gracchus, but expressed disapproval of his unconstitutional behaviour, quoting the *Odyssey* (i, 47): "So perish all who do the like again." He opposed a proposal of the tribune Gaius Papirius Carbo to allow the reelection of tribunes (131 or 130), thus increasing his unpopularity with the masses, who naturally had resented his attitude to Gracchus. In 129 he carried a measure which affected the working of the Gracchan land commissioners, probably, however, only warning them off public land held by the allies in order to avoid "international" issues. The night before he was to have spoken on the agrarian laws he died in mysterious circumstances: various persons were later suspected of murder, or even Scipio himself of suicide, but very probably he died a natural death.

An admirer of Greek literature and learning, he gathered around him the so-called "Scipionic circle," a coterie which included Polybius, Panaetius, Lucilius, Terence, and Laelius and aimed at blending the better elements of Greek and Roman life. Upright himself, he tried to revive and uphold earlier Roman standards of conduct. His thought, Stoic in philosophy, moved on traditional lines, and he tried to maintain Rome's accustomed relations toward the Italian allies and the provinces. An able soldier and orator, he acted as a moderating influence on the constitutional difficulties of his day which increased after his death. Cicero, who made him the chief speaker in his *De Republica*, regarded his era as the golden age of aristocratic government.

BIBLIOGRAPHY.—*Publius Scipio the Elder:* Polybius, iii; Livy, xxi–xxv; for the Spanish campaigns, A. Schulten, *Fontes Hispaniae antiquae*, iii, pp. 56–92 (1935).

Scipio Africanus: For the sources, of which the chief are Polybius and Livy, *see* E. Meyer, *Kleine Schriften*, ii, pp. 331 ff. (1924), and H. H. Scullard, *Scipio Africanus in the Second Punic War* (1930). *See* further B. H. Liddell Hart, *A Greater Than Napoleon: Scipio Africanus* (1926); W. Schur, *Scipio Africanus und die Begründung der römischen Weltherrschaft* (1927); H. H. Scullard, *Roman Politics, 220–150 B.C.* (1951), and "A Note on the Battle of Ilipa," *Journal of Roman Studies*, xxvi, pp. 19 ff. (1936); R. M. Haywood, *Studies on Scipio Africanus* (1933); on the "trials," P. Fraccaro, *Opuscula*, vol. i, pp. 263 ff. (1956); for portrait, M.-L. Vollenweider, "Das Bildnis des Scipio Africanus," *Museum Helveticum*, xv, pp. 27 ff. (1958).

Scipio Aemilianus: Chief sources: Polybius, xxxii–xxxix; Appian, *Punica*, 71, 98 ff., *Hispanica*, 49 ff., 84 ff.; H. Malcovati, *Oratorum Romanorum Fragmenta*, 2nd ed., pp. 122 ff. (1955). *See* also K. Bilz, *Die Politik des P. Cornelius Scipio Aemilianus* (1936); H. H. Scullard, "Scipio Aemilianus and Roman Politics," *Journal of Roman Studies*, vol. 50, pp. 59 ff. (1960). For his Spanish campaign: H. Simon, *Roms Kriege in Spanien 154–133 v. Chr.*, pp. 171 ff. (1962). (H. H. Sd.)

SCISSORS AND SHEARS, cutting instruments having two blades, usually pivoted to meet and cut as the handles at their ends close. The terms scissors and shears are nearly synonymous, although "shears" sometimes denotes larger size. Modern scissors and shears derive from two basic types: spring shears, in which the blades are connected by a C-shaped spring at the shank (handle) ends; and pivoted scissors and shears, in which the connection is made by a rivet or screw located between the handles and the blades. The spring type, probably dating to the Bronze Age, was commonly used in Europe until the end of the Middle Ages. Pivoted bronze and iron scissors and shears were used in ancient Rome and in China, Japan, and Korea. Their domestic use in Europe dates from the 16th century, but they were not in common use until the 17th century. Large-scale production of pivoted scissors and shears began in 1761, when Robert Hinchliffe of Sheffield, Eng., first used cast steel in their manufacture. Ornamental scissors with elaborately designed handles were numerous in the 19th century, and much hand-forged and beautifully decorated work was produced for the Great Exhibition in London in 1851.

By the end of the 19th century mechanical methods of production had begun to simplify styles and patterns. The blanks (unfinished pieces) for scissors are made now by high-speed forging of red-hot steel bars between the dies of drop hammers. For ordinary scissors, steel containing 0.55% carbon is mainly used; for the finest scissors and shears for trade use (such as tailors' shears and trimmers) a harder steel containing 0.75% carbon is preferred. For cutting man-made fibres, which are harder and more blunting than natural fibres, the blades of tailors' shears are sometimes made of a composite material consisting of an even harder steel on the cutting side (1.03% carbon), backed by tough iron. Many scissors, including surgical varieties, are made of stainless steel (*q.v.*), and some specialized scissors and shears are made of nonferrous alloys that will not produce a spark or interfere with magnetism (*e.g.*, for cutting cordite and magnetic tape). Very cheap scissors are made from relatively soft steel wire that is pressed cold and is not hardened. In most cases the blanks for the handles and blades of scissors are one piece, but for comfortable use of large shears and some specialized scissors the handles are made of malleable metal and are accurately positioned with, and electrically welded to, the steel blades. Pinking-shear blades may be fitted to aluminum-alloy handles to reduce weight. As the process of cold forging progresses and suitable cold-forging steels are produced, some scissors of satisfactory quality may be made from cold-forged blanks.

If scissors are to cut well, the blades should touch in two places only—at the joint and at a single spot along the blades, wherever the cutting is taking place. The blades are made to twist or curve toward one another, and, when completely closed, the points of the blades should touch. Also, both blades must be accurately tempered to equal hardness. In the finest scissors and shears, the two blanks and the screw (the nut also in the case of large shears) are coded with an identification mark early in manufacturing to ensure their being treated as a set.

The blades of scissors and shears require sharpening by experts. Each blade is passed lightly across a grinding wheel to restore the sharpness of the edge-angle. The grinding process has

to follow the twist of the blade and has to be done smoothly and continuously with an even pressure throughout the whole stroke to avoid causing any ridge or other irregularity in the edge. Inexpert treatment, such as grinding the insides of the blades, will ruin them; their cutting life may be prolonged, however, by gently rubbing the inside edges of the blades with the fingers to remove particles of material, dust, etc., and by not using scissors or shears for purposes beyond their strength. *See* also CUTLERY.

See G. I. H. Lloyd, *The Cutlery Trades* (1913); J. B. Himsworth, *The Story of Cutlery from Flint to Stainless Steel* (1953).
(A. E. HA.)

SCOLECITE, a zeolite mineral of the natrolite group. It is a hydrated silicate of lime and alumina. It commonly occurs as bundles of radiating fibres and when heated sometimes curls up like a worm; the name is derived from the Greek word meaning a worm. It is usually colourless or white and more or less transparent. It crystallizes in the monoclinic system, with angles very near those of the cube. Scolecite occurs frequently with other zeolites in the western isles of Scotland, Iceland, in the Deccan trap area in India and at Table mountain, Colo., among many other localities.

For composition, structure, etc., *see* ZEOLITE.

SCONE, a civil parish of the county of Perth, Scot., containing Old Scone, the site of a historic abbey and palace, and New Scone, a modern village, about 2 mi. NNE of Perth, near the left bank of the Tay. Pop. (1961) 3,713.

According to tradition Scone became the capital of "Pictavia" in succession to Forteviot in the 8th century and remained that of Alban after the union of Picts and Scots. In 906 Constantine III and Bishop Cellach held a council concerning the church on the Mote hill of Scone, hence also called Hill of Worship. Kenneth MacAlpin is said to have brought from Dunstaffnage the Stone of Destiny, on which Scottish kings were crowned until its removal to Westminster abbey by Edward I of England. A Culdee foundation which existed there from early times was superseded by an Augustinian house founded by Alexander I *c.* 1115. Scone remained the normal place of the coronation of Scottish kings, the last being that of Charles II in 1651. It was also the scene of many medieval parliaments.

At the Reformation the abbey was burned by a mob from Perth and the lands were given to the earl of Gowrie. On the Gowrie forfeiture in 1600 these were granted to Sir David Murray of Gospertie, afterward Viscount Stormont, to whose descendant, the earl of Mansfield, the estate still belongs. Sir David completed the mansion begun by Gowrie, but it was pulled down in 1803 to make room for the present house.

A church built on the Mote hill in 1604 was demolished in the late 18th century, except for one aisle. Scone was visited by James Edward Stewart in 1716 and by Prince Charlie in 1745. The old cross of Scone stands in the park, which also contains a racecourse.
(E. W.-M. B.-M.)

SCOPAS of Paros, sculptor and architect, was considered by ancient writers one of the greatest artists of the 4th century B.C. He is said by Pliny to have worked on the sculptures of the mausoleum of Halicarnassus (*c.* 350 B.C.) together with Bryaxis, Timotheus and Leochares, and to have carved the reliefs on one of the columns of the temple of Artemis at Ephesus; and by Pausanias to have been the architect of the temple of Athena Alea at Tegea. Unfortunately, though sculptural remains of all three buildings have survived, none can with certainty be associated with Scopas. In the case of the mausoleum, there is the difficulty of apportioning the extant reliefs to four different artists of whose style little is known. In the case of the Ephesian temple only three of the 36 columns that Pliny says were decorated have survived. And the few battered pieces that remain of the pedimental sculptures of the temple at Tegea are not necessarily the work of Scopas, for Pausanias attributes to him only the cult statues of Asclepius and Hygieia. Nevertheless, since Scopas was the architect of the temple, it is possible that he also at least sketched the designs for the pediments; and since the extant heads show a highly individual and forceful style, they may give an idea of Scopas' work. The squarish form, the deep-sunk eyes and a strongly emotional quality have, therefore, been thought to be characteristic of Scopas.

Pausanias, Pliny, Strabo and other ancient writers mention and sometimes briefly describe a number of statues made by Scopas for Asia Minor, the Peloponnesus and Attica, and some of these works have been tentatively identified in Roman marble copies and in the reliefs on Roman coins. Among them are a bronze Aphrodite seated on a goat, an Apollo Smintheus, an Apollo Citharoedus, a Maenad, a group of Eros, Himeros and Pothos, and a large group of Achilles, Nereids and Tritons.

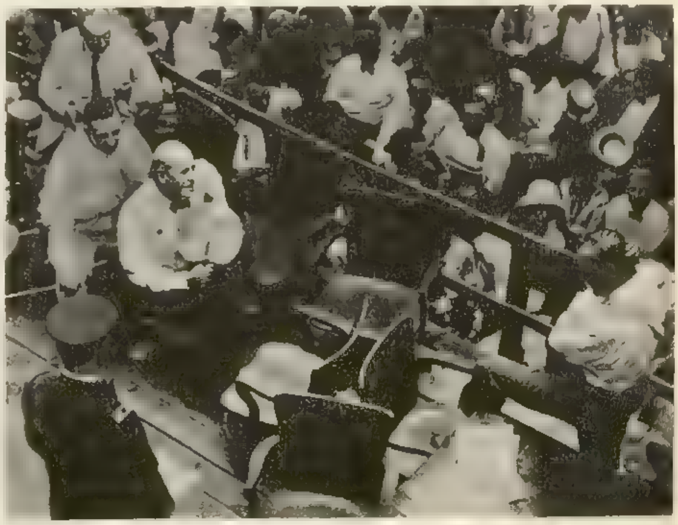
Scanty though our knowledge of Scopas is, it can be surmised from the little which remains that he, together with Praxiteles and Lysippus, were the outstanding sculptors of the middle and second half of the 4th century B.C., who imparted to their work a new grace and emotional quality and whose influence was felt for many generations.

BIBLIOGRAPHY.—J. Overbeck, *Die antiken Schriftquellen*, nos. 1149–89 (1868) for the ancient sources; G. Lippold *Die griechische Plastik*, pp. 249–257 (1950); G. M. A. Richter, *The Sculpture and the Sculptors of the Greeks*, rev. ed., pp. 269–276 (1950); C. Picard, *Manuel d'archéologie grecque*, iii, 2, pp. 1–236 (1954); P. E. Arias, *Skopas* (1952).
(G. M. A. R.)

SCOPES TRIAL, beginning July 10, 1925, at Dayton, Tenn., and ending July 21, was one of the most widely publicized legal cases in 20th-century U.S. history. The charge was that John Thomas Scopes, a teacher of science in Rhea high school in Dayton, had violated the Tennessee state law prohibiting the teaching in public schools of any theories that deny the divine creation of man as taught in the Bible. The basis for the sensational nature of the trial was laid by the increasing alarm of a Christian movement, known as Fundamentalism (*q.v.*), over the challenge of science and evolutionary theory to a literal interpretation of the Scriptures.

The legislature of the state of Tennessee was the first to pass an antievolution law, by an overwhelming majority (95–11), and the measure was signed by the governor on March 13, 1925. The essential section stated that it was "unlawful for any teacher in any of the universities, normals and all other public schools of the state, to teach any theory that denies the story of the divine creation of man as taught in the Bible, and to teach instead that man has descended from a lower order of animals."

When the American Civil Liberties union in New York city heard of the passage of the Tennessee law, the executive director Roger N. Baldwin sent a news release to leading Tennessee newspapers offering the services of the union to defend any teacher who would personally test the constitutionality of the statute by his classroom teaching. Baldwin soon received a telephone call from George Rapplyea, proprietor of a drug store in Dayton, stating that he had been discussing the offer with John Scopes. Al-



UNITED NEWSPICTURES

WILLIAM JENNINGS BRYAN (LEFT, WITH FAN) BEING CROSS-EXAMINED BY CLARENCE DARROW (RIGHT) DURING AN OUTDOOR SESSION OF THE TRIAL

though he had taught biology only briefly as a substitute, Scopes admitted to violating the law, and was indicted.

The leading lawyers on both sides were probably the most famous and the most appropriate who could have been produced. William Jennings Bryan (*q.v.*) came to Dayton to assist the prosecution. The defense was headed by Clarence S. Darrow (*q.v.*); the most famous criminal lawyer of his generation. He was assisted by Dudley Field Malone, a liberal Catholic and one of the great courtroom orators of the time, and by Arthur Garfield Hays, the outstanding civil liberties attorney of the day.

The dramatic nature of the case and the fame of the lawyers assured extended newspaper coverage. Each day that the case was in court about 175,000 words were telegraphed out of Dayton. The case was followed daily by tens of millions of readers in both the United States and Europe.

The rulings of the judge prevented any testing of the civil liberties issue of the constitutionality of the law or any testimony as to the validity of the doctrine of evolution. The sole relevant question, said the judge, was whether Scopes had actually taught the doctrine of evolution. The defense freely admitted that he had.

The case might, thus, have closed more rapidly had not Bryan made the serious mistake of permitting Darrow to get him on the stand and subject him to a long and grueling cross-examination as to his beliefs relative to the Fundamentalist attitude on science and biblical authority. The process, perhaps Darrow's most animated and sarcastic courtroom performance, was a devastating experience for Bryan; many believed it hastened his death, which occurred five days after the close of the trial. Darrow had adroitly used the judge's rulings on the irrelevance of the validity of the doctrine of evolution in the case to prevent Bryan from delivering a long oration he had prepared.

Scopes was convicted and fined $100. The defense appealed the case to the state supreme court which, in 1927, upheld the constitutionality of the 1925 law, but cleared Scopes on the ground that the lower court had exceeded its authority in fining Scopes $100. It thus turned out that the Scopes trial was more a contribution to public education on science and religion than a test of civil liberties.

BIBLIOGRAPHY.—L. H. Allen (ed.), *Bryan and Darrow at Dayton* (1925); Maynard Shipley, *The War on Modern Science* (1927); Raymond Ginger, *Six Days or Forever? Tennessee v. John Thomas Scopes* (1958); John T. Scopes and James Presley, *Center of the Storm* (1967).

(H. E. BAR.; X.)

SCOPOLAMINE (HYOSCINE) is an alkaloid drug obtained from a number of solanaceous plants including the deadly nightshade, henbane, and Jimson weed (*see* ALKALOIDS: *Detection and Preparation*). Scopolamine is the most active of the belladonna alkaloids, partly because of its greater solubility, which permits more rapid passage to the site of action. Like atropine, it has a depressant action on parasympathetic nerves and in larger doses on autonomic ganglia. It also has a marked depressant action on the central nervous system, and this is the basis of its main therapeutic uses. Combined with morphine it produces "twilight sleep," a state of analgesia and amnesia. This drug combination became popular for use during childbirth since the patient remains more or less conscious but does not usually recall later the unpleasant circumstances of the event. This and similar combinations are used for preoperative medication. Scopolamine taken orally is an effective remedy for motion sickness, probably because of its central depressant and its antispasmodic action. It also finds use in controlling the rigidity and tremor of parkinsonism. It is a toxic drug and is administered only in minute quantities. (V. E.)

SCORE (French, *partition;* Italian, *partitura;* German, *Partitur*) is the name given to the copy of a work of music containing the notation for one or many performers, and is so called from the way in which the music is divided by bars scored through the series of staves on which the music is written. In a solo work the score is the only copy of the notation. In concerted works the notation of each performer constitutes a part—*e.g.*, violin part, flute part—and a score shows the vertical alignment of each of the parts. Indications of both individual and simultaneous performance are thus provided for the principal performer, the conductor and the listener.

The word score is used in several specialized senses. The arrangement or reduction of an orchestral score for piano is commonly known as the piano score. A vocal score (of operas and choral works) consists of the vocal parts set out as in the full score, but with the orchestral score reduced for the piano. A short score is the sketch made by a composer for a full score, showing the main features set out on a few staves. An open score shows the parts, usually of a contrapuntal exercise, written on separate staves for greater clarity. A miniature score (or pocket score) is the published version of a full score reduced in size so as to be more convenient for study or following an orchestral work in performance.

The practice of writing music in score dates from the schools of polyphonic music in the early middle ages. It was replaced at the beginning of the 13th century by the choirbook—a large-size manuscript in which the soprano and alto parts usually faced each other on the upper halves of two opposite pages with the tenor and bass parts occupying the lower halves. This arrangement was considered economical since the upper parts, which sang the texts, required more space than the slow-moving lower parts. The music was read by the entire choir grouped around the choirbook set on a stand.

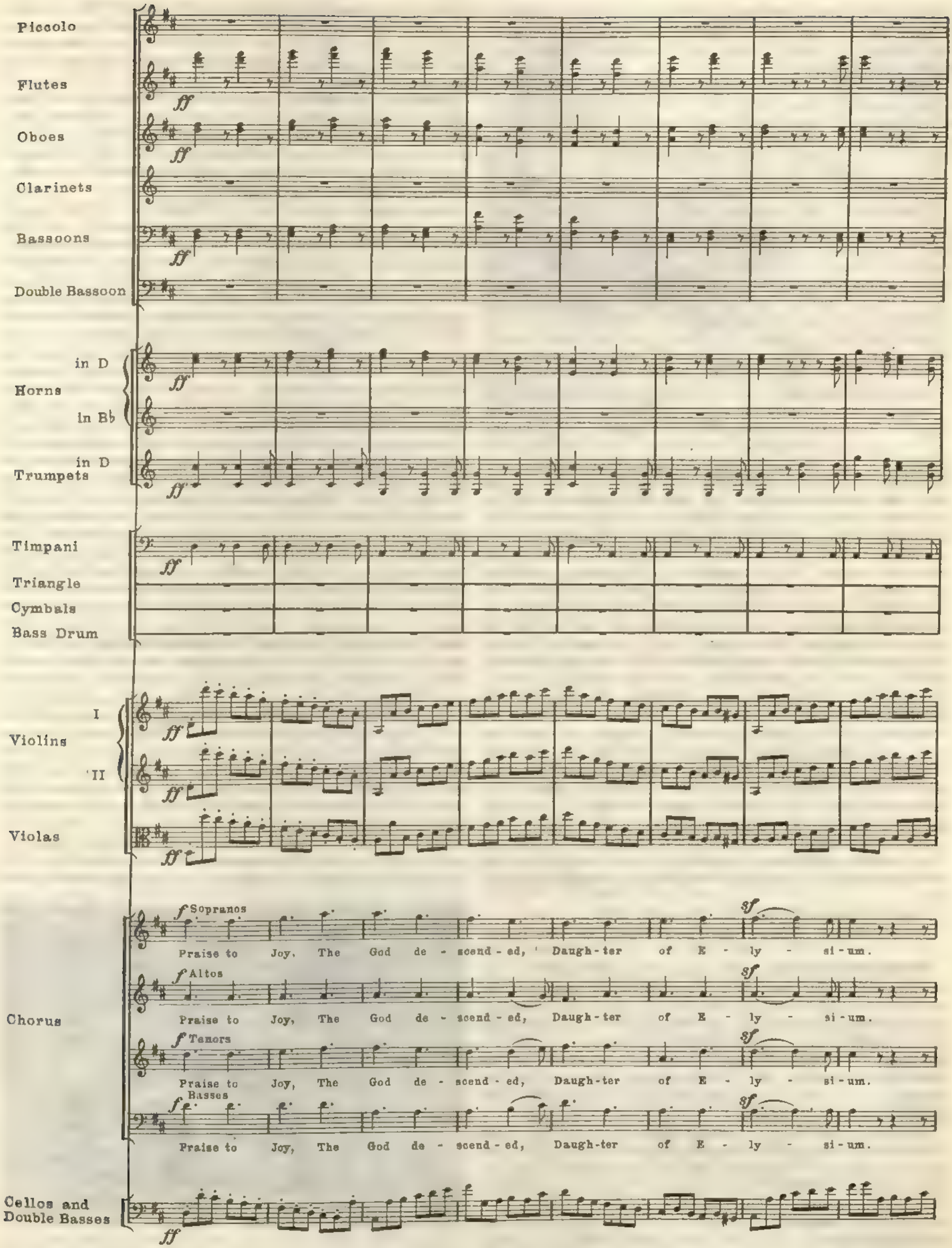
In the 15th and 16th centuries vocal and instrumental music was published in partbooks. These partbooks contained the notation of only single parts. The parts of madrigals were sometimes published crosswise on single sheets, which allowed each of the singers seated around a rectangular table to sing from his particular part.

The modern form of score, in which the bar lines are scored vertically throughout the parts, appeared in the 16th century in the madrigals of Cipriano de Rore and the orchestral music of Giovanni Gabrieli. In scores of all kinds the arrangement of the parts follows accepted schemes. In vocal works the arrangement from top to bottom is: soprano, alto, tenor, bass. Scores for unaccompanied choir in four parts are sometimes known as SATB scores. In scores of chamber works for string instruments the arrangement follows the compass of the instruments: first and second violins, viola, cello, double bass, with the part of a figured bass or a supporting keyboard instrument below. Parts for wind instruments in chamber works are placed above the strings in the same arrangement as that in the orchestral score, known as the full score.

The full score is divided from top to bottom into five main sections: woodwind instruments, brass, percussion, harps and keyboard instruments, and strings. The order followed for the woodwind is: flutes, oboes, clarinets and bassoons, with the higher and lower members of their families placed appropriately. In a work using a wide variety of instruments the woodwind section would thus be set out in this order: piccolo, flutes, bass flute; oboes, oboe d'amore, cor anglais, heckelphone; clarinet in E flat, clarinets in B flat or A, bass clarinet; bassoons, double bassoon. The order for the brass instruments is: horns, trumpets, trombones, tuba. The percussion section is headed by the part for timpani followed by the keyboard percussion instruments (glockenspiel, xylophone and vibraphone) and the instruments of indefinite pitch (triangle, cymbals, etc.), the latter requiring only a single line. The next section shows parts in double staves for harps, piano or organ. At the bottom of the score the string section is aligned as in chamber works.

In concertos the part of the solo instrument is placed immediately above the strings. In choral works the choral parts may appear either above the strings or, as in the illustration from the finale of Beethoven's *Choral Symphony*, between the viola and cello parts. In full scores of operas and oratorios solo parts appear above the choral parts. In military band scores the woodwind and brass are followed, in descending order, by string basses and percussion instruments. Brass band scores are usually arranged in the order of cornets, horns, baritones, trombones, euphonium and basses, and drums.

Score reading, which is part of the equipment of a conductor, consists of reproducing the essential features of a full score at the piano. It demands the ability to read simultaneously in the



SCORE OF THE FINALE OF THE CHORAL SYMPHONY (NO. 9 IN D MINOR) BY BEETHOVEN

alto and tenor clefs as well as the treble and bass clefs, and to transpose parts of many of the woodwind and brass instruments. Score reading of complex modern works is sometimes undertaken in the form of a piano duet.

Students of the orchestra and listeners generally find that the ear can be greatly helped by the eye in following the performance of orchestral and choral works with the score. This enables the general design of the work to be more easily grasped and ingredients of orchestral effects to be readily identified.

See also MUSICAL NOTATION.

See G. Jacob, *How to Read a Score* (1944). (E. LR.)

SCORESBY, WILLIAM (1789–1857), English Arctic explorer, scientist and clergyman, who contributed particularly to the knowledge of terrestrial magnetism, was born near Whitby, Yorkshire, Oct. 5, 1789. His father, William Scoresby (1760–1829), made a fortune in the Arctic whale fishery. The son made his first voyage with his father when he was 10 years old, but on his return he was sent back to school until he was 14. After this he was his father's constant companion, and was with him on May 25, 1806, in the whaler "Resolution," when he reached 81° 30′ N lat. (19° E long.), then the highest latitude attained by a freely navigating ship.

Scoresby then attended the natural philosophy and chemistry classes at Edinburgh university and in his voyage of 1807 he began to study the meteorology and natural history of the polar regions. In 1811 his father resigned to him the command of the "Resolution," and in the same year he married. In a voyage two years later he established the fact that the temperature of the polar ocean is warmer at great depths than at the surface. He was elected a fellow of the Royal Society of Edinburgh in 1819 and about the same time he communicated a paper to the Royal Society of London "On the Anomaly in the Variation of the Magnetic Needle." In the following year he published *An Account of the Arctic Regions and Northern Whale Fishery*, in which he gathered the results of his own observations, as well as those of previous navigators. His voyage of 1822 to Greenland, during which he surveyed and charted 400 mi. of the east coast between 69° 30′ and 72° 30′, was the last of his Arctic voyages, for his wife had died while he was away and on his return he entered the church.

In 1824 the Royal society elected him a fellow and after two years at Cambridge he took his degree (1825) and was appointed to the curacy of Bessingby, Yorkshire. The discharge of his clerical duties at Bessingby, and later at Liverpool, at Exeter and at Bradford, did not prevent him from continuing his interest in science to which he added social questions, especially the improvement of conditions in factories. When crossing the Atlantic in 1848 he made some valuable observations on the height of waves and he voyaged to Australia in 1856 in order to obtain magnetic data.

Scoresby died at Torquay on March 21, 1857.

See R. E. Scoresby-Jackson, *The Life of William Scoresby* (1861). (L. M. FS.)

SCORPIO or SCORPIUS (the Scorpion), in astronomy, a constellation and sign of the zodiac. The constellation appears rather low in the south in the early evenings of summer for observers in middle northern latitudes. The stars outline a sprawling figure that might seem to suggest a scorpion or perhaps a kite with a long tail. The brightest star, Antares, is a supergiant at the distance of something like 170 light-years. This red star has several hundred times the sun's diameter and is intrinsically 1,000 times as luminous as the sun.

The region of the heavens from Scorpius eastward to Sagittarius and northward into Ophiuchus surrounds the direction of the centre of our galaxy. This region exhibits a spectacular display of star clouds of the Milky Way, star clusters including great globular clusters, and bright and dark nebulae. In photographs the dark cosmic dust is frequently illuminated by involved stars in patches of reflection nebulae, remindful of the glows around streetlamps on a foggy night. (R. H. BR.)

SCORPION, a venomous arachnid of the order Scorpiones. Although different in size and degree of toxicity, all scorpions are similar in appearance and are easily distinguished from spiders and other arachnids by their clawlike pincers and long upturned tail, at the tip of which is the characteristic sting. Because of their poisonous qualities, scorpions have been known and feared since ancient times.

Distribution.—While scorpions are usually associated with hot, tropical countries, they are not confined to them. In Europe, several species live in Greece, Italy, Spain, and the Balkans, and one ranges as far north as southern Germany. In the Western Hemisphere scorpions are found from the United States to Patagonia. They are absent from many islands.

Most scorpion species have a limited geographic range; *Isometrus maculatus*, found in all the warmer parts of the world, and *Scorpio maurus*, ranging from the Atlantic to India, are notable exceptions. A few scorpions are found at higher altitudes; *e.g.*, in the Atlas Mountains (northwest coastal Africa).

The species found in Europe and those in the United States are known and have been described. Of the four families represented in the United States, only the Buthidae, so far, have been shown to contain dangerously poisonous species, namely two species of the genus *Centruroides*. Other less dangerous species belonging to this genus are found in the U.S. Since less than a quarter of the more than 30 species known to occur in the U.S. have been studied with a view toward ascertaining the type and virulence of their poison, caution in handling all scorpions is advisable.

General Features.—The size of scorpions varies considerably. Some species are only 13 mm. (½ in.) long, while the largest attain 175 mm. (7 in.). Although even the smallest scorpions are large enough to be easily studied under low power magnification, the identification of species is quite difficult, because of the great deal of variation in most characters. The recent families into which the scorpions are divided, as well as the genera, are more easily determined, based on structural characters such as sternum, combs, spines, rows of granules on appendages and on the surface of the body, etc. Apparently of the least importance physiologically, these features were less subject to functional modifications and remained more or less stationary throughout long geological periods. Paleozoic scorpions, for example, are very much like recent ones in appearance. Specific characters are finer and less stable.

The second pair of appendages, the pedipalps, always held forward, are powerfully developed and are provided with pincers resembling the claws of a crayfish. The first portion of the abdomen, the preabdomen, is as wide as the cephalothorax. The last five annular segments form the "tail" or postabdomen, and are much more slender than and sharply set off from the preceding seven segments of the preabdomen. At the end of the tail is a sting, somewhat curved and pointed; its enlarged base contains a pair of poison glands, which open near the tip of the sting. Because of its curvature, the sting can be used only when the tail

ROBERT H. WRIGHT

FEMALE SCORPION (HADRURUS HIRSUTUS) SHOWING TWO PAIRS OF LATERAL EYES AND ONE PAIR OF MEDIAN EYES NEAR THE ANTERIOR EDGE OF THE CARAPACE. BENEATH THE CARAPACE ARE THE CLAWLIKE CHEWING MOUTHPARTS

is raised above the preabdomen and thrust forward. On the ventral side of the abdomen, immediately behind the genital operculum, there is a pair of comblike structures, the pectines. These organs are found only in scorpions. Each pectine articulates with a chitinous plate that represents the second sternite, and each is provided with muscles and nerves. The number of teeth, or lamellae, on each pectine varies from 3 to more than 40, depending upon the species and sex of the scorpion. It is thought that the pectines perceive ground vibrations.

Natural History.—*Habits.*—Scorpions live in a wide variety of places—under stones and bark, in crevices, under dead leaves and rubbish, in barns and deserted buildings, in thatched roofs, etc. They are usually nocturnal and prefer tight quarters.

Essentially inhabitants of warm areas, they tend to become sluggish during cold periods; they can, however, withstand freezing weather for several weeks. Although often common in dry areas, scorpions hide in the ground during arid periods. When the rains come, the scorpions find their way to the soil surface and stay under loose stones. They are very sensitive to microclimatic variations, and each species appears to live and reproduce within very limited and characteristic ecological conditions.

Some of the larger species (*e.g., Palamnaeus swammerdami* of southern India) emit rasplike sounds when alarmed. The noise is produced by a special stridulatory apparatus consisting of a spinous scraper on the flat outer face of the basal segment of the pedipalps and a tubercle-studded rasp on the flat inner face of the corresponding segment of the first pair of walking legs.

Food.—The food of scorpions consists of spiders, harvestmen, many different soft-bodied insects, myriapods, and even small mice. Scorpions probably do not actively seek their food, but rather wait for the prey to come to their lairs. The prey is picked to pieces by the chelicerae, the juices and soft tissues being drawn into the mouth by the pumping action of the pharynx. A well-fed scorpion can survive for several months without either food or water.

Enemies.—Notwithstanding their poison, scorpions have enemies. In tropical rain forests of Mexico and Central America, army ants, in the progress of their raids, can quickly overpower and dismember scorpions that they come upon. African baboons have been observed catching large scorpions, tearing off their tails, and greedily devouring the rest of the bodies. Lizards, snakes, and birds are listed among scorpion predators; furthermore, many scorpions are cannibals. It has also been reported that certain North African tribes enjoy eating live scorpions. Among parasites are various mites and nematodes. Actually, the greatest threats to the scorpion are food shortage, drought, and human activities.

Reproduction and Life Cycle.—When mature, the sexes show visible differences in the relative proportions of the body. The male is more slender and has a longer tail. It also has a pair of organs used in copulation, but these are visible only when the specimen is turned over on its back. When the male encounters a female, he engages her in a "courtship dance." The male grasps the female's pedipalpal claws with his and walks sidewards or backwards as he holds on to her. In *Buthotus alticola*, this promenade is preceded by a maneuver in which the animals, while facing one another, straighten the back parts of their bodies, compress the abdomens to the ground, and extend their tails upward. The tails are alternately twined and untwined, after which the animals promenade. This performance may last several hours. Finally, the male either digs a hole (while still holding the female) or finds a retreat into which the couple enter to copulate. Special organs in the male form a temporary penis with which the sperms are inserted. After copulation, the male is often attacked and devoured by the female.

The fertilized eggs develop inside the mother, and the young are born alive. The manner in which the eggs develop depends upon the amount of yolk present. Among the Buthidae, in which there is much yolk, the eggs pass into the oviduct and develop there; the embryos consume the yolk with which they are supplied. Among the Scorpionidae, with little yolk, the fertilized eggs remain in place and become closely mingled with the maternal tissues.

GORDON S. SMITH FROM NATIONAL AUDUBON SOCIETY

FEMALE SCORPION (HADRURUS HIRSUTUS) CARRYING YOUNG ON HER BACK

At birth the young, unable to take care of themselves, mount upon the mother's back. This process may be slow and laborious, for the newborn scorpions are very weak. The young may be so numerous as to hide the body of the mother so that only her appendages and tail are visible. In 10–16 days, after their first molt, the young, which have acquired the typical scorpionlike appearance, come down off the mother's back. They remain close to her for a day or two and then scatter.

Scorpions live alone during most of their lives. They avoid each other, but when confronted, fight to the death, the victor usually devouring the victim. Growth is accompanied by molting, as in all arthropods. The old skin cracks around the carapace and is shed completely. The linings of the foregut, hindgut, and four pairs of book-lungs (situated on the ventral side of the third to sixth preabdominal segments) are also shed. The number of molts may vary, even in the same species; it is believed that *Palamnaeus longimanus* has eight and *Androctonus australis* seven.

Scorpion Poison.—Many observations on the effect of scorpion poison are unreliable for many reasons. Sources of error include a lack of knowledge of the physiological state of the victim, ignorance of the amount of poison absorbed, and the ratio of amount of poison to the total weight of the victim. Another serious source of error is that many observations refer simply to a "scorpion." There are more than 600 species of scorpions in the world, which vary greatly in degree of virulence and type of poison.

At least two general types of scorpion poison exist. One of these is local in effect and comparatively harmless to man; it is exemplified by the European *Euscorpius italicus*, the American *Centruroides vittatus*, *Vejovis spinigerus*, the formidable-looking *Hadrurus hirsutus* and *H. spadix* of Arizona and Utah, and the large *Centruroides margaritatus* of Central America. Some scorpions have reduced poison glands, and their venom has no effect. The other type is a dangerous neurotoxin that resembles the venom of some snakes; it is represented by the North American *Centruroides sculpturatus* and its close relative *Centruroides gertschi* and by the North African *Androctonus australis.*

The symptoms caused by scorpion poison of the less virulent type consist mostly in sudden sharp pain followed by numbness of the limb and local swelling. The symptoms pass within an hour or two. For a healthy non-allergenic adult man there is no more danger from this type of scorpion poison than there is from the sting of a wasp or a hornet. The neurotoxic venom has an affinity for the autonomic nervous system. The victim experiences a tightness in the throat, difficulty in speaking, restlessness, and slight involuntary twitching of muscles. Sneezing spasms may follow, accompanied by a flow of mucus from the nose and mouth. The rate of heart beat may be considerably increased. Convulsions

follow, and the extremities become bluish just before death occurs (anytime from 45 minutes to 10–12 hours after the onset of symptoms).

If administered early enough, antiscorpion serum, produced by injecting a ground preparation of scorpion terminal tail segments into horses, may reduce mortality to less than 3%.

The effect of the scorpion venom on different animals varies greatly. Hedgehogs, jerboas, and fennec foxes are nearly immune while guinea pigs and dogs are susceptible. Birds, frogs, and fishes are also sensitive.

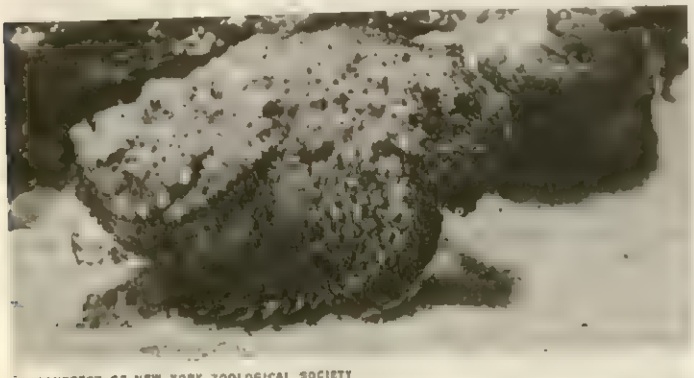
The story that scorpions commit suicide by stinging themselves when surrounded by a ring of fire is based on misinterpretation of their behaviour and has long since been discredited. Certain scorpions possess a degree of immunity to their own venoms. This is true of *Palamnaeus scaber* (formerly *Scorpio afer*), according to older statements. Other species, *e.g.*, *Androctonus australis*, may be killed by injections of their own poison only when the dose is massive, nearly 200 times greater than the one necessary to kill a guinea pig.

Classification.—The Scorpiones are divided into six families, of which the Buthidae is the largest and most important. It has more than 600 species including *Buthus occitanus*, commonly found in the Mediterranean area and in France; *Androctonus australis*, the fat-tailed scorpion of North Africa; and the neotropical genera *Centruroides* and *Tityus*. The family Diplocentridae is found in the Palaearctic region, the island of Sokotra in the Gulf of Aden, and Mexico; the Scorpionidae, in Africa, Madagascar, Asia, and Australia; the Vejovidae, mainly in Asia and America; the Chactidae, in Asia, America, southern Europe, and the Mediterranean (one small species); the Bothriuridae, in Australia.

See ARACHNIDA.

BIBLIOGRAPHY.—J. L. Cloudsley-Thompson, *Spiders, Scorpions, Centipedes, and Mites* (1958); J. Henri Fabre, *The Life and Love of the Insect*, trans. by A. T. de Mattos (1911), chap. 17 and 18, "The Languedocian Scorpion"; A. Kaestner, "Ordnung der Arachnida: Scorpiones" in *Kukenthal's Handbuch der Zoologie* (1941, vol. 3, no. 2); O. de Magalhaes, "The Scorpionic Syndrome," *J. Trop. Med. Hyg*, vol. 42 (1939), "Escorpionismo," *Bras-Méd*, vol. 56 (1942); J. Millot and Max Vachon, "Ordre des Scorpions" in *Traité de Zoologie*, vol. 6 (1949); M. Vachon, "The Biology of Scorpions," *Endeavour*, vol. 12 (1953); F. Werner, "Scorpiones," *Bronn's Klassen und Ordnungen*, vol. 5 (1935).
(C. J. Go.; M. L. G.)

SCORPION FISH, the name given to any of the spiny marine fishes of the family Scorpaenidae, characterized by a bony plate or stay across the cheek. The name is particularly used for many mostly small (a foot or less long), bottom species in the shore waters of all warm seas (*Scorpaena, Pterois*, etc.). These usually have mottled colours and are found on rough bottoms or among seaweed. In cooler, coastal waters of the North Pacific, there are larger (two or three feet long), freer-swimming, less spiny species (*Sebastodes*, etc.), commonly known as rockfish, and of greater market value. In the cold waters of the North Atlantic there is a bright red one, the rosefish, redfish, or ocean perch (*Sebastes marinus*), which is excellent eating. So far as known all members of the family are good to eat, but persons handling the tropical ones, especially, should use care, as wounds from their sharp spines may be very painful and cause swelling. The glands at the bases of the dorsal fin in the poison fish or stonefish (*Synanceja verrucosa*), of East Indian waters, can inflict severe wounds.



BY COURTESY OF NEW YORK ZOOLOGICAL SOCIETY

STONEFISH (SYNANCEJA VERRUCOSA), ONE OF THE MOST VENOMOUS OF THE SCORPION FISH

In the cool-water species, the eggs hatch within the body of the mother as a rule, broods of hundreds if not thousands of tiny fishes being spawned.

See also LION FISH; FISH: *Survey of the Bony Fishes.*
(J. T. N.; X.)

SCORPION FLY, the common name for insects of the family Panorpidae, order Mecoptera (*q.v.*), and by extension often applied to the order as a whole. The name refers to the fact that in male panorpids the abdomen terminates in a bulb-like segment which is held over the back, after the manner of a scorpion. In both sexes the chewing mouth parts are borne at the tip of a stout beak, the antennae are long and many-segmented, and there are two pairs of membranous, net-veined wings, which in many species are darkly spotted or banded. The largest genus, *Panorpa*, comprising about 170 species, is widely distributed in forested areas of the Northern Hemisphere. Adult scorpion flies, usually found on the wing in summer, feed on dead animals, especially insects, but the larvae are thought to be predaceous on minute soil animals. The immature stages of only a few species are known; the larvae resemble caterpillars, and pupation occurs in a cell in the soil.

ROSS E. HUTCHINS

A MALE SCORPION FLY (PANORPA) COMMON IN THE SOUTHERN UNITED STATES

Besides Panorpidae, the order includes the hanging flies, family Bittacidae, which hang by their forelegs and catch insect prey with their raptorial hind feet; and the tiny, blackish snow scorpion flies, family Boreidae, with greatly reduced wings.

Snow scorpion flies are occasionally seen walking over the surface of snow; both the adults and larvae are said to feed on mosses. The order is small, numbering about 350 known living species, but it is very widespread in distribution and is of great antiquity.

Scorpion flies are altogether harmless to man and serve a useful function in nature as scavengers.

See INSECT: *Classification: Mecoptera.* (G. W. Bs.)

SCOT, MICHAEL (d. *c.* 1235), Scottish scholar and mathematician whose translations from Arabic and Hebrew are a very important landmark in the reception of Aristotle in the Latin West, was famous in the Middle Ages as an astrologer and soon acquired a popular reputation as a wizard. There is no evidence as to where he obtained his master's degree, and he first appears at Toledo in 1217, where he finished translating the treatise of al-Bitrogi (Alpetragius) on the sphere. In 1220 he was in Bologna, and during the years 1224–27 may have been in papal service, as he is mentioned in several papal letters. A pluralist, he was promoted archbishop of Cashel in Ireland (May 1224), but declined the see a month later. In April 1227 Pope Gregory IX urged on the archbishop of Canterbury Michael's claims to benefices as one learned in Hebrew, Arabic, and Latin. Thereafter he was at the Sicilian court of the emperor Frederick II and was mentioned as recently dead in a poem written early in 1236.

His works are mainly undated, but those on natural philosophy seem to predominate in his earlier, Spanish period, and those on astrology in his later, Sicilian period. At Toledo, in addition to his translation of al-Bitrogi, Michael translated Aristotle's *Historia animalium* from Hebrew or Arabic. He also translated, perhaps at this time, Aristotle's *De caelo*, and he was probably responsible for the translations of the *De anima* and the commentary by Averroës which is found in the same manuscripts. There is no evidence that Scot translated Aristotle's *Physics, Metaphysics*, or *Ethics*.

In the Sicilian period he wrote three treatises on astrology: the *Liber introductorius*, a long popular work on astronomy and astrology, sometimes citing the author's own experiments; the *Liber particularis*, much shorter, dealing mainly with the reckoning of time, the calendar, and meteorology, ending with a series of questions on the relationship of earth, infinity, and heaven, posed by Frederick II and ineptly answered by Michael; and the *Physionomia*, designed to help the emperor in his judgments of men, containing a treatise on generation and an account of prognostication from dreams, complexions, and the different parts of the body. His work exhibits the worst failings of contemporary learning: prolixity, incoherence, and uncritical credulity.

Michael dabbled in alchemy, and by the Renaissance several alchemical works were ascribed to him; however, his authorship of none of these is proven. He appears in the *Inferno* of Dante (xx) among the magicians and soothsayers and has the same role in Boccaccio. He also made medical observations, but the "pills of Master Michael Scot" have nothing to do with him.

BIBLIOGRAPHY.—C. H. Haskins, *Studies in the History of Medieval Science* (1924); L. Thorndike, *History of Magic and Experimental Science*, vol. ii (1923); *Michael Scot* (1965). (A. A. M. D.)

SCOTLAND, the most northerly of the three countries of Great Britain. It is a generally cool, hilly, and in the west a wet country which stretches from the Solway Firth and the English border to the Pentland Firth in the north, with the Orkney Islands, Fair Isle, and Shetland Islands beyond, and from the North Sea in the east to the Atlantic islands of the Hebrides and St. Kilda in the west. Area 30,412 sq.mi. (78,768 sq.km.). Pop. (1961) 5,179,344. The capital is Edinburgh.

The mainland is 274 mi. (441 km.) from Cape Wrath, Sutherland, to the Mull of Galloway and 154 mi. (248 km.) from Buchan Ness to Applecross in western Ross; but it is only 26 mi. (42 km.) wide between the estuaries of Dornoch Firth and Loch Broom and 30 mi. (48 km.) between the Firths of Forth and Clyde. The indented coastline of about 2,300 mi. (3,700 km.) causes few areas to be over 40 mi. (64 km.) from the sea but in places obstructs land routes or causes expensive bridging as at the Firths of Forth and Tay. Scotland is separated from England by the Solway Firth, the Sark, Scotsdyke (an old embankment between Sark and Esk), the Esk (for 1 mi.), the Liddel, the Kershope, the Cheviot Hills, the Tweed, and a small area known as the "liberties" of Berwick. The name Scotland originated in the 11th century when the name Scotia was given to a southwestern tract settled by the tribe of Scots who had migrated from Ireland, to which land the name Scotia had previously been applied. The older name Caledonia long remained in use poetically. (A. T. A. L.)

This article contains the following sections and subsections:

For subjects concerned with the United Kingdom as a whole (*e.g.*, constitution, legislature, defense, finance and trade) *see* GREAT BRITAIN AND NORTHERN IRELAND, UNITED KINGDOM OF.

I. PHYSICAL GEOGRAPHY

1. Geology.—Scotland lies on the course of a southwesterly trending mountain chain of pre-Devonian Age, which stretches from Scandinavia to Ireland and was called Caledonian by Eduard Suess. In Scotland the northwestern front of this chain is marked by the Moine Thrust outcropping in the Northwest Highlands and extending from Durness to Skye. Its latest manifestation is the corrugation of Ordovician and Silurian in the Southern Uplands, where Charles Lapworth demonstrated the value of graptolites (*q.v.*) in determining the succession of strata.

Precambrian Below the Moine Thrust.—Precambrian rocks which occur beneath the Moine Thrust have been widely exposed in the islands and coastal belt of the Northwest Highlands. They consist of the Lewisian complex, predominantly igneous and almost wholly metamorphic, and the unconformably overlying Torridonian sediments, almost everywhere nonmetamorphic.

FIG. 1.—MAJOR PHYSICAL FEATURES, CITIES, AND TOWNS OF SCOTLAND

The igneous rocks of the Lewisian range from ultrabasic (rich in olivine and calcic feldspar) to ultra-acid (containing a high percentage of silica and yielding quartz); many have suffered in varying degree from two metamorphisms separated by a long time interval. In the Assynt district the following history has been deciphered: (1) formation of a "fundamental complex" of metamorphic igneous rocks; (2) cooling of this complex; (3) intrusion, with chilled margins, of a great suite of west-northwest basic dikes (named after the hamlet Scourie); (4) reheating accompanied by widespread, though not universal, differential movement and intrusion resulting in a second metamorphism (named after Loch Laxford). In a large proportion of Lewisian outcrops Scourie dikes are not distinguishable with certainty. This is due in part to Laxfordian metamorphism having obscured the contrast between dikes and country rock; but it may also be due in part to large-scale incorporation in the Lewisian complex of rocks later than the Scourie dikes. Certain considerable outcrops of Lewisian sediments, including magnesian marbles and graphite schists, are not cut by Scourie dikes. Such sediments in Coll and Tiree carry massive intercalations of igneous gneisses almost certainly later than themselves. After the Laxfordian high temperatures had passed away, sudden movements locally developed frictional heat sufficient to cause incipient fusion on a very small scale. This, followed by a quick cooling, produced crush rocks (*see* BRECCIA) with flinty appearance. An unusually prominent development of "flinty" crush rock follows an east-southeasterly inclined shear zone along the east coast of the Outer Hebrides.

The uncomformable Torridonian sediments probably accumulated under continental conditions. From Islay to Skye, they start with Lower Torridonian basal conglomerates (*q.v.*), grits rich in the mineral epidote (*q.v.*), grey sandstones, flags, and grey and black shales (7,000 ft. [2,130 m.]). Farther north, Middle Torridonian arkose, a type of coarse sandstone (6,000–8,000 ft. [1,830–2,440 m.]), with scattered breccias, rests among and covers over "fossil hills" of Lewisian gneiss, which are sometimes more than 2,000 ft. in individual height. Upper Torridonian sandstones, flags, and shales attain 4,500 ft.

Early Paleozoic Below the Moine Thrust.—A very regular marine succession of Durness sediments (Lower Cambrian extending into Lower Ordovician) lies with angular unconformity on gently folded Torridonian and Lewisian complex. The lithology and fauna are: basal quartzite (320 ft.); quartzite with vertical worm casts, sand-filled tubes formed by burrowing worms (*Scolithus linearis*) (270 ft.); dolomitic shales, mudstones, and dolomites, with the trilobite (*q.v.*) *Olenellus* (50 ft.); grit, with *Salterella* and *Olenellus* (30 ft.); limestones and dolomites, some fossiliferous (1,500 ft.). At Assynt plutonic (possibly early Devonian) intrusions of alkali-syenite, borolanite, etc., accompanied by sills, cut these sediments. Contact effects include dedolomitization (limestone is converted due to interaction between dolomite and igneous rock).

Moine Thrust Zone.—The Moine Thrust, traceable from Whiten Head near Durness to Skye, is of early Paleozoic date. As a result of this thrust crystalline sedimentary schists are found overlying the Cambro-Ordovician and Torridonian nonmetamorphics. The alkali-syenites of Assynt underlie the thrust; but some associated sills are intruded into the overlying schists. The suite has suffered much of the movement registered in the thrust zone, but perhaps not all. An alkali-syenite at Ben Loyal may be a late member.

Near Durness a thrust outlier (klippe) of the Moine Thrust mass (nappe) is downfaulted 10 mi. in advance of the main out-

crop. In Assynt and Glencoul, subsidiary thrusts carry 1,500-ft. slices of Lewisian for miles over Cambrian. Spectacular inversion occurs at Loch Alsh, where, as in Colonsay, disturbed Torridonian develops minute biotites (*see* BIOTITE) of Moine Thrust Age. *Schuppen* structure (*i.e.*, slices overthrust one upon another) and mylonization (*see* MYLONITE) are among the structures diagrammatically exposed at many localities.

Pre-Devonian of the Highlands Between the Moine Thrust and the Great Glen.—The sedimentary schists above the Moine Thrust extend southeast to the Caledonian Canal along the Great Glen. They are mostly quartzofelspathic or micaceous and grouped as Moine schists. With them, at Glenelg and elsewhere, are several minor outcrops of Lewisian-like igneous and sedimentary types. These are commonly interpreted as inliers of Lewisian and as interfolded and remetamorphosed with the Moines which are supposed to be unconformable. Inchbae porphyritic granite (*see* PORPHYRY), now gneissose, was intruded into the Moines before these last had suffered regional metamorphism, since adjacent indurated hornfels (*q.v.*) retains minutiae of sedimentary structures, such as grains, ripple marks (*q.v.*), and sun cracks. Garnets are widely developed in the Moine schists; and, some distance east of the Moine Thrust, sillimanite (*q.v.*) is commonly associated with granitic injection. Underlying the Moine Thrust at Tarskavaig in Skye, there is a thrust mass, the rocks of which are variously interpreted as Moines or Torridonian. Some geologists think all Moines are metamorphosed Torridonian; others regard Moine metamorphism as pre-Torridonian.

Great Glen Fault.—The Great Glen is eroded along a belt of intense shattering connected with the Great Glen fault. This fault according to good, though inconclusive, evidence is a sinistral wrench fault (*see* FAULT) with horizontal displacement of 65 mi. It appears, for instance, to shift the northern part of the Foyers granite (Loch Ness) southwestward to the Strontian district.

Pre-Devonian of the Highlands Southeast of the Great Glen Fault.—The Moine Series crosses the Great Glen fault, but the northeast, east, and southwest part of the Grampian Highlands consists of Dalradian schists, probably of post-Moine age. In the Dalradian (*see* PRECAMBRIAN TIME: *Scotland*) quartzites, limestones, and graphitic schists are characteristic members; the Schiehallion boulder bed may be glacial; pillow lavas occur at Loch Awe; basic intrusions are widespread. Current and graded bedding (the former also in the Moines) help to determine the stratigraphical order. The Dalradians overlie Moines, pitching off them toward Loch Awe and the county of Banff. Large-scale recumbent folds with fold faults (slides), the whole refolded, are typically shown at Ballachulish and elsewhere.

Zonal mapping of regional metamorphism was first carried out in the Dalradian belt. The grade varies from roofing slate to sillimanite-gneiss and is low at Loch Awe, in Banff, and along the Highland border. In the sillimanite zone, granitic permeation is characteristic. Recent research has concentrated on the study of minor folds and mineral orientation.

Middle Cambrian limestone (near Callander) and Upper Cambrian to Ordovician pillow lavas, radiolarian (*see* PROTOZOA: *Classification and Survey: Radiolarida*) cherts, and serpentine (probably as folded sheet) with fossiliferous shales outcrop discontinuously along the Highland boundary fault. At Stonehaven they are unconformably overlain by Downtonian (a division within the Devonian system, *q.v.*) sandstones with mudstones, tuffs (rocks formed by induration of volcanic ash or dust), and conglomerates (2,760 ft.), which yield fish and the fossil *Dictyocaris*.

Highland Boundary Fault.—From Stonehaven to beyond Loch Lomond runs the Highland boundary fault. It has generally functioned as a broken monocline (with complications) throwing down toward the Midland Valley (termed mid-Lowlands hereinafter). It was active during much of Old Red Sandstone times and probably earlier. In post-Carboniferous times, at Loch Lomond, it threw back, down toward the Highlands.

Pre-Devonian of the Mid-Lowlands.—The Stonehaven-Downtonian strata, strongly tilted, are followed conformably by Lower Old Red Sandstone. Similar conformity holds at Lesmahagow, where gentle anticlines expose the following succession: Graywacke (*q.v.*) and shale possibly of Wenlock age (1,300 ft.); Ludlow mudstone, graywacke, and shale (1,480 ft.); Downtonian yellow, red, and chocolate sandstone with a quartzite conglomerate and some mudstone (2,700 ft.). Pentland inliers again show a Wenlock to Downtonian succession, but here Lower Old Red Sandstone follows with strong unconformity. The Wenlock-Ludlow (*see* SILURIAN SYSTEM) rocks of both districts have yielded many brachiopods (*see* BRACHIOPODA) and mollusks, especially in the Pentland Hills; also scorpion, phyllocarids (*see* MALACOSTRACA: *Phyllocarida*), eurypterids (*see* EURYPTERIDA), and fish. The three last reappear in the Downtonian. In the Girvan district, Arenig to Wenlock emerge from beneath unconformable Lower Old Red Sandstone considerably northwest of the Southern Uplands fault.

Southern Uplands Fault.—The mid-Lowlands are bounded to the southeast by the great Southern Uplands fault. In the southwest this crosses the mouth of Loch Ryan and continues northeast till, south of Edinburgh, it turns abruptly southward. Thereafter, resuming its northeast trend, it reaches the coast just south of Dunbar. Ordovician rocks of the Southern Uplands outcrop almost continuously on its southeast side, brought against successively: for the first 30 mi., in the southwest, Girvan outcrops of Ordovician lying to the northwest; for the next 60 mi., Lower Old Red Sandstone; and for 30 mi. to the North Sea, Carboniferous and Upper Old Red Sandstone. The Sanquhar coalfield has been preserved by late backthrow down toward the Southern Uplands.

Ordovician and Silurian of the Southern Uplands and Girvan.—Closely packed isoclinal folding and rapid cross-strike change of both facies (distinct kinds of rock or biologic associations) and thickness are characteristic. The Ordovician commences with Arenig pillow lavas, probably continuing into the Llandeilo group, capped by Llandeilo radiolarian cherts (70 ft.)—the term Llandeilo is used in Lapworth's sense. Near Girvan 1,500 ft. of Arenig-Llandeilo volcanic rocks are associated with serpentine, gabbro, and granite; the whole was exposed to erosion before the deposition of conglomeratic Upper Llandeilo (830 ft., including 60 ft. interbedded fossiliferous Stinchar limestone). Southeast of Stinchar Valley, Upper Llandeilo becomes conformable to the volcanic series, and consists of mudstones and grits (500 ft.) and conglomerate (500 ft.). Farther southeast it passes to grit and graywacke (1,000 ft.) and then, near Moffat, to 20 ft. of graptolitic shale (Glenkiln). The Caradoc (including Ashgillian) at Girvan is mostly mudstones, grits, flags, and shales (2,800 ft.) with interbedded shelly and graptolitic faunas. Near Moffat this reduces to 100 ft. of graptolite shale and barren mudstone (Hart Fell). Acid lavas occur at Wrae in Tweeddale.

A general paleontological break introduces the Silurian, although without upheaval at Moffat. The Llandovery series (*see* SILURIAN SYSTEM) at Girvan consists of conglomerates, grits, flags, shales, and thin limestones (1,050 ft.) with interbedded shelly and graptolitic faunas. Near Moffat there are only 100 ft. of graptolite shale (Birkhill). The Tarannon, however, both at Girvan (2,100 ft.) and at Moffat (3,000–4,000 ft.), consists of grits, flags, and shales with occasional graptolitic intercalations. Wenlock is doubtfully represented near Girvan by conglomerates, grits, flags, and shales (500 ft.) with minor graptolitic and shelly layers. Near the English borders Wenlock conglomerates, grits, graywackes, shales, and mudstones (1,000 to 1,500 ft.) contain several graptolitic bands with occasional eurypterids. Shelly layers also occur. In the same district Ludlow mudstones with limestone nodules and grits (500 to 750 ft.) yield shells.

Ordovician and Silurian conglomerates usually contain material derived from Arenig lavas suffering erosion north of the Southern Uplands. Quartzite appears as pebbles in the Caradoc and mica-schist in the Llandovery. The graywackes frequently show graded bedding. Undated post-Silurian mineral veins occur at Leadhills. (For an explanation of the terms Caradoc, Llandeilo, and Arenig, *see* ORDOVICIAN SYSTEM.)

Devonian.—All the Scottish Devonian is continental Old Red Sandstone. In basal exposures, except at Stonehaven and Lesmahagow, it follows on older rocks with striking unconformity. In depth, however, under the mid-Lowlands, conformity with Downtonian may perhaps be the rule. Three divisions are recognized

with distinct faunas and floras. Fish have furnished the majority of determinable fossils. The Lower Old Red Sandstone is most fully preserved in the mid-Lowlands (19,000 ft., including lavas, in the county of Kincardine). Dull purplish brown sandstone is widespread, covered in Strathmore by red sandstones and marls. Mid-Lowland conglomerates attain great prominence toward the Highlands and Southern Uplands. Subalkali volcanic rocks, including basalt, andesite, and rhyolite (*qq.v.*), occupy a roughly central position in the Lowland succession (6,500 ft. in the Ochil Hills and well exposed also from the Pentland Hills to Ayrshire). In the Highlands (Oban, Glencoe, Ben Nevis) and also at Cheviot, lavas greatly exceed sediment; Stonehaven conglomerates show that early Highland volcanoes of the suite started in Downtonian times. Glencoe and Ben Nevis are famous for their cauldron subsidences. At both these localities granites and quartz diorites (*see* DIORITE) with northeasterly dike swarms are associated with the cauldrons. Granites cutting Lower Old Red Sandstone also occur at Distinkhorn (Ayr) and Cheviot. Others penetrate folded Silurian (Southern Uplands). Many undated Highland granites are probably of Lower Old Red Sandstone Age; but granite pebbles are well known in Lower Old Red Sandstone conglomerates (Glencoe, Stonehaven).

Middle Old Red Sandstone is widely developed in northeast Scotland (18,000 ft. in Caithness), extending to Orkney and the Moray Firth. It largely consists of flags, often bituminous, calcareous, ripple-marked, and sun-cracked. The Rhynie chert with wonderfully preserved plants is generally referred to this formation.

Upper Old Red Sandstone (some thousands of feet thick) is found in the mid-Lowlands and to a limited extent in northeast Scotland; it is everywhere, except perhaps in Shetland, unconformable to its predecessors. Red sandstones with some wind-rounded grains are common in its earlier parts; while paler sandstones with cornstones (concretions of impure sandy limestones) occur toward the conformable base of the Carboniferous.

Carboniferous.—The Scottish Carboniferous, marine, estuarine, and terrestrial, is typically developed in the mid-Lowlands and along the English border. These two areas of deposition were always connected across the eastern end of the Southern Uplands; but farther west (near Sanquhar) Upper Carboniferous rests on Ordovician and Silurian. In the Highlands, Carboniferous (though represented at Campbeltown, Bridge of Awe, and Morven) is scarcely known. In the mid-Lowlands and on the borders it follows Upper Old Red Sandstone conformably, though with a different set of fishes. Subdivisions in Edinburgh district are: Calciferous Sandstone series, including Cementstone group (1,000 ft.) and Oil Shale group (3,000 ft.); Scottish Carboniferous Limestone series, including Lower Limestone group (700 ft.), Limestone Coal group (1,050 ft.), and Upper Limestone group (1,050 ft.); Scottish Millstone Grit (800 ft.); Productive Coal Measures (1,500 ft.); Barren Red Coal Measures (in Ayr, 1,500 ft.). On the basis of zonal fossils the naming of these divisions has recently been revised to facilitate comparison with the standard English succession: the Calciferous Sandstone and Lower Limestone grouped together correspond with the English Carboniferous Limestone; the Limestone Coal and Upper Limestone groups equate with much of the English Millstone Grit; while the Scottish Millstone Grit is renamed Passage group, and belongs partly to the English Millstone Grit, partly to the Coal Measures. It is widely thought that the conjunction of Millstone Grit and Coal Measures in the Passage group is unconformable, and that this accounts for a plant and estuarine fish break recognized by R. Kidston and R. H. Traquair. The base of the Barren Red Coal Measures is defined for stratigraphical purposes at the top of the Skipsey Marine Band. Actually the reddening does not allow stratigraphy at all closely. It has been determined by subterranean penetration of a desert climate during New Red Sandstone times, which led to widespread oxidation of iron carbonate and sulfide, and, in Ayrshire, to local replacement of coal by calcium and magnesium carbonates.

The oil shale and two coal-bearing groups (with clayband and blackband ironstones) are economically valuable, though oil shale is little worked today. Marine limestones are scarcely represented except in the Lower and Upper Limestone groups (near Edinburgh eight beds reach the quite exceptional total of 230 ft.). A well-known 40-ft. freshwater limestone (Burdiehouse) occurs in the Oil Shale group. High quality fireclays are worked in the Passage group.

Rapid variations of thickness are characteristic, as on the two sides of the Kerse Loch fault near Patna in Ayr. Change of facies is exemplified by restriction of workable oil shale to West Lothian and its borders. Vulcanicity was common till the close of the Millstone Grit and has recently been found to extend into the Coal Measures under the Firth of Forth. Special volcanic activity reigned about the end of the formation of the Cementstone group (Clyde Plateau, Arthur's Seat, Garleton Hills). Products are distinctly, but not strongly, alkaline. They include essexitic basalts (of granular, intrusive type), mugearites, trachytes (*see* TRACHYTE), and phonolites (*see* PHONOLITE). Contemporaneous weathering of Millstone Grit basalts has given valuable bauxitic clay (Ayr). Volcanic necks filled with basaltic agglomerate are well displayed (Arthur's Seat, Campsie Fells). Some scarcely transgressive, much faulted sills of olivine-dolerite and teschenite (*see* THERALITE) may be Carboniferous (West Lothian).

Permo-Carboniferous.—Approximately at the close of Carboniferous times there was considerable folding and faulting, which developed, among other structures, many east-west faults. The movements were accompanied by intrusion of quartz-dolerite magma, affecting, more particularly, an east-west belt (Highland and Lowland) with Dundee near its northern limit and Glasgow near its southern. In most of this belt long, broad, isolated east-west quartz-dolerite dikes are characteristic; but about the Forth (Fife, Stirling, Kilsyth, Airdrie, Linlithgow, Queensferry) the dikes are associated with, or give place to, sills of identical composition. In the Lomond Hills of Fife one of these sills is pierced by Permian necks. This feature is continued over the Scottish border in Northumberland and Durham where the dike-sill association is found again east of the Eden Valley (Great Whin Sill) and is unconformably overlain by Permian sediments.

Permian.—From the Solway northwestward a train of Permian outcrops reaches intermittently across the Southern Uplands (overlying various formations from Silurian to Productive Coal Measures) and mid-Lowlands (overlying Barren Red Coal Measures) to Arran. In Ayr the Permian succession consists of lavas and tuffs (basalt and nepheline-basalt, 500 ft.) succeeded by brick-red desert sandstone (1,500 ft.). The volcanic period produced many necks there, and those in Fife are probably of the same period. Several nepheline-essexite sills (kylite) in Ayr clearly cut faulted Coal Measures and are certainly Permian. Many of the teschenite sills of Ayrshire and elsewhere in the mid-Lowlands are also, with high probability, of this date. So too are monchiquite dikes (*see* LAMPROPHYRE), common in the Ayr coalfield. Farther north, from Colonsay to Orkney, a camptonite-monchiquite suite of dikes has given rise to much discussion. They sometimes follow approximately the local direction of Tertiary dike-intrusion. A dike that may be a nontypical representative is covered unconformably by Trias in Morven; while a well-characterized example in Colonsay has given a helium figure which also supports ascription to the Permian. The alkali products of Permian igneous activity in Scotland seem to go naturally with the Exeter volcanics in England and the famous Oslo association in Norway.

At Elgin an outlier of reptiliferous sandstone (*Gordonia* beds of Cuttie's Hillock) is of marked desert character and is referred to uppermost Permian or basal Trias.

Trias.—In Arran 2,000 ft. of Permian, mostly brick-red desert sandstone and breccia, underlies 1,000 ft. of Triassic interbedded sandstone and marl with occasional nodular limestone. A mass fallen into a Tertiary volcanic neck includes the Lamellibranchia *Pteria contorta* shales (Rhaetic).

The base of the Mesozoics from Mull northward consists of Trias conglomerates and sandstones with concretionary cornstones. Red coloration is often subordinate. In western Mull 200 ft. of these rocks underlie 40 ft. Rhaetic with *Pteria contorta*. Elsewhere Rhaetic is only doubtfully distinguishable. Some of the Elgin sandstones (Lossiemouth) have yielded mid-Triassic reptiles.

Jurassic.—Low-lying outcrops of Jurassic, conformable to Trias, are preserved on the west and east coasts of the Highlands. On the west, where a cover of Tertiary lavas has furnished additional protection, a typical area, richly fossiliferous, is Skye and Raasay: Lower Lias (Broadford Beds, with limestones rich in the oyster *Gryphaea*, 340 ft.; Pabba Shales, 700 ft.); Middle Lias (Scalpa Sandstone, 240 ft.); Upper Lias (shales, 80 ft.; thinner with 8 ft. ironstone in Raasay); Inferior Oolite (mainly sandstone, 780 ft.); Bathonian Great Estuarine series (bituminous shales with thin shelly limestones, 600 ft.); Cornbrash (limestone with comminuted shells, only in Raasay, 20 ft.); Callovian, restricted (sandstone remnant in Skye); Oxfordian (shales, 120 ft. in Skye); Corallian (calcareous grits and shales, 140 ft. in Skye); Kimmeridgian (shales, 40 ft. in Skye).

At Brora on the east coast are found: Lower Lias (shales with thin coals, 80 ft.); gap in the series due to fault; Great Estuarine series (more than 80 ft. of shale, sandstone, etc., with 3 ft. Brora coal at top); Callovian, restricted (marine limestone, 5 ft.); Lower Oxfordian (shales, 275 ft.); Upper Oxfordian (gritty sandstone, 170 ft.); Corallian (limestone, 20 ft.; sandstone and shales, 200 ft.); Kimmeridgian (shales and boulder beds, 1,500 ft.). The boulders of the boulder beds consist of Middle Old Red Sandstone. Many are big; one measures 150 × 90 × 30 ft. They have fallen from an intermittently renewed, submarine fault scarp and are mixed with Kimmeridgian corals and shells. The cumulative throw of the fault exceeds 2,000 ft. In Skye another 2,000-ft. post-Corallian fault can be shown to have been planed down by erosion before Upper Cretaceous times. In Mull and on the Moray Firth the Great Glen fault seems to have stirred in Kimmeridgian times.

Cretaceous.—Transgressive Upper Cretaceous is known: *in situ* (Skye, Scalpa, Raasay, Eigg, Mull, Morven); as a block fallen down a Tertiary crater (Arran); as *remanié* flints (Aberdeen); and as glacial erratics (Leavad, Caithness). The Morven succession is: Cenomanian greensand (45 ft.); white sandstone, worked for optical glass (35 ft.); silicified chalk with the zonal cephalopod *Belemnitella mucronata* (5 ft.) (*see* CEPHALOPODA). The white sandstone consists of desert sand blown on to the Scottish shore of the Franco-Britto-Russian Chalk Sea. Desert conditions presumably reduced river pollution to a minimum. In Morven, Cretaceous is still preserved 1,600 ft. above present sea level.

Tertiary, Mainly Igneous.—The Cretaceous chalk was upraised, weathered, and silicified, and its crannies filled with desert sand (Mull). Subsequent conglomerate, lateritically weathered ash, and leaf beds among lavas imply a change to moist climate and volcanic eruptions. The plants have suggested to some early Eocene, to others Oligocene to Miocene. Subaerial basalt lavas characterize Skye, Eigg, Mull, and Morven (also Antrim). Exceptionally these lavas are columnar (Staffa). The basalt succession in Mull is olivine-rich flows (3,000 ft.) followed by olivine-poor flows (3,000 ft.). The Sgur of Eigg pitchstone seems to be an acid lava that filled a valley, the sides of which, consisting of basalt, have since been mostly eroded away. Great plutonic centres occur at St. Kilda, Skye, Rum, Mull, Ardnamurchan, and Arran. Some think that most of the lavas, still preserved, were fed from these centres. A Kilauean sink, repeatedly renewed in central Mull, was often filled by a crater lake, and lavas flowing into it developed pillow structure. Vent agglomerates are abundant in Skye, Rum, Mull, Ardnamurchan, and Arran.

The Skye plutonic centre is well known for its gabbro (*q.v.*) to granophyre (composed mainly or wholly of micropegmatite; *q.v.*) succession; Rum for its banded peridotites (*see* PERIDOTITE); Mull and Ardnamurchan for their ring-dikes. In Mull the plutonic succession is complex and begins with granophyre; in Ardnamurchan it is essentially gabbroidal. In Rum a ring-fault is associated with central and peripheral upheaval; the main recent discovery in Rum is that its granophyre (forming Mt. Orval) is earlier than its basalt lavas; Arran is specially noteworthy for the doming of its granite's roof. Arcuate peripheral folding is a feature of Mull and Skye. Cone-sheet complexes are extensively developed in Skye, Mull, and Ardnamurchan. The Skye, Rum, Mull, and Arran centres have crowded dike swarms. The general dike direction is northwest.

KENNETH SCOWEN

LOCH CAIRNBAWN, NEAR QUINAG, IN SUTHERLAND, NORTHERN SCOTLAND

Some believe that most of the Hebridean lavas were fed from dikes (fissure eruptions). It is certain, however, that the great centres were established before the dike swarms, since they locate the latter. It is also certain that most of the dikes are later than any lavas spared by erosion.

West Highland scenery has been shaped entirely since the Tertiary eruptions. The magnificent mountain and valley forms of Skye are cut in Tertiary plutonics.

Possible Pliocene gravels occur near Turriff, Fyvie, and in central Buchan, all in Aberdeenshire.

Pleistocene and Recent.—During the Pleistocene glaciation, Scotland functioned as a complex centre of dispersal in the great north European ice sheet. Scandinavian ice currents freely invaded England. Some districts were crossed by ice that traversed the bed of the sea bringing in shells and, in Caithness, Mesozoic erratics. Glacial erosion is often pronounced. Crag and tail is developed to perfection in Edinburgh and elsewhere. Rock basins are numerous in the Highlands (*e.g.*, Loch Coruisk). Some of the consequences of glacial deposition are: in the west Highlands, hummocky moraines; east Highlands, fluvioglacial gravel; mid-Lowlands, boulder clay, either flat or in drumlins, and gravel kames. In Glen Roy glacially dammed lakes are recorded by conspicuous strand lines; and throughout most of eastern Scotland glacially diverted rivers can be traced by channels now left dry.

Raised beaches up to about 100 ft. occur around Scotland but not in the Outer Hebrides, Orkneys, and Shetlands. The higher beaches are of late glacial age and were locally interfered with by glacial readvance (Loch Lomond and Mull). The most marked raised beach form is that known as the early Neolithic which varies in height from 0–35 ft. but is commonly found at 30 ft. and exhibits a temperate fauna. It rests in places on a peat or forest bed that continues below sea level. Traces of early man are found in this type of raised beach. (E. B. B.)

2. Relief and Drainage.—Scotland is often divided into three main landform regions: (1) the Highlands and islands, with the east coast Lowlands; (2) the mid-Lowlands, which include hill masses roughly between the boundary fault lines Stonehaven-Helensburgh-Bute and Dunbar-Girvan (the latter presents a much less clear boundary topographically than the former); (3) the Southern Uplands including the dales and fringing Lowlands, notably those fronting the Solway Firth and in lower Tweeddale. In each of these three divisions are marked distinctions of scenery between west and east. Yet throughout the landforms are dominated by a series of plateau surfaces dissected by erosion, especially in the west and markedly on the higher surfaces. In the mid-Lowlands and around the Southern Uplands, the plateaus rise in three main steps to summit planes of 1,500–2,000 ft. (450–600 m.) in the hills. Somewhat comparable steps are seen in the northeast and the Moray Firth Lowlands, but in the Highlands the sum-

mit planes are commonly 2,500–3,500 ft. (760–1,060 m.) and the monadnocks of Ben Nevis and the Cairngorms rise above this level to exceed 4,000 ft., while lower Highland plateaus and some valley floors may correspond with the higher surfaces of central and southern Scotland. D. L. Linton suggests that by the Cretaceous Period a low monotonous peneplain had developed cutting across various older rocks, upon which a thin cover of chalk was deposited; Tertiary uplift connected with the Alpine orogenesis raised the rocks, especially in the west and between Ben Nevis and the Cairngorms where they are higher. The main original west-east drainage pattern developed at first on the chalk, from which it was superimposed onto the diverse rocks below. There was much subsequent adjustment to underlying structure and further changes in the pattern caused by the Quaternary glaciation. The thin chalk being stripped off, the surface beneath was exposed at heights varying from 1,500 ft. in the extreme northeast to 3,500 ft. in much of the west and with a maximum of 4,300–4,400 ft. from Ben Nevis to the Cairngorms, the evidence in the present landscape being residual summits in watershed areas. Much more work is needed to determine the possible relationships and origins of the lower surfaces already noted. The Minch and the Irish Sea may have been formed by rift valley downfaulting, associated with the Tertiary basalts of Skye, Mull, and Antrim. The short, vigorous west-flowing streams and higher precipitation in the west have led to greater dissection of the erosion surfaces there. The main centres of Quaternary ice dispersion were in a westerly belt from Loch Torridon to Loch Lomond and again in Galloway, where the landscape is composed of spectacular isolated peaks, deeper corries, and wilder valleys than farther east. From this belt glaciers fanned out, altering valleys, sometimes breaching watersheds and changing the preglacial drainage pattern, gouging out many ribbon lakes, straightening and deepening the eastern glens, and depositing great sheets of boulder clay, sometimes in drumlins, on the Lowlands of the east, the mid-Lowlands, and lower Tweeddale. During the glacial phases of the Pleistocene Period the weight of the ice caused depression of the land. During the warm interglacial phases and in the post-Quaternary Period coastal submergence followed the ice melt as sea level rose; simultaneously isostatic readjustment occurred as the weight of the ice diminished on the land surface. The resulting slight uplifts produced raised beaches at several levels up to 100 ft. above mean sea level, some covering marine-cut rock benches probably formed during interglacial periods. Melt water from ice flattened many valley bottoms and temporary lake beds, carved new river channels which still form gorges, and left many dry channels and large spreads of sand and gravel. Since glacial times, normal erosion has continued to modify the landscape; while sea erosion modifies the coast lines, undercutting headlands, particularly where exposed to long-fetch Atlantic waves as around Cape Wrath and in Orkney, and builds beaches, dunes, and locally cuspate headlands.

Of the postulated original drainage pattern, the west-flowing rivers, though remaining relatively short, have enlarged their catchments eastward of the main western line of peaks. Glacial watershed breaching has also been important in forming through valleys and low passes. The postulated main east-flowing rivers now form a river network showing much adjustment to underlying structure, in places aligned northeast to southwest with the Caledonian trend, *e.g.*, the River Spey, and in places at right angles to the Caledonian trend. Similar adjustments to structure may be seen in the west coast river valleys and fjords. Glaciation has greatly influenced the evolution of the present drainage pattern. Glacial breaching of watersheds has altered the course of many a preglacial river and in some cases facilitated river capture. For example, it is postulated that because of glacial breaching of the watershed the postglacial Avon River captured the waters of the upper Don.

3. Climate.—Scotland naturally shares with the rest of the British Isles a cool west maritime climate, largely dominated by the repeated passage of easterly moving depressions and occlusions of the circumpolar front. The western fringe of the mainland and the Hebrides have mild though wet and windy winters with January mean temperatures of 40°–42° F (4.4°–5.5° C) and subtropical plants grow in sheltered gardens as far north as 58° N. The frost-free period in the west is about a month longer than in the east. The east coastal belt is subject to rather more continental influences and is colder (January means about 39° F; 3.8° C) and drier in the winter though these months are the wettest of the year. It is still colder inland, by a degree or two, and more on high ground or in frost hollows, and there is more than 30 days' snow cover above 2,000 ft. (610 m.). Spring and early summer are often relatively dry though rather cool (April means 43°–45° F; 6.1°–7.2° C). From midsummer to late summer conditions are often fairly cool, wet, and cloudy with monthly means 54°–55° F (12.2°–12.7° C) in the extreme north and 58°–59° F (14.4°–15.0° C) in the centre and south. The sunniest areas, in the east, receive 4½ hr. of sunshine per day in August, compared with 6–6½ in June. In the western coastal tract and in the western half of the mid-Lowlands mean annual rainfalls are about 40–50 in. (1,000–1,270 mm.); in hilly areas over 60 in.; in mountainous western tracts over 100 in. (2,540 mm.); and in the eastern Highlands 50–60 in. Mean annual rainfall is lower in the east coastal areas and the eastern half of the mid-Lowlands (30–35 in.), where they are not exposed to the strong rain-bearing westerlies from the Atlantic. The extreme east is remarkably dry with under 25 in. on the coast itself where spring droughts sometimes occur.

River flow fluctuates seasonally though there is rain all year; average winter flow with high precipitation and low evaporation is commonly 15 times that of summer, which comes after the drier spring and early summer and in conditions of high evaporation. A winter flood is often 300 times the low stage of a summer drought, though some of the most destructive floods have occurred after exceptional summer storms following a wet spell. This seasonal fluctuation raises problems for hydroelectric generation and also for water engineers, though Scotland in general has ample water supplies. The large area which receives more than 50 in. of rain in over 70% of years includes much of the hill country within the mid-Lowlands, and of the Southern Uplands and the west Highlands. Useful catchment areas are available both to the east coast cities, which are in the zone with less than 30 in. in at least 30% of the years, and to Glasgow and the industrial west, which are in an intermediate rainfall zone. Even in accessible areas considerable potential supplies remain untapped. (A. T. A. L.)

4. Vegetation.—The flora of Scotland is much affected by the varied climate to be found in so small a country, by the varied geology and physiography, and by the activities of man. The Atlantic climate of the west coast and islands is warm and wet, and the rocks are mainly poor and acidic. On the general covering of peat with comparatively little glacial drift beneath it grows a sour herbage of sedges with some heather and such northern shrubs as bearberry, crowberry, and blueberry. An exception is a narrow strip along the west coast of the Hebrides and some of the mainland, where shell sand of high basic content has been washed up by the ocean and blown inland; behind the marram-covered dunes is a flatter expanse of sandy soil with stable grassland and many lime-loving flowers. Tree growth is sparse on the western coasts where winds are excessive in frequency and strength, but where there is shelter birch woods grow readily, with mountain ash and willow interspersed. Vestigial oak and pine may also be found along the fjordlike sea lochs of the mainland. The tree line is low in Scotland, rising from 500 ft. (150 m.) near the west coast to around 2,000 ft. (600 m.) in central Scotland. The west Highlands and islands show vestiges of a Lusitanian flora, for example, pale butterwort, dwarf cicendia, and the moss *Myurium hebridorum*.

Glacial drift is commoner across the Grampian Hills, toward the east, and in the Southern Uplands, so heather is much commoner on the still acidic moors and may attain to an almost unbroken stand. These heather moors are an example of arrested succession, for if left to themselves they would gradually revert to woodland of birch and Scots pine, with oak in the better situations. Mankind has left very little pristine oak and pine forest in Scotland, and such stretches as remain are subject to earnest efforts toward conservation. The moors are grazed by sheep and are burnt periodically. This form of husbandry keeps the heather young and

TOM HOLLYMAN FROM PHOTO RESEARCHERS

(Left) Out shooting in the Highlands. (Above) A salmon landed from a Scottish stream

prevents regeneration of forest, but a consequence in the west is the serious spread of the bracken fern. Sheep and controlled burning on heather moors favour the density of red grouse, a valued object of sport. The upper 1,000 ft. of the hills of the Highlands, rising to over 4,000 ft., may be said to carry a more or less pristine flora of boreal mountain grassland with many saxifrages and such miniature shrubs as *Loiseleuria* and dwarf willow. Areas of serpulite grit, including Ben Lawers (3,984 ft.), have a rich Alpine flora. Some of the eastern hills of the Southern Uplands carry a grassy flora which under sheep and burning is moving toward a predominance of matgrass.

In both the Highlands and the Southern Uplands a good deal of land is being afforested, mostly with exotic conifers. Scotland, early isolated from Europe, has but three species, Scots pine, juniper, and, uncommonly, the yew. Larch was introduced about 1725, followed by Norway spruce and many west coast species from North America.

The eastern side of Scotland is highly farmed, much land being on the Old Red Sandstone; natural grassland and untouched places are few, the sea cliffs and some estuarine situations being the only sites which may be said to carry a natural flora. The central plain is in the same case, but there are a few bogs here which hold distinctive plants such as the low evergreen shrubs, marsh andromeda, and Labrador tea. It is still not unusual to find plants new to Scotland which are common in the subarctic.

5. Animal Life.—The fauna of Scotland is an enviable one for so small a country. The largest wild mammal is the Atlantic gray seal, which is more numerous off Scottish coasts than anywhere else. The Island of North Rona, less than half a square mile in extent, lying nearly 50 mi. (80 km.) NW of Cape Wrath, has a stock of thousands at the autumn breeding season. The common seal is fairly generally distributed and locally numerous. The red deer is truly wild in Scotland, one-eighth of the whole country being scheduled as deer forest, and the species ranges over another 1,000,000 ac. (404,700 ha.) of sheep ground. The total number may be almost 150,000. Roe deer are found wherever there are woods or scrub. Feral goats (goats that have gone wild) occur in many deer forests and on a few small islands. Foxes and badgers are common in Scotland. Both species prey extensively on the rabbit, an introduced animal, which, as far as the northern Highlands are concerned, was introduced as late as 1845. The wildcat has increased and widened its range in the first half of the 20th century.

Otters are common on the river systems, and on small islands they may live wholly on sea fish. The polecat is extremely rare and for some time was considered extinct. Pine martens are probably more numerous than at the turn of the century and may be expected to increase with reafforestation. Nevertheless, there are more pine martens in treeless northwest Sutherland than elsewhere in Scotland. The mountain hare is distinctly cyclic in population and is a pest on hill-sheep ground in peak years. The brown hare occurs where there is true soil and agricultural husbandry, not on the hills and peat moors.

The early isolation of various island groups has led to the evolution of distinct types of some animals, notably the voles. Differentiation has gone far enough for subspecific rank to be accorded. Thus, there are the Orkney, Raasay, and Mull subspecies of field or bank voles. The same phenomenon is evident in some birds. There are St. Kilda, Hebridean, and Shetland wrens and Hebridean starlings and thrushes. The Island of Soay in the St. Kilda group holds a pure stock of a primitive, brown, short-tailed sheep of mouflon type. The original animals are thought to have been of Neolithic origin.

Scotland lost the brown bear in the 9th century, the elk and the reindeer in the 12th, and the beaver in the 16th. These animals could not subsist adequately in Scotland now for the vegetation has so greatly changed. The last wolf was killed in Inverness County in 1743.

The bird life of Scotland is rich, though many species found farther south do not occur. Its outstanding features are the numerous seabird cliffs thickly populated by auks, fulmars, and gannets; the existence of more than 100 pairs of golden eagles; the breeding colonies of the night-flying Leach's petrel on four small, remote islets; and the sight and sounds of greenshanks and black-throated divers in the deer-forest country. Graylag geese still breed in the Hebrides, and certain places such as Loch Leven are focal points in the wintering of other northern geese.

Most Scottish rivers are run by salmon and sea trout, and many lochs hold distinct races of char. Pike are in many river systems by accident or design, but the muddy-water fish common to English rivers are for the most part absent from Scotland.

Biological exploration of the invertebrate life of Scotland is far from complete, and the life histories of many insects already known have not been determined. Summer visitors to the west Highlands appear intensely aware of the tiny biting midges.

(F. F. DG.)

II. GEOGRAPHICAL REGIONS

1. The Highlands and Islands are areas concerned predominantly with extensive pastoral farming on hilly rough grazings, with relatively small proportions of improved land, partly arable, on lower ground. Within this regional unity, however, the occurrence of subregions reflects the diversity in climate, scenery, and economy. Factors contributing to the variety are, for example, the incidence of Gaelic speech, crofting, fishing, and weaving. There are intimate relations with the richer country of the eastern Lowlands, particularly north of Glen More.

The windswept archipelagoes of Orkney and then Shetland (*q.v.*) lie northeast of Duncansby Head in Caithness. The Shetlands possess an astringent charm, with their low schistose hills amid fjords only three hours by air from Renfrew, their many crofter-fishermen with lingering cultural affinities with Scandinavia rather than Gaeldom, their slender resources of declining fishing, of tourism, and knitwear using fine local wool. The Orkney Islands (*q.v.*) have warmer soils formed from the Old Red Sandstone; the people are quick to adapt their small farms to modern techniques with their reseeded clover pastures, large beef-cattle and egg exports, and holiday catering.

The Outer Hebrides extend from Lewis-and-Harris to Barra. Gaelic is still spoken by many of the crofters. The islands are low and windswept with ice-scoured, hummocky, gneissic hills and peat-capped boulder-clay hollows partly stripped for cultivation. The good pastures on the machairs of windblown shell sand on the western shores are still in places periodically plowed in strips, allocated to crofting families by lot, and herds are pastured in summer on the high shielings. Vital subsidiary occupations are herring fishing, which is on the decline; tweed making, increasing in Harris and elsewhere; lobster catching for export by air; seaweed harvesting around Lochmaddy for processing into cattle food; and production of fertilizers.

The Inner Hebrides from Islay to Skye are somewhat similar but with distinctive mountains: the Cuillin Hills of Skye, dissected glaciated bosses of gabbro, and Ben More in Mull, a much-eroded volcanic crater complex. There are relatively good pastures on the basalt lavas of Skye, Mull, and western Morven. Mean annual rainfall rises from 40 in. (1,000 mm.) to over 80 in. near the hills, and the limit of cultivation is about 400 ft. (122 m.). North of Jura the islands and adjacent mainland areas have sporting estates and large sheep farms, and again many crofting townships, focused on well-drained, easily tilled raised beaches, but largely inhabited by older people and epitomizing the many problems of crofting. The main problems are overcrowding on small coastal holdings following the clearance of the inland glens for sheep, too few truly economic family holdings even on government estates bought for resettlement, and paucity or failure of subsidiary occupations like fishing, whisky distilling, tweed making, aluminum refining, and light engineering. Forestry Commission plantations, sporting estates, tourism, and air services have helped little, and as yet roads and public services are difficult to finance for an aging population in an area of low ratable values; but it has been possible to provide several small hydroelectric plants to serve local populations. The southern islands, with most of Argyll, have enclosed stock farms rather than small crofts, with good arable stock farming and dairying on the lower ground, yet emigration remains high; Islay lost half its population between 1880 and 1950. Air routes serve Campbeltown and Tobermory.

NOEL HABGOOD FROM PHOTO RESEARCHERS

THE HILLS OF TROTTERNISH, ON THE ISLE OF SKYE, SEEN FROM STAFFIN

Roughly bisecting the Highlands and controlled by a lateral fault line, Glen More runs northeast from Fort William to Inverness, strung with glacially gouged lochs, including Loch Ness. North of Glen More stretch the Northwest Highlands, with roughly longitudinal belts of distinctive scenery. A western coastal fringe consists of low hummocky Archean gneiss backed by stepped precipices of brown Torridonian sandstone, dissected into mountains sometimes capped by white Cambrian quartzite. Farther eastward an intermittent belt of white Cambrian limestone stretches from Durness near Cape Wrath to Strath Oykell, topped by green turf, followed by the complex belt near the Moine Thrust plane of the Caledonian orogenesis, including some notable peaks like Ben More Assynt (3,273 ft. [998 m.]), and a centre of ice dispersion in Quaternary times, as witness many ice-gouged ribbon lakes (Lochs Shiel, Fannich, etc.) and fjords (Lochs Duich, Hourn, etc.). The great central Moine schist belt stretches from Strath Halladale to Glen Urquhart and in the west from Whiten Head to Morven, with its sombre level moorlands, occasional monadnock hills, and softer glens often trending east, northeast, or southeast, with a little residual oak wood, cultivation ridges antedating the clearances, a little arable for large sheep farms of up to 30,000 ac. (12,100 ha.), and in places Forestry Commission plantations. Finally, the conical, heathery Old Red Sandstone conglomerate hills of the east merge into the Lowlands covered in warm red boulder clay and the coastal raised beaches with lighter soils. There the mean annual rainfall of as little as 25 in. (635 mm.), high sunshine especially from March to June, and long summer daylight all favour arable stock farming, with a little barley for sale, mainly on small family farms. These Lowlands are an area of greater development generally, with much of the road and rail network, the ratable value, and the population, and include the towns of the northern mainland, from Wick and Thurso through Helmsdale and Dornoch, Dingwall, and Fortrose to Inverness, the regional capital.

South of Glen More lie the Grampian Highlands, with many individual massifs such as the Monadhliath Mountains and the Cairngorms. The highest peaks overlook dissected plateaus of about 3,000 ft. which cut across the northeast to southwest trending schists of the Moine series in the north and the Dalradian series in the south, both of which were tightly folded in the Caledonian orogenesis. Again glens trending east, northeast, or southeast divide the country into blocks, sometimes of distinctive rock type, and generally more dissected into conical hills in the wetter west. The conical peaks and dark andesite lava crags of Glencoe contrast with the granite massifs, generally rounded but sometimes deeply corried as on Ben Nevis (4,406 ft.; 1,343 m.) with its crown of andesite and deep northern corries; shapely Ben Cruachan, whose granite extends under the peat and boulder clay; the wilderness of bogs and lochans of the Moor of Rannoch; the Cairngorms, Lochnagar, and the Deeside Hills with wide stretches over 4,000 ft., dropping in precipices to the northern corries; the smaller Ben Rinnes, standing between the moors and the Moray Lowlands. Again two lines of quartzose grit hills, often conical, stretch from Beinn a Ghlo and Schiehallion to the rounded Paps of Jura, and near the Highland edge from Ben Venue to Ben Chonzie and beyond. From the high precipitation area in the west, Quaternary glaciers gouged outward, forming sea-lochs, *e.g.*, Fyne, Leven, and freshwater lochs, *e.g.*, Awe, Lomond, Rannoch, before sweeping, still maintaining their powers of erosion and deposition, into the mid-Lowlands; farther east glens were broadened but high plateaus

protected by névé were less affected by ice erosion, and lochs are fewer and smaller. Large sheep farms using Border breeds, Blackface and Cheviot, have dominated farming there since the change from black cattle export in the 18th and early 19th centuries. Pastures have largely deteriorated, because of the invasion of bracken in the low winter pastures and the scarcity of shepherds for isolated hirsels (one man's charge of sheep and hill). The Forestry Commission owns much land, especially in the west, while the east, less dissected, drier, more heathery, has deer forests and grouse moors also deteriorating somewhat owing to overburning. Few medieval castles survive, but there are many 18th- and mock baronial 19th-century mansions, and estate villages like Blair Atholl, Pitlochry, or Tomintoul (at 1,160 ft.); some have grown into small towns. Tourist trade came early to towns accessible to Glasgow like Rothesay and Dunoon, but in the mid-20th century it was growing in centres like Newtonmore and Kingussie, Oban and Fort William. Main regional centres like Inverness, Perth, and Aberdeen lie rather outside the Highlands near the mouths of major straths; but Inverness is so much the capital of the north that it is included in this Highland region. It lies at the mouth of the deep, clear, swift River Ness, its late-18th-century castle and barracks overlooking wide clean streets, many churches, and peripheral factories.

2. The Northeast Region comprises the eastern Lowlands from Stonehaven to Inverness, largely influenced by Aberdeen. This region mainly forms a tract of arable stock farming, often on small family farms, with intimate relations with Highland country farther west, via small to medium-sized market towns; fishing and boatbuilding, distilling, and the holiday trade are of varying importance.

East of Inverness stretch the dry, sunny Moray Firth Lowlands with family farms practising variations on a six-course rotation, including short grass leys, but mainly concerned with arable stock farming for beef production. In the west, warm, fertile Old Red Sandstone boulder clays yield some barley often sold to whisky distilleries; the 9,500 ac. (3,845 ha.) of the Culbin sands, where about 1670–95 an estate was overwhelmed by blown sand, are now reclaimed by the Forestry Commission. East of Buckie on colder boulder clay over Highland rocks there is increased emphasis on fodder and Aberdeen Angus cattle, continued on the bleak, windswept Buchan Plateau, the northeast shoulder of Scotland, where the rotation is maintained in somewhat harsh conditions. Some coastal settlements are still active in fishing (Lossiemouth, Buckie, Macduff, Fraserburgh), a few in boatbuilding; Peterhead also processes Norwegian granite, while local quarrying is declining, and others depend more on holiday trade (Nairn, Findhorn, Burghead, Portsoy). On the coast or a little inland is a line of county and market towns with solid stone buildings lining wide streets (Nairn, Forres, Elgin, Fochabers, Banff, Turriff) and inland again a further line (Aberlour, Dufftown, Keith, Huntly) on a harsher but green plateau at 1,000–1,200 ft. with many individually held crofts, rearing beef calves sold young for fattening to Lowland farms. The hinterland of Aberdeen, the regional capital, includes the lower Don and Deeside high-grade arable and stock-farming region and markets like Banchory and Inverurie; but it embraces the whole north of Scotland for university education. South of Aberdeen the Kincardine Plateau resembles Buchan; while on the coast declining fishing villages are turning to the tourist trade. Stonehaven, the county town, lies almost on the Highland boundary fault, a reminder in turning to the mid-Lowlands region that Kincardine, like several other counties and other man-made regions, straddles the geological and landscape boundary and does not coincide with the natural region.

3. The Mid-Lowlands, containing three-quarters of Scotland's population and much of the urban and industrial scene, nevertheless include quiet arable-farming landscapes. In the east farms are oriented predominantly to cash cropping with stock fattening, in the west more to dairying. The many uplands included in the mid-Lowlands are used for hill-sheep farming, reservoir catchments, and sporting and recreational space. The urban scene in the west is dominated by the heavy industrial complex, built up around shipping and shipbuilding. The townscapes of Glasgow and the surrounding towns contain almost half the country's population. In the east is the capital Edinburgh, and considerable though lighter industry, the growing coalfields of Fife and Midlothian, and several sizable seaports. The mid-Lowlands is a core region, extending its sphere of influence in many ways over the whole country, and within which lies much diversity and contrast.

Southwest of Stonehaven lies a fertile vale, Howe of the Mearns and Strathmore, with arable stock farms producing some cereals on warm mixed Old Red Sandstone boulder-clay soils, with soft fruit on fluvioglacial sands around Blairgowrie, and a line of market towns, Brechin, Forfar, and Coupar-Angus, with textiles or other small industry. Eastward rise the andesite lava hills of the Sidlaws, falling to a dry, fertile coastal shelf of Old Red Sandstone boulder clay, with raised beaches and the cuspate headland of Buddon Ness, the shore fringed by fishing and holiday towns (Inverbervie, Carnoustie, Arbroath, and Montrose) with small industries including jute. The Sidlaws' southern slope to the Tay gives views over the firth to the compact centre and sprawling suburbs of Dundee, a city of jam and jute, once an important centre for whaling and now for government-encouraged light industries which are supplanting the formerly dominant jute industry.

From Strathmore similar vales run southwest to the Firth of Clyde, between the conical, heathery, Old Red Sandstone conglomerate hills along the Highland edge, and the line of moorland hills (1,400–1,700 ft.) of lava, continuing the line of the Sidlaw, Ochil, and Kilsyth hills, Campsie Fells, and Dumbarton Hills. Arable stock farming with cereals gives way to increased fodder production and dairying as mean annual rainfall increases westward from 30–35 in. to 45–50 in. Flanders Moss, however, remains partly unreclaimed from deep peat. Towns are often at the mouth of a Highland glen or strath tributary to them. They can be grouped as market towns and resorts (Callender, Crieff); former royal cities, remaining regional capitals (Perth, Stirling); residential towns for schools and retirement (Dunblane, Bridge of Allan); dormitories for Glasgow (Strathblane, Milngavie, Lennoxtown). On the Clyde are Helensburgh, a resort and dormitory town, the outport of Port Glasgow, and Dumbarton with shipyards and engineering beneath its ancient castle crowning volcanic crags. By the Leven flowing from Loch Lomond are textile-finishing centres and an industrial estate. Both banks of the dredged upper Clyde estuary are lined with famous shipbuilding and engineering towns; Clydebank builds ships and manufactures sewing machines. The rolling drumlins on which Glasgow is built overlie the fringe of the major syncline of Carboniferous rocks. In these rocks are the coal and blackband ironstone deposits which were the foundation of the heavy industry of the whole region. Coal production is still considerable though declining and the ironstone is no longer important; exhaustion of these resources raises many problems of transfer of redundant mining population following the reorientation of industry (see *Economy*, below).

Glasgow, with one-fifth of Scotland's population, is the commercial and industrial, though not the political, capital; its many industries include engineering of almost every kind. It is equally significant as the focus of the congeries of industrial towns comprising an eastern group (Kirkintilloch, Airdrie, Coatbridge, Motherwell, Wishaw, Hamilton) and a western group (Barrhead, Paisley, Johnstone). The eastern group is engaged in coal, iron, and steel production and heavy engineering; the western group in the manufacture of cotton thread, preserves, and sanitary ware, general engineering, and the left-bank shipyards. More than a dozen industrial estates now complement the heavy industry with varied manufactures ranging from lingerie to aero engines. The conurbation strongly influences coastal resorts in Argyll, Bute, Arran, and Ayr; while Prestwick has an international airport, seldom fog-bound. Ayr, Troon, Ardrossan, and Girvan, however, depend on the Ayr coalfield as the basis of their coal exports and industries. Production from this coalfield is still expanding, and it has its own large and small industrial towns (Kilmarnock, engineering and whisky; Darvel, lace and carpets). Links with Glasgow are strong even in farming. This is reflected in the growth of the dairy industry, including Dunlop cheese originally from remote farms, and the related development of Ayrshire cattle. These

cattle yield well from only moderate-quality improved upland pastures. Trade has developed in pedigree and attested tubercle-free stock, in bacon production, early potato crops, and strawberry growing on the light soils and in the mild climate of the west coast raised beaches. Similarly, across the volcanic heights, the central Clyde Valley produces soft and tree fruit, tomatoes, and greenhouse products from Strathaven to Carluke; upstream is the Abington-Lanark-Lesmahagow dairy belt, where cheese making has moved from remote farm to factory during the 20th century.

North of the coal towns of Coatbridge, Airdrie, and Armadale lies the peat-covered Slamannan Plateau, with abandoned coal workings and declining small towns which the introduction of light industry is reviving. North again the foundries of Falkirk, Carron, Larbert, and Bonnybridge produce domestic stoves, grates, etc. Grangemouth, a rising Firth of Forth seaport, has well-developed oil-refining and chemical industries originally based on the almost extinct oil-shale mining of the Bathgate area on the east of the Carboniferous syncline; petroleum is now imported by tanker via pipeline from Finnart on Loch Long. Eastward, Bo'ness is a small timber port once serving Linlithgow, an ancient royal town. Across the Forth, on the north of the coal basin lies Alloa, a small port with brewing, glassmaking, and engineering, and at the foot of the Ochils lie the small paper towns of Alva and Tillicoultry. Farming persists between industrial centres, dairying on the low ground, and sheep on the uplands.

South of the Firth of Tay, Fife has a series of contrasting east-west belts; a coastal fringe of good arable land, with dormitory and resort towns for Dundee; the lava hills of the eastern Ochils; the dry and sunny Howe of Fife, its boulder-clay loams and fluvioglacial sands under neat arable farms yielding wheat, barley for Markinch distillery, sugar beet for the factory at Cupar, and soft fruit and peas for quick freezing. This pattern extends to the East Neuk where ancient St. Andrews has Scotland's oldest university and the famous coastal golf links, while fishing villages cater for holiday traffic and have some boatbuilding. Farther south are lava hills in the Lomonds and the ridge east toward Fife Ness, and a third line of lavas, the Cleish Hills, and the ridge toward Kirkcaldy. The western hills and the Ochils hem in the Kinross Basin, with the famous angling of Loch Leven amid quiet farming country akin to that of the Howe of Fife. Under the coastal Lowlands fronting the Forth is the East Fife coalfield, with early exploitation around Cowdenbeath and Lochgelly behind the coal ports of Buckhaven, Methil, and Leven, but with large reserves in the east where a sharp syncline continues the Midlothian field. Important expansion in the 1950s was marked by the new town of Glenrothes, but later severe mining difficulties were encountered. Heavier boulder-clay soils are in short grass leys, with increased dairying stimulated by nearby industrial and mining towns. Dunfermline, with a former royal palace and ruined abbey, is a linen centre now turning to man-made fibres, while Kirkcaldy's coarse linen trade is dwarfed by the modern linoleum manufacture comprising most of Britain's output. Burntisland has a small shipyard and produces alumina from imported bauxite for refining at Kinlochleven. South of the Forth, Edinburgh Castle on its volcanic plug controlled a gap of only a few miles between the Pentland Hills and the port of Leith; the city retains residual functions as capital, along with varied services and industries. It spread rapidly in 1920–60 from a compact nucleus of distinctive character, now filling the Lowland gap. On nearby raised beaches and Lowland plateaus are market gardening and intensive farming for cereals, potatoes, and fodder largely for dairy herds. The breezy Pentlands give recreation space and water reservoirs; east lies a coalfield amid farming country, still expanding around Dalkeith and Gorebridge; deep seams in a sharp syncline extend to Tranent in East Lothian. Penicuik, Bonnyrigg, and Musselburgh on the Esk manufacture paper, carpets, and wire. East of the coalfield the dry, fertile East Lothian Lowland is intensively farmed for cereals, potatoes, and fodder for dairy and fattening cattle. Haddington, its small county and market town on the Tyne, with some woolen manufacturing, is expanding to take some of Glasgow's overspill population. Dormitory and resort towns fringe the coast; fishing is declining.

4. The Southern Uplands are dominated by hill-sheep farming on smooth rounded green hills; the terrain is wilder in the west. There are close relations with the dairying country of the dales and coastal Lowlands of the west, and the arable cash-cropping farms with stock fattening in the lower Tweed Basin. The many small market towns are important local centres of services and amenities, many having tweed mills or other small industries. The following paragraphs give broad indications of diversity within unity in this region.

Over the whole width of the country erosion surfaces cut across slaty rocks. These surfaces are more dissected in the wetter west where the moorland vegetation is greener and less heathery, though grassy hills occur throughout. Flocks are mainly Cheviot sheep, with Blackface on heather moors, also used as grouse moors. Most streams are valued for salmon and trout. Higher mean annual rainfall in the west (50–90 in.) and the more rugged glaciated topography, especially in Galloway's granite massifs, led to locally important hydroelectric development between the two world wars. The head of Annandale echoes this wilder scenery, with several small glacially gouged lochs and the Grey Mare's Tail Waterfall. Eastward the scenery softens to rounded hills, névé-protected in glacial times, and wide expanses of level moorland. The Lowlands around the western hills and the broad dales of Nith and Annan which penetrate them are largely under arable grass farming for dairying, now extending on to breezy plateaus up to 800 ft. None of the many former ports on Solway remain really active, but Stranraer on Loch Ryan is a packet station for Larne, Ire. Glenluce, Wigtown, and Castle Douglas are small market, service, and holiday centres. Dumfries, with growing industry (textiles, chemicals, milk canning), is spreading rapidly along the Nith. Annandale has several market towns and is an important routeway, linked south to the Carlisle Lowland and north over Beattock Summit (1,028 ft.) to Glasgow and Edinburgh.

On the rolling drumlins and fluvioglacial sands of the Tweeddale Lowlands, between the Southern Uplands and the Cheviot and the Border hills, are arable farms producing some barley and wheat as well as feeding beef cattle and heavy sheep like Border Leicesters. Intermediate plateaus at 500–600 ft., with widely improved grassland, retain peat mosses in places. Purely market towns exist (Newtown St. Boswells), but most also have textile industries with Galashiels as the centre. The tourist industry is important, especially at Peebles for the Sir Walter Scott country, Melrose and Jedburgh with ruined medieval abbeys, and at Kelso which is a renowned sheep market. Most towns and valleys retain intense local patriotism, a legacy of Border wars and raids. On higher ground in central Berwick lies the small quiet county town of Duns. (A. T. A. L.)

III. THE PEOPLE

In this age of quickened mobility of population and innovation in both material and nonmaterial culture, Scotland, like other countries, is tending to lose regional and local distinctiveness in its people. It would be futile to seek any close correlation of race, language, and religion over the country as a whole, but something can be learned of the stages leading toward the amalgam of today by the study of communities long resident in isolated localities. In such communities language is often distinctive, and on larger islands the community of each glen or sweeping bay line may have an accent and physiognomy slightly different from those of the next. Some Gaelic (*see* CELTIC LANGUAGES) is still spoken, mainly in the Hebrides and by older people on the northwest coast; in a Roman Catholic crofting islet in the Outer Hebrides one woman recently recorded some 900 proverbs and 450 unpublished songs and variants. In that area the people are mostly small, dark, and gray-eyed (though a few are tall, fair-skinned, and red-haired) and there are residuals of communal labour.

A journey in Scottish backwaters is in fact an experience of living history. This may be sensed even in the busy Scottish city. Within the modern suburban sprawl there is a preponderance of single-storied dwellings, unusual in neighbouring England, and local authority housing more often consists of flats. Both tendencies derive from a long tradition of living on one floor, in the inner

rings of tall stone tenement blocks, many of which remain today as middle-class flats, while others are occupied by manual workers living in more cramped conditions, and the oldest are condemned slums. Tenements are in turn descended from tall "lands or tenements" which filled up the former garden strips of a few central streets, sometimes, as in Edinburgh, within a wall. The high death rates associated with such overcrowding have fallen steeply but housing and health still lag behind southeast Britain. Tenement dwelling inculcated virtues and failings: the virtue of minding one's own business, sometimes perhaps carried to excess in the suspicion of new, even if friendly, contact—in the east at any rate; or the bitter feuds resulting from the need to share facilities and stairs. There remain well-marked, though fluid, social stratifications and a multiplicity of urban accents. In the countryside, Lowland Scots accents are again diversified, but with less marked social stratification, and there is particular separatism in mining and fishing communities.

TOM HOLLYMAN FROM PHOTO RESEARCHERS

SCOTTISH DANCER

Woven into the complex, flowing tapestry of the present are some of the following strands: the spread of the Neolithic farmers—mainly short, dark brunets with gray eyes, who by Roman times were speaking Brythonic Celtic and are commonly if controversially referred to as Picts; from the inflow of Beaker folk in the early Bronze Age there came an element of taller, roundheaded people; there may even be some genetic heritage from the Mesolithic hunters and fishers; around A.D. 500 came a movement of Goidelic Celtic-speaking Scots from Ireland, ultimately giving Scotland its name and apparently leading to the displacement of the Brythonic tongue, and bringing important elements of culture and communal practices discernible today; at the same time, Anglo-Saxon invasions brought less brunet, possibly taller people with techniques for clearing the damp, dense forest for open fields and long-furrow agriculture as in England; they probably intermingled with the Celts (there seems to be a remarkable parallelism between the farm-touns [towns] and the Kirk-toun of a Lowland parish, and the ordinary bal' [town] and the clachan [church hamlet] of a Highland parish); later the Norse invasions affected particularly the Orkneys, Shetlands, and Hebrides and the coastal belt of the west, northwest, and northeast mainland, and brought a fresh element of tall blonds or redheads and Norse as the dominant tongue—"Norn" survived in Shetland until the 19th century—as well as legacies of technology and farming practice such as the saeter or summer pasture (rendered in Gaelic as *seadir* and in Anglo-Gaelic as *shader*).

Norman feudalism, where it had operated, and, much later, the commercial, agricultural, and industrial revolutions resulted in rigid social stratification—tempered, however, by some sense of community in a harsh land where natural calamity could strike rulers and ruled alike. On the other hand the Reformation and the ascendancy of the initially rigid Calvinistic Presbyterianism brought also a tradition of independent thought and argument, with the acceptance of a man for his own worth as an elder of the kirk. It was perhaps no accident that at a time of technical change and increasing social tensions Robert Burns (*q.v.*) rose from a poor, small farming family of the west.

The growth of capitalism brought the Lowland sheep farmer into the Highlands as agent of the cash nexus following the military pacification in the 18th century. The clan was destroyed. Revolutionary socioeconomic changes brought a population increase at first, but later sharp depopulation was given speed and poignancy by the Highland clearances and the 1845–47 famine. In parts the Highlands have a top-heavy age structure, and in some sense a wounded and dependent apathy, which shows only local signs of being overcome from within the communities themselves. Tartans, a clan badge, bagpipe music, and some remaining chiefs are the last rallying points for residual clan loyalty, sentiment, and much sentimentality; overseas they are uneasy bedfellows with an imperfect knowledge of Burns as the focus for Caledonian societies. However, nowhere in Britain does the economic cooperative movement surpass that of the Shetlands, Orkneys, and Hebrides.

The labour-hungry 19th century saw wave after wave of initially poor and ignorant Roman Catholic Irish immigrants, who had vigour and charm and an ability to make good. On Clydeside their group cohesion is one origin of the passionate Celtic (Catholic) and Rangers (Protestant) football rivalries of today. Irish labour is attracted by hydroelectric plans and is acquiring skill; more recently West Indians have been taking the place of Irish on Clydeside, while Indian and Pakistani itinerant merchants are found, even in the Highlands and islands. (AR. G.; A. T. A. L.)

IV. ARCHAEOLOGY

1. Stone and Bronze Ages.—The oldest satisfactory evidence of men in Scotland is provided by Mesolithic flints. Some found along the southwest coast may go back to the 5th millennium B.C., but probably most belong to near the end (*c.* 2700 B.C.) of the period. Makers of microliths may be presumed in most coastal and lower river valleys. More is known about the Obanians whose cave shelters and large shell middens are found on the west coast and on Oronsay Island. They evidently used seagoing boats. Their antler and bone barbed points for harpoons suggest derivation from those of the Azilians (descendants of the Magdalenians) in France, though the relation of the sites to the raised beaches (see *Physical Geography*, above) suggests that they should be dated to the end of the Atlantic climatic phase, scarcely before 3000 B.C. Antler mattocks from the western middens and from whale carcasses near Stirling mark contact with the Baltic Ertebolle culture.

The incoming Neolithic stock raisers and agriculturists cleared the woodlands with their still ubiquitous stone axes. Their pottery and earthen burial mounds show penetration from Yorkshire to the Moray Firth. It is not yet known whether it was the first Neolithic colonists in the west who introduced megalithic chambered tombs of the gallery grave type, nor how long these tombs preceded the type with passage and chamber distinct, from Iberia and Eire, found in the Hebrides but most fully developed in Caithness and in Orkney, which in Maeshowe has the finest chambered tomb in Britain.

Late in the Neolithic Period, about 2000 B.C., communities of quite uncertain origin developed in England and also went north. The stone ax trade may have been in their hands. In Orkney they left the most complete 2nd millennium settlement known in northern Europe, at Skara Brae (*q.v.*). Quite different equipment and stone houses, scattered among fields, with poor chambered tombs and a large public building, survive from a later but still stone-using population in Shetland. In some close, probably dominant, relationship to these communities were the Beaker immigrants, whose characteristic pots and short-skulled skeletons occur in single stone-lined graves throughout eastern Scotland. Some of their pottery indicates an immediate origin on the lower Rhine. They reached the west before the chambered tombs were given up. If not themselves copper prospectors and traders in bronze implements, they must have promoted these activities. Some of the circles of standing stones are their sacred monuments.

From about 1500 B.C. there was a pause in immigration and a long period apparently of pastoral stagnation during which trade increased, especially with Ireland. The various peoples coalesced. Shorter pots, usually called "food vessels," replaced beakers in the short cist graves, often with elaborate jet necklaces. Then cremation became dominant, with various types of large urn found singly and in cemeteries; occasional beads of blue faïence seem to mark a burst of trade with the Mediterranean. Unfortunately

identifiable burials cease before 1200 B.C. and are almost entirely absent for the next 1,700 years or more. No dwellings of the makers of food vessels or of cinerary urns are yet known.

The swords and larger spearheads that, with socketed bronze axes, mark the beginning of the Scottish Late Bronze Age about 900 B.C. indicate warfare if not renewed invasions. While metal objects traded within the British Isles were increasing in number, variety, and size, a few ornaments in cached hoards suggest direct contacts between northeastern Scotland and the north European mainland both in the 7th and the 6th centuries B.C. A domestic bucket urn pottery, though associated with these ornaments, may have had a long earlier history in Ireland. The only Scottish Late Bronze Age settlement, with new pottery denoting immigrants, is at Jarlshof in Shetland, which has the longest sequence of houses and pottery in Britain.

2. Iron Age (After 500 B.C.).—Stockades and clusters of round timber houses underlie many of the Iron Age earthwork or stone forts in southern Scotland. They mark the beginning of the pattern of Celtic tribalism found in Roman times. Later there were numerous timber-laced stone forts, which often left vitrified remains when burned down. Multiple-rampart fortifications came later; some were just homesteads, as were artificial "crannogs" (defensive islands) in lochs. The Belgic and Roman invasions of England intensified northward population pressure, vividly illustrated by the hundreds of "brochs" (stone towers) peculiar to the extreme north and west, such as that at Mousa in Shetland. The few fortified settlements that can be called towns lie south of the Forth, notably that on Traprain Law near Haddington, which flourished mainly while the Romans held Britain.

The strategy of the Roman army which campaigned under Agricola from A.D. 80 remains unclear despite recent discoveries. Forts along the Highland line and in southern Scotland were held till A.D. 100, more than 12 years after the garrison of Britain was decreased and the consequent removal of the great timber buildings of the unfinished legionary fortress at Inchtuthil near Dunkeld in Perth. The Antonine frontier reorganization after 139 has left an elaborate system of forts, fortlets, and roads south of its Forth-Clyde Wall, built of turf behind a great ditch, and some forts north into the county of Perth. This was finally overthrown about 196 by the Caledonii and their allies. The few brochs south of the Forth, and subterranean chambers there, "earth-houses" or "weems" (most common in Angus, north of the Tay), may represent this alliance. But the frontier was stabilized, probably by buffer states, after Severus' punitive expeditions. To these are now ascribed a legionary headquarters at Carpow on the Tay, and a chain of great temporary camps reaching as far as the Moray Firth. Traprain Law provides evidence for Roman trade north of Hadrian's Wall in the 3rd and 4th centuries; but also, in the treasure of pagan and Christian silverware found there in 1919, for the raiding which accompanied the collapse of the empire.

Romanization continued after the legions left Britain; the first native inscriptions in Latin are on 5th- and 6th-century Christian tombstones. Recent excavations claim to have unearthed what may be actual buildings of the historic missionaries; of St. Ninian and St. Columba (*qq.v.*) at Whithorn and at Iona respectively. Finds from fortifications and from crannogs bear out close relations with Ireland and include 8th-century pottery from the Continent. Metalworking reached a very high standard, notably in the gold filigree animals on the silver brooch found at Hunterston on the Clyde estuary. The kingdoms of Scots and Picts (united about 843–844), Britons and Anglo-Saxons, are as yet distinguished archaeologically only by their sculptured monuments. Pictish sculpture, at first unexplained incised linear symbols and animals, and later cross slabs with scenes and ornament in relief, is a very remarkable artistic achievement, even if inspired by Anglo-Saxon and Roman art. One of the earliest and finest of the Anglo-Saxon free-standing crosses is at Ruthwell in Dumfries (8th century). The Viking incursions and colonization, marked by burials, some excavated settlements, and (in the 10th century) hoards of silver ornaments and Anglo-Saxon coins in the western and northern isles, led to a general impoverishment. The latest pre-Norman sculpture in southern Scotland is barbarous. (R. B. K. S.)

V. HISTORY

A. Early History

1. The Settlement of Scotland.—The earliest writings on the history of Scotland are the accounts of the several Roman expeditions; but archaeological evidence shows that many invasions and settlements took place in prehistoric times (see *Archaeology*, above). When the Romans first entered what is now Scotland, they found a tribal society dominated as far north as Aberdeenshire by an iron-using aristocracy, which had probably come fairly recently from the Continent. Farther to the north and west, the inhabitants retained a more primitive culture. The geographer Ptolemy gives the names of some of the tribes in his map; of which the most important were the Novantae in Galloway, the Selgovae in the Tweed Valley, the Votadini in Lothian and the Merse, the Damnonii around the estuary of the Clyde, and the Caledonii apparently along the Great Glen.

The Romans themselves left little mark in Scotland. Agricola, the governor of Roman Britain from A.D. 77 or 78 to 83 or 84, invaded the country, built numerous forts, and marched probably at least as far north as Strathmore before winning in 83 or 84 his great battle of Mons Graupius, the site of which remains uncertain. Although many of his forts were still occupied after his recall, they were abandoned before long, perhaps around A.D. 100. The defense of the northern frontier of Roman Britain was difficult, however, and the Romans returned in 142 to build the Antonine Wall (*q.v.*), a turf wall stretching from Bo'ness on the Forth to Old Kilpatrick on the Clyde, with forts every few miles, a few of which still survive. The Romans also held a few sites north of the wall, notably at Ardoch in Perthshire, and a network of road posts and other forts along the main valleys which cut into the Southern Uplands. This elaborate system, however, proved to be ineffective, and toward the end of the 2nd century the Romans withdrew to Hadrian's Wall (*q.v.*), running from the River Tyne to the Solway Firth, which was reconstructed and remained as the chief defense of the frontier for another two centuries.

Apart from punitive expeditions by the emperor Lucius Septimius Severus in 209 to 211, the Roman armies did not return to Scotland; but they did not ignore it. In the 4th century *exploratores* ("scouts") appear to have been stationed in southern Scotland though not in any settled bases; while the Romans seem to have established some sort of formal relations with certain of the native tribes. Client kingdoms appeared with their headquarters at Dumbarton and in Lothian; and although the eastern kingdom of Manau disappeared in the disturbances of the 5th century, the western remained to become the Dark Age kingdom of Strathclyde.

Little information survives about events among the native peoples of Scotland. Increasingly Roman sources mention the Picts; and it may be that this people, unmentioned by Ptolemy, were during the Roman period establishing a powerful kingdom with its centre in Strathmore, but nothing is known of how this was done, or about the state or history of the rest of Scotland.

During the 6th century, two other peoples appeared in Scotland. The Anglians of Northumbria spread into the Tweed Valley, the Merse, and Lothian, perhaps into the void left by the disappearance of Manau, after its ruler Cunedag (Cunedda) was removed to North Wales in the 5th century. In the early 7th century, the victories of Aethelfrith, king of Northumbria, extended the power of that kingdom well into the county of Dumfries, where remain a number of important Anglian carved monuments, notably the Ruthwell Cross. Meanwhile in the west, about the middle of the 6th century, some of the "Scotti" from Ireland settled in Argyll and the islands off its coast, to form the kingdom of Dalriada.

2. The Conversion.—Very little is known for certain about early Christianity in Scotland. Numerous lives were written of particular saints, such as St. Ninian and St. Columba, as well as of many others less well known; but of these only Adamnan's "Life of Columba" (ed. by A. O. and M. O. Anderson, 1961) was composed within a century of the life of the saint. Most were written long after, and, while they probably include traditions of historical value, these are difficult to disentangle from the legends.

It is, however, certain that Christianity first came to Scotland through contacts with Roman civilization. St. Ninian is said to have visited Rome and Gaul before establishing his church at Whithorn, traditionally in 397 (but *see* NINIAN, SAINT). The earliest evidence of Christianity in Scotland is a small group of tombstones inscribed in Latin, with Christian formulas and sometimes the Chi-Rho monogram, which are found mainly at Whithorn and Kirkmadrine in Galloway, with a few scattered examples elsewhere in the south of the country. It was possibly in the kingdom of Strathclyde that St. Patrick received a Christian upbringing before he became the apostle of Ireland.

It is uncertain how far this sub-Roman Christianity survived into the Dark Ages. Tradition has always given the main credit for the conversion of Scotland to St. Columba who came from Ireland to settle in Iona (*c.* 563 or 565). From there he is said to have converted both the Scots and the Picts. Bede, however, (*Ecclesiastical History*, III, 4) seems to imply that St. Ninian had already converted some of the Picts; and church dedications to St. Ninian are much more widespread than one would expect if he had confined his work to Strathclyde and Galloway. In any case, it is clear from legends about and dedications to many other saints that the conversion of Scotland was not a simple process.

Even less is known about the history of the church in Scotland in the following centuries. It is usually assumed that an early age of fervour was followed by several centuries of decline. Scotland was remote from the rest of Europe and developed with little contact with European civilization. It was slow to adopt such changes as the new method of calculating the date of Easter, formally adopted in northern England at the Synod of Whitby (663 or 664); and St. Margaret, who came from Anglo-Saxon England during the 11th century, was impressed by the antiquated rituals and customs of the Scottish Church. On the other hand, the church was still in existence. There is evidence of this in a large collection of tombstones at Govan which belong to the 9th and 10th centuries; in many carved crosses and other monuments; and in many small oratories and cells, remains of which can still be found in the remoter parts. One cannot be more precise than this; but it is likely that the church founded by the early saints survived in some form through the troubled period of the 8th to the 12th centuries to become the foundation on which the medieval church in Scotland was erected. In some areas at least, the parish churches of the 12th century were on sites which had long been in religious use. The church of Govan, traditionally founded in the 6th century, was in the 12th century and is today the centre of a large parish.

3. The Establishment of the Kingdom.—Meanwhile the political condition of Scotland was becoming more simple. In the late 8th and 9th centuries, there were widespread Norse raids and settlements in the north and west of the country. Orkney and Shetland, Caithness and Sutherland, became Norse provinces, as did the Western Isles, which became part of a Norse "sphere of influence" including the Isle of Man and the eastern seaboard of Ireland.

These attacks seem to have led the other peoples in Scotland to draw together in face of the new invader. In the 9th century (traditionally in 843, perhaps in 844) Kenneth I MacAlpin (d. 858), king of the Scots, became king also of the Picts. In the 11th century, the royal line of Strathclyde died out, and that kingdom also came under the rule of the king of Scots; while the Scandinavian settlements had also weakened English power in the north and Scottish influence increased in Lothian. It is hard to be sure when it was finally incorporated in Scotland: it may have been conceded by the Anglo-Saxon king Edgar between 971 and 975, but the defeat by Malcolm II (1005–34) of the Northumbrians at the Battle of Carham in 1016 or 1018 is often treated as the decisive event.

By the 11th century, therefore, apart from the regions settled by the Scandinavians, all the areas of Scotland recognized the same king. This unity, however, was superficial and nominal. The provinces of Strathclyde and particularly Lothian were very different in people and in history from the northern parts of the country, and were much more open to southern influences. These divisions were to become very obvious in the late 11th and 12th centuries; but they appear to some extent in the obscure usurpation and reign of Macbeth (1040–57) who came from the Pictish area and was opposed and defeated by Malcolm III Canmore (1058–93) with English aid.

The Scottish kingdom had been created by the 11th century, but the work of consolidation remained.

B. THE CONSOLIDATION OF THE KINGDOM, 1058–1286

This consolidation was the main achievement of the dynasty of kings descended from Malcolm III Canmore and his English wife Margaret (*see* MARGARET, SAINT). With only a brief interval after the death of Malcolm III himself, the throne remained in his family until the death of Alexander III in 1286. In 1097, Malcolm's son Edgar (1097–1107) was placed on the throne with the help of William II of England; and he was succeeded in turn by his brothers Alexander I (1107–24) and David I (1124–53). David's son Henry died before him, but by then the hereditary succession had become sufficiently established for David to be succeeded by his grandson Malcolm IV (1153–65) and he, in turn, by his brother William the Lion (1165–1214). From 1165 till 1286, Scotland was ruled by only three kings, William himself, his son Alexander II (1214–49), and his son Alexander III (1249–86), who succeeded as a child of eight and reigned until his death in a riding accident.

During this period, the dynasty had many connections with England, beginning with Malcolm III's own exile there during the reign of Macbeth and his later marriage to Margaret, the sister of the English prince Edgar the Aetheling (*q.v.*). It was with William II that Malcolm's sons found refuge during their short exile after his death; Henry I of England married Edith, a daughter of Malcolm and Margaret; while Malcolm's son David married an English noblewoman, Matilda, countess of Huntingdon, in whose right he was, before he became king, earl of Huntingdon. After David I, connections with England were less close, though David's son Henry, William the Lion, Alexander II, and Alexander III all married English wives.

1. Administrative Developments.—It was natural, therefore, that these kings should introduce into Scotland many practices which they had met in England. From the late 11th or early 12th century, Scottish kings issued documents in Latin, whose form was very closely modeled on the Anglo-Norman "writ-charter," and from this model there developed in the 12th and 13th centuries a variety of types of document which were in general parallel to those being issued in England, though the precise formulas were often different. Again, the sheriff or vicecomes appears in the 12th century as the principal royal official of a sheriffdom or vicecomitatus. Although there were earlier local officials with various titles, the sheriffdoms and sheriffs of the 12th century seem to have been new creations, modeled on English practice. These and other officials, as in England, had to render account for their receipts and expenditure of royal revenues before a board of auditors appointed to hold sessions of the exchequer. The earliest evidence for this is in some accounts rendered in the years 1264 to 1266, and 1288 to 1290, a copy of each of which has survived; but all other Scottish royal records before the early 14th century have been lost and this makes it impossible to say just when this system was introduced. It was certainly not new in 1264; and was very probably the normal practice from the first introduction of sheriffs in the early 12th century. At the same time, too, as the sheriffdoms were established, royal castles appear, normally in the hands of the sheriffs, in the chief towns.

2. The Anglo-Norman Families.—As well as bringing these innovations from England, the kings, especially David I, Malcolm IV, and William the Lion, brought to Scotland members of many of the Norman families of England. During the reign of David I, the Bruce (*q.v.*) family was established in Annandale, the De Morvilles in Ayrshire and Lauderdale, and Walter son of Alan (*see* STEWART) was settled in Renfrewshire. These brought with them others, from their English estates, who were settled on parts of their Scottish lands; while throughout the 12th century many lesser men, whose families came originally from a wide area of

northern France, but most of whom had first been settled in England, were granted estates in Scotland. This introduced a new and foreign nobility who were most powerful in the southern part of the country; while the native nobility, with whom they intermarried, remained predominant in Lennox, Fife, Angus, and further north. Gradually, the two groups fused, until by the 13th century there seems to have been little distinction between them.

With these English or Norman nobles, the Scottish kings introduced the feudal system of landholding; these settlers commonly held their lands, as they would have in England or France, by the service of so many knights. The Bruces, for example, were granted Annandale for the service of ten knights. But this service was owed only where a specific grant demanded it; most established landowners continued to hold their estates by the same obligations as before, such as the payment of food-rents of various kinds, or the performance of services, including military service; so that by the end of the 12th century the kings were owed a bewildering variety of services and payments by the landowners of Scotland.

By the reign of William the Lion, therefore, the kings of Scots had established a system of government which followed closely the lines laid down in other areas of western Europe. The innovations, however, aroused opposition from some of the native nobles who resented the new methods and new men introduced to Scotland at this period. On the death (1093) of Malcolm Canmore, those hostile to the English influences were able to put on the throne two kings of a rival line, Malcolm's brother Donaldbane (1093–94 and 1094–97) and Duncan II (1094), son of Malcolm's first marriage, before Edgar, son of St. Margaret, was established with English help in 1097. There were risings in Argyll led by Somerled in the reign of Malcolm IV; and by descendants of Duncan II in the reigns of William the Lion and Alexander II.

3. Extent of the Kings' Territorial Authority.—Moreover, although the kings of the Canmore dynasty had consolidated their authority, it was only effective within limits much narrower than those of present-day Scotland. The main centres of the king's power lay in his estates in Angus, Fife, and the Lothians, and it was there that the kings spent most of their time. Those Norman settlers who were to be the leading members of the Scottish baronage were, for the most part, granted estates in regions where the king's own personal authority was less. The Stewards (Stewarts) in Renfrewshire (a counterbalance to the native family of the earls of Lennox across the Clyde), the De Morvilles in Ayrshire, the Bruces in Annandale, the De Soulis in Liddesdale were, in effect, the instruments of royal control in areas where the crown could less easily rule by its own authority. By such means, the kings were able, either directly or through the Norman nobles, to control most of Scotland south of the Highland line.

On the borders of this area, however, there were a number of regions where native powers remained strong. Galloway, Lennox, central Perthshire, Garioch in Aberdeenshire were all areas in which the king's power was uncertain. William the Lion had much trouble with Galloway, and it was not until 1234 that the last of the native line of lords of Galloway, Alan, died, leaving three heiresses who were all married to Norman barons and thus introduced Norman influences. Even so, the area remained of uncertain allegiance as late as the wars of the 14th century. William the Lion, too, tried to extend his power at the other end of the Highland line by making his brother David earl of Huntingdon, lord of Garioch. Further to the west and north, the islands as well as the mainland of Caithness and Sutherland remained in the hands of the Norse. William the Lion made an expedition to Caithness in 1196, in which he asserted his authority over the Norse settlers. Alexander II and Alexander III both campaigned in the Norse areas, as a result of which there occurred the celebrated Battle of Largs (1263); and these efforts resulted in 1266 in the treaty by which Magnus V Lagaboeter, king of Norway, surrendered to the king of Scots all Norse possessions in the Western Isles except Orkney and Shetland.

Thus, by 1266, the kings were in theory masters of the whole mainland of Scotland; but this control was nominal in the bulk of the central Highlands, and still more so in the areas ceded by the Norse. The Western Isles in particular remained for all practical purposes independent; and under the Macdonalds, who later became lords of the Isles, acted as a separate power which entered into independent relations with England at various points during the wars of the 14th and 15th centuries.

4. Growth of the Church.—The kings, however, did not confine their innovations to the government of Scotland: they also made great changes in the church, which in St. Margaret's day (11th century) had fallen out of touch with many of the developments which had occurred in the Western Church. There were bishops, but, perhaps naturally in the still loosely organized political state of the country, no sign, so far as the evidence goes, of a clearly organized diocesan system. There seem to have been monks of a sort, or at least groups of clergy living together; some of which, mainly those in the Pictish parts of the country, are known to tradition as "Culdees" (*q.v.*) though there seems to be little difference between these and groups in other parts of the country, as at Govan, which are not so called; but regular monasticism according to the rule of St. Benedict was unknown.

The rule seems to have been introduced to Scotland by St. Margaret herself in her foundation at Dunfermline; but elsewhere in Europe many new religious orders were being founded; and, since regular monasticism was introduced so late to Scotland, it was natural that strictly Benedictine houses should be comparatively rare. Alexander I founded two houses of Augustinian Canons, at Scone and Inchcolm, while David I made foundations of the reformed orders, such as the Cistercian houses of Melrose and Newbattle, the Tironensian (from Tiron, in Picardy) house of Kelso, and the Augustinian house of Cambuskenneth. By the end of David I's reign Scotland, from being almost devoid of monasticism, was rich in houses.

The founding of monasteries, however, was expensive: it required large areas of land with which to endow the new house; and it was at first only the kings who were able to indulge in it. Therefore, the majority of Scottish houses were established in the areas where the kings themselves possessed great lands—Angus, Fife, the Lothians, and the Tweed Valley; and it is often suggested, following a comment attributed to James I that his predecessor was a "sair sanct for the crown," that David I's generosity to the church seriously impoverished the crown. This was probably not a fair comment, since signs of this impoverishment do not appear till much later, perhaps not even till the early 15th century when most European monarchs were finding themselves hard pressed for money; and for a long time David and his successors got a very good return for their investment in the form of services of all kinds performed by churchmen in the royal administration, both local and central.

Few of the nobles could afford to follow suit. There were a few important houses founded by Norman settlers, notably Dryburgh (Premonstratensian Canons) founded by Hugh de Morville in 1150, and Paisley (Cluniac) founded by Walter son of Alan (founder of the Stewart family) about 1163; but most baronial foundations were late and small. Hardly any of the native nobility founded houses of importance, unless, as is probable, the abbey of Dundrennan (Cistercian), founded about 1140, was a foundation of Fergus of Galloway. The establishment of monasteries was almost entirely an activity of the foreign-minded kings and their foreign associates.

During the 12th century also the organization of the church kept pace with developments in the secular administration of the country. As the state became organized into sheriffdoms, so the church developed a much more regular diocesan system. By 1128, there is record of bishops of St. Andrews, Glasgow, Dunkeld, Moray, and Ross; by the end of the 12th century most of the dioceses of the medieval Scottish Church were already in existence, and there is an increasing amount of evidence of the development of cathedral chapters, who were gaining the right, formally at least, to elect the bishops, and of lesser units such as archdeaconries. It is also clear that parish churches were being more regularly established and brought under the control of the bishops. Unfortunately, a very large proportion of these were "appropriated" to monastic houses in the first enthusiasm for the regular

orders; this meant that the house drew the income of the church and became responsible for maintaining its services, which it usually did by appointing a lowly paid vicar. It was not until the 13th century (as a result of the decrees of the Lateran Council of 1215) that the bishops began to make attempts to regulate the appointment of these vicars and to fix minimum stipends which they had to be paid, often in the face of violent protestations of poverty from the monastic houses. Episcopal oversight, however, was often not effective, and the problem remained until the Reformation.

5. Cultural and Economic Developments.—It might be expected that along with these developments in the church would go advances in culture and learning. No doubt these followed, but little evidence has remained. Scotland in this period produced only two monastic chronicles which have survived, the chronicles of Melrose and Holyrood: these are brief sets of annals and, although they have great historical value, they are not works of literary importance. Nor is there much evidence of manuscripts being copied or illuminated in Scotland. Attached to the monasteries were often schools; but, apart from occasional references to schoolmasters, nothing is known of their work. Scotland produced in the period one theologian of note, Adam Scotus (fl. late 12th century). There is, however, some evidence of architecture at the time which suggests that Scotland was in touch with work being done elsewhere. Fine, if small, Romanesque churches survive at both Dalmeny and Leuchars; the nave of Dunfermline Abbey is closely modeled on that at Durham; the ruins of St. Andrews Cathedral make it obvious that it was originally a fine Romanesque building; and Elgin, and particularly Glasgow, cathedrals are 13th-century churches which, though smaller, are worthy to be compared with English work of the same period. In Orkney, Kirkwall Cathedral is an example of a 12th-century church in a more primitive style.

Fewer examples of secular architecture remain: most castles were probably built of wood on natural or artificial mounds; but 13th-century work survives at Bothwell and Dirleton which seems again to compare well with English work of the period; and there are a number of 12th- and 13th-century castles in the county of Argyll and the Western Isles which, though rather more primitive in workmanship, are still notable examples of Medieval military engineering.

Contacts with England and the Continent brought new opportunities for trade; and the kings seem to have encouraged this. They deliberately founded burghs, either by creating new settlements or by conferring special privileges on existing ones. These were communities of merchants and craftsmen, though they engaged also in agriculture, which possessed a monopoly of the markets over a specified area; all buying and selling in the area had to take place in their market; and their merchants often enjoyed privileges of freedom from tolls throughout the country. Some burghs remained merely the local market; but others, particularly those with harbours, became important centres of trade. The earliest burghs recorded, which existed already in 1124, were Edinburgh, Stirling, Berwick, and Roxburgh, all in the southeastern part of the country; but many more were created, either by kings, bishops, or barons, in the 12th and 13th centuries, till they formed a network over most of the Lowland parts of the country; and foreigners were often encouraged to settle in them. No accounts survive of the extent of Scottish trade at this period, but Scottish merchants are recorded commonly enough in England; they had a wool staple in the Low Countries and traded also with Norway.

The introduction of monasteries had also some economic effects. The Border abbeys in particular developed the agriculture of the Tweed Valley; the abbey of Newbattle engaged in some coal mining; and the abbeys were normally linked to the commercial classes by having "tofts" or town houses in many of the burghs. Again, however, there is little evidence since the records which survive of the abbeys are mainly titles to their lands, and not accounts of the economic exploitation of their estates.

The impression left by the evidence is that Scotland in the 13th century was a reasonably flourishing land, though it is difficult to document this.

C. The Wars of Independence

The death of Alexander III in 1286 made immediate a problem which had been recurrent but only occasionally important during the previous two centuries: the precise relation of the Scottish crown to that of England. Twice, it seems, the kings of Scots had accepted English overlordship, once when Edgar was established by William II, and once when William the Lion, who had joined in the rebellion of Henry II's sons, was captured at Alnwick in 1174. Henry II's gains of 1174 were surrendered by Richard I in 1189 in return for money, but this left unresolved the question whether the kings of England were entitled to a general overlordship apart from that which Henry II had won as a result of the capture of William the Lion. There were remote precedents, particularly a submission by various kings to Edward the Elder in 920; and the issue was complicated by the fact that Scottish kings held lands in England for which, as English barons, they undoubtedly owed homage to the king of England. On the other hand, there was no question that Scotland was admitted to be a kingdom, albeit a kingdom with some defects of form, especially that the kings of Scots were neither crowned nor anointed. The inauguration of the Scottish kings was purely a secular ceremony, a placing on the royal seat (the "Stone of Destiny"), not an ecclesiastical consecration. In the 13th century, the Scottish kings made various attempts to gain from the pope the right to coronation and unction; but these efforts were circumvented by the English government, which was anxious to maintain any signs that the Scottish kings were not kings in the fullest possible sense. The question remained unresolved: Alexander III performed homage (1278), but, according to one version of his oath, claimed that he did it only for his English lands and saving the rights of his kingdom, while the English similarly saved the rights of their king.

1. The Disputed Succession.—In 1286, the problem became immediate. Since Alexander III's son Prince Alexander had died in 1284, the only heir was Margaret the "Maid of Norway," the young daughter of Alexander's daughter Margaret who had married Eric II, king of Norway. To her, the Scottish barons had sworn allegiance in 1284. To maintain the government till she could come to Scotland and was of an age to reign, a group of barons were appointed "guardians." For the next five years they were responsible for the welfare of the state.

Edward I made no immediate attempt to profit from the situation, and it is not likely that at this stage he had any definite aim to exploit his position as overlord. In 1289 he made the proposal that his son Edward should marry the Maid of Norway and, in course of time, rule over both kingdoms: this idea met no serious opposition. It was accepted by the guardians at the Treaty of Birgham (July 1290, confirmed by Edward in August), which, however, made elaborate provisions for the preservation of the integrity of the kingdom of Scotland, which was not to be incorporated with England but ruled as a separate realm. The plan failed, however, with the death of Margaret in Orkney in September 1290. There was now no clear heir to the throne.

The two most serious claimants were Robert de Bruce (1210–95), "the Competitor," and John de Balliol (*q.v.*); both were descended from daughters of David earl of Huntingdon, the brother of William the Lion. There were, in all, 11 other claimants, but no others seem to have been regarded seriously except, perhaps, Florence count of Holland, who was descended from a sister of David. Bruce and Balliol were each supported by strong groups of the Scottish nobility, and there were rumours of preparations for war. In this situation, it was natural to turn to Edward I and ask him to arbitrate, as he had done, for instance, in a similar dispute over the succession to Castile.

2. John de Balliol.—Edward, however, refused to act on this basis. He asserted that he was superior lord of Scotland; that the kingdom was a fief which reverted to its superior who was entitled not to arbitrate but to judge the issue. Not till all the competitors accepted these claims, and until the kingdom had been formally placed in his hands, would he act. He then assembled a court which, after long examination, pronounced in 1292 in favour of John de Balliol, as descended from the eldest of David earl of Huntingdon's three daughters. Balliol was therefore enthroned

but required to perform homage explicitly for the kingdom of Scotland and to admit Edward's rights as superior lord.

During the reign of John de Balliol (1292–96), Edward insisted on exercising these rights by hearing appeals against the judgments of Balliol's court, by summoning Balliol to answer before his own court, and, eventually, by summoning Balliol and the Scottish barons to do service in his wars in France. Balliol was unable to resist effectively; but the last demand provoked the opposition of the Scottish barons, who, in 1295, set up a commission of bishops, earls, and barons to conduct the government of the country on behalf of Balliol. These allied with Edward's enemy, Philip IV of France, and Edward retaliated by treating Balliol and the Scots as contumacious vassals. In 1296 he invaded Scotland, captured Berwick and Dunbar, forced the resignation of Balliol, and took the government of Scotland into his own hands.

Revolt, however, was soon raised, partly ineffectively by a group of barons which included Robert Bruce, grandson of the "Competitor," and future king Robert I (1306–29); and more effectively by William Wallace (*q.v.*) and Sir Andrew Moray. Wallace and Moray defeated the English armies at Stirling Bridge (1297), shortly after which Moray died; and Wallace was appointed sole guardian on behalf of the exiled John de Balliol, whose resignation was regarded as invalid and in whose name government was now carried on. Edward, however, defeated Wallace at Falkirk (1298); but found that he was now committed to the military conquest of Scotland. A series of campaigns more or less reduced Scotland to subjection, though at enormous cost in effort and money, and Edward seemed triumphant when in 1305 Wallace was captured and executed as a traitor.

3. Robert I the Bruce.—In 1306, however, Robert Bruce decided to revolt, and, after his murder of John Comyn (*q.v.*), a possible rival, at Dumfries, was crowned king at Scone in March 1306 (he had to be crowned, since the "Stone of Destiny" had been removed by Edward). At first he had little success (*see* ROBERT I); but Edward I died in 1307 and his son Edward II was incapable of waging the sort of war which had placed Edward in control of Scotland. Bruce had the support of a number of notable Scots, especially Walter the 6th Steward, James Douglas ("the Good"), and his own nephew Thomas Randolph, the future earl of Moray; and as he became more successful others joined him. By 1314 almost all the English castles in Scotland had been captured, except for Berwick and Stirling; it was Edward II's attempt to relieve the latter which led to the decisive Battle of Bannockburn (*q.v.*). With the fall of Stirling, followed by the capture of Berwick in 1318, the English had lost all their strongholds in Scotland.

The war was prolonged by Edward II's refusal to recognize Bruce's position. The Scots invaded the north of England several times; but for most of the time there was an uneasy truce during which the Scots made determined efforts to secure recognition of their independence, especially from the papacy. Eventually the government of Isabella and Mortimer, during the minority (1327–30) of Edward III, conceded what was needed, by a treaty drawn up at Edinburgh and confirmed at Northampton in 1328. The English were now prepared to recognize Bruce as king of Scots; and his triumph was sealed by a papal bull which granted the long-sought rights of coronation and unction. It was issued six days after the death of Bruce himself in 1329; and the first Scottish king to be crowned and anointed with the full rites of the church was his young son, David II (1329–71), in 1331.

4. David II.—The treaty of 1328 did not, however, put an end to the Wars of Independence: Bruce's death and the succession of his young son at the age of five gave the English an opportunity to renew their claims. The attack began in 1332 as an attempt by Edward de Balliol, the son of John de Balliol, to seize the throne with the help of a number of English nobles who had lost lands in Scotland. They won the Battle of Dupplin (*q.v.*) in August, and Balliol was crowned at Scone (September) but after a defeat in December fled to England. Edward III now gave him open support (*see* BALLIOL) and in 1333 defeated the Scots at the Battle of Halidon Hill (*q.v.*) just outside Berwick, captured Berwick itself, and went on to overrun much of Lowland Scotland. Edward de Balliol was restored, did homage to Edward for the kingdom, and, as a reward for his aid, ceded most of Scotland south of the Forth into English hands. David II was soon forced to flee to France (1334). The southern counties were now organized under an English administration with English sheriffs centred on Berwick-on-Tweed. In the rest of the country there were English garrisons holding castles for Edward de Balliol. It seems that none of these had much control outside the immediate areas of their garrisons; but the Scottish administration collapsed and does not seem to have revived for some five years.

Scotland, however, was by no means conquered. Resistance remained in the north, under Andrew Moray, a son of the elder Andrew Moray; and in the west under Robert the 7th Steward, son of Walter the Steward and Bruce's daughter Marjory, who was heir to the throne in default of direct issue to David II. In the following years there were several English invasions; but the Scottish leaders gradually recovered more and more territory. By 1340, the customs were again being collected, though at first only in the northeast; by 1341, David II was able to return to Scotland; and by 1342 English bases in Scotland were reduced to Berwick and Lochmaben.

Once again the kingdom seemed secure; but David II was captured at the Battle of Neville's Cross (1346) near Durham during an attempt to invade England, and remained in captivity till 1357. The English recovered some of the south and had to be slowly ejected by the Steward and William lord (later 1st earl) of Douglas (a nephew of Sir James Douglas). In 1356, English control was reduced to Berwick and Roxburgh in the east, and Lochmaben with lower Annandale in the west. An English expedition in that year achieved little; and in 1357 Edward agreed to release David II for a large ransom. It was not till 1384 that the Scots recovered Lochmaben; not till 1460 that they gained Roxburgh; Berwick, however, after being gained by the Scots in 1461 was finally taken by Richard duke of Gloucester in 1482 and has remained English ever since.

D. THE CONSEQUENCES OF THE WARS, 1329–1424

This long-continued struggle with England was in many ways disastrous for Scotland. The material damage caused by English attacks and by Scottish counterattacks on English-held positions was considerable; so much so that few buildings survive from before the period of the wars, especially in the south of the country. Moreover, building done in the late 14th century is often conspicuously inferior to what does remain of 13th-century work; and the surviving records of monasteries often suggest that the first half of the 14th century was a period of economic hardship. The war also involved the kings in heavy expenditure, for campaigns or to pay David II's ransom; this in turn led to great increases in taxation, especially immediately after David II's release, when the rate of the normal wool taxes was trebled and finally quadrupled.

The most serious consequences of the wars, however, were upon the position of the kings. Robert Bruce seemed by his death to have established himself with all the traditional authority of the kings of Scots; but it was long before his son enjoyed an equally secure position. Edward de Balliol existed as a rival, claiming himself to be the lawful king. Except for a short time in the 1330s, few Scottish nobles supported him; but he was there as a possible alternative to whom dissidents could turn, and a few, particularly in the far west and southwest, made a point of professing their allegiance to him. The uncertain position of David II encouraged areas not effectively under royal control to disregard it completely.

Further, King David's minority with the troubles of that time, shortly followed by his captivity, resulted in two periods in which central government ceased almost entirely to function for, together, something like 17 years. Although David made strenuous and effective efforts to regain royal control in the 1360s, he died before he could establish a tradition of strong government; neither Robert II (1371–90) nor Robert III (1390–1406) were men capable of asserting their authority effectively; while James I (1406–37) spent the first 18 years of his reign as a captive in England,

leaving Scotland to be governed in his absence by his uncle, Robert Stewart, duke of Albany, who used his considerable authority to establish his own position and that of his family rather than to restore strong central government. Thus, between 1329 and 1424, there were probably only two short periods, from 1341 to 1346, and from 1357 to 1371, in which the authority of the crown was effective.

1. Baronial Aggrandisement.—Meanwhile, the wars with England had produced important changes in the Scottish baronage. The leading supporters of Robert Bruce, particularly Sir James Douglas and his half brother Sir Archibald (*see* DOUGLAS) and Thomas Randolph, 1st earl of Moray (*q.v.*), received great grants. James Douglas got the previously royal forest of Selkirk and many other lands in the upper Tweed and elsewhere; Archibald Douglas received lands in the southwest; while Randolph got Annandale, Moray, and the Isle of Man, though he never gained effective control of the last. These grants were the natural reward for the king's most faithful supporters: Bruce could not do better than entrust the defense of these lands, many of them very open to attack, to the men who had won them from the English. The policy seemed all the sounder in that these were men whose personal loyalty to the king was beyond doubt.

Nevertheless, these grants had changed the whole balance of power in the kingdom by depriving the crown of direct control of much of the south of Scotland, which grew into something approaching a personal lordship for the Douglases. Sir James died in 1330, a year after the king; and this broke the personal link which had been the justification of Robert Bruce's action. In the course of the 14th century, several of the separate lordships in Douglas hands were united by a natural process of inheritance until, by the early 15th century, virtually the whole of southern Scotland from the middle Tweed Valley to Galloway was held by the earls of Douglas, who had achieved a power far stronger than could safely be entrusted to a subject. Within this area, the king exercised little direct control, and even less in practice than he reserved in theory; while the situation was made worse by the lawlessness natural in a border country under constant threat of raiding. The weakness of the crown was accentuated by the enormous power held by a nominal subject in a vital area.

2. Parliament.—One development, however, was helped indirectly by the troubles of the period, that of the Scots Parliament. There is no doubt that the kings in the 13th century had held assemblies of their subjects: the meeting to swear allegiance to the Maid of Norway in 1284 is an example, as are various assemblies in the 1290s. These were predominantly gatherings of the nobles; though it is likely that representatives of the burghs attended on occasion. As in England, however, it was the need to raise money which allowed the burghs to establish their position in Parliament. We first hear of burgh representatives in 1326, when Robert Bruce required money; and the rights of Parliament and of the burgh communities in particular were further strengthened during the 1360s when additional taxation was needed to pay the ransom of David II. Nevertheless, the Scots Parliament never attained the influence or prestige of that of England (*see* also PARLIAMENT: *The Parliament of Scotland*).

E. The Revival of the Monarchy, 1424–88

In the 15th century the kings made serious efforts to restore effective government, particularly to reduce the power of the greatest barons. When James I returned from his captivity in 1424 he found Scotland dominated by the family of the late regent Albany (d. 1420). Albany's son Murdac had succeeded him as regent, and other members of the family had gained great power in various parts of the country. One of King James' first acts was to arrest as many of these as he could, along with many other nobles, most of whom were soon released after this assertion of royal authority had produced its effect. Murdac, however, and his two sons were executed (1425) on obscure charges.

Now that the king's cousins, descended from Robert III's brother Robert, duke of Albany, had been disposed of, the other great baronial power was that of the house of Douglas. Ever since the reign of David II, the family had been growing still more powerful. During the regency of Albany, there is evidence in the exchequer rolls that members of the family could with impunity seize some of the customs revenue; and no attempt was made to stop their depredations or to bring them to order. James I seems to have avoided any direct clash; but his successor James II (1437–60) resorted to violence. William, 6th earl of Douglas, and his brother David were executed for alleged treason in 1440 during the king's minority. William, the 8th earl, was murdered by the king himself in 1452; and James, the last earl of Douglas, was detected in treasonable communication with England and driven into exile in 1455. He spent the rest of his life in England, from which he planned rebellions against James II and James III. Thus the power of the main line of the earls of Douglas was destroyed; though a cadet branch, the earls of Angus, remained loyal and were to become very strong in the latter part of the 15th century.

During the 15th century, too, the kings made a number of efforts to extend their effective control into the remoter parts of the country. James I made several expeditions to the north and west; but the line of the lords of the Isles, who now claimed also the earldom of Ross by marriage, remained hard to control and liable to take the law into their own hands by plundering the territories of other magnates in their area. They were the more dangerous in that they could hope to exploit English hostility to the kings of Scots. In 1462, acting as an independent authority, John MacDonald, lord of the Isles and earl of Ross, made the strange Treaty of Westminster-Ardtornish with Edward IV of England, by which they planned, with the exiled James, earl of Douglas, the partition of Scotland. Nothing came of this; and the power of the lord of the Isles was crushed after a rebellion in 1475. James III (1460–88) deprived him of the earldom of Ross, and, although he was confirmed in the title of lord of the Isles, his effective authority was severely restricted.

James III also gained finally for Scotland the two remaining groups of islands still subject to Norse control, Orkney and Shetland. These were pledged (Orkney in 1468, Shetland in 1469) by Denmark for the payment of the dowry of the queen, Margaret of Denmark, whom James married in 1469; as the dowry was never paid, the islands were formally annexed by Parliament in 1472.

1. Administrative Improvements.—During James I's captivity in England, he had learned much about English methods of government; and on his return to Scotland he tried to remodel Scottish government on similar lines. In particular, he made far greater use of Parliament than his predecessors as an instrument to introduce reforming legislation, though, like most medieval legislation, much of it had probably little effect. It is in this reign that the continuous records of the Scottish Parliament begin.

He was also impressed by the ways in which the English king's finances were organized. From the exchequer rolls, it appears that the income of the Scottish crown had been declining during the period since 1371. Further, such control as there was of the king's resources remained in the hands of the chamberlain, to whom were paid all revenues which were not at once spent locally by their collectors, on the king's warrant. This placed far too much financial power in the hands of a household official whose office had become, under the Stewart kings, a position of dignity held since 1382 first by Robert Stewart, the future duke of Albany, and since 1407 by his son John, earl of Buchan. James was probably concerned to reduce the power of an office which had become associated with the family of Albany; and he achieved this by introducing two new financial officials, who soon effectively superseded the chamberlain. The latter remained an office of some dignity, but its financial authority was gone. Routine revenues, of crown lands, customs, and the like, were to be handled by the comptroller; while extraordinary revenues and the general management of finance were given, on the English model, to the treasurer. Both these officials were working civil servants, not magnates. Thus James gained a much more effective and direct control of his revenue.

2. The Period of Minorities.—The revival of royal authority in the 15th century was real, but its effects were limited. James

I had removed the rival house of Albany; but there remained that of Atholl, descended from the second marriage of Robert II, and anxious to maintain that that king's first marriage, from which the later Stewart kings were descended, was invalid. This claim must have been the basis of the obscure conspiracy which resulted in the king's murder at Perth in 1437.

There was no general support for the conspirators, who were all captured and executed; but James II succeeded as a child of 6; and was in turn killed at the age of 29 during the siege of Roxburgh in 1460, leaving as his heir James III, aged 8. As a result, something like half of the period between 1406 and 1488 was a time in which the power and influence of government were in the hands of those who could seize the king or wield power in his name. In the late 14th century, some semblance of order had been preserved by Albany and Douglas, who were prepared to work together; but the earls of Douglas in the 15th century were not concerned to uphold authority. As a result, during the minority of James II, control of the king was disputed by the families of Crichton and Livingstone, amid a general collapse of authority; while in the 1460s the young James III came under the influence of the Boyds of Kilmarnock, whose leader Robert gained for his son the title of earl of Arran and the hand of the king's sister (1467).

Jealousy produced a reaction, and Arran and his father only escaped execution by exile in 1469. James III himself offended some of his nobles by his friendship for artists, architects, and musicians; a group revolted in 1482, in face of an English invasion, and hanged several of the king's favourites over the bridge at Lauder; and then revolted again in 1488, with the help of the king's son James, duke of Rothesay. The result was the Battle of Sauchieburn, after which James III was killed while fleeing from the field.

The repeated troubles and insurrections of the period made it difficult for the kings to assert effective control over their kingdom, and the exchequer rolls continue to give evidence of such defiance of authority as the plundering of the customs revenue by families of local power. The result, despite the efforts of the kings, was that Scotland remained a state in which great power was vested in the landed nobility. The great families of the later 15th century, the Douglas earls of Angus and the Hamiltons, together with others of more local power such as Hume and Hepburn in the borders, Kennedy in the southwest, Gordon in the northeast, could not effectively be brought to account by the crown and could only too easily take the law into their own hands when it pleased them. Their power and influence survived into the 16th century when it was to dominate the political and religious struggles of that time.

3. Scotland in the 15th Century.—Despite the wars with England, the difficulties of the Scottish government, and the internal disturbances, there is evidence that in Scotland the 15th century was a period of economic revival and development, after the more difficult times of much of the 14th century. From the late 14th century, the nobility were engaged in building on a much larger scale than before; spacious and impressive halls, as at Bothwell and Dirleton, were built, often with no regard for defense. In the church, too, the 15th century was an age when monasteries and cathedrals were much rebuilt or extended. From about the middle of the 15th century, many monasteries seem to have reorganized the leasing of their estates in a more businesslike way; numerous rentals and registers of leases survive, almost all dating from this period. Hardly any survive from before 1450. It is hard to produce statistics of trade, but there is perhaps evidence of the prosperity of the towns in the rich foundations of collegiate churches made in Stirling and Edinburgh in this period. Scotland in the late 15th century seems to have been much more prosperous than in the dark days of the 14th century.

Of culture, it is more difficult to speak. By the end of the 15th century, Scotland had three universities—St. Andrews, founded in 1410 (papal bull, 1413) and strengthened by Bishop Kennedy's College of St. Salvator in 1450; Glasgow, founded by Bishop William Turnbull in 1451; and Aberdeen, founded by Bishop William Elphinstone in 1494. Some of the teachers, such as the first rector at St. Andrews, Laurence of Lindores, were men of note; but no Scottish-trained scholar achieved international importance until the 16th century. From the late 14th century onward, Scottish writings began to be of significance. John Barbour (*q.v.*) wrote in Scots a metrical life of Robert Bruce; John of Fordun (*q.v.*) wrote in Latin the first full-scale history of Scotland, though, unfortunately, his completed version did not go beyond the 13th century, and for the 14th century there are only brief annals. His work was extended and continued by Walter Bower (*q.v.*), abbot of Inchcolm, in the 15th century, to form his *Scotichronicon;* and Andrew of Wyntoun (*q.v.*) wrote in the early 15th century a metrical history in Scots, largely based on Fordun. (For further details of vernacular literature in the 15th century, *see* ENGLISH LITERATURE: *From Chaucer to the Renaissance;* SCOTTISH LITERATURE.)

Of art there is little evidence. One of the manuscripts of Bower's *Scotichronicon* is illustrated with some striking drawings; and fragments remain of wall paintings of uncertain date and quality. All that can be said is that Scotland was not out of touch with 15th-century continental work, since James III's foundation in Edinburgh, the collegiate Church of Holy Trinity, had a painted altarpiece (two wings of which are in the National Gallery of Scotland) with portraits of the king and queen, which came from the workshop of Hugo van der Goes. Most of the remains of Scottish art in the Middle Ages perished in the years between the Reformation and the revival of antiquarian interests in the 19th century. The main evidence remaining is the buildings of the period, which do not suggest that the country was lacking in artistic achievement.

F. THE SCOTTISH CHURCH AND THE PAPACY

Since the 12th century, the Scottish Church had been in a peculiar relationship with the papacy. Since there was no archbishop in Scotland, the two nearest primates, the archbishops of York and Canterbury, had disputed the authority over the Scottish Church, which was claimed by York at least since the 11th century, when its archbishops had consecrated bishops of Glasgow. Canterbury was anxious to prevent any gain in influence by its rival and claimed that the Scottish sees were subject to it; and the claim gained colour from the personal friendship of Lanfranc of Canterbury (1070–89) and St. Margaret. The majority of the Scottish bishops, however, resisted both claims; their independence of outside jurisdiction was first recognized by the papacy in 1164 and confirmed in 1192 in a bull which described the Scottish Church as "filia specialis" (special daughter) of the papal see. From the early 13th century, the right of the Scottish Church to hold councils although they had no metropolitan was expressly recognized; and this remained the situation until 1472, when St. Andrews was erected into an archbishopric, followed in 1492 by Glasgow.

Thus in the 12th and 13th centuries, the Scottish Church had turned to the papacy as its defender against claims from York and Canterbury. From the end of the 13th century, however, the Scots lost the political support of the papacy, and, as elsewhere, there developed opposition to its financial claims. At first, indeed, Pope Boniface VIII (1294–1303) supported the Scots in their struggle against Edward I; but Robert Bruce found it more difficult to get papal approval because of his sacrilegious murder of John Comyn. Although the Scottish Church supported Bruce, it was not till 1328 that the sentences of excommunication pronounced against him in 1318 for infringing a papal truce were formally lifted. The papal schism in 1378 also affected the relations of the Scots to the papacy. Being allied to France and opposed to England, they recognized the Avignonese pope Clement VII rather than Urban VI; and in fact continued their loyalty to the Avignonese popes long after these had been deserted by France and almost all other supporters. The Scots were the last to abandon the antipope Benedict XIII, about 1418. In the meantime, all papal confirmations and provisions in Scotland had been made by Benedict XIII; this meant there was the opportunity to dispute many of these when Scotland did recognize Martin V, the pope elected (1417) by the Council of Constance, and the resulting confusion persisted for some time.

When James I returned to Scotland in 1424 he was short of money, both to pay his ransom and for the needs of his government, so naturally he objected to the heavy financial demands made by the papacy, particularly in the form of payments on appointment to benefices and offices. The whole system of payments was linked with the system of "provisions" or papal appointments to vacant offices; which both reduced the royal authority over the church and deprived local patrons of their rights. James therefore in his early Parliaments introduced a number of laws aimed at reducing papal control over appointments and at preventing the payment of moneys to Rome.

James found the Scottish Church, which had no particular loyalty to Martin V, sympathetic to his aims. In the church at large, many thinkers at the time were critical of the papacy and anxious to subordinate papal authority to the control of general councils of the church. These "conciliarist" ideas had been very common in the University of Paris at the end of the 14th century; and many Scottish churchmen had been educated there. Hence there were many conciliarists in the Scottish Church, who took their part in the movement, particularly at the later Councils of Siena (1423–24) and Basel (1431–49). From these James could count on support for his attacks on papal influence.

James also approved of conciliar demands for reform in the church. He addressed letters to the abbots of Benedictine and Augustinian houses, calling for improvements; he introduced into Scotland the Carthusians, the strictest of monastic orders, and endowed the house which he established for them at Perth. There is, however, something backward looking in this concern for monasticism; he could do little to meet the real need of the church in the 15th century, the improvement of the status and education of the parish priests.

With James' death and the accession of his six-year-old son James II, his resistance to papal demands was abandoned. It was not a policy which could be carried on in the disturbed conditions of a minority; and the conciliar movement was in any case evidently waning. His successors found it more advantageous to work with the papacy and thus through papal provisions to secure the appointments they wished. The result was that over the century ecclesiastical office became more and more a means of providing for the king's relatives and the relatives of the nobility. Not all appointments made by these means were bad. In Bishops James Kennedy (d. 1465) of St. Andrews, William Elphinstone (d. 1514) of Aberdeen, and, even on the eve of the Reformation, in Archbishop John Hamilton (d. 1571) of St. Andrews, the church had men who could do their duties well, however they gained their offices; but the general scene was not edifying and many appointments were flagrantly scandalous. (Br. W.)

G. Scotland, England, and France, 1488–1567

1. James IV and V.—James IV (1488–1513) was all that his times thought kings should be—physically impressive, active and warlike, vigorous in dispensing justice, cultured, and open-handed. Only retrospectively did his mercurial nature and his impulsiveness in battle seem great faults. He took by forfeiture the lands and title of the last lord of the Isles in 1493, extended royal administration north and west, and acted against border lawbreakers. At a time when the "New Learning" was spreading through Europe, James and Bishop Elphinstone of Aberdeen founded King's College, Aberdeen. Elphinstone probably inspired an act of 1496, designed to provide a general and legal training for the children of the governing classes. James licensed Scotland's first printers, Walter Chepman and Andrew Myllar, in 1507. Some of Scotland's greatest poets profited from court patronage: Robert Henryson, whose *Testament of Cresseid* is notable for its compassion, was simply a burgh schoolmaster; but Gavin Douglas, bishop of Dunkeld and author of a superb verse translation of the *Aeneid,* owed his preferment to his relationship to the earls of Angus; and William Dunbar, writing in magnificent latinate Scots and with consummate technical mastery, could only have been the product of a court circle.

After initial disharmony with England, James concluded a "treaty of perpetual peace" with Henry VII in 1502 and married Margaret, Henry's daughter, in 1503. James's ambition to lead a European crusade against the Turks was doomed when even the pope, Julius II, engaged in divisive power-politics and in 1511 formed the Holy League to counteract French ambitions. England joined the League; whereupon France and Scotland, both wary of Henry VIII, renewed their "auld alliance" in 1512. When Henry invaded France in 1513 James's decision to uphold his alliance by invading England, and his rash generalship in the field, led to his death and that of the flower of the Scots army at the Battle of Flodden (September 1513).

James's firm rule at home, and his international ambitions, had required great expenditure on artillery, on a navy whose greatest ship, the "Michael," cost more than £30,000, and on embassies. The crown began to feu lands, that is, give heritable possession of them in return for a large initial payment and a fixed annual money-rent thereafter. This yielded ready money but ultimately reduced royal resources during the great Europe-wide price rise which went on throughout the 16th century.

At his father's death James IV had been 15 years old: James IV's son, when he became James V (1513–42), was only in his second year. During his minority, family and personal feuds revived; furthermore, a political division became increasingly clear, between those who supported the "auld alliance" and those who drew from Flodden the moral that Anglo-Scottish antagonism benefited only France. John Stewart, duke of Albany, a grandson of James II but by residence and upbringing really a Frenchman, was regent until 1524, and French interests prevailed. Archibald Douglas, earl of Angus, who dominated the young king for the following four years, favoured England. When James began his personal rule, with a treasury drained during his minority, he found a peculiarly favourable international situation: France, the emperor Charles V, England, and Pope Clement VII all vied for Scottish support, and James played on their conflicting interests. From the pope he obtained financial concessions as the price of his continued allegiance to Rome and, after considering many brides, he made two successive French marriages, each bringing a large dowry. His first wife, Madeleine, daughter of Francis I, died shortly after their marriage in 1537, and in 1538 he married Mary, daughter of Claude de Lorraine, duc de Guise. Some of the nobles disliked James's adherence to the pope and to France; others suffered from his brutal suppression of Highland and border lawlessness, and from his acquisitiveness. The low morale of the Scots led to the humiliating defeat of a force invading England, at Solway Moss in November 1542. This disaster, added to the deaths of his infant sons, crushed James, who died in December, a week after the birth of his daughter and successor Mary (1542–67).

2. State of the Church.—It would be rash to assert that clerical immorality and ignorance were worse in 16th-century Scotland than in other lands or at other times. But the church was top-heavy with wealth, and poor and ineffective in its witness before ordinary people: by 1560 the bulk of the revenues of nearly nine-tenths of the parishes was appropriated to monasteries, cathedrals, and other institutions, and latterly a series of agreements between crown and papacy effected a share-out of church wealth that left spiritual matters out of account. Since 1487, royal control over appointments to the higher ecclesiastical offices had grown stage by stage, and in the 1530s the pope had begun to impose on the Scottish Church heavy taxation, which enriched the crown but dislocated ecclesiastical finance and embittered relations between the prelates and the people. Nobles and lairds might be attracted to the new religious doctrines that Scots scholars and traders encountered on the Continent; nobles and lairds might also be influenced by material considerations. By the 1530s they were gaining possession of much of the church's wealth on a hereditary or quasi-hereditary basis: clerics feued church lands to laymen, and laymen became collectors of church revenues and commendators of abbeys. The enriched laymen were prepared to support a Reformation that did not threaten their possessions, and perhaps some even felt compelled to take part in it to ensure that it did not so threaten. In addition, lairds relished the enhanced influence they would exercise in a church with greater lay participation.

Whether or not clerical shortcomings in the early 16th century

(BELOW) BY COURTESY OF BRITISH TRAVEL; (TOP) HOLLYMAN FROM PHOTO RESEARCHERS; (LEFT) KENNETH SCOWEN

(Above) Abbotsford, the country home of Sir Walter Scott. (Left) The ruins of Melrose Abbey, Roxburghshire. (Below) West front of the cathedral on the Island of Iona

were worse than before, awareness of them was more general, and criticism, sometimes stemming from a genuine spiritual reawakening, was more openly stated. Sir David Lyndsay, the last great Scots poet for two centuries, used his play *Ane Satyre of the Thrie Estaits* to castigate the church without saying whether reform from within was possible. Reformers from within, countenanced by Archbishop John Hamilton, who gave his name to a Scots Catechism (1552), made little headway against vested interests.

3. Mary and the Scottish Reformation.—Mary's minority allowed the tension between ecclesiastical conservatives and francophiles on one hand, and Reformers and friends of England on the other, to increase. By the Treaties of Greenwich (1543) it was agreed that Mary should marry Edward, Henry VIII's heir, but the French cause was upheld by David Cardinal Beaton and Mary of Lorraine, the queen mother, and two English invasions (1544 and 1545) hardened Scottish resistance. The murder of Beaton (1546) did not reverse the trend, for his murderers received no effective English help and ultimately had to surrender St. Andrews Castle (July 1547) to a French expedition. A third English invasion and the defeat of the Scots at Pinkie (September 1547) left English forces in occupation of southeastern Scotland: French help against them was given only on condition that Queen Mary should be removed to France. The English were ousted, but the French hold on Scotland gradually tightened until, with Mary's marriage in 1558 to the dauphin Francis (afterward Francis II), it seemed likely that Scotland would be absorbed into the French monarchy.

National resentment against the French thus combined with the work of reforming preachers to stimulate preparations for revolt. In December 1557 some leading supporters of the new faith pledged themselves, in a "Common Band," to cooperate in politics and religion. In 1559, after Mary of Lorraine determined on stern action against Reformers, they took up arms. But, despite the preaching of John Knox (*q.v.*) and the plundering of monasteries by town mobs, the decisive campaign of 1560 was fought mainly between the government's French troops and the forces sent by Elizabeth I of England to prevent French hegemony in Scotland. Mary of Lorraine died in June 1560, and by the Treaty of Edinburgh, in July, both France and England undertook to withdraw their troops.

These events were decisive politically, but less than decisive for the Reformation. A Parliament in August 1560 abolished papal authority and adopted a reformed Confession of Faith; the organization of local congregations, which had started some years before, continued, and the Reformed Church found a central legislative body in the general assembly. But the legislation of August 1560 was not ratified by the queen, and the Reformed Church could not count on establishment should she return to Scotland. Nor could it have full endowment, as so much of the ecclesiastical wealth was already in lay hands. Knox and his fellow ministers produced in the *First Book of Discipline* (1560) a noble social program, providing not only for reasonable ministerial stipends but also for education at all levels and for poor relief. But it did not receive parliamentary approval and instead of receiving the bulk of the church's revenues, as they demanded, the reformers had to be content with an arrangement (1562) whereby they shared one-third of them with the crown.

Mary's husband died late in 1560 and she returned to Scotland in 1561. She was the Catholic queen of a Protestant Scotland and also the nearest heir to the English throne, by descent from Henry VII's daughter Margaret; her return thus made difficulties for both countries. Mary's brain and political sense were probably as keen as those of most of her contemporaries; but, in striking contrast to Queen Elizabeth, she sometimes put passion before policy, and this finally destroyed her when her somewhat lax Catholicism and the strength of noble faction alone would not have done so. For her bitter disputes with Knox *see* KNOX, JOHN. For the story of her marriage (1565) with the degenerate Henry, Lord Darnley, the birth (1566) of their son James, Darnley's murder, and Mary's marriage with the adventurer James Hepburn, 4th earl of Bothwell (the main suspect), and her imprisonment, forced abdication, escape (1568), and flight to England *see* MARY (*Mary Stuart*).

H. WAR AND PEACE, 1567–1625

Amid civil war between "Queen's Men" and "King's Men"—Mary's faction and the supporters of her son the young King

James VI (1567–1625)—two out of three short regencies ended in violent death. Elizabeth had detained Mary in England, for a Protestant regency in Scotland suited her better than a Catholic queen. She therefore ultimately aided James Douglas, earl of Morton, who became regent in 1572, to achieve stability. Many nobles had favoured Mary in the struggle: that her side lost speaks much for the political cohesion that the lairds and merchants, who tended to support James, had learned in the Reformation era.

1. The Church.—After Mary's deposition the government gave full recognition to the church settlement. In 1567 provision was made for the succession of ministers of the new church to the benefices of the old as they fell vacant. In 1573 an act provided for the deprivation of anyone enjoying a benefice of the old church who had not yet recognized the new. The new church benefited by Morton's Concordat of Leith (1572), whereby the crown was to appoint bishops with the approval of the church. This measure prevented bishopric revenues from wholly following those of the monasteries into the hands of laymen.

But such a compromise was repugnant to the second generation of Reformers, the first Scots churchmen who could be truly called Presbyterian. Led by Andrew Melville, a brilliant academic who lacked the political ability to compromise, they framed the *Second Book of Discipline* (1578). They demanded the wealth of the old church for the support of the new, condemned episcopacy, and advocated rule of the church by a hierarchy of ecclesiastical courts (to include elders as well as ministers), kirk sessions, presbyteries, synods, and the general assembly. Government of church and of state should be separate, but the spiritual power could instruct the secular in the godly exercise of its duties. The implication of this was theocratic rule by a caste of learned divines. Belief in such a polity was the basis of clerical opposition, then and later, to the kings of Scots. Melville was too unsubtle in his confrontation with the monarchy, and his program held little appeal for wealthy laymen. For a time political circumstances led James, who already strongly opposed the presbyterian system, to make concessions to Presbytery culminating in the "Golden Act" of 1592, which sanctioned the Presbyterian hierarchy of church courts. This delayed but could not prevent James's eventual triumph. From 1600 the bishops he had appointed sat in Parliament, symbolic of his determination to give the law to the church as well as to the state. From 1606, with the detention in London and later banishment of Melville and others, the clerical opposition was greatly weakened. By 1610 the civil and ecclesiastical status of the bishops had been fully reestablished. Since the church courts remained in being also, the system was moderate enough for most men; and James and his advisers were sensitive enough to the opposition to the beginnings of proposed liturgical changes—the Five Articles of Perth (1618)—to go no further.

2. James VI's Achievement.—Catholic influence at court had been suspected after the execution of Morton (1581). But, in the atmosphere of the 1580s, when Protestantism everywhere seemed threatened by Spain and the other Catholic powers, a league was concluded with England (1585–86), and James, now personally responsible for policy, in the following year virtually acquiesced in the execution of his mother, who had become the focus of every plot against Elizabeth. James inherited his mother's claim to the English throne, and to inherit the throne itself he had to retain Elizabeth's friendship while yet avoiding, if possible, giving offense to Catholic Scots or Catholic powers abroad. In fairness to James it must be said that the same course of action was called for in order to preserve his Scottish throne. With unruly churchmen and factious nobles to contend with, he was fortunate in possessing infinite guile and resource. His theories of royal rule by divine right, which became rigid by the end of his life, were initially evolved as the reasonable response of a scholarly man to the beliefs of Scottish Presbyterians and continental Catholics alike that in certain circumstances tyrannicide was not a sin. In pursuing his policy of conciliation James was lenient to his Catholic nobles even when the discovery of letters and blank documents (the "Spanish Blanks" affair, 1592), showed several of them to be in treasonable conspiracy with Spain. While he flirted diplomatically with the papacy and the Catholic states, he also corresponded with Elizabeth's minister, Robert Cecil (later 1st earl of Salisbury), and was thereby enabled to succeed peacefully to the English throne (as James I) in 1603.

The era of the Scots Reformation had seen the beginnings of a fundamental reorientation of the realm and its people. Endemic Anglo-Scottish warfare had been in a sense an artificial condition, and as it ceased it was replaced by a positive unifying factor, the need to defend the Reformation. Queen Elizabeth's Act of Supremacy (1559) showed that England had again broken with Rome; the Scots did so in 1560. The greater nation drew the lesser, as Henry VII is supposed to have said it would. The process began before 1603 and continued apace. The achievement of peace and amity within Great Britain was, however, balanced by a weakening of Scotland's continental ties and a sometimes slavish following of English example. The Scots language declined—the most effective teacher of standard English being the Authorized Version of the Bible (1611)—and Scottish institutions became slightly less distinctive throughout the following centuries.

Historians have often criticized James's rule in England. But no amount of royal wisdom could have wholly averted the political and social upheavals of 17th-century England, and Elizabeth had left a legacy of trouble. James's regal qualities may be better gauged from the way in which Scotland, which he saw only once between 1603 and his death in 1625, continued to be administered as well as under any previous king, by the team of royal servants James had created. John Maitland of Thirlestane, chancellor from 1587 to 1595, was one of the first of this group. These men were lairds or burgesses, who were frequently ennobled, often by the erection of former monastery lands into a temporal lordship for them, but who remained more dependent on royal favour and therefore more submissive than members of the older noble families. Such men held offices of state, sat as judges, or served as bishops. Some bishops, like Andrew Knox who helped pacify the Western Isles, had secular as well as ecclesiastical roles. James's rule was paternalist, and he regarded it as his function to impose order on all his subjects. He looked to God for approval rather than to the ruled for consent; yet there is little doubt that that consent was widely given.

3. Government and Economic Developments.—The Scottish Parliament, which had long consisted of prelates, nobles, and representatives of the royal burghs, was strengthened after 1587 by the admission of shire commissioners representing the "barons" (lairds), who had flocked to the revolutionary Parliament of 1560 and had found their way into the general assembly. But the crown had much influence in Parliament's composition, and it could control the Committee of the Articles (*see* PARLIAMENT: *The Parliament of Scotland*) which was the sole channel through which measures could come before the full House. The Privy Council, in which royal officials were the most regular attenders, carried out royal orders sent from London. It had extensive legislative as well as administrative functions, and also considerable judicial powers, which overlapped with those of the Court of Session and the Court of Justiciary. The former court, for civil cases, had emerged from the council early in the 16th century and had, under the name of the College of Justice, been endowed with church funds through a papal concession in 1532; and justice courts for criminal cases existed in various forms until the organization of the High Court of Justiciary in 1671–72. Local justice and administration had long been largely in the hands of landowners, for regality courts, baron courts, and even sheriff courts were heritable jurisdictions. James VI, as part of his general policy of reducing the nobles' power, attempted without success to foster justices of the peace and undermine the heritable jurisdictions, but it was only by slow social changes throughout the 17th century that the old domination of the magnates over the lower classes was to some extent broken down.

Good government was one of the reasons for the increased prosperity of Scotland in the late 16th and early 17th centuries. Taking into acocunt the international inflation of the period, it still seems that the average merchant and landowner tended to become richer; and the merchant, who received cash while the landowner would be paid rentals in kind, had ready money for furnishing his

house with imported luxuries. A few innovations were made in farming (*e.g.*, the use of lime on fields close to a natural supply), and, after 1660, attempts were made to encourage some manufactures (*e.g.*, linen) and to begin others (*e.g.*, glass) by protectionist devices such as the grant of monopolies and the control of imports. But this does not indicate a dynamic economy; nor does the enumeration of the main exports—grain (in good years), hides, wool, linen, coal, and salt, sent to France and throughout the North Sea and Baltic regions—or the imports of grain (in bad years), wine and other luxury foods and manufactures, timber, iron ore, and naval stores. The Scots economy was in fact a subsistence one, essentially rural and producing above all for its own needs. Until the roads were improved in the 18th and 19th centuries it could be little else, and as such, given a good summer, it sufficed. It was national vitality, and probably pressure of increased population, rather than simple dearth, that sent Scots to England; to Ulster when it was deliberately "planted" with settlers by James VI from 1610 onward; and, in substantial numbers, as merchants and mercenaries, throughout northern Europe and beyond. The lines of future economic development were being laid in closer trading ties with the more developed economy of England: after 1603 there were still two countries in one island, with a customs barrier between them (though with a great decrease in border lawlessness through the work of James VI), but their peoples came to enjoy a common nationality and various mutual trading concessions. The educational system, which in the 18th century produced trained manpower for Scotland's industrialization, was created in the 17th in the faith that godliness would follow good learning. The Reformed Church, though lacking most of the endowments of the old, nevertheless was instrumental in creating a network of parish and burgh schools that covered most of Lowland Scotland. An act of the Privy Council of 1616, which decreed the founding of a school in each parish, and later enactments, ensured that the work went on throughout the 17th century's turmoils. Andrew Melville brought discipline and the highest standards of scholarship in divinity and oriental languages to the universities of Glasgow and St. Andrews. Aberdeen (Marischal College, 1593) and Edinburgh (the Town's College, charter of 1582) were new foundations.

I. The Age of Revolution, 1625–89

1. Charles I.—James VI's son, Charles I (1625–49), grew up in England, lacking any understanding of his Scottish subjects and their institutions. He soon fell foul of a nobility restless in a Scotland that lacked the natural focal point of a royal court. His Act of Revocation (1625), rendering void all dispositions of royal or church property made since 1540, had the laudable aim of providing for the needs of the church. Besides, it is clear that Charles intended no program of confiscation without compensation. But the nobility regarded the secretiveness of the royal administration as alarming, the increasing use made of bishops in government as an affront, and Charles himself as standing too much upon his dignity to listen to discontented subjects. The king also caused widespread anger by high taxation, by the special demands made on Edinburgh to build a Parliament House and to provide a cathedral for the bishopric founded there in 1633, and by a Spanish and a French war that were intended to further English diplomacy, but also disrupted Scottish trading ties. The aristocratic leaders of the opposition found ideal material on which to build clerical and popular support. Charles and his Scottish bishops were "Arminian" enough in theology and fond enough of ritual and splendour in church services to make plausible the (wholly incorrect) suggestion that they were ready for compromise with Rome. The new Book of Canons (1635–36) and Liturgy (1637) therefore offended by their content, as well as by being authorized by royal prerogative alone. The National Covenant (1638) astutely collected national support for the opposition's pledge to resist Charles's innovations. Condemnation of popery was written into it for the benefit of those who feared that Charles might be a crypto-Catholic; others, more sophisticated, welcomed its implicit condemnation of a royal arbitrariness with religion and private rights that was contrary to all Scottish precedent.

The Covenanters humbled Charles in two almost bloodless campaigns, the Bishops' Wars (1639–40), and left him with no alternative to asking for money from an English Parliament, in which his opponents were strongly represented. Charles had authorized a general assembly of the Scottish Church (1638) and a Scottish Parliament (1639); the Covenanters packed these meetings, scrapped all the king's innovations, and abolished episcopacy. There was, therefore, by 1641 a revolutionary situation in both kingdoms, and in August 1642 war broke out between Charles and his English opponents. Both sides sought Scots help, which was soon accorded to the English parliamentary opposition. By the Solemn League and Covenant (1643) the English promised, in return for military aid, to help preserve Presbyterian Church government in Scotland, and, so at least the Scots believed, to set it up in England. James Graham, 1st marquess of Montrose, and others who then left the Covenanting side argued that by this second Covenant, and by certain constitutional constraints they had placed upon the crown, the Scots had gone unwarrantably far beyond the aims of the first Covenant. But those of the Scots who were prepared to make common cause with the English opposition, even if the English did have a more deep-seated quarrel with their king than the Scots, had reasoned justification; for it was realistic to expect that Charles, as soon as it proved possible, would withdraw concessions made to men whom he regarded as his enemies. Personal antipathies also helped to split the ranks of the original Covenanters—notably the antipathy between Montrose, forthright and fearless but conceited, and Archibald Campbell, 1st marquess of Argyll, sincerely devoted to the cause but equally devoted to the advancement of his family. Montrose's military efforts for Charles in Scotland were crushed in 1645, and by 1646 Charles had lost the war in England, too. When Charles surrendered to the Scots Army in England, the Scots failed to reach agreement with him and handed him over to the English. The Scots contribution to the English war effort had been substantial, but not spectacular enough to leave a sense of obligation; and the English Army under Oliver Cromwell, now eclipsing Parliament in English politics, preferred Independency to Presbyterianism in the church and did not propose to honour the Solemn League and Covenant. A conservative element among the Covenanters in 1647 reached a compromise, or "Engagement," with Charles, by which they promised him help in return for the establishment of Presbyterianism in both kingdoms for three years, and went to war on his behalf: their ill-planned campaign was crushed at Preston in 1648. The clerics, who had bitterly opposed this compromise, were now able, under the leadership of a few nobles such as Argyll, to purge the Scots Parliament and Army of all tainted with collaboration with the king. The execution of Charles by the English in 1649 genuinely shocked most Scots, who were prepared to fight for his son, Charles II, once he had been constrained to accept the Covenants and once Montrose had been executed (1650). Cromwell's victory over the Scots at Dunbar (1650) gave more moderate Scots the ascendancy again, but this brought no better military result. Another, and decisive, defeat of a Scottish royalist army at Cromwell's hands came at Worcester in 1651.

2. Cromwell.—Cromwell imposed on Scotland a full and incorporating parliamentary union with England (1652). This could not enjoy the popularity of a union by consent, maintained as it was by an army of occupation, but Cromwell's administration of Scotland was efficient, and his judges, some of them Englishmen, achieved an admired impartiality. Public order was well maintained, even in the Highlands after the collapse of royalist resistance in 1654. Some nobles underwent forfeiture while others were impoverished by fines that followed campaigning expenses. The dependence of lesser men on greater, by characteristically Scottish family and feudal ties, was systematically weakened by the government. Scots commerce, disrupted by the wars, was in the worst possible condition to take advantage of economic union with England, and legislation tended to leave Scottish economic interests out of account. Cromwell did not overturn Presbyterianism, but ensured toleration for others, save Catholics and Episcopalians. Nevertheless the Scots' total failure to impose "godly" ways on England engendered serious national demoralization. The "Protest-

ers," those Presbyterians who had denounced the more moderate line of policy taken following Dunbar, remained irreconcilable with their fellows.

3. The Restoration Monarchy.—The restoration in 1660 of Charles II (1660–85) was welcomed by many moderate men of both his kingdoms. Charles had learned much from his father's fate and was prepared to forget many injuries, though his government executed some Scots, including the marquess of Argyll. The Scottish administration retained several means of raising money evolved in the previous 20 years, such as the land tax on annual valued rent, and the excise. The bishops who had enjoyed power under Charles I, and, more particularly, the ministers of the church who had exerted much influence thereafter, had sometimes been social upstarts in the eyes of the nobles: aristocratic rule was now firmly reestablished. Nobles therefore had little temptation to place themselves at the head of ecclesiastical opposition to the crown when it arose. Charles, on the whole wisely, kept his Scots servants on a slack rein, and John Maitland, earl (later duke) of Lauderdale, the chief of them from the mid-1660s, made able efforts to hold together the various political and ecclesiastical factions. But Lauderdale's distribution of government patronage, and his and his colleagues' greed, provided ample scope for criticism by a growing parliamentary opposition. The Scots Parliament provided, as never before, an arena for genuine debate.

The ecclesiastical opposition was socially and geographically restricted, but intractable. In 1662 Charles formally restored church government by bishops, but these were to act in association with synods and presbyteries, much as under James VI's compromise. Private conventicles were forbidden. Charles seems not to have been moved by rancour toward the Covenanters, who had bullied him in the early 1650s, but merely by a desire to achieve the system that satisfied most people. Many laymen accepted his system, and few nobles opposed it. Approximately 270 ministers, however—just over a quarter of the total—were deprived of their parishes for noncompliance. The Pentland Rising (1666) was easily put down and was countered by an experimental period of tolerance by the government. The Act of Supremacy (1669) reminded bishops as well as Presbyterians that episcopacy was regarded simply as a means to royal control of a church at peace, and about half the "outed" ministers accepted Indulgences (1669 and 1672) that permitted them to resume preaching legally. Persons who still persisted in attending conventicles were strong only in the southwest and to some extent in Fife, and among the small lairds and common people. These men adhered to the "Protester" position, regarding Scotland as still bound by the Covenants, and having no respect for majority opinion. In another trial of strength with the government they were defeated at Bothwell Bridge (1679). The remnant of Cameronians (*q.v.;* from Richard Cameron, who in 1680 at Sanquhar declared Charles II deposed) remained in being, meeting governmental violence with further violence, and in 1690 refused to join a Presbyterian but uncovenanted Church of Scotland (*see* below). Their brave and fanatical "thrawnness" endeared them to later generations of Scots.

Charles's brother James, duke of York, was commissioner to the Scottish Parliament of 1681, and in 1685 succeeded as James VII of Scots and James II of Great Britain and Ireland (1685–88). Most Scots showed, by taking a self-contradictory religious test oath (1681) and by the parliamentary grant to him of the excise in perpetuity (1685), that they were prepared to support him despite his Catholicism. But he showed his ineptitude by requesting Parliament to grant toleration to Catholics (1686); this caused a near-defeat in the Committee of the Articles and stirred up unprecedented opposition to royal wishes in Parliament itself. James proceeded to a limited grant of toleration (February 1687) by use of the prerogative alone; but, though he went on to make the toleration a general one (June 1687), he merely aroused resentment among his Episcopalian supporters without gaining any compensating Presbyterian gratitude. Nevertheless, although many exiled Scots were at the court of William of Orange in Holland, the collapse (1688–89) of James's regime in Scotland was entirely a result of the Revolution of 1688 (*see* ENGLISH HISTORY) in England and the landing there of William.

4. Culture.—There had been a notable scholarly outpouring under Mary and James VI: George Buchanan (d. 1582) was the greatest European latinist of his day. Thereafter, Scots cultural streams ran deep but narrow. Religion and politics received disproportionate attention, and the cleverest men tended to be forced into public action or public writing. In this context Sir Thomas Urquhart of Cromarty (d. 1660), an eccentric polymath who achieved a near-perfect translation of Rabelais, is best seen as a late survivor of the Jacobean era. Samuel Rutherford (d. 1661), a Presbyterian controversialist, appeared at his best in private letters of spiritual guidance to his friends. James Dalrymple, 1st viscount of Stair, while eminent as a statesman, was above all one of the world's greatest legal theorists, giving Scots law its philosophical foundations in his *The Institutions of the Law of Scotland* (1681). Before 1707, the "Scottish Enlightenment" was dawning. In architecture the first signs were early: the reconstruction of the Palace of Holyroodhouse in Edinburgh during Charles II's reign, and the erection of Hopetoun House in West Lothian, begun in 1698. The Advocates' Library was founded in 1682 in Edinburgh, and a physic (*i.e.,* botanical) garden had been opened there in 1670.

J. THE ERA OF UNION

1. The Revolution Settlement.—James VII (d. 1701) having fled to France, William, now William III in England and Ireland, summoned (March 1689) a Convention of Estates (really the same assembly as Parliament, but meeting less formally). It listed James's misdeeds in the Claim of Right, and gave the crown jointly to William (II of Scots; 1689–1702) and his wife Mary II (1689–94), James's daughter. A series of crises throughout William's reign exposed his total lack of interest in Scotland and placed a strain on the system which had grown up whereby the Scottish ministry took orders not only from the monarch but also from the English ministry. But William's first major decision was probably right: episcopacy was abolished in 1689, and Presbyterianism reestablished the following year. As in 1662, the decision was taken on political grounds. More than half the clergy were forcibly ejected by their parishioners, or were formally deprived;

KENNETH SCOWEN

EDINBURGH: THE ATHENIAN MONUMENT ON CALTON HILL COMMEMORATES THE SCOTTISH PHILOSOPHER DUGALD STEWART

but against this must be set purges in the opposite direction while James had been commissioner and king. Jacobite resistance, confined to the Highlands, was insignificant after the defeat of its adherents at the Battle of Dunkeld (August 1689). Nevertheless, the government used tardiness to take the oath of allegiance as the pretext for having more than 30 of the MacDonalds of Glencoe treacherously murdered in 1692. The indignation this caused was unprecedented for such an event, but much of it was synthetic, whipped up by a parliamentary opposition who were loyal to William but factious and anxious for office, and whose campaign reached its climax in 1695. The patronage available to the government helped to ensure that it was never defeated, but debates in the Scots Parliament, which had been liberated by the abolition of the Committee of the Articles in 1690, became increasingly stormy. Further serious trouble resulted from Scottish attempts to end the country's economic stagnation (*see* below) by establishing (1698) a colony at Darién on the isthmus of Panama. With Spain and the climate both unfriendly, the colony was a costly failure (1700)—it is estimated that more than £200,000 and 2,000 lives were lost—and would have been so even without English hostility and refusal to give any help. But the colony had incurred this hostility by running counter to England's need to conciliate Spain in order to check France, and Scots resented the subordination of their interests to those of William's other kingdom.

2. The Union and its Results.—William had fought one war against France (1689–97), and on his death in 1702 bequeathed another (1701–13) to his successor, his wife's sister Anne (1702–14). These circumstances, following Darién, made a Union of Scotland and England seem strategically as well as economically desirable, but negotiations in 1702 found neither side sufficiently eager for agreement. That Union was achieved in 1707 is at first sight surprising, since intervening sessions of the Scots Parliament had been in a mood to break the English connection altogether. But by 1707 England's appreciation of its own strategic interests, and of the nuisance value of the Scots Parliament, was lively enough for it to offer statesmanlike concessions to Scotland and material inducements to Scots parliamentarians to accept Union. It is the astute self-interest of the Scots politicians, led by James Douglas, 2nd duke of Queensberry, and James Ogilvy, 1st earl of Seafield, more than their economic farsightedness, that impresses historians now. After all, the constitutional farsightedness of Andrew Fletcher of Saltoun, who as a Scots anti-Unionist was on the losing side, has often been dismissed as chimerical, though ideas like his, for an informal link between the two countries, underlie the modern Commonwealth.

The Union was an incorporating one: the Scots Parliament was ended and the Westminster one increased by 45 commoners and 16 peers representing Scotland. Scotland benefited by gaining free trade with England and its colonies, by the grant of a money "Equivalent" of the share of the English national debt that Scotland would assume, and by the explicit safeguarding of its national church and legal system. The Union survived the pinpricks of an English Tory government (1710–14) not wholly sympathetic to Scotland, and sometimes prompted by mischievous Scots Jacobite members of Parliament: the imposition (1711) of a duty on linen exports; the granting of toleration to Scots Episcopalians (1712); the restoration in the Church of Scotland of individual lay patronage (1712), that had been abolished in 1690; and the refusal of the House of Lords to let the duke of Hamilton sit by virtue of his British title of duke of Brandon. After this, and after Queen Anne's death in 1714, when the Jacobites missed their best opportunity, the worst crises of the Union were past.

Eighteenth-century Scotsmen were preoccupied with the fact of Union, and often spoke of "North Britain" rather than "Scotland." Englishmen were less self-conscious, but resented the number of ambitious Scots in England. An unusual display of Scottish feeling took place after the Porteous Riot of 1736 in Edinburgh (*see* Porteous, John), when in the House of Commons all but one of the Scottish members, who were usually subservient to the government of the day, voted against the ministry's attempt to punish Edinburgh for the riot.

Politically and administratively, the Union brought Scotland absolutely into dependence on decisions taken in London. Some Scots politicians learned this lesson well, and gained a measure of real power by making themselves intermediaries between the government and Scots who sought patronage from it: Archibald Campbell, earl of Islay (later 3rd duke of Argyll) and Henry Dundas, 1st Viscount Melville, were the most notable of these Scottish "managers." Dundas' position, with East India Company patronage at his disposal and William Pitt the Younger as his personal friend, was especially strong. On the other hand, some Scots politicians made themselves disagreeable by trying to use their ministerial influence in Scotland to gain more power within the ministry. To avoid giving any single Scots politician an entrenched position, therefore, the formal channel for transacting Scottish business, the third secretaryship, was allowed to lapse in the mid-18th century, and for over a century routine government business in Scotland was normally the concern of the lord advocate, the chief law officer. Duncan Forbes of Culloden, a law officer and judge, was outstanding among administrators for his fostering of trade and industry. One of his last official concerns was the ending of heritable jurisdictions in 1747, after the failure of the rebellion of 1745 (*see* below), with compensation paid to loyal subjects who had enjoyed jurisdictions. This was probably in intention mainly a piece of patronage, since the Highland chiefs had not had to depend on jurisdictions for their power, and in the Lowlands the more important jurisdictions had fallen into desuetude.

3. The Economy.—The Scots economy had made few spectacular advances in the 17th century, in spite of encouraging legislation, exemplified in the Parliament of 1681. The linen, coal, and cattle export trades gradually increased, and an ever greater proportion of these and other exports went to England. By the 1690s many of Scotland's traditional continental trading links were declining. England's wars caused disruption; and mercantilist-minded states increased their competition and raised their tariff barriers. In good years the Scots economy was still self-sufficient. The famine of the "ill years" (1695–99) of bad weather and bad harvests gave part of the psychological impetus for Darién: in the end, Darién prompted Scots in spite of themselves toward Union. In the absence of Union, Scottish industrialization could scarcely have taken place without investment by some continental power, say France—a situation England could hardly have tolerated.

The Scottish economy in the early 18th century remained backward in many ways. English customs and excise, and land tax yields before the Union were about 38 times greater than Scotland's, while the population ratio was only five to one. The discrepancy cannot be ascribed wholly to the more primitive nature of the Scottish governmental and revenue-collecting machinery. The economic benefits prophesied by some advocates of Union did appear, but only slowly. There was no immediate influx of capital on a scale large enough to boost the stagnant Scottish economy. Gradually, however, farming was revolutionized by enclosures that ended the old "run rig" (strip) pattern of fields, by crop rotations, and by long leases granted to "improving" tenant farmers. Gradually too the cattle and linen trades, and the entrepôt trade in American tobacco that was brought to Glasgow and thence taken to the Continent, made some Scotsmen prosperous. New banks were founded, and they joined the Bank of Scotland (founded 1695) in making available the capital necessary for expansion. In 1759 the quickening pace of development was symbolized by the founding of the Carron Iron Company, near Falkirk.

K. Aspects of Change

1. Jacobitism: the Highlands.—The Jacobites were seldom more than a nuisance in Britain. An expedition from France in 1708 and a West Highland rising with aid from Spain in 1719 were abortive; bad leadership in the rebellion in 1715 (known as "the Fifteen") of James VII's son, James, the Old Pretender (*q.v.*), and divided counsels in that of 1745 ("the Forty-Five") led by the Old Pretender's son Charles Edward, the Young Pretender (*q.v.*), crippled invasions originating in France that had in any case less than an even chance of success. The government was not always sufficiently prepared against invasions, but the generalship of John Campbell, 2nd duke of Argyll, at Sheriffmuir in 1715 sufficed to

check, and that of William Augustus, duke of Cumberland, at Culloden in 1746 to give the coup de grace to a Jacobite army. The Jacobites never had full French naval and military assistance, and support in Scotland itself was limited: not many more Lowland Scots than Englishmen loved the Stuarts enough to die for them. Many politicians, especially before 1714, corresponded with the royal exiles simply as a matter of insurance against their return, and in the dying days of Stuart hopes there were fewer people than there have been since who were struck by the romantic aura surrounding Prince Charles Edward, though the Highlanders' devotion in sheltering him after the Forty-Five still overcomes the historian's cynicism. A few extreme Presbyterians were prepared to trust a Catholic king rather than an uncovenanted Protestant one. Many northeastern lairds, often Episcopalian and sometimes perhaps subconsciously uneasy at the way in which the increasing commercial prosperity of the Forth-Clyde region was overshadowing their still primarily agricultural region, were ready to turn out. But in the main the Stuarts had to rely on the clans of the Gaelic-speaking regions, and Highland support in itself alienated Lowlanders. Not all Highlanders were "out" in the Fifteen or the Forty-Five: such clans as the Campbells and Munros, Macleods, and Macdonalds of Sleat were Hanoverian because Presbyterian, or through their chiefs' personal inclinations. Many clans were, however, Catholic or Episcopalian, and favoured a Catholic monarch: they were legitimists and reasonably so, since both James VII and his son James the Old Pretender appreciated Highland problems. These were the problems of an infertile land, overpopulated with fighting men who owed personal allegiance to their chiefs and were partly dependent on plunder to maintain their standard of living. It is hard to see what in the end could have happened to this society, other than what did happen: a series of attempts by the chiefs to emulate the new capitalist agriculture of the Lowlands, thus creating an impersonal cash relationship with their tenants and leaving those who were redundant in the new economy no alternative to moving south or overseas. But the catastrophe of the Fifteen and Forty-Five made this process more rapid and more painful. This is the central fact of the situation, even though the atrocities of government soldiers and the repressiveness of government legislation did very much less than economic and social forces to usher in the new order.

Many lairds in the West Highlands and Isles, who staved off depopulation of their lands in the late 18th century by encouraging the kelp industry, were ruined when it collapsed in the decade 1815–25. Other landowners introduced sheep, sometimes forcibly removing their tenants, in the "Highland Clearances," as agents of the Sutherland family did in Strathnaver, Sutherland, about 1810–20. By the 1880s Highland tenants or "crofters" faced a new problem. Deer forests had replaced sheep runs as the most immediately profitable land use open to landowners, and high rents were asked for the land that was still worked as crofts, even though common grazing land was at the same time, in many cases, taken away. Parliamentary agitation by the crofters, who voted for the first time in 1885, and by their Lowland sympathizers, and also sporadic outbursts of violence beginning in 1882 (the "Crofters' War"), secured an act of 1886 which gave the crofters security of tenure and empowered a Crofters' Commission to fix fair rents. Unlike their Irish counterparts, the Highlanders sought not ownership of their land but the imposition of certain standards of conduct and responsibility upon their landlords. The Highland problem remained intractable: the number of Gaelic speakers steadily declined. In the mid-20th century, however, forestry, tourism, and some industrial enterprise began to diversify the economy (*see* below, section VIII, *Economy*).

(Left) Distillery producing Highland malt whisky, at Glenlivet. (Below) Sheep dog trials at Edinburgh

TOM HOLLYMAN FROM PHOTO RESEARCHERS

2. The Scottish Enlightenment.—No straightforward connection can be drawn between the Union and the exceptional 18th-century flowering of intellectual life known as the "Scottish Enlightenment." However, absence of civil strife permitted the best minds to turn, if they chose, from politics and its 17th-century twin, religion; and few of the best minds from 1707 onward were in fact directly concerned with politics. Philosophy, in which 18th-century Scotland excelled, was a proper concern for a country where for generations minds had been sharpened by theological debate. Scottish culture remained distinctive, and distinctively European in orientation. The historian and philosopher David Hume sought to remove Scotticisms from his speech, and the architect Robert Adam gained extra experience as well as income from being able to design buildings in London as well as in Edinburgh. Nevertheless, Adam drew most of his stylistic inspiration from the classical architecture he had studied in Italy, and Hume, "le bon David," was an honoured member of continental polite and intellectual society. Hume's *History of England* (1754–62) made his literary reputation in his lifetime; but it is his philosophical works, such as his *Treatise of Human Nature* (1739–40), that have caused the continuous growth of his reputation since his death, till he is seen as one of the greatest of all Scots. Adam Smith, author of *The Wealth of Nations* (1776), was the philosopher of political economy as Stair had been the philosopher of law. Henry Home, Lord Kames, may be singled out from many other significant figures to illustrate the versatility characteristic of the times. He was a judge, interested in legal theory and history; an agricultural reformer in theory and practice; a Commissioner of the Forfeited Estates (of the rebels of 1745); and a member of the Board of Trustees for Manufactures (which encouraged Scottish industries, notably linen). In poetry, there was a reaction, possibly against Union, and certainly against assimilation with England: revived interest in Scots vernacular poetry of the past was the herald of a spate of new vernacular poetry, culminating in the satires of Robert Fergusson and the lyrics of Robert Burns. Some of the greatest Gaelic poets, such as Alexander MacDonald, were writing at this time too.

The Scots educational system, its foundations so securely laid throughout the previous century, made possible, though neither it

nor any other single factor could be held to explain, this extraordinary cultural outpouring. The Scottish universities enjoyed their heyday, with Edinburgh notable for medicine and preeminent in most other subjects. Gradually the regents who taught students throughout their university course were replaced by professors specializing in single subjects. That students seldom troubled to graduate was little disadvantage in an age when appointments depended on patronage; and, not being bound by a rigid curriculum, they were able to indulge the Scot's traditionally wide intellectual curiosity by attending lectures in a variety of subjects. Scientific study was encouraged, and practical application of discoveries given due place. Francis Home, professor of *Materia Medica* at Edinburgh, studied bleaching processes and plant nutrition; and James Watt, instrument-maker to Glasgow University for a time, was there encouraged to work on the steam engine, to which he was to make crucial improvements.

3. Moderatism.—Moderatism was the name given to the climate of opinion prevailing in the 18th-century Church of Scotland. Partly in reaction to the amount of blood spilled in previous times over ostensibly religious issues, many ministers paid more attention to intellectual accomplishments or social graces than to the fruits of the spirit. At its best, this tendency produced William Robertson, historian, clergyman, and principal of Edinburgh University, a man of rare learning and tolerance. At its worst it bred callousness to the people's spiritual and material needs, and ultimately produced another reaction of opinion. This reaction was delayed because many of those who still wished to debate theology, to maintain severe moral standards, or to appoint ministers without reference to lay patrons made perhaps a tactical mistake in adhering to one or other of two secessions (of 1733 and 1761) from the Church of Scotland, thus leaving the field clear for their opponents. The older and more rigid of the Secession Churches ultimately itself split into four.

4. The American Revolution.—The American Revolution (1775–81) had important repercussions for Scots. Many Scots emigrants, especially Highlanders, were involved; some fought for Britain against the colonists. The tobacco merchants, who had played the greatest single role in the development of Glasgow and the west of Scotland, were prompt in 1776 in transferring their money to other trade routes, notably that bringing sugar from the West Indies, and to industries, notably cotton manufacture. This industry benefited from the expertise gained in linen manufacture, and, remaining especially important in the west, was Scotland's leading industry from about 1800 to 1830. Events in North America, and then the Revolution in France (from 1789) led Scotsmen to question the oligarchic basis of the British constitution. Self-perpetuating and corrupt town councils returned 15 of the 45 Scots members of Parliament; the franchise for the counties, where the other 30 commoners were elected, was based on ownership of land, and was in theory broad enough to give the lairds a say. But the feudal basis of Scots conveyancing gave antiquarian lawyers ample scope to manufacture votes in the interest of a few nobles, who thus determined the elections.

5. Politics.—Agitation for constitutional change was considered treasonable by many conservatives during the years (1793–1815) when Britain was fighting revolutionary France. Several advocates of universal suffrage, including a young Glasgow lawyer, Thomas Muir of Huntershill, were sentenced to transportation by Robert MacQueen, Lord Braxfield, in 1793. After repression had broken this first radical wave, postwar industrial depression produced another: the "Radical War" of 1820, an abortive rising of some workers in the Glasgow area. Intellectual campaigning of a more moderate sort had greater short-term success, and the *Edinburgh Review*, founded in 1802 by a group of young lawyers led by Francis Jeffrey and Henry Brougham, was influential in radical politics and in literature. Edinburgh life was particularly brilliant during the years of the war with France, when Englishmen as well as Scots, unable to study abroad, found Edinburgh University more attractive than ever. Outstanding in this period was Sir Walter Scott, although not till 1827 was he known to be the author of the Waverley novels. Scott's greatness as a novelist lay in the way he took Scottish society as a whole for his main character, and his best books are a lament for an era he knew was dying, the organic society of preindustrial Scotland.

6. The Disruption.—The French Revolution, seen as the ultimate product of the intellectualism and skepticism of the 18th century, stimulated an Evangelical reaction against Moderatism within the Church of Scotland. The Evangelicals also felt Moderatism had no answer to the material and spiritual squalor of the industrial towns. Many Evangelicals (including over a third of the ministers) broke with the Church of Scotland in the Disruption of 1843: they were determined to do without the benefits of Establishment so long as the state upheld patronage and other institutions that were believed to hinder a radical reshaping of the Church. Those breaking away, led by Thomas Chalmers, formed the Free Church of Scotland. More than the 18th-century secessions, this much bigger defection stung the Church of Scotland into reexamining its own conscience. Less beneficially, it produced a wasteful competition in building churches and made impossible Chalmers' dream of a voluntary system of poor relief run by the Church. The Scots opponents of voluntary poor relief were thus greatly strengthened; and, led by W. P. Alison, an Edinburgh medical professor whose work during epidemics had determined him to help the poor obtain better food as a means of disease prevention, this faction brought in an act (1845) making it almost compulsory for a parish to levy an assessment for poor relief. In strict statutory theory, however, the able-bodied unemployed remained outside the scope of this relief till the Poor Law Emergency Provisions (Scotland) Act of 1921.

The Catholic and Episcopalian Churches had never really gained the deepest spiritual allegiance of the Highlanders: this was shown in the later 18th century, when a conservative and morally strict brand of Presbyterianism became rooted in the northwest. Many Highlanders left the Church of Scotland in 1843 for the Free Church, where they felt they perceived more Evangelical fire. The most important development among the smaller (and mainly Lowland) Presbyterian sects was the emergence (*c.* 1800) of Voluntaryism, a belief—new in Scotland—in denying oneself as well as one's opponents the benefits or disadvantages of Establishment.

L. 19th-Century Scotland

1. The Industrial Revolution.—The Scottish Industrial Revolution was in full swing from the 1820s. Linked with this, partly as cause and partly as effect, in a way historians have not altogether disentangled, was a dramatic upsurge of population. There were perhaps about 1,000,000 people in Scotland in 1700. By 1800 there were more than 1,500,000 and by 1900 nearly 4,500,000. The manufacturing towns showed spectacular increases. Hundreds of thousands of Irish emigrants went to Scotland in the 19th century, notably during the Irish potato famine of 1846–50. In some country regions there was a population decrease as people moved to the towns, to England, or overseas. Part of the overall increase was the result of improved medical care that had lessened the ravages of epidemic diseases by the mid-19th century. Much of the food for the increased population was supplied by progressive Scottish agriculture: farming in the southeast was celebrated throughout Britain for its efficiency in the early 19th century, and the northeast became famous for its beef cattle, Ayrshire for its milking herds.

But the key advance was in heavy industry, which from about 1830 took the industrial primacy from textiles, at a time when industry as a whole had replaced agriculture as the nation's chief concern. Coal production rose, as did that of iron, with James Beaumont Neilson's hot-blast process (1828) making Scottish ores cheaper to work. Major canals, like the Forth and Clyde, completed in 1790, enjoyed a short boom before being rendered obsolescent by the railways, of which the Glasgow to Garnkirk (1831) was noteworthy for using steam locomotives (rather than horses) from the start. Above all, Scottish international trade was catered to, and Clydeside's reputation made, by the building of ships. Robert Napier was the greatest of many great Scots marine engineers.

As far as measurable factors like real wages and dietary standards are concerned, there is little agreement among historians

on the immediate effects of industrialization on the lives of Britain's workers. But there were undoubtedly deep psychological disturbances inherent in the move to squalid industrial towns and the submitting to monotonous and inexorable factory disciplines; and the Scottish Industrial Revolution, which occurred later and more swiftly than that of England, produced by the 1840s a class-conscious and militant proletariat. The workers' grievances were real enough. The controversy over voluntary or state poor relief, not resolved till 1845 (*see* above), delayed developments in other directions. Only in 1867 was a comprehensive Public Health Act passed. The overcrowded and filthy slums, consisting often of buildings unsatisfactory from the start, nevertheless acquired an awful semipermanence through the easy availability of building stone. Such neighbourhoods as the Gorbals in Glasgow, and the part of Edinburgh described in G. Bell's *Blackfriars' Wynd Analysed* (1850), became notorious. Improvement was made, for the proportion of the population living in one-room houses fell from 27% in 1861 to 8.7% in 1911, though the proportion of people (about 40%) living in two-room houses remained constant. In the first half of the 19th century epidemics had been common and deadly, cholera claiming more than 10,000 victims in 1832. Improvement in this respect was to some extent balanced by the fact that it was as late as 1897 that infant (first-year) mortality reached its peak of 138 deaths per 1,000 live births.

Development of working-class cohesiveness was uneven in different places and occupations. Cotton spinners and weavers were forming unions about the time that coal miners were freed by acts of Parliament (1775 and 1799) from being tied to their places of work. Trade unions, whose previous illegality in England was confirmed by the Combination Acts of 1799–1800, achieved the same negative status in Scotland for the first time by a series of judicial decisions culminating in 1813—a mark of public recognition if nothing else. The ban was lifted in both countries in 1824, although various legal disabilities remained. Cooperative movements ranged from the practicality of retail stores to the idealism of the community at Orbiston near Motherwell (1826–27) that was inspired by the work of the reformer Robert Owen.

Scottish Chartism was a vigorous political movement of the workers in the decade from 1838, especially in Glasgow. It stressed moral rather than physical force, and sought to pursue the long-term aim of amelioration of living and working conditions by means of education and social reform.

2. Politics and the Church.—An installment of parliamentary (1832) and burgh (1833) reform ended fictitious county votes and corrupt burgh caucuses, but it disillusioned the working classes by failing to give them the vote. As in England, they had to await the 1867 and subsequent Reform Acts. But the great bulk of the Scots middle classes were delighted with the Whigs (whose lord advocate from 1830 till 1834 was Francis Jeffrey) who had brought the reforms. The Whig Party, or Liberal Party (as it became known in the 1860s), dominated Scottish mid-19th century politics; and William Ewart Gladstone, of Scots parentage, was the great Liberal hero, whose moral dynamism and fire far outweighed in Scottish eyes his High Church Episcopalianism.

Benjamin Disraeli and the Conservatives wooed Church of Scotland opinion, in the short run unsuccessfully, by abolishing patronage in 1874. Nor did this development deter the Free Church, which still officially approved the principle of Establishment, and the United Presbyterian Church, formed by some of the sects in 1847 on an explicitly voluntaryist basis, from campaigning in the 1880s for the disestablishment of the Church of Scotland. There were general elections in 1885 and 1886. In the first of these some voters showed they feared that the Church of Scotland was in danger and that the Liberals would not save it. The following year many more erstwhile Liberals had become convinced by Gladstone's conversion to Irish Home Rule that his party was acquiescing in the destruction of Protestantism. The result was that, of the 60 Liberals elected in 1886 (out of a total of 72 members provided for Scotland in the 1884–85 Reform Act) 17 adhered to the Conservatives because of their common desire to maintain the Union between Great Britain and Ireland. The group's successors ultimately joined the Convervatives, who in Scotland are still generally termed "Unionists." The Liberals retained their majority of Scottish seats, but further defections in the 1890s, by working-class militants to the left, and by working-class and other jingoists to the right, resulted in the unprecedented event of the party's retaining only a minority of Scottish seats in the 1900 general election. In Church affairs 1900 saw the creation of the United Free Church, by union of the Free and United Presbyterian Churches on an explicitly voluntaryist basis. Some Free Church members, especially in the Highlands, broke away.

3. Economy in the Late 19th Century.—The Scots economy enjoyed the full benefits of mid-Victorian prosperity. By the 1880s, however, more difficult times had come and yet had failed to provoke an adequate response. Imports of wheat from North America from the 1870s onward, and of refrigerated meat thereafter, hit Scottish farming; but because Scots soils and climate had made Scots farmers adaptable, and ley (lea) farming common, the agricultural slump was less severe than in many other parts of Britain. Other countries were overtaking Britain's industrial lead, and Scots inventiveness was less in evidence than a century before. The Scots economy was concentrated on a dangerously narrow range of heavy industries, with shipbuilding dominant. Steel was becoming important, and because Bessemer steel was not thought suitable for ships, Scots steelmakers abandoned it almost entirely in favour of the open-hearth process. This decision narrowed the scope for steel-using industries in Scotland, and had repercussions as late as 1932 when it was the main factor behind the move of the firm of Stewarts and Lloyds to England. Because the ironmasters' demand for coal declined, a dangerously high proportion of coal production had to be exported.

4. Move to the Left.—Trade unions of skilled workers had had an uninterrupted existence since the mid-19th century. By the 1880s groups of unskilled workers were being organized. Various factors delayed the permanent organization of the miners till this period too, when there emerged from their ranks a major leader, James Keir Hardie. Failing to engage the Liberals sufficiently in support of organized labour, he helped form the Scottish Labour Party in 1888. In 1893 he created the Independent Labour Party (I.L.P.) for Britain as a whole, and this body in 1900 federated with the trade unions for the purpose of running the Labour Party (given its present name in 1906). Scottish political opinion moved left in the years before 1914, with Liberal fortunes reviving—partly because of the able leadership (1899–1908) of Sir Henry Campbell-Bannerman, a Scot—and three Labour M.P.s being elected.

M. Modern Scotland

1. World War I and After.—The war of 1914–18 had a great impact on Scots society, with 74,000 lives lost, and industry mobilized as never before in a coordinated national effort. From 1915 the Clyde Workers' Committee stubbornly opposed government action to modify established standards of pay and conditions for the duration of the war. The committee was the most active group of its kind in Britain, and this industrial militancy survived into the 1920s: the Scots trade union leaders did not, in the General Strike of 1926, lag behind their rank and file in resoluteness as did their English counterparts. The collapse of the wartime boom in 1920 began a period of economic depression in Britain, in which Scotland was one of the worst-affected regions. The war had disrupted the export pattern for staple Scottish products such as coal, and left behind it surplus world productive capacity for such traditional Scottish manufactures as steel and ships. At one point late in 1933 more than 30% of the insured workers in Scotland were unemployed. The seriousness of this is brought home when it is remembered that, already, in the decade 1921–31, nearly 400,000 people, out of a population of fewer than 5,000,000 had gone in search of work to England or overseas. Statistics are powerless to reveal the demoralization of the unemployed. The 1930s was a time of gradual improvement, with work on Clydeside's great ship, the "Queen Mary," being resumed in 1934 after more than two years' suspension.

Economic distress bred political radicalism. The Liberals were eclipsed, and in most seats the real contest was between the Union-

ists and Labour, which became Scotland's biggest single party for the first time in the election of 1922. Few Scots took their radicalism as far as did Willie Gallacher, in whose political education Lenin took a personal interest, and who was Scotland's only notable Communist M.P. The death (1930) of John Wheatley, who had been minister of health in the first Labour government (1924), deprived left-wingers in the Labour Party of a skilled leader, and counsels of moderation in the party prevented its taking any very distinctive initiative on the economic crisis. Ramsay MacDonald, a Scot who had led two minority Labour governments, agreed to form a national government in 1931. The Labour Party refused to participate, disowned MacDonald, and was heavily defeated at the polls, in Scotland as elsewhere. This Labour Party independence of mind was insufficient for the I.L.P., now largely Scottish and largely leftist in membership, and in 1932 it imitated the religious seceders of the previous two centuries by walking out of a body it had failed to dominate. Another political product of economic distress was the Scottish National Party, formed in 1934. It showed more literary than political vigour before World War II, but afterward was well run if small in scale.

2. World War II and After.—During World War II Scotland sustained some 34,000 deaths in action and 6,000 civilians were killed, many in air attacks on Clydeside. The outstanding Scot on the home front was probably Tom Johnston, a Labour M.P. who acted as secretary of state for Scotland in the wartime national government. He was active in setting up the North of Scotland Hydro-Electric Board in 1943. After 1945 better worldwide economic circumstances and a degree of government action in certain sectors (*e.g.*, agriculture) largely prevented a second postwar slump. The economy was diversified with the introduction of a substantial motor industry centred on Bathgate in West Lothian (1961) and Linwood in Renfrewshire (1963). An atomic reactor was built in Caithness and two atomic power stations in the Southwest: this helped to hasten the further decline of the coal industry. Scots industry remained vulnerable in a recession, and increasing English or overseas ownership of Scottish industrial concerns took decisions on any cutback more than ever out of Scottish hands. Slum-clearance programs and the building of new towns, and nearly full employment for most of the country for much of the period, began to bring proper minimum standards of life within the reach of almost all. Unionists in Scotland failed to match the Conservatives' dominance in England in the 1950s, and Labour strength in industrial centres was accompanied in the 1960s by signs of Liberal revival in the countryside.

3. Scottish Culture and Institutions.—Scottish culture was, for most of the 19th century, provincial in all the worst senses. There were some exceptional figures. Thomas Carlyle, historian and essayist, was a typical Scot of an older type in his philosophical bias and openness to continental influences. Robert Louis Stevenson was a novelist with a fine intuitive understanding of Scotland's past. Most other outstanding figures were in some ways isolated men, like Sir Patrick Geddes, a professor of botany who was better known as a pioneer of town planning, and Charles Rennie Mackintosh, the architect of the seminal Glasgow School of Art—men who were more acclaimed abroad than in Scotland. The 20th century brought a notable revival. While the dramatist James Bridie should be mentioned, poetry was the most fruitful field. Some writers sought a synthetic Scots language by drawing on words from all the regional dialects. Others revitalized a living language, Gaelic. Younger men, laying less stress on Scottishness, did good work in English. But the major figures were the Gaelic poet Somhairle Macghill-Eathain, and the polymath Hugh MacDiarmid. MacDiarmid (C. M. Grieve), a nationalist and Communist, was a poet of stature who wrote superior verse in a synthetic Scots based on various dialects and literary sources, in Lallans, and in English.

Scottish local government was progressively democratized. County councils were created in 1889, superseding for most purposes the commissions of supply that had functioned since Charles II's reign. As elsewhere, pressures existed for the creation of larger units, and gave rise to such legislation as the Local Government (Scotland) Act of 1929; but in the small burgh Scotland still retained an organ of truly local government that had spanned seven centuries. Demands for devolution of government within the United Kingdom led to the creation of a Scottish secretaryship (1885), which became a secretaryship of state in 1926 and transferred its offices from London to Edinburgh in the 1930s.

Scottish education retained characteristic virtues but lost its preeminence. After 1945 there was increasing subject-specialization in the upper forms of schools. Introduction of the "leaving certificate" examination (1888), normally necessary for university entrance, and subsequent regulations, and after 1945 the pressure of numbers, narrowed the doors of Scottish universities that were traditionally open to all who could stand to benefit by their teaching. In the universities, the Universities (Scotland) Act of 1889 provided a framework for higher scholarly standards at the undergraduate level at the expense of the traditional breadth of learning. These changes were the result of uncritical emulation of English models, as well as a response to the irresistible pressures of an ever more competitive world. One answer to educational challenges was the decision in the 1960s to create four new universities: at Edinburgh and at Glasgow by upgrading colleges of technology, at Dundee by giving autonomy to a university college, and at Stirling by a new foundation.

Scottish Church life remained distinctive. The Church of Scotland benefited in 1929 from union with all but a small part of the United Free Church. This was on a basis of partial disestablishment that still left the church protected by a royal oath taken at the coronation, and aided by a levy on land. The Catholic Church was the second largest, mainly because of the large-scale Irish immigration of the 19th century. These and many smaller denominations kept the proportion of active churchgoers in the population higher than in many Western European countries.

Scots law remained one of the nation's least popularly acclaimed yet most lively institutions. Appeal from the Court of Session to the House of Lords had existed since 1711 (though not specifically provided for in the Act of Union): through this door alien doctrines sometimes entered Scots law, though the danger was decreased by the creation in 1876 of Lords of Appeal, among whom there have always been Scots lawyers. Nor did legislation always take Scots law properly into account. Yet Scots law survived, and its position was recognized in 1965 by the creation of a separate Scottish Law Commission to examine legal reform. *See* also SCOTLAND, CHURCH OF; SCOTS LAW; SCOTTISH LITERATURE; and biographies of kings and notables mentioned. (J. M. SI.)

VI. POPULATION

At the census taken in 1961 the population of Scotland was 5,179,344 (females 2,696,610). This was an increase of 1.6% on the number enumerated at the previous census of 1951 and continued the series of intercensal increases recorded at every census except 1931, when a decrease of 0.8% on the 1921 figures was recorded. Since the mid-19th century the population has approximately doubled. At the 1841 census the total population enumerated was 2,620,184, which represented an increase of 1,000,000 on the total at the first official census in 1801. At the end of the 15th century it is estimated that the population did not exceed 500,000—Edinburgh having about 20,000 inhabitants, Perth about 9,000, and Aberdeen, Dundee, and St. Andrews about 4,000 each.

Since 1861 there has been a net emigration from Scotland to other countries of about 1,900,000, of whom about 1,400,000 have emigrated since 1900. In 1921–30 emigration was over 390,000, or about 8% of the population at the beginning of the decade. Because the age groups affected were in the lower range, the consequential loss was greater than the numerical total suggests. In the next decade and during World War II the outward flow of migration was checked, and in the years 1931–51 the net loss was about 220,000, mainly to England. After 1951 it averaged about 25,000 a year.

There has also been an appreciable inflow of population. At the 1931 census 346,050 of the population (7.14%) were born outside Scotland. In 1961 the number was 415,001, of whom 246,917 were born in England and Wales, 37,308 in Northern Ireland, and 39,670 in the Republic of Ireland.

If the population was evenly distributed throughout Scotland,

it would have amounted in 1961 to 170 per square mile. Actually, by far the largest part of the country had a population density of less than one person per square mile, the maximum degree of concentration being found in the highly industrial area lying approximately between the upper reaches of the Firth of Clyde and the Firth of Forth. A secondary but much smaller concentration is found along the southwestern and the eastern coasts. During the period 1931–61 there was a slight loss of population in most of the northern and border counties, and a more than compensating gain in the central industrial area. The population in the different counties is shown in the table.

Area and Population of Counties in Scotland
(1961 census)

Counties		Area (sq.mi.)	Population
Total		30,405	5,179,344
1. Northern:	Zetland (Shetland)	551	17,812
	Orkney	376	18,747
	Caithness	686	27,370
	Sutherland	2,028	13,507
2. Northwestern:	Ross and Cromarty	3,089	57,642
	Inverness	4,211	83,480
3. Northeastern:	Nairn	163	8,423
	Moray	476	49,170
	Banff	630	46,454
	Aberdeen*	1,971	321,783
	Kincardine	382	25,564
4. East Midland:	Angus*	874	278,399
	Perth	2,493	127,056
	Fife	504	320,692
	Kinross	82	6,702
	Clackmannan	55	41,394
5. West Midland:	Stirling	451	194,876
	Dunbarton	241	184,559
	Argyll	3,110	59,390
	Bute	218	15,170
6. Southwestern:	Renfrew	225	338,872
	Ayr	1,132	342,822
	Lanark*	898	1,626,424
7. Southeastern:	W. Lothian	120	92,768
	Midlothian*	366	580,329
	E. Lothian	267	52,677
	Berwick	457	22,437
	Peebles	347	14,156
	Selkirk	268	21,052
8. Southern:	Roxburgh	665	43,183
	Dumfries	1,075	88,440
	Kirkcudbright	897	28,870
	Wigtown	487	29,124

*Includes counties of cities.

Births (1951–60) were 950,142, averaging annually 95,014. There were 102,691 births (1963), a rate of 19.7 per 1,000. Births per 1,000 married women aged 16–44 (1963) were 145.2. The average family size of women 45–50 (1961) was 5.6; in 1911, 5.08. The average death rate (1951–60) was 12.1 per 1,000, a drop from 22.1 in 1861–70. The average marriage rate (1951–61) was 8.1 per 1,000; 6.9 in 1921–30. During and after World War II it rose to 9.4 (1945), being higher in the central industrial area than in the north or south. The Marriage (Scotland) Act, 1939, abolished irregular marriages by declaration, including "Gretna Green marriage" and "marriage before the sheriff," instituting a new form of civil marriage, to be contracted after publication of notice in the office and presence of registrars.

The populations of the largest towns (1961 census) were: Glasgow 1,055,017, Edinburgh 468,361, Aberdeen 185,390, Dundee 182,978, Paisley 95,750, Greenock 74,560, Motherwell and Wishaw 72,794, Coatbridge 53,825, Kirkcaldy 52,390.

During the period 1951–61 speakers of Gaelic declined by 14,469 to 80,978 of whom 1,974 spoke no English. The four principal Gaelic-speaking counties were Ross and Cromarty, Sutherland, Inverness, and Argyll. In 26 counties less than 1% of the population spoke Gaelic.

The Church of Scotland is the national church and the most influential and powerful religious body in Scotland. In 1961 members numbered 1,292,617. The Roman Catholic population in Scotland (including children) was estimated (1961) to be over 792,000, with marked concentrations in Glasgow and the neighbouring towns, where many Irish immigrants have settled. The Episcopal Church in Scotland had about 97,500 members in 1961. Other religious groups include the Baptists, Methodists, Congrega-

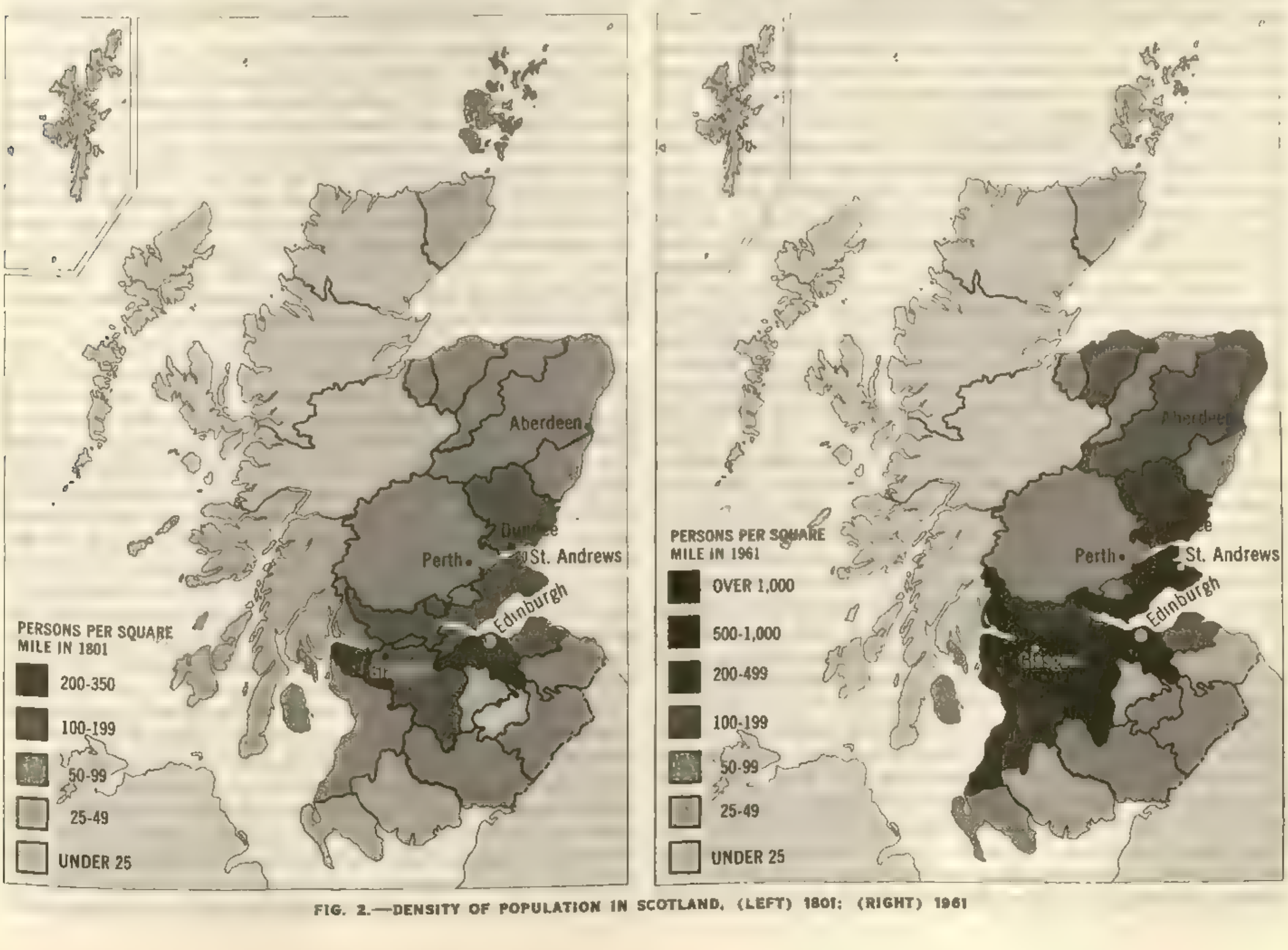


FIG. 2.—DENSITY OF POPULATION IN SCOTLAND, (LEFT) 1801; (RIGHT) 1961

tionalists, and Unitarians, while there is also a Jewish community of about 15,000, mainly concentrated in Glasgow. *See* also PRESBYTERIANISM; SCOTLAND, CHURCH OF; SCOTLAND, FREE CHURCH OF; UNITED FREE CHURCH OF SCOTLAND; SCOTLAND, EPISCOPAL CHURCH IN.

VII. ADMINISTRATION AND SOCIAL CONDITIONS

Central government administration is discussed under GREAT BRITAIN.

1. Local Government.—Scottish local government has its antecedents in the burghs, shires, and parishes of medieval Scotland. The burghs were primarily settlements of merchants and craftsmen, often with special privileges and property granted for the fostering of trade. The shire was the area over which the sheriff, as the royal representative, was responsible for law and order. The parish was centred on the church and concerned itself with the relief of the poor and with education. In the 19th century, the Industrial Revolution and the attendant huge increase in population in central Scotland brought pressing new problems of local government and at the same time a demand for more democratic control. Throughout that century there was a twofold movement: to set up town councils for new communities; and to reform the town councils of the old burghs. Elected county councils were constituted in 1889.

The present system was set up in 1929. The country is divided into 33 counties, two pairs of which—Perth and Kinross, and Moray and Nairn—are combined for certain purposes; there are also large and small burghs and four counties of cities (Edinburgh, Glasgow, Dundee, and Aberdeen); the part of the county outside the burghs is known as the landward area and is divided into districts. Local government is by town councils, county councils, and district councils. The powers of town councils vary according to the degree of importance of the burgh. In the counties of cities the town council is responsible for all the work of local government; in the 20 large burghs (broadly, burghs with more than 20,000 people) the town council performs all functions except education and valuation for rating and in a few is responsible for police; in the 176 small burghs the town council is responsible for such services as housing, water supply, and cleaning, but another large sector of local government work (including education, police, classified roads, and local health services) is the responsibility of the county council. In the landward area the county council is responsible for all the main work of local government; certain minor duties mostly concerned with amenity are entrusted to district councils. The latter may also exercise powers delegated to them by the county council.

Thus the county council as a body of representatives from the landward area of the county exercises certain local functions within that area; with the representatives of the small burghs added to it other functions are exercised within the landward area and also in the small burghs. When representatives of small and large burghs are added, the county council controls education in, and is the valuation authority for, the whole county.

Town council elections in burghs are held annually to replace the third of the council who have been longest in office. The whole of each county council is elected once in every three years. The district councils are elected along with the county council.

The cost of the services provided by the local authorities in Scotland is met by rates, government grants, and miscellaneous income, and, to meet expenditure of a capital nature, by the raising of loans. Rates are levied on the occupiers of lands and heritages on their ratable value, as shown in the valuation roll; properties are revalued every five years under the Valuation and Rating (Scotland) Act, 1956, but a new roll is made up each year to include new properties and to give effect to "material changes of circumstances" between revaluations. There are now three main kinds of government grants. Firstly, grants are still paid toward the cost of certain specific services (such as housing, roads, police, and civil defense) and are a given percentage of approved expenditure, except in the case of housing, the grant for which is largely on a unit basis. Secondly, many of the former grants for specific services, including those for education, and the health, fire, and child care services, have been replaced since 1959 by the general grant. The total general grant for Scotland is determined by reference to the latest available information of expenditure by local authorities on the relevant services, to likely fluctuations in the cost of providing them, and to the need for developing them in relation to the general state of the economy. The total grant is then distributed to the local authorities on a basis that does not relate to the actual expenditure of the individual authorities. Thirdly, there is the exchequer equalization grant, which is in aid of local authority finances generally and brings their resources up to a certain national average. The miscellaneous income of a local authority includes rents from local authority houses; fees, fines, and various charges; and profits from revenue-producing undertakings, such as transport services, water supply, markets, and slaughterhouses.

An interesting feature of burgh finance is the "common good," which originally formed the main revenue of burghs and usually consisted largely of property granted in the foundation charter or by subsequent gifts. Common good income can be used for anything that furthers the general good of the inhabitants or the dignity or hospitality of the burgh; it is available where there is no authority to levy a rate. Not all burghs have a common good. In 1947 all local authorities were enabled to set up "fee funds" from certain fees and commissions, which may be used in the same way as the common good.

In June 1963 the secretary of state for Scotland published a white paper on the modernization of local government (Cmd. 2067) suggesting that local government boundaries were becoming increasingly artificial and that many individual authorities were too small. The associations of local authorities agreed to enter into discussions on this and accepted, as a basis for consideration, a further report (*Reorganization of Local Government in Scotland*, HMSO, 1964) proposing a two-tier system of larger authorities. The secretary of state decided in November 1964 to review the question when an examination of the financing of local authorities had been completed.

Liquor Licensing.—The sale and supply of excisable liquor by retail is regulated under statute by separate licensing courts for counties and all but the smallest burghs, which grant certificates renewable annually for licensed hotels, public houses, restaurants, and off-license shops. Following the first report of the Committee on the Scottish Licensing Law (Cmd. 1217, 1960), the Licensing (Scotland) Act, 1962, fixed nationally the hours on weekdays during which liquor may be sold and introduced slightly shorter hours (not applying to public houses and off-license shops) on Sundays. Local "temperance" polls are occasionally held under statutory provisions dating from 1913 and may result in areas having a reduced number of licensed premises or none at all.

Civil Defense.—The secretary of state is responsible for most aspects of civil defense in Scotland, the departments chiefly concerned being the Scottish Home and Health departments for Scotland. The tasks of recruiting and training local divisions of the Civil Defense Corps are the responsibility of the local authorities. The strength of the corps in Scotland in the mid-1960s was over 30,000. For operational purposes the country is divided into east, north, and west zones, each under a zone commissioner, who acts under a Scottish central control with a senior minister in charge. Local authorities within the zones form groups for operational purposes. The Royal Observer Corps and the Monitoring and Warning Organization are organized in exactly the same way as in England, and the Civil Defense Corps is similarly used in peacetime emergencies; *e.g.*, floods. (*See* also GREAT BRITAIN: *Civil Defense.*)

2. Living Conditions.—*Housing.*—The grossly overcrowded and insanitary living conditions in the early years of the century were described in the report of the Royal Commission on Housing in Scotland in 1917. Local authorities had permissive power to build houses before then but did not build on a substantial scale until after 1919 when an act of Parliament following the royal commission's report made local authorities' housing powers obligatory and, for the first time, made exchequer housing subsidies available. Subsequent acts gave specific subsidies for slum clear-

ance and the relief of overcrowding, and under this impetus some 337,000 houses were built between the wars. About two-thirds of these were built by local authorities, with the result that the pattern of house building rapidly swung away from building by private enterprise. Subsidized council houses became a significant part of the social scene.

Private building for letting had been declining even before World War I as it became increasingly difficult to build new houses to let at rents that the working classes could afford and still give an attractive return on invested capital. The introduction of rent control in 1915 and the development of subsidized local authority houses hastened the decline; rent control was continued after the war. Private builders turned, with the aid of the rapidly developing building societies, to building for sale; but in Scotland the growth of owner occupation between the wars was much slower than in many parts of England, particularly the southeast and Midlands.

The more general government housing subsidies remained widely available to local authorities after World War II, but specific subsidies were introduced in 1957 to encourage the building of houses for overspill of population from congested areas and for incoming industrial workers. Other special subsidies that are now payable, *e.g.*, for flats in high blocks, and the increasing emphasis on slum clearance and redevelopment also reflected the need to take properly into account planning and economic, as well as social, objectives in housing policy, including the emergence of special needs of particular areas or of groups of people within the general housing problem.

Although government housing subsidies had been altered many times over the years, they were normally paid at a uniform flat rate to each housing authority. The Housing (Scotland) Act, 1962, represented an attempt to guide housing subsidies to those authorities with the greatest financial need; under the act subsidies are paid on a sliding scale according to a "resources" test that measures each authority's outgoings on housing account against its potential income.

The 1962 act also made loans available to nonprofit housing societies with the intention of widening the basis of housing provision by encouraging building by other agencies in addition to local authorities. This pilot scheme was followed, in the Housing Act, 1964, by the establishment of a housing corporation that, with the cooperation of the building societies, is intended to promote the building of houses for letting at cost rents or for group ownership, thus providing a source of housing to supplement subsidized council housing on the one hand and owner occupation on the other.

These schemes reflected the continued decline of private rented housing. Private building had remained at a low level, especially in the early years after World War II, when building control by licensing was strictly enforced. Licensing controls were, however, relaxed in 1952 and 1953 and abolished in 1954. Output by private builders has since increased but has been concentrated on building for sale, with virtually no building for letting, and private output has remained low in comparison with England.

The provision of new houses has remained the primary aim of housing policy, but the policy has been also to encourage the improvement of old substandard houses by improvement grants, which have been available since 1949 to private owners and to local authorities alike. From 1954 the policy has been to speed up the clearance and redevelopment of slum areas. Local authorities were required to submit three-year programs for dealing with unfit houses, and major slum clearance and redevelopment programs began. Glasgow presented a special problem of congestion, which could not be solved within the city itself, so that central development there had to be planned in conjunction with the movement of part of Glasgow's population to overspill reception areas in other parts of Scotland, including the new towns. Four of these, at East Kilbride, Glenrothes, Cumbernauld, and Livingston, are being developed under the New Towns Act, 1946. Moreover the Housing Act, 1964, gave local authorities power in certain circumstances to secure the compulsory improvement of whole areas of privately owned, substandard houses.

The rent restriction acts were relaxed in 1954 by allowing certain increases in the rents of privately owned houses, provided they had been kept in good repair. The Rent Act, 1957, allowed further increases in rent for houses in good repair and removed from control altogether any houses with a 1956 ratable value of more than £40, together with all new tenancies commencing after the act. A measure of protection of tenure for decontrolled tenants was restored on an interim basis by the Protection from Eviction Act, 1964.

In 1944 it had been estimated that 500,000 new houses were needed in Scotland. By 1964 more than that total had been built, the great majority by local authorities and the Scottish Special Housing Association (a body set up to assist local authorities by building in the areas of greatest need and financed by the exchequer). But the 1961 census had shown that more than a quarter of all Scottish households did not have the use of a fixed bath, while over 17% lacked exclusive use of water closet or toilet. Overcrowding had decreased as compared with 1951 but was still acute in some areas; in Glasgow 38% of all dwellings consisted of only one or two rooms, and in the whole of Scotland one-fifth of the total population lived at a density of more than 1½ persons per room. The increasing numbers of old people in the population, the trend to younger marriage, the desire to set up house independently of relatives, rising standards of living and of expectations had all increased the demand for separate houses of good quality so that despite the large output in the postwar years increased housing targets, rising to 40,000 houses a year in the second half of the 1960s, were announced. (*See* also HOUSING: *Great Britain.*)

Employment.—In recent years structural changes in the economy have had important effects on the pattern of employment in Scotland. Declining employment in agriculture and coal mining has been accompanied by contraction in the traditional manufacturing industries of iron and steel, shipbuilding, heavy engineering, and textiles, industries on which Scotland is still heavily dependent. Government action to attract new growth industries has helped provide fresh employment, particularly in the consumer durable and science based industries. Of 2,091,400 employees in employment in June 1963, the proportions in the main sectors of industry were: primary sector 8% (5% Great Britain); manufacturing 33% (38%); construction 8% (7%); public utilities 1% (2%); transport and communication 8% (7%); distribution 14% (13%); other services 26% (27%).

Unemployment.—Average annual unemployment in Scotland during the past ten years has varied between 2.4% and 3.8% but has remained at about twice the rate of Great Britain as a whole. The average rate during 1964 was 3.7%, compared with 1.7% in Great Britain. Heaviest unemployment occurs in the industrial belt of central Scotland, especially in Clydeside. There are also pockets of high unemployment in isolated rural areas, notably in the Highlands and islands.

3. Welfare and Health Services.—Formerly the relief of destitution was dealt with under the Poor Law, which in Scotland goes back to 1424. The act of a Scots Parliament in that year and other early acts, however, were largely concerned with merely restricting begging to those who could not win their living otherwise. But eventually a system of relief was developed based on the parish and administered largely by church authorities. The law was consolidated in the Poor Law (Scotland) Act of 1845, which remained on the statute book for nearly a century. Relief (or public assistance) consisted of "outdoor" relief, which was paid in cash, the recipients remaining at home; "indoor" relief, which was provided in poorhouses when a poor person required institutional care; and "medical" relief, also provided in poorhouses where hospital care was needed. The emphasis in Scotland was on outdoor rather than on indoor relief. In 1930 the administration of the Scottish Poor Law was transferred from about 875 parish councils to 55 large authorities made up of 31 county councils and 24 town councils of large burghs. Central supervision was exercised, in terms of the act of 1845, by a board whose duties eventually fell to the Department of Health for Scotland.

The breakup of the Poor Law in Scotland (as in the rest of the

United Kingdom) began with the Old Age Pension Act of 1908 and ended with the passing of the National Assistance Act of 1948. Under the latter act relief of poverty by cash payments is undertaken by the National Assistance Board, but the former public assistance authorities are required to provide residential accommodation for the aged and infirm who require care and attention. The act also empowers the authorities to provide welfare services for the blind, deaf, or dumb and other handicapped persons. There are well-established voluntary services for the elderly and for handicapped persons, and, in the exercise of their powers, local authorities commonly avail themselves of these services and assist the voluntary organizations financially.

Under the National Health Service (Scotland) Act, 1947, a comprehensive health service similar to that in England was set up (*see* SOCIAL SECURITY), which included: a full family doctor service with an associated pharmaceutical service; a general dental service; a supplementary ophthalmic service; all forms of hospital care and treatment, both inpatient and outpatient; specialist opinion and treatment; ambulance and blood transfusion services; arrangements for prevention of illness, care and aftercare, including care of mothers and young children, day nurseries, the visiting of homes by midwives, health visiting, home nursing and domestic help services, vaccination and immunization, and health education. A school health service is provided by education authorities under the Education (Scotland) Act, 1962, which consolidated earlier legislation. There are three main agencies for the administration of the health service. The general practitioner services are administered by 25 executive councils. Almost half of their members are appointed by the doctors, dentists, and pharmacists in each area, and the remainder partly by the secretary of state and partly by the local health authorities. There are five regional hospital boards responsible generally for hospital, specialist, and allied services, the actual day-to-day management of hospitals being entrusted to 79 boards of management. Regional board members are appointed by the secretary of state, and board of management members by the regional boards. The 31 county councils and 24 town councils of large burghs as local health authorities are responsible for the remaining services. All services are available to every member of the public without any insurance qualification. Charges to patients were introduced in 1951, 1952, and 1956 covering part of the cost of the general dental service, the pharmaceutical service, the supplementary ophthalmic service, and the supply of certain appliances to hospital outpatients; and for certain services, such as domestic help, local health authorities may make such charges as they consider reasonable, having regard to cost of the service and the person's means; in 1965 the Labour government abolished charges for pharmaceutical services and announced its intention to remove the other charges in due course. Otherwise the services are free of charge. The greater part of the costs is met by the exchequer. A further source of income is derived from weekly contributions under the National Health Service Contributions Acts, 1957 to 1961. Since 1959 the amount expended by local authorities on local health services has ceased to be directly grant aided but is taken into account, together with expenditure on education, child care, etc., in the general grant distributed to local authorities.

For national insurance and industrial injuries insurance services, Scotland is administered with the rest of Great Britain through the Ministry of Pensions and National Insurance. Under the National Insurance Act, 1946, which came into full operation in 1948, everyone over school-leaving age who has paid the required contributions can benefit from the scheme, which includes payments in case of unemployment, sickness, and confinement. Cover against industrial injuries and diseases arising from the nature of the employment is provided under the National Insurance (Industrial Injuries) Act, 1946. (*See* also SOCIAL SECURITY.)

4. Justice.—By the Act of Union of 1707 the legal system then existing in Scotland was preserved, and the civil and criminal jurisdiction of the Scottish courts has remained separate and distinct from that of England. The supreme civil court is the Court of Session, which is composed of 16 judges and sits in the old Parliament House in Edinburgh. It is divided into an outer house consisting of eight judges, and an inner house consisting of two divisions of four judges. Each division of the inner house sits as a court of appeal (the first presided over by the lord president and the second by the lord justice clerk) to hear appeals from the outer house or the Sheriff Court (*see* below). A right of appeal from the decision of the Court of Session normally lies to the House of Lords. The supreme criminal court of Scotland is the High Court of Justiciary, of which the judges are the same as those of the Court of Session. It sits not only in Edinburgh but also on circuit in certain cities and burghs. Trials before the High Court are always trials by jury, the Scottish jury consisting of 15 persons in criminal cases. The verdict may be returned by a majority and may in cases of reasonable doubt be one of "not proven." Since 1926 the High Court, sitting as a court of criminal appeal, has been empowered to hear appeals in certain cases tried by jury in either the High Court or the Sheriff Court. There is no appeal in criminal cases to the House of Lords. Scotland is divided for local judicial purposes into 12 sheriffdoms, for each of which the crown appoints a sheriff and a number of sheriffs substitute. The sheriffs of the county of Lanark and of the Lothians and Peebles and the sheriffs substitute throughout Scotland are whole-time resident judges presiding over the Sheriff Courts, a Scottish institution of great antiquity. The other ten sheriffs are part-time officers, chosen from among leaders of the Scottish bar. The sheriffs are mainly concerned with the hearing of appeals from their sheriffs substitute in civil cases and the performance of certain administrative duties. The Sheriff Court has a civil jurisdiction virtually concurrent with that of the Court of Session (a notable exception is that it cannot deal with divorce), and it also has a wide jurisdiction in criminal cases. Minor criminal offenses may be tried in the Burgh Police Court, presided over by magistrates appointed (except in the case of one stipendiary in Glasgow) by the town council from among their own number, or, mainly in county areas, before justices of the peace appointed by the secretary of state on the recommendation of local advisory committees.

Responsibility for the institution of criminal proceedings in the High Court and the Sheriff Court rests with the lord advocate, who is assisted in the discharge of this function by the solicitor general, four advocates-depute, and procurators fiscal stationed throughout Scotland. The crown office in Edinburgh, of which the permanent head is the crown agent, is the central department concerned with this work. Prosecutions in the Burgh and Justice of the Peace courts are instituted by prosecutors specially appointed by the town council or the justices. (*See* also SCOTS LAW.)

Police, Prisons, and Borstal Institutions.—Before 1857 the provision of police forces in Scotland was governed by adoptive acts. In that year the Police (Scotland) Act made the establishment of efficient police forces in every county compulsory. In burghs, most of which already maintained police forces, their establishment was made compulsory in 1892. The Local Government (Scotland) Act, 1929, transferred to county councils responsibility for policing all burghs other than large burghs then having their own force and limited to burghs with a population of 50,000 the right to establish their own forces in future. The Police (Scotland) Act, 1946, empowered any two or more police authorities to amalgamate their police forces and administer them by a joint committee, and it also empowered the secretary of state to amalgamate police forces compulsorily if he considered this to be in the interests of efficiency. The law relating to police administration was consolidated by the Police (Scotland) Act, 1956. There were 31 forces in Scotland in 1964, with strengths ranging from 19 to 2,729; the total establishment was 10,729. The final report of the royal commission on the police, issued in 1962, recommended some reorganization, including a reduction in the number of separate police forces and an extension of the responsibility of powers of the secretary of state. Some of the recommendations have been implemented in the Police Act, 1964. (*See* POLICE.)

Under the Prisons (Scotland) Act, 1877, the administration of prisons in Scotland was transferred from the local authorities to the state, and prison commissioners for Scotland were established. The administration passed in 1929 to the Prisons Department for Scotland and in 1939 to the Scottish Home Department.

The legislation governing the administration of prisons and Borstal institutions was consolidated under the Prisons (Scotland) Act, 1952. There are eight prisons, two of which have sections set aside for the detention of Borstal inmates, and five separate Borstal institutions, one of which has a small prison section attached. A person of 16 and under 21 years of age may be sentenced to undergo a period of Borstal training; a person of not less than 14 but under 21 may be sentenced to a term of three months in a detention centre; there are two senior detention centres for the reception of male persons aged 17 and under 21. There are no detention centres for offenders of 14 and 17 years of age and none for females of any age. The Criminal Justice (Scotland) Act, 1963, prohibited the imposition of imprisonment on any person under 21 and introduced instead a new form of detention for persons aged 17 but under 21—detention in a young offenders' institution. The first two young offenders' institutions were opened in January 1965. Young offenders may also be sent to approved schools or remand homes instead of to young offenders' institutions. *(See also* BORSTAL SYSTEM; JUVENILE DELINQUENCY; GREAT BRITAIN: *Constitution and Government.)*

5. Education.—Since 1929 the local administration of education has been a responsibility of the county councils and of the town councils of the four cities (Aberdeen, Dundee, Edinburgh, and Glasgow). The Education (Scotland) Act, 1962, which consolidated earlier Scottish educational legislation, laid on these 35 education authorities the duty of making adequate provision for educational needs and conferred on the secretary of state for Scotland powers and duties of supervision and administration, which are exercised through the Scottish Education Department.

Types of Schools.—There are three main classes of schools in Scotland, all subject to inspection: public schools, managed by education authorities (almost 95% of the total); grant-aided schools, conducted by voluntary managers who receive grants direct from the Scottish Education Department; and independent or private schools, which receive no assistance from public funds but are required to be registered. In the mid-1960s there were about 3,200 public schools, with more than 870,000 pupils, almost 50 grant-aided schools (with over 22,000 pupils), and just over 140 independent schools (with about 18,000 pupils). About one-fifth of the public schools had fewer than 25 pupils, while those with less than 100 pupils amounted to almost half the total number. Just over 700 of the public and grant-aided schools provide secondary education and, of these, almost 400 provide certificate courses extending over at least four years. Primary and secondary education is available free in public schools. The functions of education authorities include the medical inspection of school children, the provision of milk and midday meals, the conveyance of children living at a distance from school, and the award of school bursaries.

Primary education is from 5 to about 12 years of age. The Schools (Scotland) Code prescribes the broad general lines which the curriculum should follow, the framing of detailed courses being left to the school authorities, subject to the approval of her majesty's inspectors of schools. Pupils are allocated, usually between the ages of 11½ and 12½, to the various types of secondary course according to their fitness to profit from them, assessed on teachers' estimates of attainment and, in most areas, the results of verbal reasoning and attainment tests. Regard is also paid to the wishes of the parents and any exceptional circumstances. Secondary schools are of two main types: junior secondary schools providing noncertificate courses extending to three years and senior secondary schools providing certificate courses of four, five, or six years. In each type the courses are intended to provide a general education but are differentiated to suit the varying needs and abilities of the pupils; they include academic, commercial, technical, domestic, and rural courses. There are a number of comprehensive or omnibus schools in which both noncertificate and certificate courses are provided. Most schools are attended by both boys and girls. The courses in senior secondary schools normally lead to presentation for the Scottish Certificate of Education which was awarded and issued by the Scottish Education Department until 1965, when these functions passed to the Scottish Certificate of Education Examination Board. A student securing the requisite number of passes obtains exemption from the entrance examinations of Scottish and some other universities and also admission to or exemption from the examinations of a number of other professional and examining bodies. Special schools are provided by education authorities (mainly day schools) and by voluntary bodies (residential schools) for children who are physically or mentally handicapped or maladjusted. There are 26 approved schools in Scotland to which the courts under the Children and Young Persons (Scotland) Act, 1937, may send children under the age of 17 who need care and training; all but two of these schools are under voluntary management.

Further Education.—The majority of young people enter employment direct from school, most of them at the age of 15. Since 1942, those who wish to enter certain industries have in many parts of the country been given the opportunity of attending full-time prevocational courses. Once they are in employment young people may supplement practical experience by attending evening classes or part-time day-release classes. Apprenticeship training schemes are in operation in various industries. Advanced instruction in various technical and commercial subjects, in the arts, and in domestic science is provided in the colleges, known as central institutions, which are under the management of independent bodies of governors and function to a considerable extent on a regional basis. Three of these are agricultural colleges under the general administration of the Department of Agriculture and Fisheries for Scotland and the remaining ten are under that of the Scottish Education Department. Many of the central institutions award associateships or diplomas. Others present students for national diplomas, or for certificates awarded after three- and five-year courses in part-time classes. For certain courses some of the central institutions are affiliated to local universities and prepare students for degrees.

A number of people (more than 100,000) attend cultural and recreational classes for adults. Through the generosity of the 11th marquess of Lothian, Newbattle Abbey College was opened as a residential college of adult education under a representative governing body in 1937. Facilities for leisure-time occupation for young people over school age are provided largely by voluntary

THE TIMES, LONDON

STUDENTS OUTSIDE OLD COLLEGE, EDINBURGH UNIVERSITY. THE BUILDING DATES FROM 1789 AND WAS DESIGNED BY ROBERT ADAM

associations and the churches, with assistance from education authorities, who also make some direct provision.

All teachers in public and grant-aided schools in Scotland are expected to hold one of the certificates issued by the secretary of state on successful completion of a course of professional training. Courses of training last from one to three years according to the nature of the certificate and the degree or other qualification already held.

Universities.—There are five universities in Scotland: St. Andrews (1410), Glasgow (1451), Aberdeen (1494), Edinburgh (1583), and Strathclyde (1964). By the Universities (Scotland) Act, 1858, as modified by that of 1889, the government of the four ancient universities on the academic side is entrusted to the university senate, and as regards finance and general administration to the university court. Within this framework, the University of St. Andrews, under an act of 1953, enjoys a constitution differing in some respects from those of the other universities because its teaching activities are divided between Dundee and St. Andrews. The local town councils, including Dundee in the case of St. Andrews University, are represented on university courts. The Universities acts do not apply to the University of Strathclyde, which was formed from the Royal College of Science and Technology and the Scottish College of Commerce and has a charter which lays down its constitution. Although the universities receive parliamentary grants they are autonomous with their own endowments.

The titular head of each university is the chancellor, elected for life by the general council. The office of vice-chancellor is usually held by the principal of the university, who is appointed in the case of St. Andrews, Glasgow, and Aberdeen by the crown, and in Edinburgh by a body of curators of patronage. The principal is also president of the university senate, and the rector, elected triennially by the matriculated students, is the president of the university court. In Strathclyde the chairman of the university court is elected annually by the court from among such of its members as are not members of the university's staff. The court is the supreme authority in regard to finance and the regulation of the duties of professors and lecturers. The senate consists of the principal, all the professors, and readers and lecturers specially elected, its main responsibilities being teaching and discipline. The general council, comprising the chancellor, the members of the university court, professors, lecturers and readers, and all graduates, meets twice a year to consider proposals for alteration or improvement in the arrangements of the university. In Strathclyde this function is undertaken at the annual meeting of the general convocation which comprises the chancellor, the members of the university court, and representatives from both inside and outside the university. The domestic affairs of each faculty are administered by a dean. The faculties consist of arts, science, divinity, law, medicine, engineering, and architecture.

In 1901 Andrew Carnegie founded the Carnegie Trust for the universities of Scotland, the annual income of which (about £200,000) is devoted partly to the payment of students' fees but mainly at the present day to assisting the universities in their development projects and in fostering research.

The Royal College of Surgeons, the Royal College of Physicians of Edinburgh, and the Royal Faculty of Physicians and Surgeons of Glasgow jointly confer registerable medical qualifications. *See* also the appropriate sections of EDUCATION, HISTORY OF; ELEMENTARY EDUCATION; ADULT EDUCATION; UNIVERSITY; TEACHER TRAINING; TECHNICAL EDUCATION. (X.)

VIII. THE ECONOMY

In the middle decades of the 20th century, Scotland's population was a little over 5,000,000, or about 10% of that of the United Kingdom; almost 75% of the people live in the mid-Lowlands, compared with 50% a century earlier. Of the working population of just under 2,100,000, about 35% were in manufacturing industries, of whom half were in metals, engineering, and vehicle manufacture; 35% in distribution, administration, and professional and financial services; 16% served public utilities; and 10% were in primary production (farming, forestry, fishing, mining, etc.).

Unemployment tended persistently to be about twice the figure for the United Kingdom; the annual average unemployment percentage (1964) was 3.7 in Scotland and 1.7 in the U.K. as a whole, suggesting, it has been said, a "peripheral economy."

A. PRODUCTION

1. Agriculture.—In the mid-1960s more than 15,000,000 of Scotland's 19,069,568 ac. (7,717,454 ha.) were used for agriculture, but over two-thirds were in rough grazing, only 883,000 ac. (357,350 ha.) in permanent grass, and just under 3,500,000 ac. (1,416,450 ha.) in arable (almost 50% in temporary grass, 25% under oats, mainly for fodder, and the rest under fodder crops, *e.g.*, turnips, or cash crops, *e.g.*, potatoes, wheat, barley). Cattle totaled over 1,989,000, all attested as free from bovine tuberculosis (incoming animals being subject to control). Dairy cattle were mainly Ayrshire and British Friesian; beef cattle included Aberdeen Angus, Shorthorn, Galloway, and West Highland or Kyloe, along with various crosses. About 8,500,000 sheep included Blackface (on heathery hills), South Country and North Country Cheviots (on grassy hills), and Border Leicester, Suffolk Downs, and other English breeds (on Lowland farms), as well as crosses with hill sheep. Pigs numbered 439,000 and were of growing importance; poultry, totaling almost 8,500,000, were of particular importance to small farmers in the counties of Aberdeen and Orkney; production in the artificial environment of battery and deep litter houses is increasing in importance. About 61,000 recorded agricultural holdings included some comprising several farms; about 28,000 were regarded as full-time holdings (over 30 ac.; 12 ha.); the rest, crofts and other small holdings, provided only part-time occupation to the holders. The main regional types of farming are: (1) arable stock farming in the eastern Lowlands, often on large farms, sometimes with important cash cropping (potatoes, wheat, barley, and locally sugar beet); (2) dairying, largely based on arable with temporary grass, often on small to medium-sized family farms, on the Lowlands and the lower plateaus in the southwest and the western mid-Lowlands, and spreading elsewhere; (3) hill-sheep and cattle-rearing farms, often large units operating on very poor land, in the Southern Uplands and in the Highlands, where they are also run by crofters and other small holders.

The present state of farming in Scotland owes much to the agricultural revolution effected mainly during the 18th century. There was a great gain in technical efficiency and a general rise in standards of living. Early in the 18th century an almost medieval runrig or rundale open-field system was seen in decadence, with many signs of imminent collapse. This type of open-field system generally consisted of a constantly cultivated infield periodically subdivided into a number of strips according to the needs of the community, and of intermittently cultivated outfields periodically left fallow or brought under the plow by communal decision and effort and similarly divided into strip fields by lot. It is certain that in decadence there was fragmentation of individually held strips, which contributed to the eventual breakdown of the system. Agricultural improvements lagged behind those taking place in England, save for patchy, piecemeal changes by individual improvers, and standards of living were generally poor. By 1800, and certainly by the end of the Napoleonic Wars, which accelerated the changes, the more productive Lowland areas had been largely enclosed and improved, creating landscapes ancestral to those of today, with estates or mansion, park, and wood, and substantial tenant farms with large rectangular fields, hedged, fenced, or walled. Farmsteads and workers' cottages were sited as centrally as possible amid the fields and commonly some distance from villages, which supplied service functions and a little small-scale manufacturing but took little direct part in the farming.

In the Highlands comparable enclosures followed the breakup of the runrig system, strained by population increase in a harsh physical environment. The notorious Highland clearances, probably inevitable but too often brutal in execution, took place in the century following the Jacobite rising of 1745. Many peasant farmers found themselves landless as land was taken over for sheepwalks or enclosed into larger farm holdings. Related events were the abolition of heritable jurisdictions (a hallmark of the clan system), the pacification of the Highlands, the building of

FIG. 3.—MAJOR AGRICULTURAL AREAS OF SCOTLAND

military roads, the coming of a cash economy, and landlordship on the English model. These changes left the landscape much as we now see it, with sporting estates and mansions, large sheep farms divided into hirsels (each in the charge of a single shepherd often living in an isolated cottage), and, in contrast, the often overcrowded crofting townships. These settlements were groups of cottages each with individual strips of arable and improved land, which were open to the township as a whole for winter grazing, and rights were held to common grazing on the hill land. The crofting problem was complicated by the famine of 1846–48 and by the continual emigration of young people to industrial Britain and overseas. This trend continued into the 20th century, despite the reforming acts of 1886 and later years, which brought security of tenure, fixed rents, enlarged holdings, improved cattle breeding, roads and piers, home industries, and the resettlement of some families on estates bought by the government. The Crofters (Scotland) Act of 1955 encouraged reorganization of holdings for greater efficiency, with assistance by special grants, while a permanent Crofters Commission was set up to speed and smooth negotiations for such a reorganization and had powers to control the reletting of crofts and to terminate the tenancies of absentee crofters. Concurrently, several other forms of development have improved conditions locally: Forestry Commission plantations; a fish-curing plant; hydroelectric projects and aluminum refineries at Kinlochleven and Lochaber, near Fort William; harvesting and processing of seaweed; a renewal of interest in cattle rearing on hill land, publicized as cattle ranching; a newsprint factory in the Fort William area; various experiments in small-scale manufacturing; and some increase in catering for tourists.

In Scotland as a whole the Napoleonic Wars not only speeded agricultural change but also increased industry, commerce, and facilities for agricultural finance and marketing. The mid-Lowlands region was especially affected and was transformed from an underdeveloped to a highly developed part of Britain. With peace came more difficult times for farmers, with the competition of increased trade from overseas, and strain and restlessness among their workers. With the flood of overseas cereals in the late 19th century came more widespread changes: decreased extent of tillage; increased grass and stock; marked rural depopulation; the breaking up of some estates, coupled with an increase of owner-occupied farms and of working, as against gentlemen, farmers, although technical standards remained high. The early 20th century saw the spread of formalized agricultural education, the intensive farming effort of World War I, the boom in food crops; then, with peace, harder times again for farming, the establishment of new agricultural research organizations, and the purchase of estates by the government to set up small holdings to give needed opportunities to the younger generation and especially to ambitious farm workers. Following the general economic slump of the 1930s came government subsidies on wheat (1932), barley and oats (1937), and on beef. World War II brought guaranteed prices and assured markets for the chief products; subsidies for plowing up land, fertilizers, and reseeding grassland; intensified advisory services; and, for a time, emergency powers of control over, and even ejection of, inefficient farmers. After the war came new expansion programs (1947 and 1952). Legislation dealt with production and marketing and the fixing of minimum and overtime wages, and hill farming (1946), livestock rearing (1951), and security of tenure for tenant farmers (1949). The government makes an annual review of agriculture, then fixes guaranteed prices for cattle, sheep, pigs, milk, eggs, wheat, barley, oats, rye, potatoes, sugar beets, and wool. Grants or subsidies are available for plowing up marginal land, drainage, water supplies, bracken eradication, farm cheese making, and the raising of hill sheep, hill cattle, calves, and attested herds.

2. Forestry.—From a predominantly wooded landscape, deforestation at several phases of history had produced an almost treeless landscape by 1800. Landowners' reafforestation, amenity woodlands, and shelter belts, along with scrub oak and birch, exceeded 1,000,000 ac. (about 405,000 ha.) by 1900, but World War I brought widespread felling of useful timber. The Forestry Commission, set up in 1919, established many plantations, which unfortunately had not matured by World War II, and further reduction of reserves of standing timber resulted. Of the target of a national planting scheme (1946) of 5,000,000 ac. (2,023,500 ha.) of productive woodland on private and Forestry Commission land to be realized by the end of the century, at least half was to be in Scotland. In the mid-1960s private woodland totaled about 470,000 ac. (190,000 ha.); commission-owned woodland totaled about 660,000 ac. (267,000 ha.), employing just under 4,000 people. The commission buys estates for planting, mainly with Scots and Corsican pine, Sitka and Norway spruce, European, Japanese, and hybrid larches, and Douglas fir, according to the particular site. Since 1945 policy has been notably broad based, including agricultural interests, and in the late 1950s the emphasis switched from the strategic importance of timber production to the significance of its role in the development of integrated land use. This trend continued into the 1960s. Farmers use pastures above the forests in conjunction with improved low-lying fields, sometimes in forestry holdings (where the farmer agrees to carry out and is guaranteed at least 150 days' forestry work per annum). The social aspects also are now stressed. Rehabilitation of depopulated communities has been achieved, so that new schools and better services can economically serve the increased population; new villages have been formed and National Forest parks (Argyll, 58,000 ac.) demarcated.

3. Mining.—Coal is the chief mineral resource. Mining rights were granted in a 12th-century charter to William Oldbridge of Carriden, West Lothian, and in 1291 to the abbot and convent of Dunfermline, but coal was not widely used in houses until the late 16th century, when coal exports were beginning to be important. The Industrial Revolution brought great expansion for both export and internal use in homes and factories. By 1910–13 Scotland was producing around 15% of the United Kingdom's total output, but the 1920s brought loss of overseas markets, lower domestic demands because of the general slump, industrial unrest, and competition from other fuels. In 1939 production was 30,500,-

000 tons, falling to 21,400,000 in 1945—about 10% of the U.K. total—mined by 12% of the mining manpower. Nationalization in 1946 brought surveys of the problems, and in 1950 the National Coal Board's (NCB) plan for coal (for Great Britain). The Scottish coalfields incur high running cost because of the presence of steep folding, volcanic intrusions, thin seams, and excessive water. These conditions are marked in the central field on which so much heavy industry was based, and output has decreased. This decline, coupled with expansion in the eastern fields of Fife and Midlothian, raised social and economic problems reaching beyond the industry. But by 1960 overproduction had succeeded shortages, owing largely to competition from oil, and the NCB's plan was revised.

In Scotland coal of economic importance occurs in the Limestone Coal Group and in the Productive Coal Measures. The main coalfields' resources, current development, and problems may be summarized as follows. In the coalfield of the East Fife-Midlothian syncline the steep beds of western Midlothian and around Kelty and Lochgelly in Fife are largely exhausted; but there are large new mines and reserves in the flatter beds, including the deep Limestone Coal Group farther east, around Dalkeith in Midlothian, and the new town of Glenrothes in Fife. However, at Glenrothes unforeseen technical difficulties underground have brought a premature end to a major development project, but development of the town is continuing with associated light industry. Coal from this field is mainly highly volatile, weakly coking, or noncoking, in seams three to six feet (one to two metres) thick, probably with large reserves under the Firth of Forth. In the West Fife-Clackmannan syncline the coal is similar, the main reserves lying in the eastern flank. This field continues under the Forth into the declining but historically important central field lying mainly in the county of Lanark. There the thick upper coals of the Productive Coal Measures are almost exhausted and the remainder too thin or rendered dangerous by water from old workings, while the Limestone Coal Group, deep under the Productive Coal Measures and the Millstone Grit, is not exploitable at present. A factor that has contributed to the decline of the central field is the mid-20th-century tendency to close the small, old, inefficient pits and to open larger modern ones, mainly sited in the east, where the main reserves are found at depth, especially extending from Stirling under the Forth. There are, however, accessible seams and considerable reserves in a belt from northeastern Glasgow to Denny, and appreciable output of anthracite, special coals, and some strongly and medium-coking coals, which, mixed with Durham coking coal, are very important to the Scottish steel industry. Therefore, demand may call for considerable capital expenditure to exploit the known reserves. In the Ayr and Dumfries field weakly coking and noncoking coals are declining in the central area, with expansion on the margins and considerable reserves under the Permian rocks in the Mauchline Basin.

4. Fisheries.—Fishing is more important proportionally in Scotland than in the United Kingdom as a whole, but there has been a steady decline in the number of fishermen. In the mid-1960s there were just under 3,000 Scottish vessels, including about 160 steam and motor trawlers, with over 10,500 fishermen, of whom roughly one-quarter were part-time and crofter fishermen. Landings by British and foreign vessels exceeded 6,000,000 cwt. (305,000 metric tons), principally haddock, cod, whiting, skate, plaice, herring, mackerel, sprat, and shellfish (lobster, crab). Landings are greatest in the Aberdeen fishery district, well ahead of Leith, Fraserburgh, Lossiemouth, Peterhead, and Buckie on the east coast, Loch Broom, Ayr, Oban, Stornoway, and Campbeltown on the west coast, and Orkney and Shetland. Salmon and sea-trout angling is an important tourist attraction, while in 1963 netting employed over 1,600 men in the season, with a catch of 1,900 tons, including about 350 tons by rod and line.

There are government fishery officers at ports and on fishery cruisers; research stations at Aberdeen and (for fresh water) at Pitlochry; the head office of the United Kingdom Herring Industry Board; and a committee of the White Fish Authority for Scotland and Northern Ireland. Current trends include a change from steam to motor vessels, with grants for new vessels or for conversion from coal to oil, and a scheme for the encouragement of full-time fishing in the Outer Hebrides.

5. Power.—Electricity is distributed by two public boards responsible to the secretary of state for Scotland. The North of Scotland Hydroelectric Board (1943), employing over 3,000 people, operates over the area roughly north and west of the Firths of Clyde and Tay; the South of Scotland Electricity Board (1954), employing about 12,500, operates over the remainder of the country. Electricity is generated by steam using coal, oil, or atomic energy, by waterpower, sometimes by diesel engines, and experimentally by gas turbines, including the plant using peat at Altrabreac.

There has been a great mid-20th-century drive to develop the hydroelectric power potential of the Scottish Highlands. The aim has been to supply power to all the farms and crofts, stimulate industrial development, indirectly increase agricultural production, and generally stem the emigration of the rural population from the Highlands. The heavy rainfall in this area and high incidence of ribbon lakes, which act as reservoirs, are great advantages for hydroelectric development; but most large lakes are relatively low in altitude, streams are variable in volume, and catchment areas are often small so that tunnels must be excavated through watersheds to increase the drainage into the reservoirs. In this way, the natural catchment area (6½ sq.mi.; 17 sq.km.) of Loch Sloy was increased to 31 sq.mi. (80 sq.km.); the water dammed back in Loch Sloy Reservoir travels 2 mi. (3 km.) by tunnel through Ben Vorlich to the power station (1950) on Loch Lomond. The larger Highland schemes, of which Loch Sloy is an example, contribute electricity to the grid system in central Scotland. The Tummel-Garry scheme and the Mullardoch-Fasnakyle-Affric scheme are similar, and both involve the construction of tunnels to increase the catchment areas. Smaller schemes, such as Glen Lussa and Morar, supply the needs of the local population. In the mid-1960s only about 10% of Scotland's potential had been realized: capital costs of construction are high and would presumably be higher as less easily exploitable sources were tapped, and on current costing thermal power is a little cheaper.

Scotland shares with the rest of the U.K. in the development of nuclear power stations. In the mid-1960s three were functioning: Dounreay fast-breeder reactor station; Chapelcross, associated with a plutonium factory; and Hunterston station.

The Scottish Gas Board in the mid-1960s employed about 9,500 people at over 100 works, producing 41,000,000,000 cu.ft. (1,161,000,000 cu.m.) of gas (84% coal gas, 15% water gas). Trends since nationalization in 1949 have included the closing of about 100 small gasworks as supplies became available from larger, more economical works by linkage mains. In the early 1960s a Lurgi high-pressure complete gasification plant at Westfield, Fife, marked the beginning of a supergrid system for the mid-Lowlands, designed to link with local grid systems.

6. Industries.—To understand the problems of Scotland's economy as a whole, it is necessary to look at it, and at industry in particular, from the historical viewpoint. The Scots Parliament's last few years were far from inactive in economic affairs, yet at the Union of Parliaments (1707) the economy was still largely medieval, and in decadence as a result of the wars of independence, involvement in the struggle for power between France and England, and the religious wars (from the late 13th to the late 16th centuries), and latterly of several famines between 1695 and 1702, and the halting of projects for economic improvement after the disaster of the largest, the Darién scheme (1695–1700), which was partly caused by English opposition. In 1707 Scotland was in no position to take advantage of the new access to the markets of England and the colonies. But by 1750 trade was flowing: for a time Glasgow merchants had a near monopoly of the Virginia tobacco trade and also joined in the slave trade, thus encouraging developments in agriculture and banking. The latter had unique features adapted to a poor country in which capital grew largely through reliance on borrowers' and guarantors' moral character, and in consonance with current religious thought. As the Industrial Revolution began to affect textiles, Scotland was well placed, with its large flax and linen hand-spinning and weaving industries

(the encouragement of which followed the Union and was the original function of the British Linen Bank), its large supplies of pure water, and many waterfalls useful in the waterpower phase. Linen factories, followed by cotton, became widespread, especially in the mid-Lowlands. With the introduction of coal-based steam power and more centralized factories, sites near the mid-Lowlands coalfields were favoured, and the linen industry had vigorous survivors up to the 1960s, notably in Fife and Angus, and also added the daughter industry, jute. Woolen factories for spinning, weaving, and knitwear, often high-grade or luxury goods, are found in many of the larger Scottish towns, being particularly important in the Border country, and handloom Harris tweeds are an important export commodity from the Hebrides. Paisley is a prominent centre for the cotton industry; there are also about 30 cotton factories (including two survivors from the 18th century at Doune in Perth and Catrine in Ayr).

The early dominance of the textile industry was followed by a decline initiated by the dislocation of the American Civil War of 1861–65, leaving Lancashire henceforth dominant in Britain's cotton industry. Displaced capital tended to move toward heavy industries, iron and steel, heavy engineering, and shipbuilding, which were to dominate a second phase of the Industrial Revolution in Scotland, felt notably in the mid-Lowlands. Glasgow became the principal centre in commerce, shipping, coal mining, textiles, chemicals, and engineering, which has now evolved into the great Clydeside conurbation. The discovery of how to use blackband ironstone both as a fuel and an ore, Neilson's hot-blast steel plant (1828), Naysmyth's steam hammer (1825), the coming of railways from 1830 on, steam propulsion in ships (William Symington in 1802, Robert Fulton in 1803, and Henry Bell in 1812), the building of ships of iron and, later in the century, of steel, all marked a phase of inventiveness, confidence, and increasing industrial development. The Clyde became and remains one of the world's major shipbuilding districts, and there were smaller but locally important shipyards on the east coast (Aberdeen, Dundee, and Leith), and engineering there and in many other towns. The Clydeside industrial complex has persisted despite the decline in coal output from the central field and the fact that the blackband iron ores, on which the iron industry was originally based, are little worked, having been replaced by imported ores. The Clyde Valley is one of the major steel-producing regions in Britain. Heavy steel products are typical, and only about a fifth of the total output is exported, as Scottish production is oriented to the home demand of the major, essentially heavy industries, notably shipbuilding, marine engineering, and boilermaking.

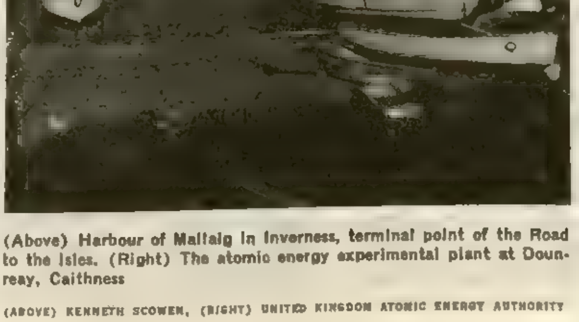
The depression of the early 1930s brought unemployment throughout the country for about one worker in ten; the unemployment ratio was considerably worse in the heavy industrial areas (since known successively as depressed areas, special areas, and development areas). This situation emphasized the need to diversify Scotland's economy and gave rise to a drive to increase the country's participation in the manufacture of automobiles, aircraft, radio, electrical apparatus, and modern light engineering and mass production industries. Special advisory bodies were set up, now consolidated as the Scottish Council (Development and Industry). Scotland's first industrial estate, Hillington, in the county of Renfrew, was opened in 1938, and by 1939 there were 70 factories, including a Rolls-Royce aero-engine factory. By the mid-1960s more than 20 industrial estates and 30 individual new plants of converted armaments works housed over 350 industrial tenants on an average floor space of over 40,000 sq.ft. (3,700 sq.m.). Products included tractors, earth-moving and agricultural machines, typewriters, computers, refrigerators, and other electrical and electronic apparatus (the last fostered by special research facilities), watches and clocks, textiles and clothing, and numerous food products.

Car manufacturing is once more carried out in Scotland by branch factories of firms of the English Midlands, encouraged by the government to move new production developments to central Scotland. Exports are substantial, and Scotland has attracted over 60% of those United States firms that have set up factories in Britain since World War II. Despite all these developments, however, the traditional heavy industries persist as the basis of Scotland's economy. Shipbuilding remains at a high level, even if lower relative to world production and latterly with a serious falling off in new orders. Major reorganization and reconstruction in the steel industry broadly favours traditional sites. In exports textiles, whisky, and the newer industrial estate products all play a vital part. Yet up to the 1960s unemployment persisted at a higher level than in England. There is growing conviction, indeed growing evidence, that in certain respects Scotland's economy is merely peripheral to the main flow of prosperity in Britain. Yet, as Scottish nationalists point out, Scotland is not lacking in resources. It is more nearly self-supporting in food than England and its farming less dependent on imported fodder. In industry the relative lack of experience in, and concentration on, mass production may prove a less grave disadvantage as automation increases. Twenty years' real achievement and experience provided material for a review of Scottish industry, which was undertaken in the 1960s by the Scottish Council. (A. T. A. L.)

7. Banking.—The monopoly granted to Scotland's first bank, the Bank of Scotland, was not renewed when it expired in 1715. The resulting freedom to issue notes, which in those days was synonymous with banking, made possible the formation of large banks and the spread of branches throughout the country. As a result, the provision of banking services had been concentrated in the hands of a few relatively large and stable institutions even before 1844, when the Bank Charter Act began the process of absorption and amalgamation among the small country banks in England.

The Scottish banks still retain the right of note issue, but the Bank Notes (Scotland) Act, 1845, obliged them to hold gold coins as cover against notes actually in circulation in excess of a stated maximum, which has not since been increased and now forms a very small proportion of the total circulation of the banks. With the withdrawal in the 20th century of gold coin from the country's currency, Bank of England notes became the principal covering requirement. These issues, however, still make it possible to maintain branches in small communities where, without them, the provision of banking services would be too expensive.

The following banks now operate in Scotland: Bank of Scotland



(Above) Harbour of Mallaig in Inverness, terminal point of the Road to the Isles. (Right) The atomic energy experimental plant at Dounreay, Caithness

(ABOVE) KENNETH SCOWEN, (RIGHT) UNITED KINGDOM ATOMIC ENERGY AUTHORITY

(1695), Royal Bank of Scotland (1727), British Linen Bank (1746), National Commercial Bank of Scotland Ltd. (1810), and Clydesdale Bank Ltd. (1838). (X.)

B. Transport and Communications

1. Roads.—Roads are classified as trunk roads, a responsibility of the secretary of state for Scotland, about 1,950 mi.; and as Class I, 4,800 mi., Class II, 3,900 mi., Class III, 6,300 mi., unclassified, 11,400 mi., all these being the responsibility of local authorities (mainly counties since 1929). In 1963 work began on a motorway between Glasgow and Edinburgh. The network comprises: (1) the mid-Lowlands, dense, well surfaced, in places overcrowded, with a close, largely radial pattern around Glasgow, and a persisting concentration of routes through Edinburgh's Lowland gap; (2) the Southern Uplands, a much wider pattern, though still well surfaced; in lower Tweeddale a fairly close network; (3) east of the Highlands, dense and well surfaced around Aberdeen, linking with good east coast routes, with central Grampian routes (Glen Shee-Devil's Elbow; Glen Garry-Strathspey). Beyond single strands follow glens to the west coast, sometimes single lane only, with passing places; minor roads are poorly surfaced. In the southwest Highlands, the roads Loch Fyne-Oban and Glasgow-Loch Lomond-Glencoe-Fort William are spectacular.

The mid-Lowlands firths are a considerable barrier; the Forth has a bridge at Kincardine and another at Queensferry (completed 1964); the Firth of Clyde as yet has no road bridges, though two tunnels were constructed at Whiteinch in the early 1960s to relieve congestion on ferries and Glasgow's bridges; a bridge across the Firth of Tay was scheduled to be completed by 1966. The present pattern is largely the legacy of the 18th century turnpike acts, with minor roads long maintained under earlier statute labour acts, and of military road building in the Highlands by General George Wade and others.

2. Railways.—Administered under the Scottish Region of British Railways since 1947, Scottish railways include: (1) in the mid-Lowlands a close network, especially around Glasgow, with more east-west lines than elsewhere, those between main centres bearing heavy traffic; (2) in the Southern Uplands a looser mesh, including busy main lines from Glasgow and Edinburgh to the south seeking low crossings (Beattock, the highest, 1,028 ft. [313 m.]); (3) in the north, continuous lines up the east coast Lowland to Wick and Thurso, complemented by the Perth-Inverness route through Drumochter Pass (1,500 ft.; 457 m.) and lines southwest from Dingwall to Kyle of Lochalsh on the west coast, Stirling-Oban and Glasgow-Mallaig, all to small Highland ports.

In the 18th century horse trams took coal downhill from pithead to port; the first act authorizing such a horse railway (Kilmarnock-Troon) was in 1812, and the 191¼ mi. of line was approved by Parliament in 1840. This was followed by rapid development and consolidation into the North British Railway (1845, mainly on the east coast), the Caledonian (1845, mainly west and central), the Great North of Scotland (1846, mainly around Aberdeen), the Glasgow and South Western (1850, sharing the Clyde sailings with the Caledonian and North British), and the Highland (1865). The Railway Act of 1921 amalgamated these private lines into two groups: the London, Midland and Scottish; and the London and North Eastern railway companies, until nationalization in 1947. As elsewhere the 1950s–60s saw the closing of many branch passenger lines and more frequent services on remaining lines, using diesel to replace steam engines.

3. Shipping and Canals.—The principal ports are Glasgow, a major port of the United Kingdom, Leith, Grangemouth, Dundee, Greenock, Aberdeen, Ardrossan, and Methil. The ports are mainly on the firths penetrating the richer mid-Lowlands region. The east coast ports were more important in the 15th to 17th centuries, but the west coast gained with the increase in Atlantic trade and the industrial development of the Glasgow area.

The British Transport Commission owns and operates the Caledonian Canal, the Crinan Canal, and the Forth and Clyde Canal. They are economically unimportant, mainly carrying relatively small quantities of coal, coke, and liquid fuel; all run at a deficit. The Monkland and Union canals, now closed to navigation, are maintained for industrial water supplies and are operated at a small profit.

4. Air Services.—Prestwick is an international airport (controlled by the minister of transport and civil aviation) used by most airlines with North Atlantic services and connected with several European capitals. It was developed by Scottish Aviation Ltd. and taken over and developed further by the Air Ministry during World War II; in the early 1960s it was extended to accommodate large jet aircraft. In British European Airways' (BEA) internal services, Renfrew is the main base for Campbeltown, Islay, Tiree, Barra, Benbecula, and Stornoway; Aberdeen and Wick serve Orkney and Shetland, and Edinburgh and Inverness are included in the network. Renfrew and Edinburgh have services to London and the south.

Commercial civil aviation began in the 1930s (Inverness-Wick-Kirkwall, 1934; Newcastle-Aberdeen-Stavanger, 1937). By 1939 the framework of the present system was based on Renfrew and Aberdeen, with an early and fine tradition of an ambulance service bringing urgent cases from remote landing strips to Renfrew, Inverness, Aberdeen, and Wick. This service is still maintained by BEA and augmented in emergency by Royal Navy and Royal Air Force helicopters. Future developments may include a more economical scale of service, to enable still more remote areas to benefit from air transport.

5. Postal and Telecommunications Services.—These services are integrated with those of the rest of the United Kingdom (*see* Great Britain and Northern Ireland: *Transport and Communications: The Post Office*). Postal services are maintained in Scotland by head post offices in the main cities and post and sub-post offices throughout the rest of the country. Of the telegram and telephone services, the former is receding in use and constant availability as the latter becomes more widespread. The main flow of telephone calls is between the main cities of central Scotland, with busy trunk lines from Glasgow and Edinburgh to the Midlands of England and to London. However, automatic and subscriber trunk dialing already extends to quite small rural exchanges. In the mid-1960s, 5 of the Shetland Islands, 6 of the Orkneys, and 15 of the Western Islands were served by radio-telephone.

The country is covered by national and regional sound radio services of the British Broadcasting Corporation; truly local broadcasting is as yet in its infancy. Scotland also has a television network of BBC and commercial transmitters. (A. T. A. L.)

See also references under "Scotland" in the Index.

Bibliography.—*Geography:* E. B. Bailey, "The Interpretation of Scottish Scenery," *Scot. Geogr. Mag.*, v. 50 (1934); A. Geikie, *The Scenery of Scotland*, 3rd ed. (1901); A. Gregory, "Climate and Water Supply in Great Britain," *Weather*, v. 14 (1959); D. L. Linton, "Problems of Scottish Scenery," *Scot. Geogr. Mag.*, v. 67 (1951), and "Morphological Contrasts of Eastern and Western Scotland," in R. Miller and J. W. Watson (eds.), *Geographical Essays in Memory of Alan G. Ogilvie* (1959); Crofters Commission, Annual Report for 1959 (HMSO, 1960); F. F. Darling (ed.), *West Highland Survey, an Essay in Human Ecology* (1955); J. B. Fleming and F. H. W. Green, "Some Relations Between Country and Town in Scotland," *Scot. Geogr. Mag.*, v. 68 (1952); A. Geddes and F. D. N. Spaven, *The Highlands and Isles* (1949); A. G. Ogilvie (ed.), *Great Britain: Essays in Regional Geography* (1928) and "Physical Features and Scenery," in *Scotland* (1947); Secretary of State for Scotland, *Review of Highland Policy*, Cmnd. 785 (HMSO, 1959); Scottish Home Department, *A Programme of Highland Development*, Cmd. 7976 (HMSO, 1950); A. C. O'Dell and K. Walton, *The Highlands and Islands of Scotland* (1962); the following regional geologies: *Scotland: Northern Highlands* by J. Phemister, 3rd ed. (1960); *The Grampian Highlands* by H. H. Read and A. G. MacGregor, 2nd ed. (1948); *Scotland: Tertiary Volcanic Districts* by J. E. Richey, 3rd ed. (1961); *The Midland Valley of Scotland* by M. MacGregor and A. G. MacGregor, 2nd ed. (1948); *South of Scotland* by J. Pringle (1948). Ordnance Survey, 1/625,000 Map Series, Sheet I, especially Land Classification, 1945; Types of Farming, 1944; Land Utilization, 1944; Vegetation, 1953; Geological Map of Great Britain prepared by the Geological Survey, Sheet 1, 2nd ed. (1957); Rainfall, Annual Average 1881–1915, Sheet 1, 1949, compiled by the Department of Health for Scotland and the Ministry of Town and Country Planning from isohyets and statistics of the Meteorological Office. F. Fraser Darling, *Natural History in the Highlands and Islands* (1947); D. N. McVean and D. A. Radcliffe, *Plant Communities of the Scottish Highlands* (HMSO, 1962); E. V. Baxter, L. J. Rintoul, and J. Morton Boyd, *The Birds of Scotland*, 2 vol., 2nd ed. (1964); J. H. Burnett (ed.), *The Vegetation of Scotland* (1964).

The People: J. A. MacCulloch, *The Celtic and Scandinavian Religions* (1948); J. M. Reid, *Scotland, Past and Present* (1959); S. Piggott (ed.), *The Prehistoric Peoples of Scotland* (1962).

History: Bibliographies, works of reference, sources: G. F. Black, *A List of Works Relating to Scotland* (1916); H. W. Meikle, *Scotland: a Select Bibliography* (1950); P. D. Hancock, *A Bibliography of Works Relating to Scotland, 1916–50*, 2 vol. (1960); C. S. Terry, *A Catalogue of the Publications of Scottish Historical . . . Clubs and Societies . . . 1780–1908* (1909), continued for the period 1909–1927 by C. Matheson (1928); A. H. Dunbar, *Scottish Kings . . . 1005–1625*, 2nd ed. (1906); W. J. Watson, *The History of Celtic Place-Names of Scotland* (1926); G. F. Black, *The Surnames of Scotland* (1946). Much of the writing on Scottish history has taken the form of introductions to collections of documents. Particularly useful are the introduction to vol. i of the *Regesta Regum Scottorum*, ed. by G. W. S. Barrow (1960), for Scotland in the 12th century, and the introductions to the Scottish Record Office series of the *Exchequer Rolls of Scotland*, 23 vol. (1878–1908), for the period from the 14th to the end of the 16th century; also A. O. Anderson, *Early Sources of Scottish History, A.D. 500–1286*, 2 vol. (1922); W. C. Dickinson, G. Donaldson, and I. A. Milne (eds.), *A Source Book of Scottish History*, 3 vol., 2nd ed. (1958–61), covering the period from the earliest times to 1707. *See* also *The Scottish Historical Review* (1904–in progress) and the publications of the Scottish History Society, the Stair Society, and the Society of Antiquaries of Scotland.

General histories: P. H. Brown, *History of Scotland*, 3 vol. (1909–12); A. M. Mackenzie, *The Rise of the Stewarts, 1329–1513* (1935), *The Scotland of Queen Mary . . . 1513–1638* (1936), *The Passing of the Stewarts, 1639–1748* (1937), *Scotland in Modern Times, 1720–1939* (1947); R. S. Rait and G. S. Pryde, *Scotland*, 2nd ed. (1954); R. L. Mackie, *A Short History of Scotland*, rev. ed., ed. by G. Donaldson (1962); W. C. Dickinson, *Scotland from the Earliest Times to 1603* (1961); G. S. Pryde, *Scotland from 1603 to the Present Day* (1962).

Special periods: V. G. Childe, *The Prehistory of Scotland* (1935), *Scotland Before the Scots* (1946); R. W. Feachem, *A Guide to Prehistoric Scotland* (1963); Sir G. Macdonald, *The Roman Wall in Scotland*, 2nd ed. (1934); H. M. Chadwick, *Early Scotland* (1949); F. T. Wainwright (ed.), *The Problem of the Picts* (1955); A. C. Lawrie, *Early Scottish Charters Prior to A.D. 1153* (1905); R. L. G. Ritchie, *The Normans in Scotland* (1954); on the Wars of Independence *see* G. W. S. Barrow, *Robert Bruce* (1965); R. L. Mackie, *King James IV of Scotland* (1958); G. Donaldson, *The Scottish Reformation* (1960); John Knox, *History of the Reformation in Scotland*, 2 vol., ed. by W. C. Dickinson (1949); D. H. Willson, *King James VI and I* (1956); D. Mathew, *Scotland under Charles I* (1955); A. V. Dicey and R. S. Rait, *Thoughts on the Union Between England and Scotland* (1920); G. S. Pryde (ed.), *The Treaty of Union of Scotland and England, 1707* (1950); Sir C. A. Petrie, *The Jacobite Movement*, 3rd ed. (1959); G. P. Insh, *The Scottish Jacobite Movement* (1952); H. W. Meikle, *Scotland and the French Revolution* (1912); G. Donaldson, *Scotland: James V–James VII* (1965).

Special topics: R. S. Rait, *The Parliaments of Scotland* (1924); Lord Cooper, *Select Scottish Cases of the Thirteenth Century* (1944); J. D. Mackie and G. S. Pryde, *Local Government in Scotland* (1935); W. M. Mackenzie, *The Medieval Castle in Scotland* (1927), *The Scottish Burghs* (1949); J. and J. A. R. Mackinnon, *The Constitutional History of Scotland from Early Times to the Reformation* (1924); R. K. Hannay, *The College of Justice* (1933); A. R. MacEwen, *History of the Church in Scotland, 397–1560*, 2 vol. (1913–18); J. Dowden, *The Medieval Church in Scotland* (1910), *The Bishops of Scotland* (1912); J. R. Fleming, *A History of the Church in Scotland, 1843–1929*, 2 vol. (1927–33), *The Story of Church Union in Scotland . . . 1560–1929* (1929); J. K. Hewison, *The Covenanters*, 2 vol. (1908); G. D. Henderson, *Religious Life in Seventeenth-Century Scotland* (1937); I. F. Grant, *The Social and Economic Development of Scotland Before 1603* (1930); M. Plant, *The Domestic Life of Scotland in the Eighteenth Century* (1952); W. H. Marwick, *Scotland in Modern Times; an Outline of Economic and Social Development Since the Union of 1707* (1964); G. P. Insh, *The Company of Scotland, Trading to Africa and the Indies* (1932); H. Hamilton, *The Industrial Revolution in Scotland* (1932), *An Economic History of Scotland in the Eighteenth Century* (1963); A. W. Kerr, *History of Banking in Scotland*, 4th ed. rev. by F. H. Allan (1926); L. C. Wright, *Scottish Chartism* (1953); T. Ferguson, *The Dawn of Scottish Social Welfare . . . from Medieval Times to 1863* (1948), *Scottish Social Welfare, 1864–1914* (1958); A. R. B. Haldane, *The Drove Roads of Scotland* (1952); W. C. Mackenzie, *The Highlands and Isles of Scotland*, rev. ed. (1949); E. S. Haldane, *The Scotland of Our Fathers: a Study of Scottish Life in the Nineteenth Century* (1933); J. Kerr, *Scottish Education: School and University, from Earliest Times . . . to 1913* (1913); H. M. Knox, *Two Hundred and Fifty Years of Scottish Education, 1696–1946* (1953).

Administration and social conditions: M. R. McLarty (ed.), *A Source Book and History of Administrative Law in Scotland* (1956); Sir David Milne, *The Scottish Office and Other Scottish Government Departments* (1958); *Report of the Royal Commission on Scottish Affairs*, Cmd. 9212 (1952–54); *Local Government (Scotland)* (HMSO, 1958); J. Mill, *The Scottish Police* (1944); W. M. Gloag and R. C. Henderson, *Introduction to the Law of Scotland*, 6th ed. (1956); RIPA (Edinburgh Branch), *A Guide to the Scottish Legal System* (1958).

Economy: A. K. Cairncross (ed.), The Scottish Economy, Department of Social and Economic Research, University of Glasgow, *Social and Econ. Studies* No. 2 (1954); *Agriculture in Scotland*, the Report of the Department of Agriculture in Scotland (HMSO, 1964); Forestry Commission, *Annual Report for the Year Ending 30th Sept. 1963* (HMSO, 1964); I. F. Grant, *Economic History of Scotland* (1934); L. J. Saunders, *Scottish Democracy, 1815–40* (1950); J. E. Handley, *Scottish Farming in the 18th Century* (1953), *The Agricultural Revolution in Scotland* (1963); A. Lamont, *Scotland, the Wealthy Nation* (1952); W. H. Marwick, *Economic Developments in Victorian Scotland* (1936); *North of Scotland Hydro-Electric Board: Annual Report and Accounts, 1963* (HMSO, 1964); *South of Scotland Electricity Board: Annual Report and Accounts, 1963* (HMSO, 1964); C. A. Oakley, *Scottish Industry*, Scottish Council (Development and Industry) (1953); C. J. Robertson, "Scotland, a peripheral economy?" *Tijdschrift voor Economische en Sociale Geografie*, vol. 44 (1953); Scottish Council (Development and Industry), *Natural Resources in Scotland* (1961); *Central Scotland. A Programme of Development and Growth*, Cmnd. 2188 (HMSO, 1963); *Development and Growth in Scotland, 1963–64*, Cmnd. 2440 (HMSO, 1964).

SCOTLAND, CHURCH OF. The Church of Scotland claims to be "part of the Holy Catholic or Universal Church," reformed in the 16th century but continuous with the church of the apostles. This article outlines its history and indicates its position in the national life. It is established by law in Scotland and its polity is Presbyterian. (*See* further SCOTLAND: *History*; PRESBYTERIANISM.)

Early Period.—Christianity was first brought to what is now southwest Scotland during the last years of the Roman occupation of Britain. According to tradition, St. Ninian, a Briton by birth but trained on the continent of Europe, founded a church at Whithorn (*c.* 400) and did missionary work beyond the sphere of Roman influence. The next great name is that of St. Columba, who settled at Iona *c.* 563. He was only one of many missionaries from Ireland at that period, and there is controversy as to how much of the Christianizing of the country is due to him, to them, or to Ninian and his successors. At all events the church in Scotland until the 12th century was the church in its Celtic form (*see* CELTIC CHURCH), affected by influences from the Roman Church in England but cherishing its native customs, partly in opposition to external pressure and partly because for generations Scotland was comparatively isolated by the raids and settlements of the pagan Norsemen.

The Medieval Church.—The assimilation of this church to the general pattern of the church of the West was part of a change which came over Scotland when its kings in the first half of the 12th century formed their kingdom after the fashion of the Anglo-Norman kingdom of England. The rudiments of parochial and diocesan organization were greatly developed, and monastic orders from England and the continent were introduced and lavishly endowed, especially by David I (*q.v.*). There was as yet no Scottish metropolitan, and the English archbishops of Canterbury and of York, particularly the latter, claimed that the Scottish bishops were subject to them. In 1192, however, the Scottish church was declared "a special daughter" of the Roman see, subject only to the pope. Not until 1472 did St. Andrews become an archiepiscopal see, followed by Glasgow in 1492.

During the medieval period the church in Scotland shared in the features common to the Western Church as a whole. It derived stimulus from monastic revivals and the coming of the friars: the universities of St. Andrews, Glasgow, and Aberdeen were founded in the 15th century: its outstanding figures were statesmen as well as clergy, and though little is certain about its education, its theology, or its devotional life there is evidence that these were far from negligible and that the medieval church in Scotland preserved the essentials of Christian faith and practice, however much these were entangled with notions and customs that were rejected at the Reformation (*see* REFORMATION: *Scotland*).

The medieval church had glaring defects. About 85% of its parishes were "appropriated" to cathedrals, bishoprics, abbeys, and other institutions, with the result that the bulk of the parochial revenues went to these higher levels, leaving a pittance for the vicar who did the parochial work. Another abuse was the violation of church discipline through scandalous appointments to high office in the church. Responsibility for this lies chiefly on the

crown, but the connivance of the papacy cannot be ignored. The statutes of the Scottish provincial council in the decade before the Reformation are evidence for the abuses that abounded; but they also testify to the existence of a body of opinion favourable to moderate reform.

The Reformation and After.—Had there been a strong monarchy in Scotland the outcome might have been a reformation like that in England or in Scandinavia. The Scottish crown, however, weakened by a succession of royal minorities, passed in 1542 to the infant Mary, queen of Scots, who was sent to France and betrothed to the dauphin, later Francis II, and from 1554 Scotland was governed by her mother, a Frenchwoman, Mary of Lorraine. The Scottish Reformation therefore was in large part a national movement for Scottish independence against the domination of France, and at the same time a revolution in opposition to the crown.

The earliest Scottish reformers, such as Patrick Hamilton, were under Lutheran influence. George Wishart made contact with the Swiss reformers, but the Calvinistic tone of the Scottish Reformation was due to John Knox (*q.v.*). Knox's admiration for Calvin and Geneva is evident in his Scots Confession, in the Book of Common Order, often known as Knox's Liturgy (*see* Common Order, Book of), and in the Book of Discipline (later called First Book of Discipline) which set out a scheme for a godly church and commonwealth. The Reformers held a Parliament in August 1560 which abolished the authority of the pope in Scotland, adopted the Scots Confession, and forbade the celebration of Mass. The First Book of Discipline, which included a plan for education and the care of the poor, presupposed the handling by the Reformed church of the great wealth of the medieval church. This was abhorrent to the nobility and gentry, who in various ways had obtained control of much of the church's wealth even before 1560, and was never approved by Parliament. Although the absent queen Mary Stuart sanctioned the meeting of Parliament, the religious question was reserved, and this legislation, never confirmed by her, was ratified by the regent in the name of her infant son, James VI, after her abdication in 1567.

After the breach with Rome there followed a century and a quarter during which it was uncertain whether the Reformed church was to be episcopalian or presbyterian in government. The Reformers recognized no order superior to the ministry of the Word and sacraments, but episcopacy was not a vital issue in the Reformation. Four of the Scottish bishops in 1560 supported the change and three of them continued to serve under the General Assembly, a body partly clerical partly lay, while five superintendents were also appointed. After the abdication of the queen an act was passed giving the superintendents authority to admit men to benefices, which meant in practice that when the Romanist holders of benefices died each was succeeded by a minister of the Reformed church. Similarly, by the Concordat of Leith (1572) ministers were to be appointed to the bishoprics. They were to have the same powers as superintendents and be subject to the General Assembly. Nomination of these bishops lay with the crown. This concordat was accepted by the General Assembly only as an interim measure. Nevertheless it was accepted, even by John Knox, and it was not till after his death (1572) and after Andrew Melville (*q.v.*) returned to Scotland from Geneva in 1574 that controversy broke out.

Melville, not Knox, should be called the originator of Scottish Presbyterianism, for it was he who converted many Scottish churchmen to the view that the bishop in the New Testament was the pastor of a particular congregation having no authority over other ministers, so the functions of supervision should no longer be exercised by bishops or superintendents, but should be taken over by a body called a presbytery, composed of ministers and elders. Eventually this view was approved, and a beginning was made in 1581 when 13 presbyteries in the Lowlands were organized. King James VI, however, was determined that bishops should rule the church and that he should appoint and rule them. The struggle was long and keen between the divine right of presbytery and the divine right of kings. At first the king had to yield from time to time. In 1592 he was compelled to assent to an act which gave parliamentary sanction to the system of presbyterian courts. Yet there were still parts of the country where ministers and kirk sessions and presbyteries were lacking, and the act left the king a measure of control over the church, for it reserved to the crown the right to appoint the time and place for the meeting of General Assemblies. From that time the king's policy steadily advanced, and after he had become James I of England he secured most of what he wanted. By 1610 he had persuaded the General Assembly to accept bishops who should indeed be subject to its censure but have a real authority in their dioceses, and for them he secured consecration at the hands of English bishops. No doubt his intention was that in time presbyteries and General Assemblies should be allowed to lapse.

The story of the ecclesiastical policy of Charles I and later kings, and the opposition to it which found expression in the covenants and the resultant civil war and persecutions, is long and complicated. Briefly, there were two periods, 1610–38 and 1661–89, when the Stuart kings were successful in imposing episcopal government on the lower presbyterian courts. At the end of each period, the final rejection of episcopacy was primarily due to its identification with royal absolutism in church and state and with a hankering after Roman Catholicism. Under pressure from the Scottish Parliament, William of Orange agreed not only to the abolition of episcopal government as "contrary to the inclinations of the generality of the people ever since the Reformation" but also in 1690 to the establishment of presbyterian church government, "that is to say the government of this Church by kirk-sessions, presbyteries, provincial synods and general assemblies," and to the ratification of the Westminster Confession (*q.v.*), which had been approved by the General Assembly in 1647. Nothing was said of the divine right of presbytery or of the perpetual obligation of the covenants of 1638 and 1643 whereby the Scottish Presbyterians had pledged themselves to uphold their own ecclesiastical system (*see* Covenanters). For that reason the extreme Covenanters remained outside the established church, Presbyterian though it was, as the convinced Episcopalians and Roman Catholics did for other reasons.

18th Century.—An act for securing the Protestant religion and Presbyterian church government was included in the Treaty of Union in 1707 (*see* English History), but five years later the united Parliament in London passed an act for the restoration of patronage in Scotland, which, although not the cause of, was an important factor in, the troubles which issued in the First Secession of 1733 under Ebenezer Erskine and in the Second Secession of 1761 under Thomas Gillespie, founder of the Relief Presbytery (*see* Presbyterianism). Another factor was the difference of policy and outlook between the two parties in the church which came to be known as Moderates and Evangelicals. Moderatism was the Scottish version of what is commonly known elsewhere as the Enlightenment. In reaction from religious wars and persecutions, men everywhere saw the virtue of tolerance and benevolence. Fanaticism was to be repudiated and everything brought to the bar of reason. The application of such ideas to religion did not necessarily lead to heresy, but it minimized the distinctively Christian affirmations in the sphere of doctrine and laid a greater emphasis on moral teaching. The Evangelicals held firmly to the traditional Calvinism of the Westminster Confession, sometimes in an intellectual dryasdust fashion and sometimes in a warmer evangelical fashion.

During the predominance of Moderatism in the later 18th century the Secession and Relief churches continued to flourish owing to the increase of population, the continued operation of patronage, the desire for a warmer fellowship and more stirring preaching than was often to be found in the parish church, and the rise of an independent middle class, which found more scope in the nonestablished churches.

The Evangelical Revival.—Despite the draining away of evangelical fervour into the Secession and Relief churches, the Evangelical party in the Church of Scotland was ready to take advantage of the change of atmosphere brought about by the French Revolution. There was a revulsion from the cult of reason toward the traditional, the romantic, and the mystical, and in some circles

the revolutionary idea of the rights of man produced a democratic, even secularist, ferment. It began to be suggested that one of the rights of man was to choose his religion and pay for it, the state neither establishing nor endowing any church.

By the turn of the century the evangelical revival in the Church of Scotland was making progress, aided indirectly by the movement led by the brothers Robert and James Haldane who afterward left the Church of Scotland. Leaders of talent and rank such as John Erskine (d. 1803), friend of George Whitefield, and Sir Henry Moncrieff Wellwood (1750–1827) worked for the evangelical revival, and then Andrew Thomson with his magazine the *Edinburgh Christian Instructor* began to build up a party of younger men who after his death in 1831 looked to Thomas Chalmers (*q.v.*) as their leader.

The Disruption and After.—Once again patronage became an issue, but only as part of a larger program. The Evangelical party, which dominated the General Assembly from 1834 onward, realized that Scotland had changed in two generations. Population had increased and was more than before concentrated in the larger cities and towns, where more parishes, churches, and schools were needed. Chalmers and his followers were distrustful of the democratic trends in the agitation for the Reform Bill of 1832 and did not favour popular election of ministers; yet they could only hope to hold the people by some relaxation of the patronage laws which would give the congregation a control of the choice of its minister. To effect this the General Assembly passed the Veto Act, in terms of which if a majority of the male heads of families in a parish disapproved of the patron's presentee, this should be sufficient reason for the presbytery to reject him.

The act was conservative in intention; but when it was challenged the law courts declared it *ultra vires* and the patrons' rights were upheld. To the Evangelicals this intervention of the law was an invasion of the church's spiritual jurisdiction over the settlement of pastors, and thus the question of patronage merged in the deeper question of spiritual independence. Again and again cases came before the courts, and whereas the assembly maintained its Veto Act the courts instructed presbyteries to disregard it.

More than one political leader in Scotland framed measures designed to provide a way out of the impasse, but the government in London was indifferent, and though in 1842 the General Assembly by a two to one majority adopted the Claim of Right setting forth the church's complaints about the encroachments of the civil courts, nothing was done. The majority of the Evangelical party then began to prepare to end their relationship with the state. This took place in May 1843, when over a third of the ministers, followed by about the same proportion of the laity, "went out" to form the Free Church of Scotland (*see* SCOTLAND, FREE CHURCH OF).

The Disruption, as this event was called, was a severe blow to the Church of Scotland. It still held the loyalty of many persons of all ranks, however, and soon new leaders arose to vitalize its parochial system and guide it into fresh service. Nevertheless the Disruption, by reducing the size and prestige of the established church, accelerated changes which were already in the offing In 1845 the state took over the management of the poor law previously vested in kirk sessions, and in 1872 the act for compulsory state education brought the church's control of the parochial schools to an end.

The later 19th century was a period of change. Innovations by Robert Lee (1804–68), designed to make worship more ordered and dignified, aroused controversy but gradually won acceptance. The growth of congregational organizations was largely due to the vision and perseverance of Archibald Hamilton Charteris (1835–1908), who was particularly active in the organization of women's work. A new emphasis on youth made the Sunday school an essential congregational agency and led to the founding of uniformed organizations of which the first was the Boys' Brigade. The Scoto-Catholic movement, as it is sometimes called, for improvements in worship, distinct from but related to that of Robert Lee, aimed successfully at beauty and symbolism in church furnishing, observance of the Christian year, especially Christmas and Holy Week, more frequent communion, and a willingness to contemplate closer relations with episcopal churches.

Reunion.—Patronage was abolished by act of Parliament in 1874, and though the bitterness engendered by a subsequent disestablishment campaign remained a barrier between the Church of Scotland and the United Free Church (*see* UNITED FREE CHURCH OF SCOTLAND), the atmosphere was so changed that in 1909 the two churches entered into "unrestricted conference." The prime demand of the United Free Church was freedom from state control, and that was found to be compatible with the national recognition of religion which was what the Church of Scotland chiefly valued in its relation to the state. The outcome was the framing of Articles Declaratory of the Constitution of the Church of Scotland in Matters Spiritual, their presentation to Parliament in the form of a bill which was passed in 1921, followed by the act of 1925 which dealt finally with the endowments of the Church of Scotland, and then, on that basis, by the union of the two churches in 1929 which achieved a Church of Scotland united, national, and free. A minority, however, stood out and continued under the name United Free Church.

The Mid-20th Century.—More notable than this union as a testimony to the vitality of the Church of Scotland is the way in which it adapted itself to changing conditions. Church extension had been going on in Scotland for two centuries, but in the 20th century the parochial system was challenged in a more insistent way by slum clearance, the growth of new suburbs and even of new towns, as vast numbers of the population were rapidly rehoused. The implications of this were realized by John White (1867–1951), minister of the Barony of Glasgow, whose persistent efforts over 20 years to meet the situation made the church as a whole conscious of its significance so that from 1950 to 1960 new churches were opened at the rate of one every month. Parallel to this extension went a readjustment of agencies which sometimes involved the closing of redundant churches. During the 30 years that followed the union of the United Free Church with the Church of Scotland nearly a quarter of the parishes or charges were merged or linked with other parishes in the 60 presbyteries.

The legislation associated with the union of 1929 threw increasing financial responsibility on congregations, particularly in the country parishes; and, despite the decline of rural life in general, they responded well to the challenge, though the average liberality of the total membership is still low.

The General Assembly's Foreign Mission Committee was formed in 1825 and Alexander Duff went to Calcutta in 1829 as the first official missionary of the church. By the second half of the 20th century the Overseas Council, comprising the Foreign Mission Committee, the Colonial and Continental Committee, the Church and Israel Committee, and the Committee on Inter-Church Aid and Refugee Service, was sending missionaries, chaplains, and other workers to share in the life and witness of indigenous churches in many lands.

The Church of Scotland, however, is in a missionary situation in Scotland itself. Although nearly half the children born in the country are baptized in infancy in the Church of Scotland, only a quarter of the total population (or over a third of those aged more than 20) are communicant members, while regular attenders at worship are estimated at about 7% of the total population. For many years there has been a decline in the number of children attending Sunday school and Bible classes.

The older methods of evangelism have continued in operation but there have also been experiments tending away from the large meeting toward intensive work among people in their homes and places of work. Parish missions and area missions, prepared by preliminary surveys and training of workers, have become common, and there have been approaches to special groups through school chaplains, industrial chaplains, and the Scottish Christian Industrial Order. The Iona community, founded by George MacLeod in 1938 and integrated into the organization of the church in 1951 (*see* IONA), has strongly influenced methods of evangelism as well as worship and the witness of the Christian in society. Another movement more widely spread is "Tell Scotland," begun as a radio mission, which sponsors visitation missions, house

churches. Bible study groups, as well as more spectacular efforts such as the "Kirk Weeks," held in imitation of the German *Kirchentag*.

The "Tell Scotland" movement and many other practical initiatives are ecumenical in the sense of being backed by other Protestant denominations in Scotland, but the latter are comparatively so small in membership that the Church of Scotland as a whole is little affected by them and does not feel the need of ecumenical support at the parochial level. Nevertheless the Church of Scotland is a member of the World Alliance of Reformed Churches (World Presbyterian Alliance), of the British Council of Churches, and of the World Council of Churches and has contributed to the leadership of these bodies. From 1932 conversations with the Church of England (in conjunction since 1954 with the Episcopal Church in Scotland) took place at intervals, but no important moves toward union or intercommunion were accepted.

The new stress on evangelism kindled interest in adult Christian education and the task of the laity in church and community. Not only did the number of elders increase by about a quarter after 1929, but there was a renewed sense of the spiritual potentialities of the eldership. The admission of women as elders has been proposed, but indifference and opposition, not least on the part of many women, have prevented it. Women, however, are often more active in their membership and more alive to the mission of the church than men are, largely because of the Woman's Guild, which is well organized and well led at every level and has a branch in almost every parish. Most of the deaconesses in the Church of Scotland are engaged in parochial work. Women are not eligible for the ministry, but deaconesses when suitably trained may be presented to presbyteries for license as preachers of the Word.

The interest of the church in the poor and handicapped has continued throughout the centuries in changing forms; but in the 20th century its machinery for giving effect to that interest has been greatly increased. The Committee on Social Service has under its charge many agencies, including homes for children of different ages, hostels for young men and for young women, eventide homes, as well as providing for special groups such as unmarried mothers.

The position of the church today presents peculiar, even paradoxical, features. Although it represents so large a part of Scottish Protestantism, its proportion of the total population is but moderate, and, if the test be devoted and effective membership, its position is weak. Having lost its statutory connection with education and the poor it is no longer at the centre of social life and can be ignored, as it is by many. On the other hand it is clearly the national church. The lord high commissioner, representing the sovereign, is an honoured guest at its annual General Assembly, the annual report of the assembly's Church and Nation Committee touches all aspects of the nation's life, and British governments sometimes pay attention to the General Assembly's utterances. Even Scotsmen who show no interest in the church's worship and teaching are concerned about it as an institution. When through Anglican-Presbyterian conversations its government seemed to be in question, strong general interest was aroused. In that sense the people feel the church belongs to them; it is a part of Scottish history, the product of a distinct national character as well as part of the Holy Catholic Church.

See also references under "Scotland, Church of" in the Index.

BIBLIOGRAPHY.—The only comprehensive modern work is J. H. S. Burleigh, *A Church History of Scotland* (1960). Older standard histories: J. Cunningham, *The Church History of Scotland*, 2nd ed., 2 vol. (1882); G. Grub, *An Ecclesiastical History of Scotland*, 4 vol. (1861); A. Bellesheim, *History of the Catholic Church of Scotland*, 4 vol. (1887–90).

Period Histories: A. B. Scott, *The Rise and Relations of the Church of Scotland* (1932); J. A. Duke, *The Columban Church* (1932); J. Dowden, *The Medieval Church in Scotland* (1910); D. E. Easson, *Medieval Religious Houses, Scotland* (1957); A. R. MacEwen, *A History of the Church in Scotland*, 2 vol. (ends at 1560) (1913–18); John Knox, *History of the Reformation in Scotland*, ed. by W. Croft Dickinson, 2 vol. (1949); D. Hay Fleming, *The Reformation in Scotland* (1910); G. Donaldson, *The Scottish Reformation* (1960); David McRoberts (ed.), *Essays on the Scottish Reformation* (1962); J. K. Hewison, *The Covenanters*, 2 vol. (1908); G. D. Henderson, *Religious Life in Seventeenth-Century Scotland* (1937); A. F. Mitchell, *The Westminster Assembly* (1883); A. J. Campbell, *Two Centuries of the Church of Scotland, 1707–1929* (1930); H. Watt, *Thomas Chalmers and the Disruption* (1943); J. R. Fleming, *A History of the Church in Scotland, 1843–1929*, 2 vol. (1927–33); John Buchan and George Adam Smith, *The Kirk in Scotland, 1560–1929* (1930); Augustus Muir, *John White* (1958).

Important General Works: J. Dowden, *The Bishops of Scotland . . . Prior to the Reformation* (1912); D. Patrick, *Statutes of the Scottish Church, 1225–1559* (1907); H. Scott (ed.), *Fasti ecclesiae Scoticanae: the Succession of Ministers in the Church of Scotland from the Reformation*, rev. ed., 9 vol. (1914–51); M. B. Macgregor, *The Sources and Literature of Scottish Church History* (1934).

Special Topics: J. Walker, *The Theology and Theologians of Scotland Chiefly of the 17th and 18th Centuries*, 2nd ed. (1888); C. G. McCrie, *The Confessions of the Church of Scotland;* G. W. Sprott and T. Leishman (eds.), *Book of Common Order* (1868); W. McMillan, *The Worship of the Scottish Reformed Church, 1550–1638* (1931); W. D. Maxwell, *A History of Worship in the Church of Scotland* (1955); M. Patrick, *Four Centuries of Scottish Psalmody* (1949); D. MacGibbon and T. Ross, *The Ecclesiastical Architecture of Scotland from the Earliest Christian Times to the 17th Century*, 3 vol. (1896–97); G. Hay, *The Architecture of the Scottish Post-Reformation Churches, 1560–1843* (1957); H. Cowan, *The Influence of the Scottish Church in Christendom* (1896); A. L. Drummond, *The Kirk and the Continent* (1956); J. MacKinnon, *The Social and Industrial History of Scotland*, 2 vol. (1920–21); S. Mechie, *The Church and Scottish Social Development, 1780–1870* (1960); G. D. Henderson, *Church and Ministry: a Study in Scottish Experience* (1951); I. M. Clark, *A History of Church Discipline in Scotland* (1929); J. T. Cox (ed.), *Practice and Procedure in the Church of Scotland*, 5th ed. (1964); H. J. Wotherspoon and J. M. Kirkpatrick, rev. ed. by T. F. Torrance and R. S. Wright, *A Manual of Church Doctrine According to the Church of Scotland* (1960). (S. Me.)

SCOTLAND, EPISCOPAL CHURCH IN, an independent province, of Scottish origin, within the Anglican Communion. Ecclesiastical history in Scotland during the first century after the Reformation is confused, with alternating periods of episcopalian and presbyterian supremacy (*see* PRESBYTERIANISM: *History of the Presbyterian Churches: Scotland*). After the Restoration in 1660 the two coalesced into a modified episcopacy, which might have united church and nation had not the two parties fallen apart after the accession of William and Mary, when Episcopalians were in conscience unable to take the oath of allegiance to them. Thus Presbyterianism was established as the national religion (1689). The Episcopal Church is the direct descendant of those churches that remained loyal to the episcopal tradition, and its bishops are the direct successors of those consecrated to Scottish sees after the Restoration.

Loyalty to the exiled Stuarts resulted in the implication of the Episcopal Church in the Jacobite risings of 1715 and 1745, after which penal laws reduced it almost to extinction. (*See* SCOTLAND: *History*.) The repeal of these laws in 1792 consequent on the death of Prince Charles Edward in 1788 marked a turning point, and the church began to revive. It was organized in seven dioceses; canons were enacted; churches and mission halls were built; schools were founded—signs of an astonishing recovery. An outstanding factor in this development is the increasing part taken by the laity in the affairs of the church, not only in financial matters through the Representative Church Council but also in matters liturgical and canonical by their membership of the Provincial Synod, the supreme legislative body, as authorized in 1961. The Provincial Synod is presided over by the primus, elected by the seven bishops from their number. The seven sees are: Aberdeen and Orkney; Argyll and the Isles; Brechin; Edinburgh; Glasgow and Galloway; Moray, Ross and Caithness; St. Andrews, Dunkeld and Dunblane. Communicants number about 56,000, clergy about 370.

The Scottish Communion Office, based on the liturgy in the service book imposed on Scotland by Charles I in 1637, took shape in 1764. In the 1920s the revision of the whole Prayer Book was undertaken and a complete Scottish Prayer Book was produced in 1929. For the most part this is an improved and amplified version of the Prayer Book of 1662, save for the liturgy which itself was revised slightly (*see* COMMON PRAYER, BOOK OF).

The Episcopal Church has undertaken missionary responsibilities overseas. It provided the endowment when the diocese of St.

John's in South Africa was founded in 1871; ever since, it has made substantial contributions annually to the work of the diocese, and a number of priests have gone out from Scotland to work there. Shortly after, responsibility was also accepted for the district of Chanda in India. A number of other missionary agencies are supported on a smaller scale. After World War I the new Social Service Board undertook moral welfare work, work among fisher folk, youth work, and the provision of "eventide" homes. The theological college is in Edinburgh.

See F. Goldie, ***A Short History of the Episcopal Church in Scotland*** (1951 [1952]). (T. Ha.)

SCOTLAND, FREE CHURCH OF, came into existence fully fledged on May 18, 1843, when 203 members of the General Assembly of the Church of Scotland marched out of St. Andrew's Church, having read and laid on the table a protest asserting that it was impossible for them, in view of the "recent encroachments of the civil courts, and the still more recent sanction of these decisions by the legislature," to hold a free assembly of the church. They therefore protested

> that it shall be lawful for us, and such other commissioners as may concur with us, to withdraw to a separate place of meeting, for the purpose of taking steps, for ourselves and all who adhere to us—maintaining with us the Confession of Faith and standards of the Church of Scotland as heretofore understood—for separating in an orderly way from the Establishment, and thereupon adopting such measures as may be competent to us, in humble dependence on God's grace and the aid of His Holy Spirit, for the advancement of His glory, and the extension of the gospel of our Lord and Saviour, and the administration of the affairs of Christ's house, according to His Holy Word.

In orderly procession they marched to Tanfield Hall over half a mile away, where they and delegates from all sympathizing congregations constituted themselves into the first General Assembly of the Church of Scotland, free. Thomas Chalmers (*q.v.*) was elected moderator by acclamation. Of the vast amount of business transacted in a crowded fortnight the outstanding item was the Act of Separation and Deed of Demission on Tuesday, May 23, 1843, to which 396 ministers and professors publicly appended their names (others being added later, bringing the total to 474) by which they renounced all claim to the benefices they had held in connection with the establishment, declaring them to be vacant, and consenting to their being dealt with as such, thus voluntarily surrendering churches, professorships, manses, and glebes and an income from the state calculated at over £100,000 per annum. This was the real act of disruption—a disruption not from their brethren in the ministry but from the state. It would, to their mind, have been a still more "glorious" disruption had all the ministers of the Church of Scotland swelled their ranks in Tanfield Hall. Though the General Assembly of 1842 had passed the Claim, Declaration and Protest, popularly known as the Claim of Right, by a majority of over two to one, it was only a minority of four to seven of the ministers that followed Chalmers to this, its logical conclusion; the proportion among both office-bearers and ordinary members was distinctly higher.

Ten Years' Conflict Before the Disruption.—Though the tangled story of the years 1833–43 belongs chronologically to the article Scotland, Church of, it is necessary to recall the salient points which brought the Church of Scotland to such an impasse. The root of all the bitterness was the reimposition of the law of patronage in 1712. What the Evangelicals were fighting for were the rights of the Christian people in the calling of their minister and the right of the church to initiate new work to areas of increasing population. A remedy was found in two assembly acts of 1834, the Veto Act and the Chapel Act. For some years the remedy seemed effective. The great bulk of the church rejoiced in their smooth operation. However, some patrons began to feel that both acts impinged on their privileges. An appeal to the law courts by one patron and his vetoed presentee proved successful. This encouraged other patrons to present cases not dissimilar, but making progressive inroads on the powers hitherto claimed and exercised by the courts of the church.

Both church and law courts hardened their positions in the ten years 1833–43. The General Assembly, which was at first content to accept patronage while tempering its operation, demanded in 1842 its abolition as the only way out. The Court of Session, while at the outset recognizing a coordinate jurisdiction of the church in spiritual things, came to deny that jurisdiction in all cases in which it had not been conferred by definite statute. Some of the judicial dicta help to explain the passionate earnestness of the final paragraph of the Claim of Right, which thus begins:

> And, finally, the General Assembly call the Christian people of this kingdom, and all the churches of the Reformation throughout the world, who hold the great doctrine of the sole headship of the Lord Jesus Christ over His church, to witness, that it is for adherence to that doctrine, as set forth in their Confession of Faith, and ratified by the laws of this kingdom, and for the maintenance by them of the jurisdiction of the office-bearers, and the freedom and privileges of the members of the Church from that doctrine following, that this Church is subjected to hardship, and that the rights so sacredly pledged and secured to her are put in peril.

When the final appeal to Parliament failed, the church had to prepare for an uncertain future; for months, committees had been busy preparing plans, gathering money from willing givers, for every branch of work to which it had been committed in its undivided state.

Early Days.—The basis of the support of the Free Church was the Sustentation Fund, the principles and details of which had been elaborated by Thomas Chalmers himself—a central fund for the maintenance of ordinances, from which every minister in a charge would receive an equal dividend. In the very first year the 583 ministers received at the rate of £105. Throughout the whole course of its history this fund grew with the growing church. But the stupendous financial burden facing its first assembly arose from the grim facts of the situation. There were no churches for their services of worship, no houses for their ministers to live in, no places for the training of the ministry, few schools for their children to be educated in, no mission premises abroad, for though every overseas missionary save one had intimated his adherence to them they had to do so at the expense of buildings and equipment. Special building funds for churches, for manses, for a college, for schools and schoolhouses, and for foreign mission stations were launched and entrusted to various leaders. Everything had to be done on the simplest scale, but before Chalmers died, in 1847, he could see every activity of the church adequately housed or in process of being so. It was an unprecedented exhibition of Christian liberality, for the great bulk of the funds came from loyal members, though occasional donations from outside sympathizers helped to swell the total.

After the first difficulties had been overcome, the church embarked on a career of orderly, if less sensational, progress. It made the minimum of changes in constitution and general setup, being eager to prove that the former constitution, if rightly understood and fully operated, was adequate for the religious needs of the land. The reintroduction of deacons ensured the progress of the Sustentation Fund. The addition of the Claim of Right of 1842 and the Protest of 1843 to the subordinate standards of the church ensured that the claim to spiritual independence should not be ignored or minimized. The church was fortunate in securing a succession of strong leaders both at home and abroad. In domestic affairs the outstanding leaders of its earliest days, Chalmers, R. S. Candlish, and William Cunningham, were succeeded by Robert Rainy, who, for nearly two generations, was its trusted guide and pilot and who left two younger ministers, Archibald Henderson and Alexander Martin, to carry on his tradition into the United Free Church. It produced notable preachers, teachers, and evangelists at home, and in the foreign field it led in educational work. Alexander Duff in Calcutta, John Wilson in Bombay, William Miller in Madras, and Stephen Hislop in Nagpur inaugurated and developed Christian colleges which, though criticized by some as placing education before evangelism, did much to develop an India permeated with Christian values. For Africa, the institutions at Livingstonia and Lovedale under Robert Laws and James Stewart did a similar service.

The emergence of the Free Church stirred hopes of union among those already outside the Church of Scotland. So 1852 saw the accession of one smaller body, the Original Secession Church

(Associate Synod of Original Seceders), and 1863 saw the beginning of wider negotiations with the two larger denominations in Scotland, the United Presbyterian Church (*q.v.*) and the Reformed Presbyterian Church (*see* CAMERONIANS). With the latter they were successful and a union took place in 1876; with the former they failed after ten years of patient and hopeful negotiation through the stubbornness of a minority within the Free Church which threatened legal action if they were forced to unite with a professedly Voluntary church (*i.e.*, one which opposed the establishment principle and held that the church should be maintained only by voluntary support of its members). The two results of the ten years were a Mutual Eligibility Act of 1874, by which a minister of either church might be called to a congregation in the other, and a great mass of documents which materially shortened the later successful negotiations initiated in 1896.

Theological Developments.—The mind of the Free Church was not absorbed only in questions of ecclesiastical polity. The colleges of the church were the home of the most active interest in theological and historical questions. There was a stormy period, however, when extreme conservatism appeared to gain a victory. In 1870 W. Robertson Smith (*q.v.*) was ordained to the office of professor of Oriental languages and Old Testament exegesis in the Free Church College, Aberdeen. There from the first he advocated views which though now widely accepted were then regarded with apprehension. In 1876 a committee of the Free Church reported so adversely on his writings that Smith demanded a formal trial. The indictment failed; but a vote of want of confidence was passed, and in 1881 he was removed from his chair. This event, however, was no adequate expression of the real mind of the church. During the last quarter of the 19th century the Free Church continued to be the most active, theologically, of the Scottish churches.

The college chairs were almost uniformly filled by theologians and historians of progressive views, inspired more or less by A. B. Davidson of New College, Edinburgh. A. B. Bruce was appointed to the chair of apologetics and New Testament exegesis in the Glasgow College in 1875; Henry Drummond was made lecturer in natural science in the same college in 1877 and became professor in 1884; and George Adam Smith was called to the Hebrew chair in 1892.

Attempts were made between the years 1890 and 1895 to bring all these professors except Davidson (similar attacks were also made on Marcus Dods, afterward principal of the New College, Edinburgh) to the bar of the assembly for unsound teaching or writing; but in every case these were abortive, the assembly never taking any step beyond warning the accused that their primary duty was to teach and defend the church's faith as embodied in the confession. In 1892 the Free Church, following the example of the United Presbyterian Church (1879), passed a Declaratory Act relaxing the stringency of subscription to the confession, with the result that a small number of ministers and congregations, mostly in the Highlands, severed their connection with the church and formed the Free Presbyterian Church of Scotland, on strictly orthodox lines.

Meanwhile other changes were taking place. The standard of parochial and congregational activity was raised, the use of instrumental music was sanctioned, and special attention given to the promotion of edification, order, and reverence in public worship. And the establishment principle was almost entirely abandoned.

Incorporation in United Free Church.—During the last four or five years of the 19th century the Free and United Presbyterian churches devoted their energy to the arrangement of an incorporating union. The synod of the United Presbyterian Church resolved in 1896 to "take steps toward union," and next year the Free Church assembly responded by appointing a committee to confer with a committee of the other church. The joint committee discovered a "remarkable and happy agreement" between the doctrinal standards, rules, and methods of the two bodies, and with very few concessions on either side a common constitution and common "questions and formula" for the admission of ministers and office-bearers were arranged. A minority, always dwindling, of the Free Church assembly protested against the proposed union and threatened if it were carried through to test its legality in the courts; nevertheless the union was carried through. The supreme courts of the churches met separately for the last time on Oct. 30, 1900, and on the following day the union was completed, and the United Free Church of Scotland (*q.v.*) entered on its career.

The Free Church Minority ("Wee Frees").—The protesting and dissenting minority at once claimed to be the Free Church. Denied admission to the assembly hall, they adjourned to another building, where they held, so they claimed, the remaining sittings of the assembly. At their next assembly (1901) it was stated that there had adhered to their antiunionist program 25 ministers and at least 63 congregations. The character of the church is indicated by the fact that its office-bearers were the survivors of the minorities of the church who had protested against the disestablishment resolution, against the toleration of any tinge of higher criticism in the theological chairs, and against the use in public worship of organs or of "human hymns." The congregations were mainly in the Gaelic-speaking districts of the Highlands and islands. Many held high hopes of considerable accessions if they won the lawsuit which they had entered into (*Bannatyne* v. *Overton*) claiming that they were the rightful owners of the properties which had belonged to the Free Church and had been carried into the United Free Church. But there was little change of allegiance when the House of Lords (1904) gave decision in their favour. Nevertheless the church continued to hold its own in the areas of its original congregations.

In the mid-1960s its General Assembly comprised 15 presbyteries, of which 12 were in Northern Scotland and the islands. It had foreign mission stations in South Africa, Peru, and India. Under its overseas missions it had six congregations in Canada and the United States. For the training of its ministers it maintained a strong theological college in Edinburgh. *See* also PRESBYTERIANISM; SCOTLAND, CHURCH OF; UNITED FREE CHURCH OF SCOTLAND.

BIBLIOGRAPHY.—J. R. Fleming, *The Church of Scotland 1843–1929*, 2 vol. (1927–33); P. Carnegie Simpson, *The Life of Principal Rainy*, 2 vol. (1909); J. H. S. Burleigh, *A Church History of Scotland* (1960).
(X.; H. WA.)

SCOTLAND YARD, the headquarters of the metropolitan police of London, Eng., was originally an area just off Whitehall, and took its name from the palace, dating back to the reign of Edgar (959–975), which stood on the site. Edgar gave King Kenneth of Scotland a piece of ground lying beside the new palace of Westminster for his residence, requiring him to make a yearly visit to London to pay homage for the kingdom of Scotland. The palace built there by Kenneth remained the property of the Scottish kings, where they stayed when visiting London, but in the reign of Henry II, after the rebellion of William of Scotland in 1173–74, it was "resumed into the king of England's hands." Margaret, the widow of James IV of Scotland and sister of Henry VIII of England, appears to have been the last Scottish sovereign to have made use of the palace, living there for a considerable time after the death of her husband in 1513, but by the time Elizabeth I came to the throne it was a neglected ruin.

By James I's reign the site, which had come to be known as "Scotland," was divided into two yards, known as Great Scotland Yard and Middle Scotland Yard, and was used for government buildings. These included the official residence of the surveyor of works to the crown, which was at one time occupied by Inigo Jones and his successors Sir John Denham and Sir Christopher Wren, who had offices there and designed a house in Scotland Yard for Sir John Vanbrugh. This house was built from the ruins of buildings in Whitehall destroyed by fire in 1697. John Milton, when appointed secretary for foreign tongues under Oliver Cromwell's Commonwealth in 1649, soon found it convenient to take up official residence at Scotland Yard, where he lived until he moved to Petty France in 1651. Other notable inhabitants of Scotland Yard include Susannah Cibber, wife of the actor and playwright Theophilus Cibber and the sister of Thomas Arne, the composer; and Thomas Campbell, the Scottish poet.

Early Police Connections.—The first association of police with Scotland Yard appears to have been in 1662, when the first police or improvement commissioners for London and Westminster set up offices there, but these commissioners were more concerned with paving, lighting, and cleaning of streets than with police duties as they are understood today. A further association occurred when Col. Sir Thomas de Veil, the first of the Bow Street Magistrates, had an office in Scotland Yard, where he acted as an agent for memorials to public departments before moving to Bow Street about 1735. But the direct association between Scotland Yard and the metropolitan police started when the force was formed in 1829 by Sir Robert Peel, the home secretary of the time. The newly formed force was to police the area in and immediately around London—with the exception of the City of London, with its area of just more than one square mile, which was not included but which was to form its own police force ten years later—and the headquarters of the metropolitan force were at no. 4, Whitehall Place, with a police station at the rear, entered from Scotland Yard itself. Although the proper official address of the headquarters was no. 4, Whitehall Place, from the very beginning it was always referred to as Scotland Yard by the press and others.

At the time of their formation, the metropolitan police were placed under the leadership of two joint commissioners, Richard Mayne (made K.C.B. in 1851) and Col. Charles Rowan (made K.C.B. in 1848), the former "a sensible lawyer" and the latter "a military man conversant with the details of the police system in Ireland." At first the new force met with fierce and bitter opposition from many quarters and was referred to by many uncomplimentary names, including the "raw lobster gang," "Peel's bloody gang," and "the blue devils," but under the wise and able control of Mayne and Rowan they succeeded in overcoming this opposition and by the time of Mayne's death in 1868 (Rowan had retired in 1850) the force had gone far to establish the reputation and popularity which it enjoyed thereafter.

Building Development.—As the work and duties of Scotland Yard increased, adjoining premises were taken over and by the early 1880s the headquarters consisted of a number of separate buildings which, apart from the inconvenience, were extremely overcrowded. Important records and papers were crammed into cupboards and piled high on landings, and the provision of a new headquarters had become a pressing necessity. It was generally agreed that the headquarters should be situated in the immediate vicinity of the home office and the seat of government, since the home secretary is directly responsible to parliament for police matters. The only acceptable site was one on reclaimed land made available by the construction of the Thames Embankment, which was opened in 1870. On this site of about 70,000 sq.ft., Colonel Mapelson had begun to build a national opera house on a very ambitious scale. The foundation stone was laid in 1875 by the duke of Edinburgh and the intended opera house had been partly constructed, between £80,000 and £90,000 having been spent on it, when the scheme fell through for lack of further financial support.

A private offer was made in 1880 to the deputy surveyor of the police to buy the site as it stood for £25,000, and after very lengthy discussion the site was finally purchased in 1886. Richard Norman Shaw was entrusted with the designing of the building and his plan, although adopted, was the subject of a great deal of disagreement in official and other circles. By late 1890 the building was completed, faced up to the top of the second floor with 2,500 tons of granite quarried and dressed by the convicts of Dartmoor, and continuing upward in red brick with Portland stone dressings. On May 11, 1941, a high explosive bomb fell on one corner, causing extensive damage and resulting in many tons of masonry and millions of index cards from the registry two floors above falling into the commissioner's room. No lives were lost, but had the building been less substantial far greater damage would have been done.

The move of all the headquarters staff to the new premises took place in Dec. 1890, and the building was officially designated New Scotland Yard. Scotland House (to become known later as New Scotland Yard South), occupied by the receiver for the metropolitan police district, to the south of New Scotland Yard and connected by an archway over the road, was added several years later, from Shaw's designs. At the same time were added the wrought-iron entrance gates, which had been designed by Thomas Elsley as an exhibition piece and found by Shaw at Elsley's shop, standing unused.

By 1935 the position at New Scotland Yard had once again reached a point where some extension or re-siting was very necessary, and it was decided by the commissioners of the office of works to incorporate a new police building in the development scheme for Richmond Terrace, Westminster, and a site extending northward from New Scotland Yard was agreed upon.

The reproduction of Shaw's design in the new building was found unacceptable on the grounds that it would be difficult to reconcile it with further developments in the immediate vicinity and that New Scotland Yard itself was considered to be wasteful in planning by modern standards. A building more in relationship with modern ideas was required, and indeed it was even suggested that Shaw's building should be demolished and replaced by buildings with better accommodations and which were more in harmony with a general development scheme. This suggestion was, however, dropped when questions of finance and the remoteness of such a general scheme were raised.

W. Curtis Green was the man selected to design the new extension, which soon began to take shape. At the outbreak of World War II in September 1939, however, the building was still 14 months from completion. Under increased pressure of war duties, it was found necessary to move some personnel into the unfinished building, and in early 1940 parts of it were occupied while work was still in progress on the upper part of the structure. By December 1940 the building was ready for complete occupation, but certain parts of it were, as a wartime necessity, used for purposes other than police functions.

The section of the fighting services known as combined operations took up its headquarters in the building in 1940, and occupied part of it until 1945. It was not until the latter year that the building became completely available for its original purpose. By mid-century it housed the immense telephone switchboard of Scotland Yard, the information room with its emergency 999 service, the forensic science laboratory, the criminal record office, the fingerprint and photography departments, and other branches of the criminal investigation department. At first it was decided to name the new building Richmond House, as the first step in the Richmond Terrace Development Scheme, but on reconsideration it was thought that the title of Scotland Yard had come to have such a special meaning for the metropolitan police that its use should be retained.

The new building was accordingly designated New Scotland Yard North, and the official name of Shaw's building was altered to New Scotland Yard Central. Scotland House, the receiver for the metropolitan police district, became officially New Scotland Yard South. By the 1960s, New Scotland Yard Central was inadequate, and it was planned to re-site by 1967 the whole metropolitan police headquarters at the east end of Victoria St., near the Houses of Parliament.

Apart from the 175 police stations and other police buildings in the metropolitan police district, there are the public carriage office, the police garage and lost property office at Lambeth, the aliens' registration office in Theobalds Road, and offices at Walton Street, Kensington, and Ramillies Place, Westminster. All are working branches of Scotland Yard's headquarters.

Organization.—At its commencement in 1829, the extent of the metropolitan police district controlled from Scotland Yard was approximately 120 sq.mi., with an estimated population of less than 1,500,000 persons; for this area there was a police force of 1,000. Ten years later, in 1839, the area was extended to 699.42 sq.mi., and the strength of the police had been increased to 3,350 for a population of nearly 2,250,000. No further alteration was made to the size of the district until 1947, when, by a readjustment of police boundaries under the Police Act of 1946, it was extended to cover 734.88 sq.mi. By the early 1960s the strength of the

force had increased to over 18,000, for a population exceeding 8,000,000. The London Government Act, 1963, increased the metropolitan police district by about 51 sq.mi., to a total of 786 sq.mi.

At the top of the administration of Scotland Yard and in charge of all the police of the metropolis is the commissioner, who is appointed by the crown on the recommendation of the home secretary. Directly under the commissioner is the deputy commissioner, and below them at headquarters there are six main departments: administration ("A"); traffic ("B"); criminal investigation ("C"); organization ("D"); secretariat ("S"); and the legal department ("L"). The first four of these departments are each headed by an assistant commissioner. Each assistant likewise is appointed by the crown on the recommendation of the home secretary. The commissioner, deputy commissioner, and assistant commissioners are also justices of the peace for Greater London and the counties of Surrey, Hertfordshire, Essex, Berkshire, Kent, and Buckinghamshire, but their functions as justices are limited to executive police matters.

The assistant commissioner at the head of the administration department is responsible for discipline, distribution, and other administrative matters in the force, and his department includes the women police and mounted branch. Not the least of his duties is the control of all police arrangements in connection with the many ceremonies in London.

The traffic department deals with all traffic matters, including supervision of the public carriage office (with its licensing and control of cabs and their drivers, and the drivers and conductors of buses, coaches, and trolley vehicles in the metropolitan area), police transport, the police motor driving school, and the lost property office.

The criminal investigation department deals with all aspects of criminal investigation, including the criminal record office, fingerprint and photography sections, the company fraud squad, the flying squad, special branch, the metropolitan police laboratory, and the detective training school.

The organization department is concerned with civil defense, supplies and equipment, first aid, buildings and housing, recruitment and training, police dogs, medical services, communications including the information room, and welfare.

The secretarial department is under the control of the secretary, who is not a police officer but head of the civilian staff of the metropolitan police force, appointed by the home secretary on the recommendation of the commissioner. His department employs more than 4,000 civilians in all its secretarial work, including its vast registry, pay and pensions, and the press and information department.

The legal department is headed by a solicitor. The solicitor to the metropolitan police district is appointed by the home secretary on the recommendation of the commissioner and is also the prosecuting solicitor to the quarter sessions within the metropolitan area.

In addition to these main departments, a small research and planning branch was established, responsible directly to the commissioner, through the deputy commissioner, for keeping the whole organization under constant review, particularly with regard to manpower and future planning. (R. M. H.; X.)

SCOTS LAW. At the union of the Parliaments of England and Scotland in 1707, the legal systems of the two countries were very dissimilar. Scotland, mainly in the preceding century, had adopted as a guide much of the Roman law, as developed by the jurists of Holland and France. But it is a fallacy to suppose that the law of Scotland is founded on the law of Rome: the Scots only turned to Roman law when there was a gap in their own common or customary law. There is, however, a considerable infusion of Roman law, not least in the field of nomenclature and in the emphasis on principle rather than precedent. English lawyers, on the other hand, had forgotten or refused to acknowledge the debt owed to Rome by both common law and equity. At the union the law of Scotland had been set forth in the *Institutions* (1681) of Lord Stair, a masterpiece of lucidity and orderly arrangement. In England there was little guidance through a maze of precedents and forms of pleading beyond the difficult pages of Coke. Perhaps the most important distinction was the fact that Scotland, unlike England, did not separate the administration of equity and law.

The Scottish conception of equity is not, like the English, that of a system running parallel to the common law (*see* EQUITY). It is rather: (1) a few fairly simple rules aimed at supplementing the law in order in avoid hardship; and (2) the relegation of certain remedies to the class of equitable remedies, which the court has a large discretion to grant or withhold. The word "equity" in the law of Scotland has always retained its original meaning. The Scottish outlook upon this whole topic places Scots law clearly alongside the continental and not the English system.

Historical Development.—The period following the union has been characterized by the assimilation of Scots and English law. One main cause of assimilation is that much of the existing law of Scotland depends on statutes applicable to both countries. The House of Lords, consisting in its legal aspect until 1876 exclusively of English lawyers acting as the supreme court of appeal from Scotland, had a tendency to apply English law in Scottish appeals, and in some cases seems to have forgotten the distinction between its legislative and its judicial functions. Another cause of assimilation is the influence of text writers, some of whom, such as G. J. Bell, have tended to treat English law as though it were the law of their own country. The citation, too, of English authorities in court (partly explicable by the dearth of modern Scottish textbooks), despite judicial remonstrance, has had considerable effect. Not surprisingly the greatest assimilation has been in the field of mercantile law. In other fields the systems are still widely separated.

Courts of Law.—The system of Scottish courts is completely different from that of the English and here again is closer to the continental pattern. The supreme Scottish court (the House of Lords, which is not a native court, being for the moment disregarded) is the Court of Session, instituted by King James V in 1532 probably upon a French model. The court has two main functions. It has original jurisdiction in a very wide range of cases, which is exclusive in a few matters such as actions affecting status (*e.g.*, divorce and bastardy); in its appellate capacity it hears appeals (by reclaiming petition) from the nine Court of Session courts of first instance (called compendiously the Outer House), each presided over by a lord ordinary, and also from the sheriff's courts. The appellate court (Inner House) sits in two divisions, the first and second, presided over respectively by the lord president of the Court of Session (*see* LORD JUSTICE-GENERAL) and the lord justice clerk (*q.v.*). All the judges have the courtesy title of "lord," but are not on that account peers.

While the judges of the Court of Session are traditionally judges of both fact and law, in the early 19th century the civil jury was introduced, less because it was wanted in Scotland than because the House of Lords was weary of the great number of appeals which (since the decision of a jury is not in the ordinary sense appealable) would be drastically reduced by the change. From the Inner House appeal lies in many cases to the House of Lords as of right and not, as in England, by leave. The right of audience in the Court of Session is possessed exclusively by members of the faculty of advocates.

The lower civil courts are first the sheriff courts, which are ancient courts distributed by counties; then the burgh police courts and justice of the peace courts. Sheriff Courts are manned by (1) the sheriff principal and (2) the sheriff substitute. The former (with administrative as well as judicial functions) is in civil matters an appellate judge only and hears appeals from the sheriffs substitute of his jurisdiction. The sheriff substitute's absurd title conceals his great importance. For except in divorce and one or two other less important matters his jurisdiction almost exactly coincides with that of the Court of Session. As a local and comparatively inexpensive court the Sheriff Court is more popular than the Court of Session. It will be seen that the Sheriff Court (which has also a wide criminal jurisdiction) cannot for a moment be compared with the English County Court, with its very limited civil jurisdiction and complete lack of criminal jurisdiction. The Justice Court has a limited jurisdiction in small debt actions.

The dean of Guild Court has a quaestorial jurisdiction in questions of building in the towns.

It should be added that the Court of Session has absorbed the functions of certain ancient courts, the Court of Exchequer, the Admiralty Court, the Teind (or Tithe) Court, and the Commissary Court, which dealt with questions of marriage law and executry, while the judges have by statute been given separate duties in a Lands Valuation Appeal Court, a Registration Appeal Court, and an Election Petition Court.

The Scottish Land Court, established in 1911, has jurisdiction in a wide range of matters relating to agriculture. Questions between landlords and tenants of agricultural holdings may be brought before it by judicial process or, by agreement of parties, in lieu of arbitration. It also deals with questions referred by the secretary of state for Scotland.

Land Tenure.—The law of land tenure has been complicated by the survival of much of the feudal system. In theory the crown is the ultimate owner of land, which passes from it to others in the first place through a grant. The grantee or "vassal" may make a subaltern grant and so can successive grantees, by what is called subinfeudation, a process forbidden in England by the statute Quia Emptores (1290). The grantee can also sell his land so as to take himself out of the descending chain and put a purchaser in his place. As in other systems, land can be used to provide security for a loan to the owner of it. This is done by a bond with a disposition of the land in security, or by an outright conveyance together with a "back letter" from the lender entitling the owner to recover the land on repayment of the loan. The word mortgage is not a Scottish term. The consequence of unrestricted subinfeudation has been that the rights of subject superiors—the superior is the person who feus or grants the land—have bulked much more largely in Scotland than in England. A series of statutes, culminating in the Feudal Casualties Act, 1914, abolished all payments exigible from the vassal except feu duty, and at mid-20th century it might be said that a feu, the normal tenure in Scotland, was equivalent to an English freehold, subject, in cases where there was a feu duty, to a perpetual rent charge. There is, however, a very important distinction in the system of registration of title, only imperfectly developed in England. In Scotland, since the establishment of the Register of Sasines in 1617, all deeds relating to land may be recorded in that register, and it has long been established that a purchaser, or a lender on heritable security, is entitled to trust to the register, and is not affected by any conveyance or burden not there recorded.

The law of house leases differs markedly from the law of England in a number of ways. For example, Scots law demands in most cases that notice of intention to remove be given even if the lease itself provides for termination at a particular date. Failing this the lease will be continued by what is called tacit relocation. Also, the landlord is bound to provide and to maintain premises reasonably fit for the purposes of the let. Failure to repair may be met by retention of the rent by the tenant. On destruction of the premises both parties are freed from their obligations. Local rates in Scotland, formerly payable partly by the owner of property and partly by the occupier, are payable only by the occupier.

The law of agricultural leases is very different from the law of town leases and is complex. For example, a tenant is entitled to compensation for disturbance if given notice to quit, for improvements made by him, and for damage done by game. A tenant is entitled, on conditions, to bequeath his lease. Further specialties apply in the case of the so-called small landholder and crofter by virtue of various statutes. The Crofters Act, 1955, established a crofters' commission to promote the welfare of the crofting communities.

The Rent Restriction Acts, running from 1915 onward, are designed to prevent increase of the rents of small and medium houses and to give security of tenure. While no doubt serving these purposes, the acts led to a natural reluctance on the part of owners to spend money on what, with the diminished purchasing power of money, was an unprofitable investment. This in turn led to a serious deterioration in the condition of property covered by the acts. Consequently, in the 1950s, legislation considerably relaxed the control of rents and enabled owners to recover more readily the possession of their property (*see* LANDLORD AND TENANT).

Rights in Land.—Apart from its theoretical ownership of all Scottish land the crown is an important owner of heritable property; *i.e.*, land and what is built upon or planted in land. Its rights are known as the *regalia*. Some are *majora* and are inalienable. The *minora* or lesser rights may be alienated.

The foreshore of the sea may be alienated but not the public rights to anchor boats, embark, and disembark. Mention may be made of particular *regalia minora* such as the rights to salmon fishing, to precious metals, and to highways. Possession of heritage has some important consequences. Thus a bona fide possessor who unjustifiably takes the produce of the land need not make it good, in contradistinction to the possessor in bad faith, who must pay penal damages. Again a bona fide possessor is entitled to recompense for improvements.

The individual owner of land is entitled in theory to use it as he likes, but there are numerous statutory restrictions and disabilities. Among common-law rights, the right of support of land in its natural state by the land of one's neighbours is important. Surface or percolating water may be impounded but water in a stream may not be prevented from descending to lower proprietors. Where the lands of several proprietors adjoin a loch, they have a joint right in it. The right to fish in inland waters and the right to take game are incidents of property. It is open to a proprietor of land to confer upon neighbouring proprietors what is called a servitude; *i.e.*, the right to use the grantor's land in some way (a positive servitude) or the right to restrain the grantor from using his land in some way (a negative servitude). Examples would be, respectively, a right of way given to X over A's land and a right to X of preventing A from building above a certain height on his, A's, land. A public right of way, unlike a servitude, exists for the benefit of the public and not for the benefit of a particular piece of land. When land is held on a title *ex facie* valid for 20 years recorded in the register, an unchallengeable title is acquired by prescription.

Husband and Wife: Children.—The enfranchisement of women has placed husband and wife, as in England, upon an almost equal footing in matters of property. Gone are the days when on marriage the wife's whole movable property became her husband's and his assent was needed to any transaction affecting her heritage. In the law of marriage and divorce, too, the two systems have been gradually drawing closer. Thus the celebrated irregular marriage of Scotland has all but disappeared, though irregular marriage through "cohabitation with habit and repute" remains valid. "Gretna Green" is no more. Yet in fairness it should be realized that the irregular marriage was as binding as the regular and found a warranty in the canon law. Its abolition, many think, was a misfortune. In the law of divorce, on the other hand, English law has come closer to Scots law. In Scotland, divorce on the grounds of adultery and desertion was allowed to either party from shortly after the Reformation (1560). Today divorce is granted for adultery and for desertion by either spouse for three years; as also on the ground of cruelty, unnatural vice, and incurable insanity, as in England. One very important difference between the two systems is that in England "recriminations," *i.e.*, the answer by a respondent that the petitioner has also committed adultery, is a bar to a divorce unless the court exercises discretion. In Scotland recrimination is no defense and a party, though himself or herself guilty of adultery, is entitled to divorce when the other party has committed adultery. Cross divorce is the solution.

In Scotland a decree of divorce dissolves the marriage at once, a contrast with the English rule. In both countries the crown may interest itself in the proceedings as, for example, to discover collusion or, in England, concealment of adultery by the petitioner. In Scotland this intervention is made by the lord advocate. It is rare, probably because the law does not recognize recrimination. There is no queen's proctor. For the matrimonial offenses of adultery and cruelty (including habitual drunkenness) the

courts grant, if sought, the lesser remedy of judicial separation. This, if allowed, is normally accompanied by an award of aliment. No divorce or separation decree will be pronounced unless the court is satisfied that the best arrangements possible have been made for any children under 16 years of age. For property division on divorce, until 1964 the guilty spouse was treated as notionally dead and the legal rights exigible on death (see *Succession*, below) were similarly exigible on divorce. From 1964 the successful pursuer in the action of divorce might ask the court for an order for payment to him (or her) by the defender of a capital sum, or a periodical allowance, or both. Any order for a periodical payment is variable on change of circumstances, and terminates on remarriage or death of the pursuer.

Children.—Scots law divides minority into: (1) minority proper, from 12 to 21 for girls, 14 to 21 for boys; and (2) pupillarity, up to 12 and 14 respectively. Pupils act by their tutors, normally the father. Minors act with the consent of their curators, normally the father. A pupil can never act for himself but a minor, if he has no father, need have no curator and can legally treat himself as if he were major. The independence thus allowed to young people is an outstanding feature of the law. But 16 is the lowest age for marriage. Disadvantageous transactions by minors and pupils can be set aside within four years of attaining 21.

Intestate Succession.—Scots law never admitted absolute freedom of bequest where there was a surviving spouse or child. A distinction was drawn between heritable and movable property. A widow has a legal right in movable property (*jus relictae*); to one-half of her husband's movable property if there is no surviving child; to one-third, if there is. A widower since 1881 has a similar right in the movable property of his wife (*jus relicti*). Children have a legal right in the movable estate of each parent, one-half if there is no surviving spouse, one-third if there is. Legal rights cannot be excluded or limited by will, and attempted disregard of this may be successfully challenged in the courts.

In heritable property the widow had a legal right (terce) to a liferent (use and enjoyment for life) of one-third. A widower generally had a right of courtesy which gave him a liferent of the whole of the wife's estate. The ancient distinction between heritable and movable property, however, as well as the legal rights of terce and courtesy, were abolished by the Succession (Scotland) Act, 1964. Under this statute, if a husband or wife dies intestate the surviving spouse may claim the dwelling house (belonging to the deceased) or in certain cases its value up to £15,000; and furniture and plenishings up to the value of £5,000. If there is no issue the surviving spouse can also claim £5,000; and, if there is issue, £2,500. If the surviving spouse, however, is left a legacy he or she cannot take both this and the benefits under the act. If a woman dies intestate leaving illegitimate children those who survive take the whole of her estate.

Testate Succession.—No legacy made orally is valid unless it is of £100 Scots (£8 6*s.* 8*d.*) or less. This apart, a will must be made by a writing signed and witnessed as a deed (by two witnesses) or by a writing entirely in the testator's handwriting and signed by him. When a testator cannot write, the will may be signed for him by a solicitor, notary, justice, or parish minister.

The validity of a will is, in Scots law, not affected by the subsequent marriage of the testator. With considerable differences of detail, the law of wills is not markedly different from that of England. However, the existence in Scotland of "legal rights" to succeed introduces a distinctive complication. (*See* also WILL.)

Executors and Trustees.—These are persons nominated by the deceased (executors nominate) or appointed by the court (executors dative) who alone are entitled to administer the estate and pay creditors and legatees. Sanction to act is conferred by confirmation; the term probate is not used in Scotland. Any other person dealing with the estate is called a *vitious intromitter* and may, unless relieved by the court, incur liability for the debts of the deceased. A special summary process is competent in the case of small estates of a few hundred pounds in value. The heir of an executor does not succeed to his office. The same persons are frequently executors and trustees in estates of any size and especially where life interests have to be carried on. The law of trusts, contained principally in the Trusts (Scotland) Act, 1921, resembles the law in England in many respects, but the conception of holdings in trust as a sort of property running parallel to legal property is quite foreign.

A Scottish trust is normally constituted by an act of its creator, but it may also arise through implication from circumstances. Writing is not in all cases necessary. Trustees act by a majority. They may assume new trustees and may resign if they provide for the continuance of the trust. A trust may be terminated with the consent of grantor and beneficiaries, but a trust taking effect on its creator's death may not be terminated even with the beneficiaries' consent if its purpose is alimentary; *i.e.*, for the maintenance of the beneficiary. Such a protected trust may be created in favour of a man.

Contract.—There are three possible forms of contract: it must be writing (*e.g.*, for sale of land); it may be made orally but must be proved by writing (*e.g.*, contracts of an unusual nature); or it may be made and proved in any way (*e.g.*, sale of goods). The specialties of the document under seal are unknown, and consideration (something of value given in exchange) is unnecessary. Where writing is needed merely by way of proof and none is available, the party desiring to prove the contract may have the other put on oath and called upon to admit the contract. If this fails only part performance can save it. The only drawback to a contract lacking consideration is that it must be proved by writing. It is a peculiarity of Scots law that when a third party, X, is deliberately benefited by a contract between A and B, X has a right to sue A or B if either of them fails to perform the obligation. Where one party has been precluded through some cause beyond his control from performing his obligation, the law of Scotland enforces the return of any advance payment that has been made. (The law of England which refused this had, by mid-20th century, come into line.) Mention may be made of the negative and short prescriptions, so closely though not exclusively connected with contract. Very shortly, a person who fails to enforce his rights under contract for 20 years is prevented by the long negative prescription from doing so. Shorter periods such as three and six years apply to particular contracts, but these in the usual case merely add the need of writing to the burden of proof, without extinguishing the rights.

A useful conception is that of quasi contract. This means the obligation imposed, not by express agreement, but by force of circumstances. One example is *negotiorum gestio*, which means the obligation to make good expense incurred in intervening to help another in an emergency.

Sale of Goods.—The Sale of Goods Act, 1893, provides a short code of law for Scotland and England. Although many differences were reconciled in the act, a few, of varying importance, remain, such as the law pertaining to the effect of breach of warranty and condition. In English law the remedies of a buyer depend upon whether the seller's failure is concerned with a warranty or a condition. In Scotland the question is whether the seller's failure is a material failure. As regards forms, except in the case of hire-purchase (installment) contracts, there is no need for writing. (*See* also SALE OF GOODS.)

Movable Property.—Whatever has never had an owner may be taken and converted into property by *occupatio*, subject to some restrictive laws. Nondomestic animals are an example, even game, but poached game may be forfeited. *Accessio* is another mode: whatever is affixed to heritage or grows in it tends to become the property of the owner of the heritage. (Land itself cannot be acquired by *occupatio* since land in nobody's hands reverts to the crown.) But most property is acquired by transactions such as sale or gift or succession. By other transactions such as liferent, hire, deposit, and loan something short of property is acquired, usually possession, which is a concomitant of property. These transactions have their counterpart in heritage (immovable property), and liferent is of special importance there. A liferent cannot be created in favour of a person as yet unborn as, for example, by a bachelor in favour of his legitimate grandchild, who would, if this were attempted, become the owner.

Liferent confers rights upon the holder so extensive as to render

him virtually owner for life. Thus the liferenter is liable for discharge of burdens of all kinds upon the property not in the nature of capital payments; while he is entitled to use the liferented property as he pleases, except that he may not dilapidate it nor dispose of it for a period longer than the duration of his life. One important feature of the Scots law of property is the stress laid upon possession of the subject of a transaction. With some exceptions no security over movables can be created without actual or constructive delivery. Several important differences from English law follow. The bill of sale is unnecessary and unknown and the "floating charge" is incompetent. Security, too, by way of deposit of deeds is out of the question. Gift is, in accordance with the same principle, effected only by transfer of the thing given. And before 1893 the property in goods sold only passed on actual delivery.

Delict.—In modern usage delict signifies a wrong in its civil aspect and closely corresponds to tort in Anglo-American law. The basic rule of Scots law is that whoever injures another in his fortune, his person, or his good name either intentionally (delict) or through negligence (quasi delict) is liable to pay such damages as will, so far as practicable, restore the *status quo ante.* This is known as reparation. Damages are therefore given for injury, pain, damage to property, interference with trade, defamation, wrongful imprisonment, and the like. Injury to feelings, as through bereavement, is met by an award of solatium. This occurs where near relatives (father, mother, children) or husband or wife are killed.

The law of reparation rests upon the existence of culpa or fault and no liability attaches because of mere ownership of an object which causes damage. It is, too, only for the foreseeable results of negligence that damages are given. An important body of case law illustrates the liability of employers for the acts of their servants. In the law of defamation no distinction is drawn between written and spoken words. Utterance to the defamed person himself constitutes publication. It is not necessary in any case to prove special (particular and ascertainable) damage. There is no such thing as criminal libel, and holding up to mere ridicule is not actionable. Whatever the preunion law of Scotland, the crown was in postunion times immune from liability for wrongs done by its servants. In 1947, however, the Crown Proceedings Act introduced liability in most cases, thus probably restoring the Scots common law.

Civil Procedure, Diligence, and Bankruptcy.—All these differ greatly from their English counterparts. Civil procedure in the Court of Session and Sheriff Court is similar in many respects. If A has a claim against B it will first be made by the pursuer's (A's) solicitor against the defender, B. If the allegations are disputed or the case is not settled extrajudicially, the pursuer serves a summons on the defender, setting forth: (1) the sum or other remedy sued for; (2) a "condescendence" or statement of facts (averments) on which his claim is based and which he proposes to prove by evidence; (3) numbered pleas-in-law which set forth his legal contentions on the facts averred. The defender, if he means to defend the case, lodges defenses specifically answering the pursuer's averment and in turn stating his pleas-in-law. When the summons and defenses (known as "the pleadings") have been "adjusted" (*i.e.*, elaborated and revised), they are embodied in a document called a "closed record"; and at the subsequent inquiry evidence is restricted to proving the facts averred therein. Much importance is attached to relevancy; *i.e.*, the averments made must, if proved, entitle the party to the remedy sought. If they would not, then the case may be dismissed outright. If the case proceeds there is no opening speech except in a jury trial. The evidence may be heard by a judge or a sheriff substitute sitting alone (a proof); or in certain types of case, *e.g.*, action of damages for personal injury, by a judge and jury. In the Sheriff Court, the jury (7 in number as against 12 in the Court of Session) only hears actions of damages for more than £50 by employees against their employers, usually in respect of injuries sustained in accidents arising out of their employment. In a proof, when the evidence is complete, the parties' legal representatives make their submissions on the evidence to the judge, who may give an immediate judgment or take the case to *avizandum, i.e.*, for consideration, when he will give his judgment later. In a jury trial the speeches are made to the jury, who are then charged by the judge on the law applicable to the case. They retire to consider their verdict, and they will assess any damages. In the Sheriff Court the case is usually conducted throughout by a solicitor, though counsel (advocates) may appear. In the Court of Session only advocates, who must be instructed by solicitors, prepare the pleadings and have a right of audience. Both branches of the legal profession for centuries have provided free legal advice for poor persons in civil and criminal matters. After 1949, however, in civil proceedings, state provision for legal fees was available to persons of moderate means, often subject to a contribution by the assisted person.

Appeal from the decision or verdict lies: (1) in the Sheriff Court, first to the sheriff principal and thence to the Inner House of the Court of Session; or direct to the Inner House. In a jury trial, appeal is to the Inner House. (2) In the Court of Session, appeal lies from the lord ordinary, *i.e.*, the judge who first hears the case, to the Inner House. From the Inner House, appeal is to the House of Lords, whose decisions are binding on all inferior Scottish courts.

Diligence is the Scottish equivalent of execution and it takes several forms. Thus, for example, poinding (pronounced "pinding") means taking movables to satisfy the decree of the court. Arrestment involves directing the debtor of the judgment debtor to pay the latter's creditor direct. Adjudication means taking heritage in satisfaction.

Bankruptcy (technically "sequestration") is the state of officially recognized insolvency which follows usually upon unsatisfied diligence. It affects all property of the bankrupt in the British Isles. Various types of arrangements with creditors may avert the disaster. Statutory safeguards prevent abuse by way of fraudulent preference and the like. A bankrupt suffers civil disabilities ceasing on his discharge, which depends upon his conduct and upon satisfactory payments to the creditor. (*See* BANKRUPTCY.)

Criminal Law.—A crime is a wrong which, even if it only directly injures an individual, is held to be harmful to the community. Prosecution is therefore almost always a public function carried through by a public prosecutor. (*See* SCOTLAND: *Administration and Social Conditions: Justice.*)

Proof of a guilty intention on the part of the accused, as well as of an overt act, is necessary before a conviction can be obtained except where statute makes an exception or where the court has admitted the principle of vicarious responsibility. Such circumstances, therefore, as youth (no crime under eight), insanity, or compulsion will, subject to certain rules, render conviction impossible. The question of the responsibility of the insane is no more satisfactorily settled than elsewhere, but Scots law is more lenient in cases of insanity than English law. The English distinction between felony and misdemeanour is not made.

Nomenclature differs considerably from the English, and the same word (like embezzlement) may point to different offenses. Burglary and housebreaking are not distinguished; manslaughter is culpable homicide; receiving is reset; concealment of pregnancy and not concealment of birth is the crime and so on. The criminality of suicide lacks satisfactory warranty and D. Hume, the prime authority, shies away from pronouncing even upon attempt. In modern practice the public prosecutor may, but rarely does, prosecute an attempt as a breach of the peace. Almost always such cases, which usually involve insanity, are dealt with administratively after medical examination. The incidence of crime is different north and south of the border and the popular attitude to certain crimes also differs. Thus blackmail is not common in Scotland. And for many years in the earlier part of the 20th century Scottish juries could scarcely be brought to convict of murder.

Criminal Courts and Procedure.—Criminal procedure is either "solemn," *i.e.*, commenced by an indictment and tried before a judge and jury; or "summary," commenced by a complaint and tried before a sheriff or magistrate sitting alone. The High Court of Justiciary, the supreme criminal court, consisting of

judges of the Court of Session sitting as Lords of Justiciary, sits with one judge and a jury for trials; three or four of its judges decide appeals against conviction or sentence in cases tried either on indictment or summarily. The Sheriff Court tries cases in either manner. For summary cases, unless sentence is prescribed by statute, the sheriff may impose fines up to £150 and imprisonment up to six months. On indictment, he may sentence up to two years' imprisonment; and, if he considers that the crime requires more severity, he may remit the prisoner to the High Court for sentence. The burgh police courts and county justice of the peace courts can hear cases summarily and impose fines up to £25 and imprisonment up to 60 days.

Where a crime has been committed, the police report their findings to the district procurator fiscal, who may, and in all cases to be tried on indictment must, consult the crown counsel, at the Crown Office, in Edinburgh: they in turn being instructed by the crown agent. It is for the procurator fiscal or crown counsel to decide if a prosecution is to be brought, and, if so, in what court and whether on indictment or summarily. Proceedings are at the instance of the lord advocate or the procurator fiscal; the police do not prosecute. All investigation is private and potential jurors cannot be prejudiced by reading reports of earlier hearings, as could happen in England. An accused has no right to elect for trial by jury. A constable or citizen who sees a crime may arrest the offender without warrant; otherwise a warrant is obtained from a magistrate or sheriff.

Solemn Procedure.—In cases of serious crime the procurator fiscal presents a petition to the court setting forth the charge, and on this a warrant is granted. The accused, on arrest, is brought before the court and asked if he wishes to make a declaration. (Declarations, formerly common, are now rare.) Thereafter he may be granted bail. An indictment is served on him, specifying the charge and listing the witnesses and productions by which the crown proposes to prove the charge. At a pleading diet he intimates whether he pleads "guilty" or "not guilty." If "not guilty," the next stage is the trial. A jury of 15 is empaneled and the indictment read to them. There are no opening speeches: the evidence is led, first for the prosecution and then for the defense. The accused must be present throughout and is presumed innocent until proved guilty, the onus being on the crown to establish the charge beyond reasonable doubt. The proceedings are recorded in shorthand. At the conclusion of the evidence, the prosecutor addresses the jury first, the defense next, and the judge sums up. The verdicts, which can be by a majority, are "guilty," "not guilty," and "not proven"—this last when grave suspicion attaches to the accused but there is insufficient evidence to warrant conviction. There can be no retrial, even on a verdict of "not proven." The crown has no right of appeal against an acquittal.

Summary Procedure.—In minor offenses, a complaint is served on the accused narrating the charge, *e.g.*, careless driving. For certain statutory offenses, *e.g.*, traffic offenses, pleas of guilty can be made by letter and the accused be dealt with in his absence. If the plea is "not guilty," the evidence is heard by the sheriff substitute or magistrate, and the same presumptions and rules of evidence apply as in solemn procedure. After the parties have addressed the judge on the evidence he gives his verdict and sentences the offender. First offenders may not be imprisoned unless the court is satisfied there is no other appropriate way of dealing with them. If a fine is imposed, the court may allow it to be paid by installments and may not sentence to imprisonment for nonpayment without first inquiring into the accused's means.

Juveniles.—Special courts hear charges against children (aged 8–14), and young persons (14–17); and also deal with juveniles in need of care or protection, or who are refractory. If the accused is not legally represented his parent, guardian, or other relative may assist his defense. If found guilty he may be admonished; or put on probation; or fined (his parent is fined if he is a child); or detained for up to 28 days; or sent to approved school (up to age 17) or borstal (16–21). The court's paramount consideration is his welfare.

BIBLIOGRAPHY.—Viscount Stair, *Institutions of the Law of Scotland*, ed. by J. S. More (1832); J. Erskine, *An Institute of the Law of Scotland*, ed. by J. B. Nicholson (1871); G. J. Bell, *Principles of the Law of Scotland*, 10th ed. by W. Guthrie (1899), *Commentaries on the Law of Scotland*, 7th ed. by J. M'Laren (1870); *Green's Encyclopaedia of the Laws of Scotland*, 16 vol., 3rd ed. by J. L. Wark *et al.* (1926–33); W. M. Gloag and R. C. Henderson, *Introduction to the Law of Scotland*, 6th ed. by A. D. Gibb and N. M. L. Walker (1956); W. M. Gloag, *Law of Contract*, 2nd ed. (1929); A. D. Gibb, *Preface to Scots Law*, 4th ed. (1964); T. B. Smith, *A Short Commentary on the Law of Scotland* (1962); J. J. Gow, *The Mercantile and Industrial Law of Scotland* (1964); J. D. B. Mitchell, *Constitutional Law* (1964); D. M. Walker, *The Scottish Legal System: an Introduction to the Study of Scots Law*, 2nd ed. (1963).

Criminal Law: D. Hume, *Commentaries on the Law of Scotland Respecting Trial for Crimes*, 4th ed. (1844); J. H. A. Macdonald, *Practical Treatise on the Criminal Law of Scotland*, 5th ed. by James Walker and D. J. Stevenson (1948).

Comparison of Scots and English Law: Lord Bankton, *Institutes of the Law of Scotland*, 3 vol. (1751–53); J. Paterson, *Compendium of English and Scots Law*, 2nd ed. (1865); J. W. Brodie-Innes, *Comparative Principles of the Law of England and Scotland: Courts and Procedure* (1903); H. Burn Murdoch, *Notes on English Law as Differing From Scots Law* (1924); A. D. Gibb, *Law from Over the Border* (1950); C. d'O. Farran, *The Principles of Scots and English Land Law* (1958).

Dictionaries: W. Bell, *Dictionary and Digest of the Law of Scotland*, 7th ed. by G. Watson (1890); A. D. Gibb and A. W. Dalrymple, *Scottish Judicial Dictionary* (1946). (A. D. G.; J. I. S.)

SCOTT, ALEXANDER (*c.* 1525–*c.* 1585), Scottish poet, one of the "last of the makaris" of the 16th century, a writer of lyrical freshness and technical skill. Nothing is known of his life, though he seems to have been familiar with Edinburgh and Dalkeith, and he is probably the "old Scott" referred to by Alexander Montgomerie, his younger poetic contemporary, in a sonnet dated *c.* 1584. All his extant poems, about 35 in number, are contained in the Bannatyne manuscript (1568; in the National Library of Scotland). Since Allan Ramsay printed a selection of his verse in his collection of old Scots poetry, *The Ever Green* (1724), Scott's reputation as a genuine minor lyric poet has been maintained. He left an amusing burlesque, *"The Justing and Debait vp at the Drum betuix William Adamsone and Johine Sym."* He also wrote a ceremonial and highly alliterative poem, *"Ane New Yeir Gift to the Queen Mary, quhen scho come first Hame, 1562,"* which gives an interesting reflection of early Reformation Scotland. His best work, however, is to be found in his love lyrics; these show a striking range of mood, from the tender to the coarse, and an admirable metrical suppleness and variety.

BIBLIOGRAPHY.—Scott's poems were edited by J. Cranstoun, Scottish Text Society (1896); by A. K. Donald, Early English Text Society (1902); and by A. Scott, Saltire Society (1952). *See* also C. M. Maclean, *Alexander Scott, Montgomerie, and Drummond of Hawthornden as Lyric Poets* (1915). (E. G. M.)

SCOTT, CYRIL MEIR (1879–), English composer and poet known for his piano and orchestral music, was born at Oxton, Birkenhead, on Sept. 27, 1879. In 1891 he began to study the piano at Frankfurt am Main, where he returned in 1895 for three years' training in composition with Ivan Knorr. During this period he became a friend of the writers Stefan George and Charles Bonnier and began to write poetry about 1900. He published translations of Baudelaire and two volumes of poems in 1910 and 1912. His early musical compositions, which he later destroyed, were performed at Liverpool, Manchester and Darmstadt; his reputation was established with a piano quartet (1901) and Second Symphony (1903). Under the influence of Claude Debussy and of oriental music, he later developed a freer harmonic and rhythmic style, notably in his sonatas for piano and for violin and piano. In 1913 *La Princesse Maleine*, the third of his overtures on plays of Maeterlinck, was given in Vienna. Between 1915 and 1948 he wrote several orchestral and choral works, including concertos for piano, harpsichord, violin and cello, a symphony, and a setting of Keats's *La Belle Dame sans Merci*. His opera, *The Alchemist*, was given at Essen in 1925. His chamber works include trios, quartets, quintets and numerous short pieces for the piano. His songs, written between 1903 and 1939, are settings of poems by himself, Ernest Dowson, Christina Rossetti, R. M. Watson and others. At the beginning of his career Scott established a reputation as one of the principal English composers known on the continent. His critical works include *The Philosophy of Modernism in Its Connection With Music* (1926) and *The Influence of Music on*

History and Morals (1929). He published his memoirs, *My Years of Indiscretion*, in 1924.

See A. E. Hull, *Cyril Scott: Composer, Poet and Philosopher* (1921).

SCOTT, DUKINFIELD HENRY (1854–1934), English paleobotanist, the leading authority of his day on the structure of fossil plants, especially those of the Carboniferous Period, was born in London on Nov. 28, 1854. His early education was by private tutors. In 1876 he graduated from Christ Church, Oxford. He became financially independent when his father died. Soon thereafter, he decided to study botany under Julius von Sachs (*q.v.*) at Würzburg, where, in 1882, he received his doctorate. He taught at University College, London (1882), and the Royal College of Science (1885); in 1892 he became honorary keeper of the Jodrell Laboratory at Kew. He was elected a fellow of the Royal Society in 1894. He left Kew in 1906 and settled at Oakley, where he died on Jan. 29, 1934.

Scott's early work was in plant anatomy, but his interests turned to paleobotany when he wrote, with William Crawford Williamson (*q.v.*), three memoirs on the organization of the fossil plants of the Carboniferous rocks (1895–96). In 1897 he described the complex cone of *Cheirostrobus*. His paper on *Lepidocarpon* (1901) showed that some Paleozoic lycopods bore "nascent" seeds, and the paper on *Botrychioxylon* (1912) revealed the presence of secondary wood in a Paleozoic fern. In 1904 Scott and F. W. Oliver announced their discovery that the seed *Lagenostoma* belonged to the fernlike *Lyginodendron*, thus bringing to light a new class of seed plants for which the name Pteridospermae was proposed. He wrote *Studies in Fossil Botany* (1900; 3rd ed., 1920–23) and *Extinct Plants and Problems of Evolution* (1924).

See notices in *Ann. of Bot.*, vol. xlix, pp. 823–840 (1935), containing a list of Scott's 154 publications, and in *Nature*, vol. cxxxiii, pp. 317–319 (1934), vol. clxxiv, pp. 992–993 (1954). (J. W. Tr.)

SCOTT, DUNCAN CAMPBELL (1862–1947), Canadian poet and man of letters, a member of the so-called "Confederation group" (*see* CANADIAN LITERATURE [ENGLISH]), was born in Ottawa on Aug. 2, 1862. He was the son of a Methodist minister, a Lincolnshire man who had emigrated to the U.S. in 1834 and to Canada in 1837. Scott was educated at various secondary schools in small communities in Ontario and Quebec and for two years beginning in 1877 at the newly founded Wesley Academy at Stanstead, Que. He then entered the civil service in the Department of Indian Affairs. From 1913 until his retirement in 1932 he was administrative head of the department. He did not begin to write poetry until about the age of 25 when he came under the influence of the Ottawa poet Archibald Lampman, with whom and with Wilfred Campbell he collaborated in writing a weekly literary causerie *At the Mermaid Inn*, which ran in the *Toronto Globe* from February 1892 to July 1893. On Lampman's death in 1899 Scott edited his friend's poems and wrote a memoir of him. Between 1893 and 1947 he published numerous volumes of verse: *The Magic House* (1893); *Labour and the Angel* (1898); *New World Lyrics and Ballads* (1905); *Via Borealis* (1906); *Lundy's Lane and Other Poems* (1916); *The Green Cloister* (1935); *The Circle of Affection* (1947). Among his prose writings are two volumes of short stories, *In the Village of Viger* (1896) and *The Witching of Elspie* (1923); two biographies, *John Graves Simcoe* (1905) and *Walter J. Phillips, R.C.A.* (1947); a one-act play and literary essays. Many of his poems and stories deal with Indian subjects and with the life of the habitant, but his love poetry and his nature poetry combine intensity and fire with accurate observation and a classical precision of form. He died in Ottawa on Dec. 19, 1947.

BIBLIOGRAPHY.—See *Memoir* by E. K. Brown in *Selected Poems of Duncan Campbell Scott* (1951); E. K. Brown, *On Canadian Poetry* (1945); D. Pacey, *Ten Canadian Poets* (1958); W. P. Percival (ed.), *Leading Canadian Poets* (1948); C. F. Klinck and R. E. Watters (eds.), *Canadian Anthology* (1955); A. J. M. Smith, "Duncan Campbell Scott," R. L. McDougall, *Our Living Tradition* (1959). (A. J. M. S.)

SCOTT, SIR GEORGE GILBERT (1811–1878), English architect of the Gothic revival and eminent Victorian, was born on July 13, 1811, at Gawcott, Buckinghamshire. He was apprenticed to a London architect and designed the first of his many churches in 1838, but his real artistic education dates from his study of A. W. N. Pugin's works on medieval architecture. The first result of this new study was the Martyrs' Memorial at Oxford of 1841.

Scott's international success in the competition for the new Nikolaikirche in Hamburg in 1844, with a design in 14th-century German Gothic, launched a career, the eminence of which now seems somewhat disproportionate to his architectural talents. Thereafter, however, he was one of the most prolific as well as the most successful exponents of the Gothic style throughout the High Victorian period. Among his best-known and most original works are the Albert Memorial (begun 1862–63) and St. Pancras Station Hotel (designed 1865), both in London. The restoration of long-neglected medieval cathedrals and abbeys, which was one aspect of the Gothic revival, was a controversial issue even in the 19th century, and Scott's treatment of such famous monuments as Ely, Salisbury, and Lichfield cathedrals, as well as Westminster Abbey, has been regarded with mixed feelings by subsequent generations. His own account of this work can be read in his *Personal and Professional Recollections*, published in 1879. Scott was knighted in 1872. He died in London on March 27, 1878.

See K. Clark, *The Gothic Revival*, 2nd ed. (1950).

SCOTT, SIR GILES GILBERT (1880–1960), English architect, designer of numerous buildings in the "traditional" style of the 20th century, was born in London on Nov. 9, 1880. Like his grandfather, Sir George Gilbert Scott (*q.v.*), he was a church builder, his greatest individual commission being for the new Anglican cathedral in Liverpool. The building of this massive Gothic structure in red sandstone, begun in 1904, spanned Scott's entire working life. He was knighted after the consecration ceremony in 1924.

Among his many secular works were the new Cambridge University Library (designed 1933), the new Bodleian Library, Oxford, and the new Waterloo Bridge, London. In 1944 he received the Order of Merit and in the same year was chosen as architect of the new House of Commons chamber, a simplified version of the one destroyed in World War II; the building was completed in 1950. Scott died in London on Feb. 8, 1960.

SCOTT, JAMES BROWN (1866–1943), U.S. jurist, publicist, and educator, one of the leaders in the international peace movement of the 20th century, was born of American parents on June 3, 1866, at Kincardine, Ont. Graduating from Harvard in 1890, he subsequently studied at Berlin, Paris, and Heidelberg. He devoted his life to expounding and advocating the application of law and justice between nations through arbitration and judicial settlement. Scott is remembered for the important part he played in establishing the Permanent Court of International Justice (1921) and the Academy of International Law (1914) at The Hague. He was one of the founders of the American Institute of International Law (1912) and of the American Society of International Law and its *Journal* (1906). This publication, of which he was editor in chief, was the first English-language periodical of its kind. He was author, among other works, of *Cases on International Law* (2nd ed. 1922), *The Spanish Origin of International Law* (1933), and *Law, the State and the International Community* (1939).

Dean of the Los Angeles (University of Southern California) and University of Illinois law schools and professor at Columbia and Georgetown (1896–1940), Scott was also editor of several casebooks, as well as of the *American Casebook* series, which was begun at his suggestion in 1908. During 1906–10 he acted as solicitor and technical adviser to the U.S. Department of State. He was delegate to numerous international conferences, including the 1907 Hague peace conference, where he urged the establishment of an international court of arbitration. Scott also served as secretary (1910–40) of the Carnegie Endowment for International Peace and as director of its division of international law (1911–40). He died on June 25, 1943, at Annapolis, Md.

BIBLIOGRAPHY.—George A. Finch, "James Brown Scott, 1866–1943," *American Journal of International Law*, 38:183–217 (1944); Carnegie Endowment for International Peace, *Year Books* (1912–40) and *Summary of Organization and Work, 1911–1941* (1941). (E. H. Fi.)

SCOTT, ROBERT FALCON (1868–1912), British sailor and explorer, leader of the second expedition that reached the

South Pole, where he arrived on Jan. 17, 1912, only to find that Roald Amundsen had preceded him by little more than a month. Scott was born near Devonport, Eng., on June 6, 1868. Educated at Stoke Damerel and Stubbington House, Fareham, he became a naval cadet on HMS "Britannia" in 1881 and in 1883 a midshipman on the "Boadicea." Appointed full lieutenant by 1891, he was chosen in 1900 to lead the National Antarctic Expedition, taking up his duties in the "Discovery" with the rank of commander. Scott proved both an able and resourceful leader and a competent scientific investigator; on his return in 1904 he was promoted captain and the following year he was awarded the commander of the Royal Victorian Order. He continued in the navy, commanding successively the "Victorious," "Essex," and "Bulwark," until in 1909 he announced his intention of organizing another Antarctic expedition, for the purpose of continuing the work of the "Discovery" and of reaching the South Pole. Backed financially by the British and dominion governments, the expedition sailed in June 1910 in the "Terra Nova," and in November 1911 Scott, with four companions, began his southern sledge journey.

Delayed by bad weather, they reached the pole, where they found that they had been beaten by Amundsen. Sickness, insufficiency of food, and severe weather made traveling very slow on the return journey, and on Feb. 17, Petty Officer Edgar Evans died under the strain. A month later Capt. L. E. G. Oates, who was too ill to travel farther, walked out into a blizzard, hoping, by his sacrifice, to save his companions. But the weather prevented any possibility of advancing much farther, and four days later the party pitched the final camp. On March 29 Scott made this last entry in his diary:

> Everyday we have been ready to start for our depot *11 miles* away but outside the door of the tent it remains a scene of whirling drift. . . We shall stick it out to the end, but we are getting weaker, of course, and the end cannot be far. It seems a pity but I do not think I can write any more.

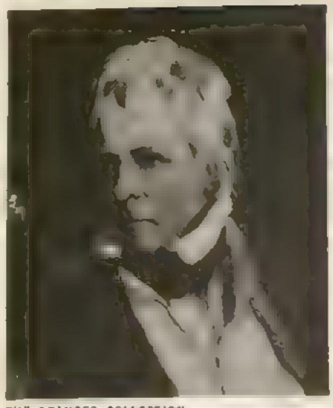
On Nov. 12 a search party found Scott's tent, containing the bodies of Scott, E. A. Wilson, and Lieut. H. R. Bowers, valuable geological specimens from the Beardmore glacier, and Scott's records and diaries, which gave a full account of the journey. A Mansion House fund was raised to provide for surviving relatives, and a polar research institute was founded (1920) in Cambridge to commemorate Scott's name and work. The rank and precedence of the wife of a knight commander of the Bath were conferred on Captain Scott's widow.

See also ANTARCTICA: *Exploration and Discovery: Quest for the South Pole.*

BIBLIOGRAPHY.—R. F. Scott, *The Voyage of the "Discovery"* (1905), and *Scott's Last Expedition,* his journals, arranged by Leonard Huxley (1913, 1957); Lord Mountevans, *South with Scott* (1921, rev. ed. 1948); *British Antarctic Expedition (Terra Nova) 1910–13: Scientific Results* (1914–64); S. Gwynn, *Captain Scott* (1929). (H. G. Ko.)

SCOTT, SIR WALTER, BART. (1771–1832), Scottish novelist, poet, and man of letters, the founder of the historical novel, and the first British novelist to become a great public figure, was born at Edinburgh, on Aug. 15, 1771, the son of Sir Walter Scott, a solicitor and writer to the signet, whose portrait is drawn by his son as Saunders Fairford in *Redgauntlet;* and of Anne Rutherford, the daughter of a distinguished professor of medicine at Edinburgh University. Scott liked to think of his remoter ancestry as including picturesque Border adventurers as well as feudal dignitaries, and he looked back with pride to Walter Scott of Harden, "Auld Watt," a 16th-century Border character who figured in ballad and story and whose grandson, Walter Scott, first laird of Raeburn, was the father of a Walter Scott ("Beardie") who was well known in Teviotdale and was Scott's greatgrandfather. In his memoir of his early life, written in 1808 (and included in J. G. Lockhart's *Life;* see *Bibliography*), he emphasized the colourful aspects of his Border ancestors, about whom as a child he had heard stories from his paternal grandmother and his uncle, Thomas Scott. Thus from early childhood his imagination was nourished by stories of an exciting past which conflicted with the staid Calvinist Whiggery of his father.

Early Life.—An attack of poliomyelitis at the age of about 18 months left him lame, and, after vainly seeking a cure from a variety of physicians, his parents sent him to stay for some time



THE GRANGER COLLECTION

SIR WALTER SCOTT. DETAIL FROM AN OIL PORTRAIT BY SIR EDWIN H. LANDSEER, 1824

with his paternal grandfather at the Border farmhouse of Sandyknowe. In his fourth year he spent about a year at Bath, again in search of a cure, before returning first to Edinburgh and then to Sandyknowe, where he divided his time between country expeditions (for he was by then healthy and vigorous in spite of his lameness) and avid reading of whatever poetry, history, and drama he could lay hands on. He had a remarkably retentive memory, and astonished visitors by his eager reciting of reams of poetry.

In 1778 Scott entered the high school at Edinburgh, and three years later, under the headmaster, Alexander Adam, attained fluency in reading Latin. He left in the spring of 1783 and spent six months with his aunt Janet at Kelso, where he further developed his love of the Border country and where attendance at the local grammar school improved his Latin. There, too, he read avidly "ten times the usual quantity of fairy tales, eastern stories, romances, &c" (*Memoir*) and discovered and delightedly read Bishop Thomas Percy's *Reliques of Ancient English Poetry.* His explorations of the neighbouring countryside developed that "love of natural beauty, more especially when combined with ancient ruins, or remains of our fathers' piety or splendour," which was to reflect itself in his writing.

In the autumn of 1783 Scott matriculated at the Old College of Edinburgh, and attended the usual Latin and Greek classes, though he never mastered the latter language. In May 1786 he began a five-year legal apprenticeship with his father but continued his miscellaneous reading and his country rambles. In 1788 he began to attend law classes at the university as well as classes in history and moral philosophy, and, having served his solicitor's apprenticeship, he decided to enter the higher branch of the legal profession and become an advocate. Between 1788 and 1792 he made a variety of friends among his fellow students (notably William Clerk; Adam Ferguson, son of the famous professor of moral philosophy; and John Irving). It was at Ferguson's house that Scott had met Robert Burns on the latter's visit to Edinburgh in the winter of 1786–87. He worked hard at his law studies but also led a sociable life, attending literary and debating societies and walking and dining with his friends. In July 1792 he qualified as an advocate.

It was during his student days that Scott's inherent love of conviviality, which was an important aspect of his character, first manifested itself. All his life he wanted to have his place in a hierarchy of social life in which equals as well as superiors and dependents played their recognized and accustomed parts. This wish was bound up with his sense of community, which, in turn, was bound up with his sense of history. He never practised extensively at the bar, but he learned a great deal about human nature and also about traits peculiar to the Scots character from his connection with the law.

The outbreak of the French Revolution appalled Scott by its wholesale scrapping of tradition and ruthless break with the past, and when war with France developed he hailed enthusiastically the enrollment of volunteers. In 1797 he eagerly joined the new cavalry corps, the Royal Edinburgh Volunteer Light Dragoons. He continued to indulge his taste for rural excursions and explored further into the Border country than he had yet done, making a notable visit to Liddesdale in the autumn of 1792 (a visit repeated annually for the next seven years) on which he got to know the country people and collected ballads. About the same time he was influenced by the new German Romanticism (*q.v.; see* also GERMAN LITERATURE), which had been introduced to Edinburgh by Henry Mackenzie (*q.v.*). The English "Gothic" novel also cast its spell. Scott's first published work, *The Chase, and William and Helen: Two ballads from the German of Gottfried*

Augustus Bürger, which appeared anonymously in 1796, shows his interest in the ballad temporarily turned toward the fashionable German vogue, as does his poor translation of Goethe's romantic tragedy, *Götz von Berlichingen* (1799).

Marriage.—By this time Scott had been through the one passionate love affair of his life, his unsuccessful courtship of Williamina Belsches, who was of a higher social class than that in which Scott's family moved. The affair seems to have begun in the early 1790s, and after many ups and downs it was ended in October 1796 by Miss Belsches' engagement to a rich young landowner, William Forbes. There is no doubt that this failure in love upset Scott deeply and permanently haunted his imagination. But his temperamental resilience, his active life, and his literary interests enabled him to avoid excessive brooding on his failure, and the publication of his two German ballad translations was soon followed by his ardent activities with the volunteers. He continued his eager reading of a great variety of literature and history throughout 1797, and in September made one of his characteristic expeditions through Tweeddale and over the Border to end up in the little watering place of Gilsland. There he met Charlotte Carpenter (Margaret Charlotte Charpentier), daughter of a former head of the military academy at Lyons, and of his much younger and lively wife. Mme Charpentier had run off to England with a young Welshman, had eventually come under the "protection" of Lord Downshire, and had returned to France when Lord Downshire married in 1786. Charlotte (b. December 1770) was educated in a French convent until 1787, when she went to England. She was baptized into the Church of England in May 1787 and apparently lived thenceforth with the family of Charles Dumergue, a successful French dentist in London, under the general supervision of Lord Downshire. When she met Scott she was visiting Gilsland with her companion, Jane Nicolson. She and Scott soon became engaged and were married in Carlisle on Dec. 24, 1797. It was on both sides a marriage of affection rather than passion; Charlotte had a gay and lively disposition but shared none of Scott's deeper interests. If, as Scott himself later admitted, their mutual feeling "fell short of love in all its forms," this was really what he preferred; "folk who have been nearly drowned in bathing rarely venturing a second time out of their depth," as he put it.

Literary Beginnings.—Scott and his bride set up house in Edinburgh. He was making a little money at the bar, his wife had brought him a few hundred pounds, and he had an allowance from his father. In 1798 they took a summer cottage at Lasswade, on the River Esk some six miles from Edinburgh. In April 1799 Scott's father died, while Scott and his wife were on a visit to London. Scott's meeting with M. G. ("Monk") Lewis (*q.v.*) in Edinburgh, 1798–99, and in London in the spring of 1799, encouraged another splash in the Gothic wildness of German Romanticism, which further postponed the emergence of his true literary interest and talents.

These at last manifested themselves with his edition of *The Minstrelsy of the Scottish Border* (3 vol., 1802–03). His view that the texts available to him represented oral corruptions of the original compositions led him to try to restore many of them to what he considered something closer to their original form; his methods of conflating, emending, "regularizing," patching, or simply adding to his texts are not today approved. Sometimes his efforts resulted in powerful and impressive poems, showing a sophisticated savouring of romantic detail which immediately betrays the hand of the "improving" editor. The last paragraph of his long introduction provides an important clue to the links in Scott's mind between patriotism, history, and literature: "By such efforts, feeble as they are, I may contribute somewhat to the history of my native country; the peculiar features of whose manners and character are daily melting and dissolving into those of her sister and ally. And, trivial as may appear such an offering, to the manes of a kingdom, once proud and independent, I hang it upon her altar with a mixture of feelings, which I shall not attempt to describe."

Scott had by now a wide circle of friends, including George Ellis, Richard Heber, that extraordinary self-taught prodigy John Leyden, James Hogg (*q.v.*), William Laidlaw, another ballad enthusiast, Lord Dalkeith, Lady Anne Hamilton, and many more. In 1804 he brought out an edition of the medieval romance *Sir Tristrem*, to which he added a "fytte" (or book) of his own composition. At last he found that he could not resist the temptation to turn minstrel in his own right, and what had begun as a ballad on the subject of a goblin page suggested to him by Lady Dalkeith grew into a full-length narrative poem, *The Lay of the Last Minstrel* (1805).

His appointment as sheriff-depute of the county of Selkirk in 1799 added £300 a year to his slender earnings at the bar and, with his father's estate, brought his income to respectable proportions. The duties of his sheriffship were light. The lord lieutenant of the county of Selkirk, who considered that he was spending too much time with the volunteers in Edinburgh, invoked the long-neglected rule that a sheriff should reside for at least four months of each year within his own jurisdiction, with the result that in 1804 Scott rented the pleasant house of Ashiestiel, on the south bank of the Tweed a few miles from Selkirk. In the same year his financial situation improved with the death of his uncle Robert Scott, who left him some property which he sold for £5,000. When the Scotts moved from Lasswade to Ashiestiel in the late summer of 1804 they had three children, two daughters (the elder of whom was to marry J. G. Lockhart; *q.v.*) and a son, Walter (b. 1801); a second son (b. December 1805) completed the family.

The Poems.—*The Lay of the Last Minstrel* was an immediate success and ran into many editions. It was followed by a spate of other narrative poems—*Marmion* (1808), *The Lady of the Lake* (1810; the most popular), *Rokeby* (1813), *The Lord of the Isles* (1815), and others. *The Lay* shows the influence in its rhythms of Coleridge's *Christabel*, but before long it settles down into the trotting octosyllabic couplets (varied occasionally with interspersed shorter lines rhyming with each other in ballad-metre style) which represent the norm of Scott's narrative verse. Scott wrote fast and without much artistic conscience; so long as he could keep the rhymes and the verse movement going he was content. He is at his best at describing settings (as in the opening of *The Lady of the Lake*) and sometimes in presenting fierce and rapid action; in love scenes and with sentiment generally he is at his worst. The description of the Battle of Flodden in *Marmion* has a fine heroic vigour and finely conveys a sense of doom. But sooner or later patches of mechanical padding are inserted to carry the verse on at all costs. The narrative is as a rule studiedly objective, but in *Marmion* Scott introduced preliminary passages of personal reminiscence. Scott also had a strangely melancholy lyric strain, which he learned from folk literature, found in occasional incidental lyrics introduced into the narrative poems as well as scattered throughout the novels.

Abbotsford and Publishing.—The success of *The Lay* meant that henceforth literature was to play an increasingly important part in Scott's life. His literary, legal, and social activities became increasingly lively; he was at the centre of a brilliant Edinburgh society. His country home, Ashiestiel, was only rented, and in the spring of 1812 he moved to a cottage on the small estate of Abbotsford (*q.v.*), near Melrose, and began the buying of land and the building of a country house there which was to absorb so much of his money. In 1806 he was appointed clerk of session, which meant his being in Edinburgh when the Court of Session sat. The arrangement was that the salary of £1,300 should go to the retiring clerk during his lifetime, but Scott would draw it on the former's death (Scott began drawing the salary in 1811). The expense of Abbotsford and other commitments necessitated his making ever more money. In 1802 Scott had given a loan to the Kelso printer James Ballantyne to enable him to remove his printing press to Edinburgh. In 1805 he became a sleeping partner in Ballantyne's firm, whose fortunes he at once began to push, trying to make it a condition with publishers of his own work that Ballantyne should do the printing. His own work by now included not only his poems but an increasing amount of miscellaneous and editorial writing: in 1808 his 18-volume edition of Dryden appeared, followed by an edition of Captain Carleton's *Memoirs* (1808) and of the Somers Tracts (1809–15), a 19-volume

edition of Swift (1814), and many other works. Publishing was soon added to printing, and James Ballantyne's irresponsible younger brother, John, was taken in as a partner. The firm started by publishing *The Lady of the Lake*, but Scott's enthusiasm for antiquarian, historical, and miscellaneous publications at this time led to the issuing of works such as *The Edinburgh Annual Register*, John Jamieson's *History of the Culdees*, H. W. Weber's edition of *Beaumont and Fletcher*, *The Poetical Works of Anna Seward*, among many others, most of which sold badly. From about 1810 almost everything that Scott wrote was in order to make money to meet financial obligations already incurred.

By 1813 the Ballantyne firm was moving rapidly to financial disaster. Faced with bankruptcy, Scott approached the publisher Archibald Constable (*q.v.*), who saved the situation by taking over £2,000 worth of Ballantyne's stock, on the condition that the publishing firm be wound up. A few months later an overdraft of £4,000 from the duke of Buccleuch again averted crisis, and by 1817 Scott was able to pull the printing business out of its indebtedness. Throughout this period Scott was leading a highly active social as well as literary life; he was still paying for Abbotsford as well as for enormous quantities of antiquarian relics for his collection there; and he was giving financial help to his brother Tom. His annual income was now £2,800 apart from the considerable profits of his writing. But he continued to spend more than he earned. The enormous success of *The Lady of the Lake* emboldened him to ask for and receive from Constable extravagant terms for *The Lord of the Isles*.

The Novels.—But already he was working on *Waverley*, originally projected in 1805 but left with only the opening chapters written. The greater verve and dash of Byron's narrative poems threatened to oust Scott from his position as supreme purveyor of this kind of literary entertainment, and he was looking for another vein. In the autumn of 1813 he accidentally came across the unfinished manuscript of *Waverley*, and completed it. It was published anonymously in July 1814 by Constable, with John Ballantyne & Co. doing the printing; Constable got half the profits, Scott the other half. Scott preserved his anonymity because he felt that it was not quite becoming in a clerk of session to write novels, then considered a rather frivolous form of literature. Leaving Edinburgh to buzz with excitement and speculation about *Waverley*, Scott set off at the end of July on a visit to the northern and western islands in the yacht of the lighthouse commissioner (R. L. Stevenson's grandfather). This six-week trip provided him with Scottish scenes he was later to use in novels. In April 1814 Constable asked Scott for an article on "Chivalry" for *Encyclopædia Britannica*, which he had just bought. "Chivalry" finally appeared in 1818, "The Drama" in 1819, and "Romance" in 1824.

Waverley was followed by *Guy Mannering* (1815), "the work of six weeks at Christmas"; *The Antiquary* (1816); two series of *Tales of My Landlord* (1816–19), which included *Old Mortality* and *The Heart of Midlothian*; *Rob Roy* (1818); a third series of *Tales of My Landlord* included *The Bride of Lammermoor* and *The Legend of Montrose* (1819). It was only after writing these novels of Scottish history that Scott, driven by financial need and the need to satisfy the public appetite for historical fiction that he had created, turned to themes from English history with *Ivanhoe*, *The Monastery*, and *The Abbot* (all 1820). *The Pirate*, set in the Orkneys, and *The Fortunes of Nigel* were published in 1822. Later novels included *Quentin Durward* (1823), *St. Ronan's Well* (1824), and, also in 1824, the last and one of the best of his Scottish novels, *Redgauntlet*. As a historical novelist dealing with medieval England, or France, or Germany, or the crusaders' Palestine, Scott showed a flair for highly coloured picturesque incident and situation. He also exhibited in varying degrees a sense of the poetry of history and an ability to project in terms of character and action something of the life and manners of the feudal ages. But he becomes a more complex and mature novelist the nearer he comes to his own time and country. *Ivanhoe* and *The Talisman* (the latter in *Tales of the Crusaders*, 1825) are colourful, somewhat theatrical, novels of rather obviously stylized period characters and action. *Quentin Durward*, *Kenilworth* (1821), and *The Fortunes of Nigel*, set respectively in the 15th, 16th, and 17th centuries, show a greater awareness of the complexities of the human situation and less disposition to be content with colourful surface action arising from the histrionic attitudes of somewhat cardboard figures who speak an invented antique dialect. Scott was nearly always better with minor and often low-class figures than with conventional heroes and heroines who had to act out the love story which the reader demanded but which never engaged Scott's real interest. He wrote too fast, and he wrote simply to entertain, and his novels often show a lack of artistic conscience. In dealing with picturesque aspects of the distant past, as well as on the one occasion, in *St. Ronan's Well*, when he dealt with the surface of life in his own day, Scott's imagination worked perfunctorily and did not draw on its deepest sources of inspiration. His real claim to greatness as a novelist rests on the "Scotch novels."

In dealing with the recent past of his own country Scott was able to find a fictional form to express the deep ambiguities of his own feeling about Scotland. One side of him welcomed the union with England and the commercial progress and modernization which it promised to bring; the other regretted bitterly the loss of Scotland's independence and the steady decline of its national consciousness and traditions. In novel after novel—most specifically in *Waverley* and *Rob Roy*—he introduces a polished and elegant representative of 18th-century British civilization, brings him for a short period into sympathetic contact with a representative of the older, heroic tradition, and then makes it clear that this representative of the heroic tradition is at best without function and at worst totally degenerate, so that the hero retires at the end to the "refined" modern society from which he came, enchanted with his adventure yet sobered into a realization that Highland chiefs are after all little more than freebooters in the modern world, where their tradition can have no place. In *The Heart of Midlothian*, one of the greatest of his novels, Scott looked at his country a generation after the Treaty of Union of 1707 and found characters and incidents which provided an "objective correlative" for his emotions about it. In *Redgauntlet* he showed with moving particularization how in late 18th-century Scotland (the period of his own childhood) the heroic Jacobite emotions of an earlier generation could find no realistic outlet in action. In *Old Mortality* he went to the 17th century to show the bitter meaning of the last full-scale flare-up of a really fighting Scotland. Scott's true heroes are not knights-at-arms but lawyers, farmers, merchants, and simple people who go about their business oblivious to the claims of the past.

Scott became a novelist by bringing his antiquarian and romantic feeling for Scotland's past and for a heroic age into relation with his sense that Scotland's interest demanded that she break with the past and look forward to a prudently commercial "British" future. Civilization must be paid for by the cessation of the old kind of individual heroic action: Scott welcomed civilization, but he also sighed after individual heroic action. This ambivalence of feeling gives life to his best novels. It also explains why Scott's finest dialogue is in the Scots speech he put into the mouths of his humbler characters and not in the stilted diction of his nominal heroes and heroines. The heroic characters in the end become histrionic or even ludicrous, and the humble or the prudent remain: Rob Roy is dismissed as a ranting exhibitionist and Bailie Nicol Jarvie, the canny merchant, survives into the future. And in *The Heart of Midlothian* the obviously heroic actions turn out to be merely criminal and sordid, and true heroism is found in the behaviour of Jeanie Deans, a simple Scottish lass. Scott's attempt to set himself up as the laird of Abbotsford, maintaining an almost feudal position with the proceeds of his finally disastrous publishing speculations and adorning a modern country house with an extraordinary collection of historical and pseudohistorical objects, symbolizes the ambiguity of his attitude.

Literary Success and Financial Failure.—It was not until 1827 that Scott formally revealed his authorship of the Waverley novels, but most people had guessed the truth. When he visited London in 1815, he was regarded as the greatest living Scotsman. Later that year he visited Flanders and the field of Waterloo, pub-

lishing his account of the trip as *Paul's Letters to His Kinsfolk* (1816). His arrangements with his publishers were becoming more involved; he was always in need of money, always arranging for money for books not yet written. Ballantyne and Co. still had stock they wanted taken off their hands, and Scott was using this in negotiations. He negotiated with William Blackwood and his London associate, John Murray (*qq.v.*), as well as with Constable, with the Ballantynes playing a confusing middleman's part. The first series of *Tales of My Landlord* was published by Murray, but *Rob Roy* and subsequent novels went to Constable. Both Scott and the Ballantynes seem to have played one publisher against the other. From 1818 Constable was publisher of the novels and the Ballantynes were Scott's agents. Scott, living expensively, and continually expanding his Abbotsford estate, was always asking Constable for payment in advance; he also raised loans from Ballantynes' printing business. Every new expense—travel, his daughter's dowry, the purchase of his son's commission—was met by bills discounted on work still to be done. Scott never guessed that all this time Constable and Co. was itself on the verge of bankruptcy.

In the years 1817 to 1819 Scott had frequent attacks of painful stomach cramp, but he went on doggedly with his writing. In February 1818 he was involved in the ceremonious discovery of the long-concealed regalia of Scotland. He received a baronetcy in March 1820, and in April his daughter Sophia married J. G. Lockhart in Edinburgh. In July 1821 he went to London for George IV's coronation, and in the summer of 1822 he was active in stage-managing the king's visit to Scotland. Just as Scott's Scottish poems and novels had made the Highlands popular haunts for tourists, so the king's appearance in a kilt led to a wave of kilt wearing and tartan enthusiasm. Scotland as it appears to the tourist and the foreigner is largely Scott's creation. In 1824 Abbotsford was finally completed and lavishly furnished, and in the Christmas and New Year season of 1824–25 Scott threw a great housewarming party. It was a happy season, and the last of its kind for Scott. In December 1825 he heard of the failure of Hurst and Robinson, Constable's London agents, and the subsequent chain reaction involved bankruptcy for both Constable and Ballantyne in January 1826. Scott's intimate involvement in both firms, so long his secret, now became public property; he was personally responsible for the Ballantyne liability. He faced the position with courage, surrendering to a trust his property and the profits of work both done and contracted for, and determined to work off the vast debt with his pen. He was allowed to keep the house and home farm at Abbotsford and his official salary of £1,600 a year. His wife died in May 1826.

Last Years.—Scott's chief nonfictional works in this last phase of his life were *The Life of Napoleon* (1827) and *Tales of a Grandfather* (1828), stories from Scottish history. His novels he kept on doggedly writing almost until the end with inevitably deteriorating quality. He resigned his clerkship of session in 1830 and was given a retirement allowance of £800 annually. The trustees allowed him to keep for himself the profits from *Tales of a Grandfather* and other nonfictional works, and he was far from living a life of poverty. But his health was breaking. His rash engagement in politics in Jedburgh on the antireform side in March 1831 was followed by a paralytic stroke in April; it proved temporary, like an earlier and less severe one in February 1830, but it was the beginning of the end. The government put a frigate at his disposal for a Mediterranean cruise in search of health, but the voyage—from October 1831 to June 1832—was marked by further strokes, and he was brought home to die. The sight of his beloved Border country briefly revived his awareness, and he died at Abbotsford, after a long period of semiconsciousness, on Sept. 21, 1832. He had no dying words for posterity, his alleged last words to Lockhart—"Be a good man"—being, as H. J. C. Grierson has shown, a deliberately pious fiction suggested to Lockhart by a lady correspondent. He was succeeded in the baronetcy by his son, Walter, on whose death (1847) the title became extinct.

Of the progress of Scott's tangled financial affairs, all that need be said is that the publisher Robert Cadell, after a variety of maneuvers, took over both Scott's liabilities and Scott's copyrights and made a fortune as a result. "I have paid Sir Walter Scott's debt and the debt of the estate of Abbotsford," Cadell wrote in 1848, "in a word a trifle of £7,600 and what is more surprising . . . the demand for his work continues." It continued until well into the 20th century, when critics of the novel began to demand of a novelist a greater awareness of the demands of fictional art than Scott had ever had the time to develop.

Scott invented the historical novel, and he invented the tourist's view of Scotland. Later readers and critics fastened on the high romantic aspect of his view of the past and ignored that disturbing double vision that resulted from the tension between his love of "the crowded hour of glorious life" and his shrewd awareness of its impossibility and even undesirability in the modern world. It is that double vision that modern criticism sees as the source of his greatest strength as a novelist.

See also ROMANTICISM; NOVEL; and references under "Scott, Sir Walter" in the Index.

BIBLIOGRAPHY.—*Bibliography and Editions:* G. Worthington, *Bibliography of the Waverley Novels* (1930); W. A. Ruff, *Bibliography of the Poetical Works* (in *Transactions of the Edinburgh Bibliographical Society*, I, 1937–38); J. C. Corson, *Bibliography of Sir Walter Scott, 1797–1940* (1943), a bibliography of books and articles about Scott and his writings. There are many collected editions of Scott's poems and of his novels. Of the former, the principal are the *Poetical Works* (1820), the *Poetical Works*, ed. by J. G. Lockhart (1833–34), both in 12 vol., and the *Poetical Works*, ed. by J. L. Robinson (1904). Editions of the Waverley novels include the Border Edition, 48 vol. (1892 ff.), the Centenary Edition, 48 vol. (1898); and the Edinburgh Edition, 48 vol. (1901–03). The novels and tales, with prefaces and notes by Scott, appeared in 48 vol. (1829–34). *Miscellaneous Prose Works* appeared in 30 vol. (1834–1871). *The Minstrelsy of the Scottish Border* was ed. by T. F. Henderson, 4 vol. (1932).

Correspondence and Biography: The standard edition of Scott's letters is that by H. J. C. Grierson *et al.*, 12 vol. (1932–37). J. G. Tait and W. M. Parker's edition of *The Journal of Sir Walter Scott* (1950) provides a better text than that of David Douglas (2 vol. 1890), although Douglas' annotations remain valuable. J. G. Lockhart, *Memoirs of the Life of Sir Walter Scott, Bart.*, 7 vol. (1837–38), a finely rendered picture of Scott and his environment, is not altogether trustworthy. It should be corrected by H. J. C. Grierson, *Sir Walter Scott, Bart.: a New Life Supplementary to, and Corrective of, Lockhart's Biography* (1938); John Buchan, *Sir Walter Scott* (1932) is derived from Lockhart but contains interesting critical perceptions. Hesketh Pearson's *Sir Walter Scott* (1954) is lively, and includes the findings of modern scholars, but is not an outstanding original contribution.

Criticism: Among critical works on Scott are: E. Muir, *Scott and Scotland* (1936); J. T. Hillhouse, *The Waverley Novels and Their Critics* (1936); W. L. Renwick (ed.), *Sir Walter Scott Lectures, 1940–48* (1950); D. Daiches, "Scott's Achievement as a Novelist," in *Literary Essays* (1956); D. Davie, *The Heyday of Sir Walter Scott* (1961); A. Welsh, *The Hero of the Waverley Novels* (1963). (D. Ds.)

SCOTT, WINFIELD (1786–1866), U.S. Army officer who served as a general in three wars and was the Whig candidate for president in 1852. Born near Petersburg, Va., on June 13, 1786, he entered the College of William and Mary in 1805. He was commissioned a captain of artillery in 1808 and fought on the Niagara frontier when the War of 1812 broke out. Although captured during this campaign, he was exchanged in January 1813 and fought through the rest of the war, rising to the rank of major general.

He took part in the battle of Chippewa (July 5, 1814) and was twice wounded in the battle of Lundy's Lane (July 25). After the war Scott remained in the military service, studied tactics in Europe, and eventually rose to become commanding general of the U.S. Army, a position he held from 1841 to 1861. In 1838 he supervised the removal of the Cherokee Indians from Georgia and other southern states to reservations west of the Mississippi River. Because of his emphasis on the military proprieties he was nicknamed "Old Fuss and Feathers," but the nickname should not be allowed to obscure his solid military talents.

During the Mexican War (*q.v.*; 1846–48) Scott commanded a seaborne invasion of Mexico that captured Veracruz late in March 1847. After victories at Cerro Gordo, Contreras, Churubusco, Molino del Rey, and Chapultepec, he entered Mexico City on Sept. 14, thus effectively ending the war. Scott was recalled in January 1848 by Pres. James K. Polk, with whom he had not been on good terms, to face a court of inquiry appointed to look into criticisms made of Scott's conduct of the campaign. The criticisms were later withdrawn and Scott was honoured by appoint-

ment to the brevet rank of lieutenant general.

A prominent Whig, Scott had been mentioned as a presidential candidate as early as 1839. At the Whig convention of 1848 his name was placed in nomination but he was defeated by Gen. Zachary Taylor. Four years later Scott won his party's nomination but lost the election to Franklin Pierce, mainly because the Whigs were divided on the slavery question. Scott remained a popular old soldier, and at the outbreak of the Civil War he commanded the U.S. Army until November 1861 when age forced his retirement. He died at West Point, N.Y., on May 29, 1866.

See also references under "Scott, Winfield" in the Index.

See Charles W. Elliott, *Winfield Scott: the Soldier and the Man* (1937).

SCOTTISH LITERATURE. The term Scottish literature is difficult to define satisfactorily. It must include writing in Gaelic (*see* below); writing in Scots (*i.e.*, Lowland Scots or Lallans); writing in standard English; and writing in every possible degree of English and Scots combined. There is a tendency for the nationality of the Scottish writer to make itself felt, even when he is not using his own language. To this the English prose of Thomas Carlyle is a witness, as is also the Latin verse of George Buchanan; and even where the writer is not of fully Scottish blood it will sometimes be found that Scottish characteristics assert themselves, as they do in Byron. Although, therefore, works in Scots must claim most attention, especially in the period before 1600, works wholly or partly in English will also on occasion require consideration. The work of many Scottish writers forms part of English literature (*q.v.*); *see* also such general articles as BALLAD; SATIRE; etc., and articles on many of the writers mentioned.

IN SCOTS AND ENGLISH

Medieval Period.—Apart from poems of doubtful provenance and authorship (*e.g.*, the 13th-century romance *Sir Tristrem* with which Thomas the Rhymer's name is linked) and a few fragments quoted by later writers (such as the well-known *Quhen Alysandyr our King wes Dede*, included by Andrew of Wyntoun in his *Cronykil* of *c.* 1420), the earliest extant literature belongs to the second half of the 14th century. No doubt much earlier literature has been lost, as neither John Barbour's *Bruce* (completed 1376) nor *The Pistill of Susan* (written *c.* 1360; perhaps Northern English rather than Scottish; ascribed to Huchown of the Awle Ryale, but *see* HUCHOWN) can be called fumbling or primitive in technique. It is clear that these poems represent two distinct traditions, the former being in octosyllabic rhyming couplets, and having affinities with the French *chanson de geste*, and the latter having that alliterative rhyming stanza with concluding "bob" and "wheel" which apparently reached Scotland through the Middle English alliterative writers of the north and northwest and which became a characteristic Scottish poetic form. To the first tradition belong Andrew of Wyntoun's *Orygynale Cronykil* (completed *c.* 1420; publ. 1795), a metrical history of Scotland; and *The Buik of Alexander* and *Legends of the Saints*, which belong to Barbour's time and region if not to Barbour himself. The second tradition developed romance and fantasy rather than history, and includes Richard Holland's *Buke of the Howlat* (written *c.* 1450); the anonymous *Golagros and Gawane* (printed 1508); *Rauf Coilȝear* (written *c.* 1480); and shorter pieces such as *The Gyre-Carling*, William Dunbar's *Kynd Kittok* (if it is by Dunbar), and Robert Henryson's *Sum Practysis of Medecyne*. The rather dry, economical, unadventurous style of John Barbour (*c.* 1325–95), suited though it was to his own patriotic narrative, had little influence on later Scottish poetry; the *Bruce*, national epic in kind if not in value, is an isolated poem. The more flamboyant alliterative tradition, however, which in Scotland was readily employed not only for serious romance narratives but also for delightful "dremys and drevillings" such as *The Buke of the Howlat*, and for the wild ribaldry of *The Gyre-Carling* or *Sum Practysis*, proved more congenial and had a lasting effect well beyond the medieval period. Satire and fantasy, always strong elements in Scottish poetry, were well served by "thir mokking meteris and mad matere." The alliterative style was also valuable in forging a link between popular and sophisticated verse, which in Scotland was not broken to the same extent as it was in England after Langland's "rum ram ruf" gave way to Chaucer's decasyllabics.

The Makaris.—The great period of the makaris, or Scottish Chaucerians as they have been inadequately but inescapably labeled, may be taken to be *c.* 1425–1550. It includes the four chief pre-union poets: Robert Henryson (fl. 1475), William Dunbar (*c.* 1460–*c.* 1520), Gavin Douglas (*c.* 1475–1522), and Sir David Lyndsay (*c.* 1485–*c.* 1555); and to them must be added the author of *The Kingis Quair* (probably James I of Scotland, 1394–1437) and the author of *Schir William Wallace* (traditionally Harry the Minstrel, or Blind Harry, fl. 1470–95). Scottish poetry has never been so confident, so dexterous, or so varied as in the century that produced Henryson's *Testament of Cresseid*, Dunbar's *Tua Mariit Wemen and the Wedo*, Douglas' *Eneados*, and Lyndsay's *Ane Satyre of the Thrie Estaits*. The rough, swarming, precarious culture which nourished this literature has often the appearance of a rejuvenated medievalism rather than of a premature Renaissance. A fable by Henryson closes with its "moralitas"; Dunbar and Douglas continue and develop the conventional and centuries-old dream-allegory; Lyndsay is a reformer who remains within the church. What was not medieval, however—and it is an important element if we are to speak of rejuvenation—was the keen linguistic consciousness, the desire to expand and enrich the vernacular resources by all available means, which followed the political self-determination of Scotland in the 14th century and accompanied the Scottish writer's literary pride in the 15th. The elaborate "aureate" style in which the makaris often indulged has been criticized as artificial and excessive; but such excesses, which have their parallel in Elizabethan England, were a necessary stage in the development of a literary medium. The term "aureate" itself tells only half the story; Dunbar's *Goldyn Targe* cannot be called typical either of him or of his century's poetry. When Gavin Douglas, justifying his borrowings from other tongues in his translation of Virgil, speaks of his desire for "fouth of langage," he is revealing a very Scottish preference. Fouth—copiousness—demanded the encouragement of anti-aureate and popular elements as well as the aureate and the classical; of alliterative as well as Chaucerian tradition. Douglas' powerfully atmospheric descriptions of scenery and weather are almost bursting with words, and the etymological and idiomatic texture is far too rich for the restrictive implications of "aureate diction" to hold any meaning. This is even truer of works like *The Cursing of Schir John Rowll* and *The Flyting of Dunbar and Kennedie*, where commination and *estrif* induce the most uninhibited fouthiness of effect. (For the continuity of this

BY COURTESY OF (LEFT, THE TRUSTEES OF THE NATIONAL LIBRARY OF SCOTLAND, (RIGHT) THE BODLEIAN LIBRARY, OXFORD

TITLE PAGES FROM THE PERIOD OF THE MAKARIS

(Left) Robert Henryson's "The Morall Fabillis of Esope," printed in Edinburgh, 1571; (right) "Ane Satyre of the Thrie Estaits" by Sir David Lyndsay, printed in Edinburgh, 1602

quality, *see* also *The Complaynt of Scotlande; Polemo-Middinia;* Sir Thomas Urquhart *passim;* Hugh MacDiarmid's *Scots Unbound.*)

Chaucer was to the makaris the eagerly acknowledged "rose of rethoris all," and there is no doubt that they gained from him an ideal of poetic utterance, a rhetoric or a diction, and metrical forms like the decasyllabic couplet used in *Schir William Wallace* and the Troilus stanza in rime royal (*q.v.*) used in *The Kingis Quair,* which taken together go far toward confirming the term Scottish Chaucerians. Scottish poetry, with its highly developed uncouthness, was fascinated by Chaucer's subtleties, just as its love of the incongruous almost succumbed to Chaucer's good sense and order. Nevertheless, his was only one influence among many. Both Chaucer and Dunbar were closely connected with the court circles of their day, and they addressed a courtly and educated audience; yet Dunbar refers with relish to farcical anonymous fantasies like *Colkelbie's Sow* which enjoyed widespread popularity beyond the court, and the grave Gavin Douglas shows familiarity with *Rauf Coilzear.* Lyndsay goes further toward meeting "John the Commounweill" (the allegorical representative of the common weal) halfway, and declares "... to colzearis (colliers), cairtaris (carters), and to cukis (cooks); /To Jok and Thome my rhyme sall be direckit." If the two extremely popular poems of Jock and Tom's festival merrymaking, *Christis Kirk on the Green* and *Peblis to the Play,* did in fact originate from James I or James V, as variously surmised, it would be another remarkable instance of this social give-and-take, perhaps not unexpected in a small country that had been only partly and haphazardly feudalized. "A man's a man for a' that" is a recurring statement in Scottish literature, representing a point of rest between outspoken social criticism on the one hand (Henryson, Lyndsay, Burns, MacDiarmid), and all the degenerations of the couthy and the pawky on the other hand (*Whistle Binkie* and Kailyard, for which *see* below). Apart, therefore, from questions of verse technique, we must supplement the Chaucerian influence by two additions: the vehement, practical, propagandist social satire of a Lyndsay, and (perhaps as a counterpoise) the extravagant daftness of a *Lichtounis Dreme,* with its impractical but equally Scottish occupation of "rostand straberries at ane fyre of snaw."

The golden period of the makaris had something of an aftermath in the accomplished but limited love poetry of Alexander Scott (*c.* 1525–*c.* 1585), Alexander Montgomerie (*c.* 1545–*c.* 1611), and William Fowler (1562–1612); the Italian translations and moral reflections of John Stewart of Baldynnis (*c.* 1550–*c.* 1605); the impressive plainness and sincerity of Sir Richard Maitland (1496–1586); and single outstanding lyrics from Alexander Hume (*c.* 1560–1609) and Mark Alexander Boyd (1563–1601). Here may also be mentioned the only surviving vernacular drama apart from Lyndsay's *Satyre*—the anonymous *Philotus* (*c.* 1600), a complicated romantic comedy, coarse in detail but vigorously written and certainly amusing.

The 17th Century.—With the union of the crowns in 1603 and the transference of the court to London, Scottish writing became increasingly anglicized; nor was much of it, in this century, of the first quality. Although James VI himself showed that he was capable of turning out a lyric in Scots, the new poets wrote almost exclusively in English. The best of these, William Drummond (1585–1649), was as "well-languaged" as his English contemporary Samuel Daniel, but his lyrics lack intensity and force, and one would gladly attribute the quiet un-English *Polemo-Middinia* to him, if only for its bold macaronic gusto. The lyrics of Sir Robert Ayton (1570–1638) and James Graham, marquess of Montrose (1612–50), are graceful without being distinguished. The sententious and indefatigable Sir William Alexander (*c.* 1567–1640) accomplished nothing that Fulke Greville and George Chapman had not done better. Vernacular poetry was kept alive in popular songs and ballads, and such literary verse as does guard the thin continuity of tradition between the makaris and Allan Ramsay is equally popular in character, with leanings toward the rustic, the homely, and the humorous (*see,* for example, Robert Sempill's *Life and Death of Habbie Simson, the Piper of Kilbarchan*). Whether grim and tragic as in the older ballads, or amusing as in *Habbie Simson,* this tradition was of the heart rather than the head, and indeed the absence of an intellectual and critical quality has been a besetting weakness in later Scots verse. The divorce of intellect and feeling was a natural result of the loss of Scots as a complete communicative instrument.

At this point it will be convenient to deal with the somewhat unspectacular history of Scottish prose, which covers (in its extant examples) the approximate period 1450–1630. The earliest work, dated 1456, is Sir Gilbert Hay's *Buke of Bataillis, Buke of the Order of Knichthood,* and *Buke of the Governaunce of Princes;* but these are translations. The first original literary prose appears in the theological writing of John Ireland (fl. 1480–90). The prose of these works, written in "the commoune langage of this cuntre," has not yet developed great variety or flexibility of construction. Early in the 16th century there was an attempt at a more plain, less latinized prose in Murdoch Nisbet's version of Wycliffe's New Testament, but the language is too southern to be quite successfully idiomatic. Some English influence is seen also in the 16th-century historical writings of John Bellenden, John Leslie, and Robert Lindesay of Pitscottie; and to a greater extent in works like John Knox's *History of the Reformation in Scotland* (*c.* 1567). These writers, however, had a good command of narrative, and Pitscottie and Knox had also the gift of style, which has made passages like Pitscottie's account of the Linlithgow apparition and Knox's interviews with Mary Queen of Scots justly famous. Almost as vivid as Knox's writing is the autobiographical *Historie of the Lyff of James Melvill,* which covers the second half of the 16th century. Standing by itself is the remarkable anonymous *Complaynt of Scotlande* (1548–49), interesting both as the work of an earnest and well-read Scottish patriot and as the first work of a Scottish experimenter in prose style, ranging from the most aureate "oncouutht exquisite termis" to the most common and onomatopoeic "domestic Scottis langage." The need to reach English readers is seen in the writings of James VI: the manuscript of his *Basilikon Doron* (completed 1598) is in Scots, but when published in 1599, it had been translated into English. William Drummond's *Cypresse Grove* (1623) is in careful, grave, mannered English; and not much Scots prose occurs after that date. A general justification for this turning to English may well be seen in the lucid and forceful 18th-century prose of David Hume, Adam Smith, James Boswell, and Tobias Smollett; though on the debit side must be placed the extreme difficulty of all attempts to revive vernacular prose, in a tradition so drastically severed. One particular justification is Sir Thomas Urquhart (1611–*c.* 1660), "the free'st spoke Scot of any," whose translation of Rabelais (1653) and extraordinary original works display a polymathic ornateness that would have cracked the molds of Scots as it did those of English.

The 18th Century.—While Scottish prose writers were preparing to challenge English on its own terms, and while the union of 1707 was suggesting an even closer cultural binding of the two countries than had been possible a century before, a contrary impulse was making itself felt in poetry. Almost in the year of the union appeared James Watson's first volume of *Choice Collection of Comic and Serious Scots Poems* (1706), which contained (among English pieces) vernacular poetry, from *Christis Kirk on the Green* and *The Cherry and the Slae* to recent verse like *Habbie Simson* and *The Blythsome Bridal.* This was followed by Allan Ramsay's *Ever Green* (1724) and *Tea-Table Miscellany* (1724 ff.), and by the later collections of David Herd, John Pinkerton, James Johnson, and George Thomson. These anthologies testify to a new national consciousness which began with a deliberate invocation of past achievements and eventually produced original work in the tradition of these achievements. As political identity was lost, cultural differences were increasingly recognized as significant. Educated speech was slowly but surely following written prose into anglicization, but in poetry, the heart of the language was still felt to beat, however faintly, as it was also in common speech and in the old songs and ballads which were being reprinted. Ramsay's own poetry, like that of Robert Fergusson (1750–74) and Robert Burns (1759–96), flourished on a union of these elements: the racy vigour of Scots speech, the musical lilt of the

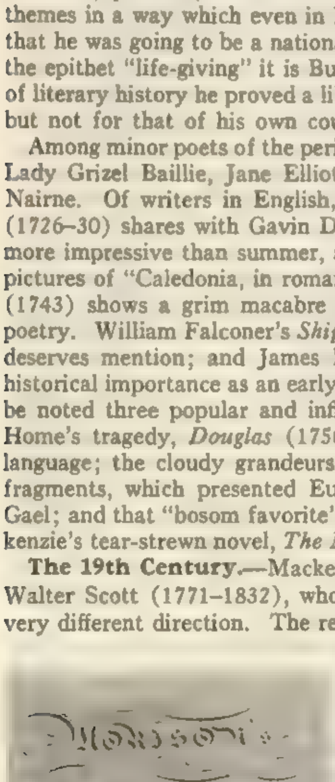
songs, and the poetic forms and techniques, such as the Habbie Simson stanza, the humorous elegy, and the verse epistle, with which recent poets like the Sempills and William Hamilton of Gilbertfield had helped to solder the vernacular tradition. This new poetry lacked the range and also the intellectual power of the makaris' verse. Nevertheless, its qualities were notable and infinitely attractive: in the love songs "Duncan Gray" and "O, Wert Thou in the Cauld Blast"; in the voice given to the joy and pathos of "randie gangrel bodies" in "The Jolly Beggars"; in the unembittered descriptive tartness of Fergusson's "Auld Reikie" and "Butterfly." Burns excelled in both satires and love lyrics; in the former he carried on where Dunbar left off, and in the latter he crystallized the floating anonymous songs in forms at once popular and authoritative. Many traditions meet in Burns and his poetry reaffirms, perfects, and consolidates existing poetic modes and themes in a way which even in his own lifetime made it apparent that he was going to be a national bard. If any poet has deserved the epithet "life-giving" it is Burns, and yet by one of the ironies of literary history he proved a liberator for the poetry of England, but not for that of his own country.

DETAIL FROM A MINIATURE BY ALEXANDER REID; PHOTO BY COURTESY OF THE TRUSTEES OF THE SCOTTISH NATIONAL PORTRAIT GALLERY

ROBERT BURNS, ABOUT 1795

Among minor poets of the period, some of the best were women: Lady Grizel Baillie, Jane Elliot, Lady Anne Lindsay, and Lady Nairne. Of writers in English, James Thomson in *The Seasons* (1726–30) shares with Gavin Douglas the ability to make winter more impressive than summer, and produces some heavily drawn pictures of "Caledonia, in romantic view." Robert Blair's *Grave* (1743) shows a grim macabre relish not uncommon in Scottish poetry. William Falconer's *Shipwreck* (1762), admired by Burns, deserves mention; and James Beattie's *Minstrel* (1771–74) has historical importance as an early Romantic poem. Here must also be noted three popular and influential works of the time: John Home's tragedy, *Douglas* (1756), Scottish in theme but not in language; the cloudy grandeurs of James Macpherson's Ossianic fragments, which presented Europe with a vision of the noble Gael; and that "bosom favorite" of the young Burns, Henry Mackenzie's tear-strewn novel, *The Man of Feeling* (1771).

The 19th Century.—Mackenzie hailed the early works of Sir Walter Scott (1771–1832), who was destined to turn prose in a very different direction. The re-emergence of a vernacular prose, mainly within the novel and the short story, is one of the distinguishing features of this period. The writers may be divided into two groups: those who restricted Scots to dialogue and used English for the narrative, and those who employed a form of Scots, or at least a scotticized English, throughout. The former included Scott himself, James Hogg (1770–1835), Susan Ferrier (1782–1854), George Macdonald (1824–1905), Mrs. Margaret Oliphant (1828–97), William Black (1841–98), and R. L. Stevenson (1850–94); in the second and smaller group are John Galt (1779–1839), David M. Moir (1798–1851), and (in the 20th century) Lewis Grassic Gibbon. A book like William Alexander's *Johnny Gibb of Gushetneuk* (1871) contains so much dialogue as practically to belong to the scotticizers; it is, however, of interest mainly to philologists. The advantage of Galt's method in his *Annals of the Parish* (1821), of casting the book as the reminiscences of a Scots-speaking person, was that it secured a homogeneous tone, and the exact but not exaggerated Scots idiom is admirable. Scott, on the other hand, has the wider range that comes from contrast between English and Scots (as, most obviously, in "Wandering Willie's Tale" in *Redgauntlet*), though the faults of his English writing, and the restriction of his brilliant Scots dialogue to minor characters, sometimes make the result patchy and unsatisfactory, even if the great sweep of the story carries the reader through.



BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

TITLE PAGE OF A 1793 EDITION OF JAMES THOMSON'S "THE SEASONS," A SPLENDID EXAMPLE OF SCOTTISH PRINTING AND ENGRAVING

A novel that begins "Come oot o' the gutter, ye nickum!" (George Macdonald's *Sir Gibbie,* 1879) is evidently of another type, and it may in fact be claimed as one of the many forerunners of the so-called Kailyard School of Fiction, which appeared toward the end of the century, and included Ian Maclaren (the Rev. John Watson, 1850–1907), S. R. Crockett (1860–1914), and James Barrie (1860–1937). (At the same time Fiona Macleod [William Sharp, 1856–1905] in his Celtic stories was creating the "Kailyard of the Gael.") The sentimental viewpoint that vitiates Maclaren's *Beside the Bonnie Brier Bush* (1894) and weakens Barrie's *A Window in Thrums* (1889), the cosy hypocrisies, the idealization of village and humble life, were pitilessly exposed in George Douglas Brown's *House with the Green Shutters* (1901). Brown's brutal tale, nevertheless, had its own exaggerations, those needed for its theme of tragic melodrama, and it had no more than a temporary effect in "sticking the Kailyarders like pigs," as one critic hoped. There is something in Kailyard, with its evasion of reality and its tendencies toward the sentimental and the pawky, still acceptable to the Scottish popular temperament.

The history of poetry in this period is very nearly a blank. Two poems of the supernatural keep up an ancient Scottish tradition: James Hogg's *Bonny Kilmeny* (1813) and William Bell Scott's *Witch's Ballad* (1875); there are some good vernacular lyrics by Scott, Allan Cunningham, and George Macdonald; but for the most part, the century after Burns saw merely the proliferation of the maudlin and the jocose in well-worn verse forms, the distortion and coarsening of Burns's gaiety and pathos, and the popularizing of everything in him that was less than first-rate. The widely read mid-century anthologies of verse called *Whistle Binkie: a Collection of Songs for the Social Circle* (1846) were the poetic equivalent of Kailyard, and their popularity is a comment on the poetic taste of the age. Indeed, during the second half of the century more interesting poetry was being written in English, by James Thomson (1834–82), Alexander Smith (1830–67), and John Davidson (1857–1909). What we feel to be lacking in Kailyard and *Whistle Binkie*—the impact of the contemporary world, of science, of industrialization, of the city, even (one must add) of ideas—gives strength and power to Smith's *Glasgow* (1857), to Thomson's *City of Dreadful Night* (1874), and to Davidson's *Testaments* (1901–08).

The 20th Century.—Prose fiction has developed, with minor variations, mainly within the Scott-Stevenson tradition. The "descriptions of the physical" on which Stevenson prided himself (though it was by no means his only gift), and the reliance on action rather than on psychological subtleties which we find in Scott, are continued in varying degrees in John Buchan (1875–1940), Neil Munro (1864–1930), and R. B. Cunninghame Graham (1852–1936), the last-named having shown in his *Scottish Stories* (1914) a nice use of Scots and a fine command of the macabre. More original are the picaresque, witty, and often fantastic narratives of Eric Linklater (1899–); the passionately written *Scots Quair* of Lewis Grassic Gibbon (James Leslie Mitchell, 1901–35), with a remarkable stylistic attempt to sink Scots idiom and vocabulary into the body of a richly descriptive English; the Celtic visionary quality in Neil Gunn (1891–); and the sharp character studies of Robin Jenkins (1912–).

Drama, revived to some extent by Barrie (*e.g., Dear Brutus,* 1917), discovered a more distinctively native voice in James Bridie (O. H. Mavor, 1888–1951); his entertaining plays, on an unusually wide variety of themes, combine the humorous, the ethical, and the supernatural in a very Scottish manner.

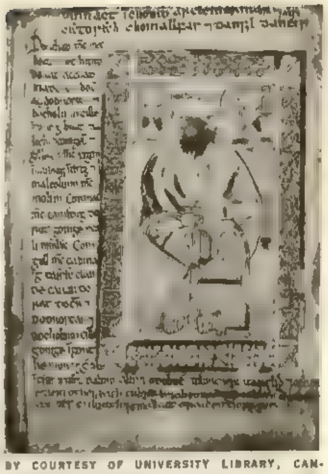
It is to poetry, however, that the term revival is more properly applied. The first steps toward a Scots verse that would more closely represent real experience were made by Stevenson (in *Underwoods*, 1887), Marion Angus (1866–1946), Charles Murray (1864–1941), Violet Jacob (1863–1946), Pittendrigh Macgillivray (1856–1938), Helen B. Cruickshank (1886–), and Alexander Gray (1882–). In the second decade of the 20th century the explosive personality of Hugh MacDiarmid (Christopher Murray Grieve, 1892–) made itself felt, as the centre and motive force of what has been variously called the Scottish Renaissance movement, the Lallans revival, and Synthetic or Plastic Scots. This movement, by re-expanding the Scots vocabulary after its post-Burns stagnation, complacency, and "hameliness," aims to restore the intellectual prestige of Scots and to create a medium capable of dealing with the objects and ideas of the modern world as well as with the perennial sources of lyricism. MacDiarmid himself has written the best lyrics since Burns; his sesquipedalian, polyglot, encyclopaedic, polemical, metaphysical verse has not aroused the same enthusiasm, but is interesting and significant. Other members of the movement include the humorous Robert Garioch (R. G. Sutherland, 1909–), the learned Douglas Young (1913–), the love poet Sydney Goodsir Smith (1915–), and the wry observer William Soutar (1898–1943). Of poets using English, Edwin Muir (1887–1959), with his meditative myth-haunted verse is the most distinguished; Norman MacCaig (1910–), W. S. Graham (1917–), George Mackay Brown (1927–), and Iain Crichton Smith (1928–) write with craftsmanship and force. In a class by itself is the delicate experimental poetry of Ian Hamilton Finlay (1925–). The Scottish sense of separateness is too real to be swamped, but English is dominant in prose writing and in educated speech, and the Scottish poet of the 20th century has been therefore like a shaman who must try out many disguises until he finds the one in which he can best prophesy and persuade. (E. G. M.)

MARK GERSON-CAMERA PRESS

HUGH MACDIARMID (CHRISTOPHER MURRAY GRIEVE), FOREMOST SCOTTISH POET OF THE EARLY 20TH CENTURY

IN GAELIC

It must be remembered that English or Scots was for long the language of a minority only in a Gaelic-speaking nation, and, although less than 2% of the population speaks Gaelic today, Gaelic literature still flourishes in Scotland. The early history of Scottish Gaelic literature, like that of the language, cannot be adequately considered in isolation from the literature and language of Irish Gaelic (*see* CELTIC LANGUAGES; IRISH LITERATURE). Kenneth Jackson demonstrated in his Sir John Rhŷs memorial lecture (1951) that Scottish and Irish Gaelic were identical until the 10th century, that they proceeded to diverge from the 10th to the 13th centuries, but that the language which he conveniently calls "Common Gaelic" survived as a living tongue until the 13th century. The conservatism of the literary classes ensured that there was a common corpus of literature even after the spoken languages diverged. Thus a distinct Scottish Gaelic period in literature can hardly be said to begin much before the 15th century. The Scottish parliament passed a law against the bards in 1457, when it was decided to take an inquisition of all "sornares, bairdes, maister-full beggers or feinziet fuiles." This law and some later ones were no doubt directed against the strolling bards, but governmental disapproval of the bards and of the Gaelic language and way of life grew; and the policy of centralization, and the gradual de-Gaelicization of the titled and moneyed classes, eventually deprived the Gaelic poets of their patrons, bringing both loss and gain to Gaelic poetry. But before looking at the literature of the modern Scottish Gaelic period we must take a glance at what lies behind it.



BY COURTESY OF UNIVERSITY LIBRARY, CAMBRIDGE

ILLUMINATED PAGE FROM "THE BOOK OF DEER," AN EARLY MANUSCRIPT CONTAINING GAELIC PASSAGES

Earliest Manuscript Remains.—The earliest document containing Gaelic matter which Scotland can claim is *The Book of Deer*, preserved in the University Library, Cambridge. This manuscript contains portions of the Gospels in Latin, written in an Irish hand, with illuminations. Inserted in the margins and blank spaces are later notes and memoranda partly in Latin, partly in Gaelic. The six Gaelic entries seem to have been made between 1131 and 1153, although some may be copies of earlier, perhaps traditionally preserved, grants of land and other privileges made from time to time to the monastery of Deer, Aberdeenshire. The scribes seem to have been trained, though somewhat imperfectly, in the Common Gaelic tradition of the period. Another strangely isolated survivor is the Islay charter of 1408. This charter, written in Gaelic, was granted by Donald, Lord of the Isles, to Brian Vicar Mackay of Islay. It is witnessed, and was very probably written, by Fercos MacBetha, one of the famous family of hereditary physicians of that surname.

These MacBeths, Beatons, or Bethunes had in their possession a number of important medical manuscripts, some of which are in the collection of Gaelic manuscripts in the National Library of Scotland. The O'Conachers, or McConachers, of Lorne were also hereditary physicians whose manuscripts are in part preserved. These manuscripts translate and paraphrase the works of such men as Bernard Gordon, Galen, Avicenna, and Hippocrates, while in their extra-medical digressions they show acquaintance with the works of Aristotle, Socrates, St. Thomas Aquinas, and others. Such various subjects as metaphysics, astronomy and astrology, theology, and music are treated. In medical matters the scribes sometimes give their own observations. These manuscripts date mainly from the 15th, 16th, and 17th centuries.

There are many other manuscripts, belonging to the common Scots-Irish tradition, which contain versions of heroic tales and sagas, poetry, ecclesiastical writings, and miscellaneous matter. One such, the Glenmasan ms. (no. liii in the National Library collection) was probably written *c.* 1500, but may derive from a 13th-century original. It contains the famous story of Deirdre (*q.v.*) and the sons of Uisneach. A part of the manuscript, containing the story of *Mesca Ulad*, dates from 1538. There is also a fragment of the most famous of Irish epics, the *Táin Bó Cúailnge* (ms. lix), and a Gaelic version of Lucan's *Pharsalia* (ms. xlvi), besides poetry written by Irish and Scottish bards.

But from the purely Scottish Gaelic point of view, the most important manuscript from this period is the *Book of the Dean of Lismore*, an anthology of verse compiled between 1512 and 1526 by Sir James MacGregor, dean of the diocese of Lismore in Argyllshire, and his brother Duncan. The poems fall into three main groups: (1) poems by Scottish authors; (2) poems by Irish authors: and (3) Ossianic ballads. The Irish poems belong mainly to the period 1200–1500. In the third group there are 29 poems, of which 4 belong to the Cú Chulainn cycle rather than the Ossianic one; these poems are attributed to various authors, for example, Oiséan, Fearghus File, Caoilte Mac Rónáin, and Conall Cearnach.

The poems by Scottish authors are interesting historically, linguistically, and poetically; their dates of composition range from *c.* 1310 to *c.* 1520. They give valuable evidence concerning aspects of contemporary society, and, being written in a semiphonetic spelling, partly based on the spelling of contemporary Scots, they

provide some evidence as to the pronunciation of Gaelic in the dean's time. Peculiarly Scottish features of pronunciation, morphology, and grammar can be detected. Many have a Perthshire setting, or are written to a MacGregor chief, but there are also, for example, poems to MacDougall of Dunollie, Macleod of Lewis, MacNeill of Gigha, and others. The bard best represented is Fionnlagh Ruadh, bard to John, chief of Clan Gregor (d. 1519). There are also three poems by Giolla Coluim mac an Ollaimh, probably a poet at the court of the Lord of the Isles. He laments the misfortune which had befallen this proud family in the late 15th century. Other poets are the dean of Knoydart (late 15th century), Duncan Campbell, the Good Knight of Glenorchy (who died at Flodden, 1513), Giolla Críost Brúilingeach, and two women, Aithbhreac Inghean Coirceadail (mid-15th century) and Isabella, countess of Argyll (probably the wife of the first earl). These last two are satisfying poetically, as is also the work of Giolla Coluim and of Giolla Críost.

The 16th and 17th Centuries.—Very little Gaelic poetry from the 16th century has survived, and most of it survived in oral tradition until the mid-18th century when it was written down. We can point to a few poems like *An Duanag Ullamh*, composed *c.* 1556 in honour of Archibald, fourth earl of Argyll; the lovely lament *Griogal Cridhe*, composed *c.* 1570; *Oran na Comhachaig*, in part no doubt composed in the second half of the 16th century; some poems by Sir John Stewart of Appin; and some by Niall Mór MacMhuirich. It is certain that the poetry recorded in the *Book of the Dean of Lismore* was not an isolated outburst; poetry was an integral part of the society of the time. Much must have been lost, both of the work of the professional bards and of the popular songs. Songs in the nonsyllabic, accented measures survive, again orally, from the earliest years of the 17th century, and it must be concluded that these forms have a previous history in the 16th century. This is the tradition which produced the work songs (*e.g.*, waulking songs used as an accompaniment to the fulling of cloth) better known from 17th- and 18th-century examples.

In 1567 the first book printed in Gaelic in Scotland appeared: Bishop John Carswell's *Foirm na Nurrnuidheadh*, a translation of John Knox's Liturgy. Carswell is modest about his ability to write Gaelic, but in fact his prose is in classical Common Gaelic, and resembles that of the 17th-century Irishman Geoffrey Keating. His prefatory epistle makes lively and pleasant reading.

The 17th century is a period of great interest in Gaelic Scotland. The political, ecclesiastical, and social structures of Scotland were gradually changing, and so was the relationship between the central government and the Gaelic area. Enough survives of the period's Gaelic poetry to show that there was a large number of poets of great talent. It is probably not a coincidence that the 17th century is also the great age of the waulking songs and of *ceòl mór*, or the classical music of the bagpipes. There seems to have been a diffusion of artistic talent such as is scarcely matched in any other period in Scottish Gaelic history. Some of the poetry and prose is contained in three 17th-century manuscripts. The first two are the *Black* and the *Red Books of Clanranald*, written by members (*e.g.*, Cathal and Niall) of the MacMhuirich family, hereditary bards to the MacDonalds of Clanranald. They were probably written for the most part in the 17th century, but contain poems by earlier representatives of this family of bards. Much of the two manuscripts is taken up with genealogical matter, and with an account of the Montrose wars and of the achievements of Alasdair MacDonald (Mac Cholla Chiotaich). From 18th-century accounts it is known that the MacMhuirich family had compiled other manuscripts, most of which were carelessly destroyed, although ms. xlviii in the National Library of Scotland is apparently a MacMhuirich poetry manuscript. Other hereditary bardic families were the MacEwens, bards to the Campbells, and the Ó Muirgheasáins, bards to the Macleods of Harris and Dunvegan.

The other important 17th-century Gaelic document is the *Fernaig Ms.*, compiled by Duncan Macrae of Inverinate between 1688 and 1693. This contains *c.* 4,200 lines of verse, mostly political and religious. The poets represented belong to the 16th and 17th centuries, although little work by the major 17th-century poets is given. The compiler's Jacobite and Episcopalian leanings are reflected in his anthology.

The two best-known poets of the 17th century are Mary Macleod and Iain Lom. The former, known as Màiri Nighean Alasdair Ruaidh (*c.* 1615–*c.* 1706), was closely associated with the households of the Macleods of Harris and Dunvegan. She seems to have been a member of the household of Rory Mór Macleod (d. 1626), and to have acted as nurse to successive generations of Macleod chiefs. Those for whom she had the greatest affection were probably Sir Norman Macleod of Bernera and Iain Breac Macleod of Dunvegan. Her earliest poem of certain date is the elegy for Roderick Mackenzie of Applecross (d. 1646), and one of her latest is the elegy for Sir Norman Macleod of Bernera (d. 1705). Her poems show deep personal emotion. Her imagery, though restricted in range, is sincere and telling, and her style is fresh and natural. She inherits much of the imagery of the bardic poets, but places it in a new setting, and her metres are the comparatively new strophic ones rather than the strict syllabic metres of the classical bards.

John Macdonald, known as Iain Lom (*c.* 1620–*c.* 1710), was related to the chiefs of the Keppoch Macdonalds, and took an active interest and an active part in the events of his time. His life spanned an eventful period in Highland history, and his poetry is full of comment on events. He composed poems about the battles of Inverlochy (1645), Auldearn (1645), and Killiecrankie (1689), a lament for the marquess of Montrose (d. 1650), a poem on the restoration of Charles II in 1660, several poems dealing with the Keppoch murder of 1663, an elegy for Sir James Macdonald of Sleat (1678), a song to William and Mary (1688), and a song bitterly opposing the union of the parliaments in 1707. He was a royalist and a fervent Macdonald panegyrist. As a Macdonald he lashes the Campbells; as a Roman Catholic he does the Presbyterians less than justice. But the breadth of his interest in national affairs is notable. His versification is less melodious than Mary Macleod's, and he has none of her tenderness and nostalgia. But he has that compression and concentration of expression lacking in so much later Gaelic poetry. This may derive in part from the style of the classical bards, as the directness of his utterance may from popular or folk poetry. He can achieve more at times by understatement than can the 18th-century poets by a plethora of epithets.

We can do little more than mention some other 17th-century poets of whose works interesting fragments have survived. Donnchadh MacRuaraidh was bard to Macdonald of Sleat, and his best-known poem consists of four calm resigned verses composed on the day of his death (*c.* 1630). The work of Alasdair MacKenzie (d. *c.* 1642) and of Murdoch Mackenzie (fl. 1650), both of Achilty, is represented in the *Fernaig Ms.* Alasdair is the *laudator temporis acti*; Murdoch sometimes shows lively interest in action and movement, as in his *An Làir Dhonn.* Roderick Morison, usually known as An Clàrsair Dall (the Blind Harper), was born in Lewis and became harper to Iain Breac Macleod of Dunvegan. Little of his work survives, but the strong texture and poetic intensity of his *Oran Mór MhicLeòid* and his *Creach na Ciadain* are remarkable. Dorothy Brown and Silis na Ceapaich (*c.* 1660–1729) are women poets of great talent, and with them may be grouped Mairearad Nì Lachainn (b. *c.* 1660) and Catriona Nic Ghilleathain (fl. 1700). Dorothy Brown's poem on Alasdair Mac Colla, and Silis na Ceapaich's laments for Lachlan MacKinnon and for Alasdair of Glengarry (d. 1721) are moving and artistic utterances. Eachann Bacach Maclean would be honoured for the one poem *A' Chnò Shamhna* (1648), even if no more of his work survived.

Four other poets mark, in various ways, the transition from the poetry of the 17th century to that of the 18th. They are Lachlan MacKinnon (Lachlann Mac Theàrlaich Oig), John Mackay (Am Pìobaire Dall), John Macdonald (Iain Dubh Mac Iain 'Ic Ailein), and John Maclean (Iain Mac Ailein). All four were born *c.* 1655–65. Much of their work lacks the economy of the best 17th-century poetry, and also its richness of imagery. John Macdonald's *Oran nam Fineachan* is not the worst in a series of tedious jingles which can be called clan verses, and which were popular with

18th-century poets. John Maclean shows a great interest in early Gaelic legend, and he composes amusing verses in a mock-heroic style. He also uses a mixture of verse and prose, as in his *Crosanachd Fhir nan Druimnean*. John Mackay's *Coire an Easa* is significant in the development of Gaelic nature poetry. Finally, bardic poetry continued to be composed into the 18th century by two members of the MacMhuirich family, Niall and Domhnall.

The 18th Century.—The 18th century has eclipsed the 17th partly because of the brilliance of some individual achievements, and partly because it is from the viewpoint of the 18th century that scholars often begin their retrospect of earlier Gaelic poetry. Practically no secular poetry in Gaelic was printed before 1751, and the bulk of earlier verse was recovered from oral tradition after that date. Interest in it was part of the antiquarian revival.

Much of the inspiration of Gaelic printing in the 18th century can be traced to Alexander Macdonald, or Alasdair Mac Mhaighstir Alasdair (*c.* 1700–*c.* 1770), who published a Gaelic vocabulary in 1741 and the first Scottish Gaelic book of secular poetry, *Aisheiridh na Sean Chánoin Albannaich* (*Resurrection of the Ancient Scottish Tongue*), in 1751. On both political and literary issues he showed virtues of independence and leadership which mark him out as a man apart among Gaelic poets of the century. During the '45 rising he was granted a captain's commission by Prince Charles Edward, and he seems to have regarded the rising as a crusade for the Highlands, or even for the Gaelic way of life. He rallied his fellow-Highlanders to the prince's cause with such poems as *Brosnachadh nam Fineachan Gaidhealach* (*Incitement to the Highland Clans*), and *Hi ri rì tha e tighinn*, a song of welcome to the prince. His poetical range is wide. His songs on the seasons must have been composed within 15 years of Thomson's *Seasons* (1726–30). Another notable nature poem is his *Oran do Allt an t-Siùcair*. His *Urnuigh do'n Cheòlraidh* (*Invocation of the Muses*) contains a mature statement of the aims of his art. His masterpiece, and longest poem, *The Birlinn of Clanranald*, is an extravaganza which is ostensibly a description of a voyage from South Uist to Carrickfergus. A fine poem in a quite different mood is the love poem *Moladh Mòraig* (*The Praise of Morag*). He also composed drinking songs and bitter satires. He has a rich vocabulary and a strong sense of technique.

Duncan Macintyre, or Donnchadh Bàn (1724–1812), was influenced in various ways by Macdonald. He published his poems in 1768, and signs of this influence can be seen in the subject matter and the metres, and sometimes in the phraseology of certain poems. But his individuality is not lost. The two men were very different in nature and in training. Macdonald had received a formal education; Macintyre could neither read nor write. Macdonald was ambitious, independent, and restless; Macintyre had none of these uncomfortable qualities. Macintyre fought on the Hanoverian side at the battle of Falkirk (Jan. 17, 1746), and later sang the praises of George III in *Oran do'n Rìgh*. He composed other soldiering songs, and various other set pieces on the clans, the bagpipes, and the Gaelic language, but his true subjects are wildlife and wild nature—the haunts and the habits of the red deer. He had been a forester on the Perthshire-Argyllshire borders in early manhood, and this is the setting of his greatest poems, *Moladh Beinn Dóbhrain* (*The Praise of Ben Dorain*) and *Oran Coire a' Cheathaich* (*Song to Misty Corrie*). These are remarkable for their emotional closeness to nature, their minute and objective detail, and their personal and lyrical quality. His most famous love song is addressed to his wife, *Màiri*. He is the best-loved of the Gaelic poets.

There were several other poets of note in the 18th century. John MacCodrum (*c.* 1693–1779) was appointed bard to Sir James Macdonald of Sleat in 1763, and his elegy for Sir James is one of his best poems. He had a good memory for genealogy and poetry, and was a known wit. He composed much humorous and satirical poetry. Rob Donn Mackay (*c.* 1714–78) was the bard of the Reay country, and used his gifts of observation and satire in commenting on the life of his district. Characters such as the miserly brothers of Rispond have a quality of universality. There is much social satire in his verse, and some delicate feeling for nature, as in *Cead Fhir Bhloguis do'n Fhrith*. William Ross (*c.* 1762–91) is the Romantic poet of the group. His poetry is emotional, subjective, and at times sentimental. An unhappy love affair occasioned several of his best poems, such as *Feasgar Luain* and *Oran Eile*. Another well-known poem is *An Suaithneas Bàn* (*The White Cockade*), composed on hearing of Prince Charles Edward's death in 1788. David MacKellar published his hymn on the Creation in 1752, but the greatest composer of Gaelic religious verse in the 18th century was Dugald Buchanan (1716–68), whose poems were published in 1767. His *The Day of Judgement* and *The Skull* are impressive and sombre, and show considerable imaginative power. Buchanan assisted Stewart of Killin in preparing his Gaelic translation of the New Testament (1767). Ewen MacLachlan (1773–1822), librarian of King's College, Aberdeen, transcribed many Gaelic manuscripts, including the *Book of the Dean of Lismore*, translated seven books of Homer's *Iliad* into Gaelic verse, and wrote Gaelic poems on the seasons.

19th and 20th Centuries.—A considerable amount of periodical literature was published in the 19th and 20th centuries. *An Gaidheal* is the lineal descendant of some of these periodicals. It has been joined by a quarterly, *Gairm* (1952 ff.). These periodicals contain the main volume of the Gaelic prose that we possess.

In addition there is a quantity of translation of religious works. There is a regular tradition of such translations since the 17th century. A translation of Calvin's Catechism was issued in 1631. Fifty of the psalms, together with the Shorter Catechism, appeared in 1659, Kirke's Psalter in 1684, and the Synod of Argyll's Psalter in 1694. Kirke's Irish version of the Bible in Roman type appeared in 1690. James Stewart's Gaelic edition of the New Testament was published in 1767, and John Stewart of Luss and John Smith of Campbeltown published a translation of the Old Testament (1783–1801). The complete Gaelic Bible was issued in 1807, and the first standard revision was that of 1826. Meantime many other religious translations were appearing, of works by Richard Baxter, Thomas Boston, John Bunyan, Philip Doddridge, and others, as well as volumes of sermons.

Some notable original prose works are Lachlan Maclean's *Adam and Eve* (1837), the Rev. Norman MacLeod's *Caraid nan Gàidheal* (1867), Donald Mackechnie's *Am Fear Ciùil* (1904), Angus Robertson's *An t-Ogha Mor* (1913), and the writings of Donald Lamont (1874–1958).

There was a dearth of vital poetry in the 19th century, when much laboured and mannered verse was composed. But there are partial exceptions, in some of the work of John MacLachlan of Rahoy (1804–74), Evan Maccoll (1808–98), Neil MacLeod (1843–1913), and Mary Macpherson (1821–98). John Morison (1790–1852), the poet-blacksmith of Rodel, composed some elevated and imaginative religious poetry. William Livingstone (1808–70) is full of rebellion and resentment, and he tried, without much success, to clothe Scoto-Norse history in poetic and dramatic dress. Angus Robertson's *Cnoc an Fhradhairc* (1940) is a hard poem to read, but it is rewarding in parts. Later a new movement to free Gaelic poetry from oppressive traditional shackles was inaugurated by Somhairle MacGhill-Eathain's (*q.v.*) *Dàin do Eimhir* (1943), George Campbell Hay's *Fuaran Slèibh* (1948) and *O na Ceithir Airdean* (1952), and Ruaraidh MacThómais' (Derick Thomson's) *An Dealbh Briste* (1951). In a style partly traditional and partly new is James Thomson's *Fasgnadh* (1953). Interesting new work in poetry has been published by Donald MacAulay and Iain Crichton Smith from 1956 onward.

Little of permanent value has been done in the drama. The art of the short story has been cultivated in periodicals and in Gaelic radio programs, notably by Finlay J. Macdonald, Colin Mackenzie, and Iain Crichton Smith. A collection by Smith, *Bùrn is Aran*, appeared in 1960. No fully adult Gaelic novel has appeared. The art of essay writing, however, has been pursued with distinction by Donald Lamont and others.

Collections of Gaelic Poetry.—From the mid-18th century onward a number of important collections of Gaelic poetry were made. One of the earliest was the Rev. Alexander Pope's (*c.* 1739). Jerome Stone collected Ossianic ballads and other verse before he died in 1756. The Rev. James Maclagan continued to collect material throughout the second half of the century, and his

manuscripts are in Glasgow University Library. The Rev. Donald MacNicol's collection also belongs to this period. Dr. Hector Maclean of Mull's manuscript dates from *c.* 1768. In 1776 Ronald Macdonald's *Eigg Collection* was published, and in 1786 John Gillies', based on Maclagan's. The notable enterprises of James Macpherson (*q.v.*), his *Fragments of Ancient Poetry* (1760), *Fingal* (1762), and *Temora* (1763), stem from this new collecting zeal, and in turn reinforce it. Macpherson's work is partly based on genuine Gaelic ballad sources, which can be elucidated, and partly constructed from borrowed English and Irish sources, and from his own imagination. But it is probable that we owe to Macpherson the collection and in some cases the preservation of valuable oral and manuscript material, as for example the dean of Lismore's manuscript. The Rev. Patrick Macdonald's *Gaelic Vocal Airs* was published in 1784. Important collections of the early 19th century were those of Alexander and Donald Stewart (1804), Patrick Turner (1813), and Alexander Campbell's *Albyn's Anthology* (1816–18). The movement continued with John MacKenzie's *Sàr-Obair nam Bàrd Gaelach, or the Beauties of Gaelic Poetry* (1841; new ed. 1904), Donald Macpherson's *An Duanaire* (1868), and the work of the two giants among 19th-century collectors, John Francis Campbell of Islay (1822–85) and Alexander Carmichael (d. 1912). Campbell's four volumes of *Popular Tales of the West Highlands* were published in 1860–62, and his *Leabhar na Féinne*, a great collection of heroic ballads, in 1872. The first installment of Carmichael's great work, *Carmina Gadelica*, was published in two volumes in 1900 (see *Bibliography*). The Rev. Alexander Cameron's *Reliquiae Celticae* (1892–94) made available in print much valuable manuscript material, such as the *Fernaig Ms.*, parts of the *Books of Clanranald*, and some of the Edinburgh manuscripts, with Ossianic and other material from various manuscript collections. A collection of Gaelic proverbs was published, with an English translation, by Donald Macintosh in 1785, and supplemented and enlarged by Alexander Nicolson in 1881; a new edition, edited by Malcolm MacInnes, appeared in 1951. Frances Tolmie's fine collection of Gaelic folk songs was published in 1911, and Margaret Fay Shaw's *Folksongs and Folklore of South Uist* in 1955. Beginning in 1872, there is the valuable series of *Transactions of the Gaelic Society of Inverness*, which has published much material from manuscript and oral sources. The periodical *Scottish Gaelic Studies* was launched in 1926.

(D. S. T.)

BIBLIOGRAPHY.—*General Works:* J. H. Millar, *A Literary History of Scotland* (1903); G. G. Smith, *Scottish Literature: Character and Influence* (1919); J. Speirs, *The Scots Literary Tradition* (1940); S. G. Smith, *A Short Introduction to Scottish Literature* (1951); *Scottish Poetry: a Critical Survey*, ed. by J. Kinsley (1955); K. Wittig, *The Scottish Tradition in Literature* (1958).

Scots and English: (a) Anthologies: *Specimens of Middle Scots*, prose and verse, ed. by G. G. Smith (1902); F. Urquhart (ed.), *Scottish Short Stories* (1932; 3rd rev. ed., 1957); R. L. Mackie (ed.), *A Book of Scottish Verse* (1934); M. M. Gray (ed.), *Scottish Poetry from Barbour to James VI* (1935); H. MacDiarmid (ed.), *The Golden Treasury of Scottish Poetry* (1940); *The Green Garden*, verse from the 15th to the 18th centuries, ed. by J. Fergusson (1946); J. M. Lindsay (ed.), *Modern Scottish Poetry: an Anthology of the Scottish Renaissance, 1920–1945* (1946); J. W. Oliver and J. C. Smith (eds.), *A Scots Anthology* (1949); D. C. C. Young (ed.), *Scottish Verse, 1851–1951* (1952); N. MacCaig (ed.), *Honour'd Shade: an Anthology of New Scottish Poetry* (1959); J. M. Reid (ed.), *Scottish Short Stories* (1963). (b) History and Criticism: T. F. Henderson, *Scottish Vernacular Literature*, 3rd ed. (1910); W. Geddie, *A Bibliography of Middle Scots Poets* (1912); J. H. Millar, *Scottish Prose of the 17th and 18th Centuries* (1912); C. M. Grieve, *Contemporary Scottish Studies* (1926 ff.); A. J. Mill, *Mediaeval Plays in Scotland* (1927); W. L. Renwick and H. Orton, *The Beginnings of English Literature to Skelton, 1509*, 2nd ed. (1952); D. Craig, *Scottish Literature and the Scottish People, 1680–1830* (1961).

Gaelic: (a) Anthologies: *An t-Òranaiche*, a Gaelic songster ed. by A. Sinclair (1879); A. M. Sinclair (ed.), *The Gaelic Bards, 1411–1765*, 2 vol. (1890–92); *The Glenbard Collection* (1890–95); *Carmina Gadelica*, A. Carmichael's collection of hymns and incantations, with parallel translations, 5 vol. (1900–54); M. C. Macleod (ed.), *Modern Gaelic Bards* (1908–13); A. and A. Macdonald (eds.), *The Macdonald Collection of Gaelic Poetry* (1911); W. J. Watson (ed.), *Rosg Gàidhlig: Specimens of Gaelic Prose* (1915), and *Bàrdachd Ghàidhlig: Specimens of Gaelic Poetry, 1550–1900* (1918); *Highland Songs of the Forty-Five*, ed. and trans. by J. L. Campbell (1933); *More West Highland Tales*, transcribed and trans. by J. G. MacKay (1940–60); G. R. D. McLean (ed.), *Poems of the Western Highlanders* (1961). (b) Editions of individual authors and mss.: J. Carswell's translation of Knox's Liturgy, ed. by T. Maclauchlan (1873); *Dain Iain Ghobha: the Poems of John Morison*, 2 vol., ed. by G. Henderson (1893–96); *Songs and Poems by Rob Donn*, ed. by H. Morrison (1899); *The Spiritual Songs of Dugald Buchanan*, ed. by D. Maclean (1913); *The Fernaig Ms.*, ed. by M. Macfarlane (1923); *The Poems of Alexander Macdonald*, ed., with parallel translations, by A. and A. Macdonald (1924); *Gaelic Songs of Mary Macleod*, ed., with parallel translations, by J. C. Watson (1934); *Gaelic Songs by William Ross*, ed., with parallel translations, by G. Calder (1937); *Ewen MacLachlan's Gaelic Verse*, ed. by J. Macdonald (1937); *Scottish Verse from the Book of the Dean of Lismore*, ed., with parallel translations, by W. J. Watson (1937); *Heroic Poetry from the Book of the Dean of Lismore*, ed., with parallel translations, by N. Ross (1939); *The Songs of John MacCodrum*, ed., with parallel translations, by W. Matheson (1938); *The Songs of Duncan Ban Macintyre*, ed., with parallel translations, by A. Macleod (1952); *Prose Writings of Donald Lamont*, ed. by T. M. Murchison (1960); A. M. Mackenzie (ed.), *Orain Iain Luim* (1964). (c) History and Criticism: N. MacNeill, *The Literature of the Highlanders* (1892; new ed. 1929); M. Maclean, *The Literature of the Celts* (1902; new ed. 1926), and *The Literature of the Highlands* (1904; 2nd ed. 1925); J. S. Smart, *James Macpherson: an Episode in Literature* (1905); D. Mackinnon, *A Descriptive Catalogue of Gaelic Mss. in the Advocates' Library, Edinburgh* (1912); D. Maclean, *Typographia Scoto-Gadelica* (1915); J. L. Campbell, *Gaelic in Scottish Education and Life* (1945); K. Jackson, "Common Gaelic," in *Proceedings of the British Academy*, vol. xxxvii (1951); D. S. Thomson, *The Gaelic Sources of Macpherson's "Ossian"* (1952). (E. G. M.; D. S. T.)

SCOTTSBORO CASE, an incident in the history of U.S. criminal law that had international repercussions in the 1930s and resulted in two important decisions of the U.S. supreme court affecting criminal procedure. It originated in a fight between white and Negro youths on a freight train in Alabama on March 25, 1931. The white youths were thrown off the train, and in revenge they reported that the Negroes had raped two white girls who were also on the train. A mob formed in Paint Rock, Ala., and took nine Negroes off the train when it arrived there. The Negroes were arrested and jailed in Scottsboro, Ala., the county seat. Their trials began on April 6 and were completed in three days; eight were sentenced to death, the ninth to life imprisonment.

Communist organizations saw this case as an opportunity to dramatize oppression of the Negro in the U.S.; they took over the defense and gave the case international publicity. The executions were stayed pending appeal. In Oct. 1932 the supreme court reversed the convictions on the ground that the defendants had not had adequate representation by counsel (*Powell* v. *Alabama*, 287 U.S. 45). This decision stated an important new constitutional rule—that defendants in capital cases in state courts must be represented by counsel.

New trials were held in Alabama in 1933. Though one of the women involved testified this time that no rape had occurred, all of the defendants were re-convicted and sentenced to death. Again the convictions were appealed, and in April 1935 the supreme court again reversed the convictions, this time because Negroes had been systematically excluded from grand and trial juries in the county where the trials occurred (*Norris* v. *Alabama*, 294 U.S. 587). The defendants were again indicted, this time by a grand jury on which a Negro served.

After the first four trials resulted in convictions the state dismissed indictments of the others. Two of the prisoners were paroled in 1944 and another in 1951. The fourth, Haywood Patterson, escaped from prison in 1948 and two years later published his story of the case in *Scottsboro Boy*. The supreme court decisions in this case are landmarks in U.S. constitutional law and have had great effect in improving standards of criminal justice in the states. (C. H. P.)

SCOUT. The term "scout" generally is used in its military sense to indicate a member of a command who operates some distance ahead of the main body of troops for the purpose of obtaining information concerning the strength and position of the enemy. The scout is often described as "the eyes and ears of the army." Formerly scouts worked on foot or as horsemen, but in recent times they have done so from lightly armoured motor vehicles sometimes known as "scout cars." Navies have used small, fast craft for scouting purposes and aircraft also frequently act in this capacity. A scout is usually a uniformed member of the

armed forces, although in the 18th and 19th centuries—particularly in the American Indian wars—local inhabitants, with a knowledge of the language and local conditions, were frequently employed clandestinely and called scouts. Today such men are termed spies or intelligence agents, although the difference between these and true scouts may often be difficult to define. In World War II the term scout largely fell into disuse, the term patrol—indicating two or more men in a scouting role—being substituted. It is considered tactically unsound to use single scouts or single patrolmen.

Although scouts have been employed in all wars, the best known were probably those employed in the American Indian wars of the 1860s and 1870s. They were not uniformed members of the U.S. Army but were former fur trappers who had acquired a detailed knowledge of the country and agreed to serve the Army. Occasionally Indians were used as scouts to aid the troops against an enemy tribe. Among the best-known American scouts was William ("Buffalo Bill") Cody, who was flamboyant and colourful and had a gift for showmanship, although his actual importance probably was not as great as some of the others. Another well-known scout was Christopher ("Kit") Carson, who served as a guide for John C. Frémont, the famous "pathmarker" of the American West. Not so well known but possibly of greater importance was Jim Bridger, who had spent the greater part of his life in the mountains and on the Great Plains. Another was Frank Grouard, a mystery man of whom Gen. George Crook is said to have remarked that he would rather lose half his command than Frank Grouard. *See* also BOY SCOUTS; GIRL SCOUTS.

(E. I. S.)

SCRANTON, a city of northeastern Pennsylvania, U.S., the county seat of Lackawanna county, is 134 mi. W.N.W. of New York city and 110 mi. N.N.W. of Philadelphia. It is situated in the Lackawanna river valley on the western fringes of the Pocono mountains at an elevation of 752 to 1,770 ft. above sea level. Pop. (1960) city 111,443; standard metropolitan statistical area (Lackawanna county) 234,531. For comparative population figures *see* table in PENNSYLVANIA: *Population*.

Permanent settlement of the valley dates from 1788 when it was known as Slocum Hollow. In the next few years a gristmill, a sawmill and a charcoal furnace were built along the Lackawanna (Indian for "stream that forks"), but there was little development except for names: the village was called Slocum Hollow, Unionville, Harrison and Scrantonia before becoming Scranton in honour of the New Jersey family which established the Lackawanna Iron and Coal company there in 1840. George W. Scranton (1811–61) and his brother, Selden, bought a large tract (most of the present downtown business section) for $8,000 and began to smelt iron from local ores, using an anthracite hot-blast process. Persistent efforts by the two Scranton brothers and timely financial aid by their cousin, Joseph Hand Scranton, made the venture a success. By 1850 a rolling mill, a nail factory and a steel-rail works were in operation and transportation facilities had been provided. Scranton was incorporated as a borough in 1853 and chartered as a city in 1866.

Even before the iron ore was exhausted, the iron industry was overshadowed by the anthracite industry which has had a profound impact on the history of Scranton. The ethnic diversity of the area and the rapid population growth stemmed from the need for labour for the mines and allied industries. The first wave of immigration brought skilled English, Welsh and German miners but they were soon followed by the Irish, Austro-Hungarians, Poles, Russians and Italians. The wives of these immigrants provided a labour force that made possible the development of silk, apparel and other industries; Scranton has long been noted for its production of Nottingham lace.

Like the other cities of the anthracite coal region in northeastern Pennsylvania, Scranton has been forced to readjust its economy to the decline in the anthracite industry. Coal production valued at $196,908,000, or two-thirds of the combined total of mining and manufacturing in 1948, dwindled to approximately one-fourth of the total value a decade later. During the same period the percentage of workers employed in producing anthracite dropped from 33% to 14%, creating chronic unemployment of 15% to 20% of the total labour force. The "Scranton Plan" was developed to provide more jobs through industrial expansion. In 15 years after World War II over 30 plants and 3,000,000 sq.ft. of new floor space were made possible by the combination of community contributions, private financial and state aid. These new plants were either leased or sold to industries moving into the area, and the net income was reinvested in more plants. Textile, apparel, electronics and metal products manufacturers are the leading employers.

The economic readjustment brought on by the end of the war and the decline in the anthracite industry was also eased by Scranton's natural position as the centre of a large retail, wholesale and jobbing trade. Railroads, bus lines and airlines (at the Scranton-Wilkes-Barre airport, 6 mi. S. of the city) provide communication with all urban centres in the eastern United States. Three federal highways, and the northeastern extension of the Pennsylvania turnpike and the Pan-Canada international highway system provide the road transportation network of the area.

Educational facilities include the University of Scranton, established in 1888; Marywood, established in 1915; Lackawanna Junior college, organized as the Scranton Business college in 1894, becoming a junior college in 1959; a campus of the Pennsylvania State university commonwealth campus system (of two-year colleges), opened in 1921; the state school for the deaf; and the International Correspondence school, largest institution of its kind in the world. The public library has branches throughout the city and there is a county historical society.

The fine arts form a prominent part of the city's cultural life with a ballet guild, a philharmonic orchestra and an art league. The Everhart Museum of Natural History, Science and Art, located in Nay Aug Park, provides a focus for the appreciation of the fine arts. The park also contains zoological and horticultural gardens and a model mine open to the public. The Pocono mountains provide recreational opportunities as do the many mountain lakes and state parks within the area. (R. D. WI.)

SCRAP METAL. Used metals may be as important as ores as a source of metals and alloys for industrial use. The relative importance of scrap metal increases as the supply of metal in use accumulates and the supply of ores diminishes or the average grade declines. Scrap metals are particularly important in the production of steel, copper, lead, aluminum and zinc. Smaller amounts of tin, nickel, magnesium and precious metals are also recovered from scrap.

There are two distinct kinds of metal scrap: "old scrap" is metal that has been discarded after use and is returned to the metallurgical industry for reprocessing; "new scrap" is metal from manufacturing operations that has not been used in an end product and is therefore still in the process of manufacture. Metals produced from scrap usually are known as secondary metals as distinguished from primary metals, which are produced directly from ores. No such distinction is made for steel and precious metals. Steel is produced from a combination of scrap iron, pig iron and iron ores. Old scrap is gathered from widely scattered sources by scrap-metal or junk dealers and sold to secondary smelters. New scrap is concentrated in industrial areas and is usually purchased directly by secondary smelters. Many of the larger metal producers who produce secondary as well as primary metals buy back the new scrap from manufacturers to whom they sell primary metal.

Scrap usually is processed by sorting, blending and remelting to produce alloys similar to or more complex than those from which the scrap was derived. Old scrap in particular ordinarily is contaminated with foreign materials. Organic materials such as wood, plastic, paint or fabric can be burned off. Metallic impurities that are beneficial or inert usually are retained in the secondary ingot. Undesirable metallic impurities are diluted to tolerable proportions by the addition of pure metal or are removed by refining, *e.g.*, zinc is readily removed by distillation; lead can be refined by melting with a suitable flux. The price differential between scrap and refined metal permits only simple refining methods to be used. (H. W. S. C.)

SCREAMER, the name given to the three species of birds constituting the South American family Anhimidae, of the order Anseriformes. Although distantly related to ducks and flamingos, and like them gregarious and essentially aquatic in habits, they have marked anatomical differences. Perhaps the most noteworthy are the absence of a hooklike projection of the rib, the uncinate process—a deficiency unique among modern birds, but characteristic of *Archaeopteryx* (*q.v.*) and many reptiles—and the presence of a layer of small bubbles in the skin. The intestines

YLLA FROM RAPHO GUILLUMETTE

CRESTED SCREAMERS (CHAUNA TORQUATA)

are extremely long, the toes unwebbed and the bend of the wing is armed with a pair of spurs present even in new chicks.

The three species measure from 28–36 in. in length and are much alike in general appearance. All are black or gray above and pale gray or white below. The head is rather small and the bill chickenlike; the feet are disproportionately large. One species (*Anhima cornuta*) of tropical South America south to Bolivia and the Mato Grosso has a slender, recurved spine, three to five inches in length, protruding from the forward edge of the crown. The other two species, *Chauna chavaria* and *C. torquata,* are both crested. The former inhabits northern Colombia and western Venezuela; the latter southern Brazil and Argentina south to Mendoza and Buenos Aires.

Screamers are nonmigratory and frequent marshes and inundated grasslands where they subsist on vegetable matter. At times they may be found in flocks numbering thousands. They often sing in unison, their voices resembling loud trumpeting. A singular habit is that of soaring for hours in circles at immense altitudes, often uttering melodious cries.

The two to six white eggs are deposited in a shallow nest of rushes on marshy ground. Both sexes incubate the eggs and attend the young. (E. R. Be.)

SCREW, a cylindrical or conical member with a helical groove and ridge on it. (For the screws used to propel ships and airplanes, *see* Naval Architecture: *Resistance and Propulsion;* Aircraft Propulsion: *Propellers.*) The helix is found in a wide variety

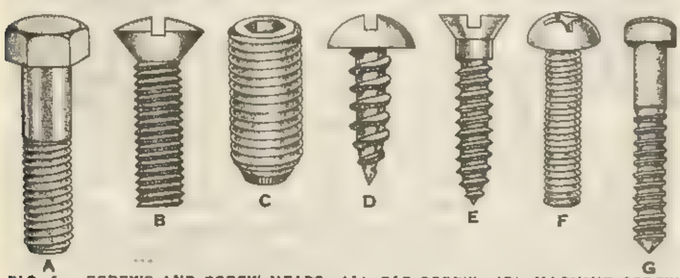
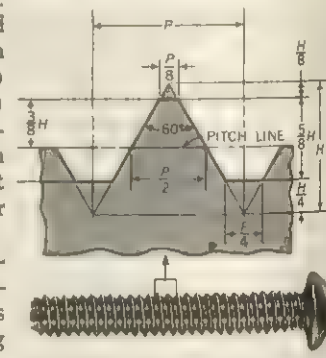


FIG. 1.—SCREWS AND SCREW HEADS: (A) CAP SCREW, (B) MACHINE SCREW WITH OVAL HEAD, (C) SET SCREW WITH HOLLOW HEAD, (D) SELF-TAPPING SCREW, (E) FLAT-HEAD WOOD SCREW, (F) PHILLIPS HEAD SCREW, (G) LAG SCREW

of applications, such as screw pumps (the first by Archimedes) and boring tools for metal, wood, soil, etc. In particular, a screw (without modifying adjective) is a threaded fastening device, without a nut; common examples are the cap screw, machine screw, set screw, wood screw, and lag screw (fig. 1). The cap screw is machine finished and made in sizes of about $\frac{3}{8}$ in. and larger; it generally has a hexagonal head, but may also be made with a hollow head such as shown on the set screw in fig. 1C.

Machine screws also are usually finished all over but are made in the smaller sizes—up to $\frac{3}{8}$ in. in diameter; various types of heads, nearly always slotted for a screw driver, are used, and patented heads, such as the Phillips head (fig. 1F), are common. Set screws are used to prevent relative motion between two parts which tend to slide over one another, as in fastening a small pulley to a shaft; they are obtainable with several different types of heads and points. Self-tapping screws, fig. 1D, are hardened screws designed to tap their own threads as they are turned into the hole; they are used to join thin sheets of metal, plastics, die castings, etc. Some types are applied with a screw driver, some with a hammer, and they are frequently made with a Phillips head. Wood screws are made in various diameters and lengths, in using the larger sizes, a hole is drilled first in order to avoid splitting the wood and to make the turning easier. A lag screw is a large wood screw, often used to fasten machinery and other heavy objects to wood. It usually has a square or hexagonal head so that it can be turned by a wrench.

Screw Threads.—A thread on the outside of a cylinder is an external or male thread; cut on the inside of a hole, as in a nut, it is an internal or female thread. The purpose in using threaded members may be: (1) to join two or more parts together; (2) to transmit power or force; (3) to obtain a mechanical advantage, that is, a large force from a small one; or (4) to transmit motion either rotary to linear or linear to rotary.



BY COURTESY OF BRITISH STANDARDS INSTITUTION B.S. 1580: 1962

FIG. 2.—SECTION (ABOVE) OF SCREW (BELOW) SHOWING BASIC FORM OF THE ISO/UNIFIED THREAD: P REPRESENTS PITCH AND $H = 0.866\ P$

Until about 1841 each manufacturer using threaded fastenings designed the threads for his own convenience. Fastening threads in the form of a sharp V had disadvantages: the sharp crests, being weak, broke off in small pieces, and the sharp root shape resulted in a high stress concentration that led to failure of the bolt. Failure of a screw or bolt resulted in delay, inconvenience, and expense because of difficulty in obtaining a replacement that would fit.

Trying to overcome these disadvantages, Sir Joseph Whitworth in 1841 proposed a fastening thread with rounded crests and roots with an angle between the sides of the thread of 55°; this became the British standard Whitworth thread used for general engineering. In 1864, in the U.S., William Sellers proposed a thread with flat crests and roots and a 60°-thread angle; this later was accepted as the U.S. standard. Toward the end of World War II a 60°-thread Unified Standard was recognized by the U.S., the United Kingdom, and Canada. This thread form (fig. 2) was accepted by the International Organization for Standardization (ISO). Publication of international standards for screw-thread systems in inch and metric units began in 1958.

Thus there is international agreement on some degree of standardization of the form of thread and general recognition of the desirability of international interchangeability. By interchangeability is meant, for example, that when a large number of bolts and nuts are made to a certain size and standard, any one of the nuts fits properly on any one of the bolts. In the Unified screw-thread standard there are several classes of fits possible. Class 1 is the loosest and cheapest, having the largest tolerances and allowances. Class 2 is about the average fit obtained by ordinary

production methods and is therefore commonest. Class 3, the closest fit and most expensive, would not be used without reason; the specified allowance between bolt and nut is zero. Intermediate classes of fits may be obtained by, for instance, using a class 3 nut with a class 2 bolt.

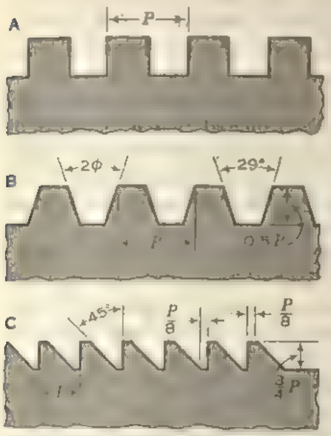
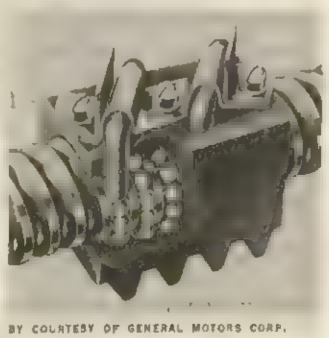


FIG. 3.—POWER SCREW THREADS: (A) SQUARE THREAD, (B) ACME THREAD, AND (C) BUTTRESS THREAD. P, PITCH; ϕ, PRESSURE ANGLE; AND 2ϕ, THREAD ANGLE

The largest diameter of a screw thread (*e.g.*, the outside diameter of an external thread) is called the major diameter; the smallest diameter (*e.g.*, the root diameter of an external thread) is called the minor diameter. The top of the external thread is known as the crest, while the crest of an internal thread mates with the root of an external thread.

There are several standard series of fastening threads. The coarse-thread series, with the fewest threads per inch for a particular size, is recommended for general use. The fine-thread series, which has more (and therefore finer) threads per inch, was developed for use in the automotive industry where strength and vibration resistance are important, and where the internal threads are in steel (not cast iron or other brittle material). The extra-fine-thread series of the American standard came about largely in response to the needs of the aeronautical industry. This thread has the maximum root diameter and strength; it is suitable for thinner walls; it is more resistant to loosening from vibration. Fastening screws with these threads are likely to be made of a heat-treated alloy for maximum capacity.

Threads for Transmitting Power.—One of the three forms of threads shown in fig. 3 is usually used for transmitting power or motion. Since the square thread must be cut on a lathe and the acme thread may be cut with dies, the square thread is more expensive to produce. Not only is the acme thread less expensive, but it permits use of a split nut with shims between the halves so that wear may be compensated for by removing shims. Ball-bearing screws (fig. 4) are used where it is advantageous to have a relatively low input effort and where arrangements with zero backlash are desirable. A worm driving a worm gear may be classified as a screw, often with a thread shape similar to the acme thread of fig. 3.



BY COURTESY OF GENERAL MOTORS CORP.

FIG. 4.—BALL-BEARING SCREW: BALLS ROLLING ON EXTERNAL AND INTERNAL THREADS COMPLETE THEIR CIRCUIT THROUGH TUBES AT TOP

The pitch P of a thread is the distance between corresponding points on adjacent thread profiles—in inches in some countries, millimetres in countries using the metric system. The number of threads per inch (or per millimetre) is the reciprocal of the pitch. The lead of a thread is the axial distance that the helix advances in one turn, or that a nut will move axially in making one turn. In a single-threaded screw, the lead is equal to the pitch; if two threads run parallel, the lead is twice the pitch (fig. 5). When high mechanical advantage is the prime requisite, a single-threaded screw is best. When high efficiency is essential, as in some worm drives, a large lead (*i.e.*, multiple threads) is necessary. An exception is the ball-bearing screw, which is said to have an efficiency of about 90% at very small leads. However, the efficiency of a screw-and-nut combination depends not upon the magnitude of the lead but upon the slope of the thread at the pitch helix, which is the tangent of the lead angle. The lead angle

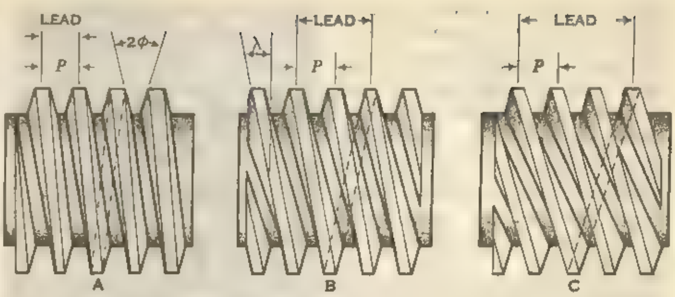


FIG. 5.—DIAGRAM SHOWING RELATION OF PITCH TO LEAD. ϕ, PRESSURE ANGLE; 2ϕ, THREAD ANGLE; AND λ, LEAD ANGLE

λ is the angle between a plane perpendicular to the axis of the screw and a line tangent to the pitch helix. The pitch helix has somewhat different definitions in different applications, but it is approximately at the mean diameter of the thread. The efficiency also depends upon the thread angle, the degree of lubrication, rubbing speed, materials, alignment, and workmanship in general. Some of these factors are accounted for by the coefficient of friction.

A value of the coefficient of friction equal to 0.05 is easily obtained in well lubricated worm drives; a value 0.15 corresponds to a situation at slow rubbing speed with the surfaces merely damp with lubricant.

Methods of Cutting Screw Threads.—The methods of cutting external threads may be described in six categories: turning, die threading, chasing, milling, rolling, and grinding.

Turning, a process still used occasionally, employs a single-point cutting tool of the same shape as the space between threads. The member to be threaded (that is, the work) is mounted in a lathe. The change gears on the lathe are arranged so that the work and the lead screw of the lathe turn with the proper relative speed to give the desired number of threads per inch on the work. If the work and lead screw turn at the same angular speed, the number of threads per inch on the work is the same as that on the lead screw; fewer threads are cut if the work turns slower than the lead screw, more if the work turns faster.

BY COURTESY OF (LEFT) NATIONAL ACME COMPANY (RIGHT) GREENFIELD TAP & DIE DIVISION OF UNITED-GREENFIELD CORPORATION

FIG. 6.—(A) THREAD-CUTTING DIE; (B) CIRCULAR THREAD CHASERS ON REVOLVING HEAD

Dies for cutting threads are made with internal cutting edges of the proper shape, size, and pitch (fig. 6A). Manual operation of thread dies is common for cutting small pipe threads on the job, but the dies are also used in machines, including automatic thread-cutting machines. On small threads and at slow speed, the die can be depended on to follow its own helical path with reasonable accuracy. On large threads and at high-production speeds, the relationship between die and work must be positively maintained.

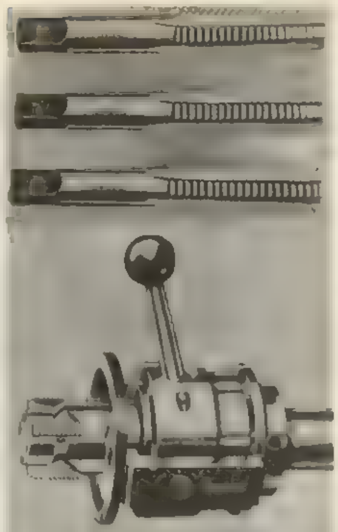
For machine cutting of threads, the rate of production is improved if the die automatically opens, allowing the work to be directly withdrawn (not unscrewed).

Chasing is the same in principle as die threading. The difference is in the configuration of the cutting edges. In one type, the cutting edges are on small cylindrical tools, as in fig. 6B; in another type the cutting edges are on flat pieces mounted with the tops of the cutting edges tangent to the root of the threads being cut.

Milling of threads is done on a standard milling machine, or one especially built for this work, using a milling cutter with a profile the same as the space between threads. In one type of machine, the cutters rotate on their axes and revolve about the work. Milling is useful in producing large, accurate threads.

Rolling produces threads by rolling the work between flat or cylindrical dies. Because the rate of production is high, the

process was first used to manufacture cheap stove bolts, but it has been improved until bolts and screws of superior quality can be produced in this manner. The die has hardened parallel ridges whose profile is the same as the desired thread profile and whose slope produces the correct lead of the screw. The shank to be threaded has a diameter approximately equal to the pitch diameter. As the blank is rolled between the dies under great pressure, it is indented and metal flows to the root of the die grooves (forming the crest of the thread). The cold working of steel during this process improves its strength, and the resulting surface finish is excellent.



BY COURTESY OF GREENFIELD TAP & DIE DIVISION OF UNITED GREENFIELD CORPORATION

FIG. 7.—TAPS: (ABOVE) HAND TAPS FOR CUTTING INTERNAL THREADS AND (BELOW) COLLAPSING TAP

The accuracy that may be obtained depends upon the accuracy of the dies and of the machine. When cylindrical dies are used, there are three of them, and of course the forces exerted by the rollers must be sufficient to cause the metal of the work to flow completely into the valleys of the roller threads.

Grinding can also be used to make threads by grinding with a wheel whose form matches the space between the threads. Grinding usually is used as a finishing operation, especially after heat treatment, for producing threads in hardened steel, and where cost is not a primary consideration.

Internal Threads.—Internal threads may be cut on a lathe or on a milling machine or ground by methods similar to those described for external threads. However, the most common method is by means of a tap (fig. 7). Hand taps may be adapted to low-speed machine work. The taper tap is used first, and it is the only one needed for a through hole. If it is desired to have good threads all the way to the bottom of a blind hole, the taper tap is followed by the plug tap and then by the bottoming tap. There are different kinds of collapsible taps suitable for high production: after they have completed the tapping, they automatically collapse and withdraw quickly. (V. M. F.)

SCREW PINE, the popular name for plants of the genus *Pandanus*, which are shrubs or trees of peculiar habit, having a main stem and a few branches at the ends of which is a tuft of long, stiff, narrow leaves closely arranged in three strongly twisted lines. The stem forms stout roots some distance above the ground; they grow obliquely downward to the soil and, because of the decay of the lower part of the stem, the plant is often supported merely by these strong prop roots. The ripe fruits are borne in very large spherical or cylindrical heads, which are often extremely hard. The genus is the principal one of the family Pandanaceae, a small family of monocotyledons, which is widely distributed through the tropics of the Old World, especially in the islands of the Malay Archipelago and of the Indian and Pacific oceans.

EDGAR AUBERT DE LA RUE

SCREW PINE (PANDANUS) GROWING ALONG THE COAST OF THE ISLAND OF EPI, THE NEW HEBRIDES, PACIFIC OCEAN

SCRIABIN, ALEKSANDR NIKOLAEVICH (1872–1915), Russian composer of piano and orchestral music, was born in Moscow on Jan. 6, 1872 (new style; Dec. 25, 1871, old style). He was trained as a soldier at the Moscow cadet school from 1882 to 1889 but studied music at the same time, taking piano lessons from S. Zverev. In 1888 he entered the Moscow Conservatory where he studied the piano with V. I. Safonov and composition with S. Taneev and A. Arenski. By 1892, when he graduated from the conservatory, he had composed the piano pieces which constitute his op. 1, 2, 3, 5, and 7. His first recital in St. Petersburg in 1894 led to a friendship with M. P. Belyaev, who not only published nearly all his later works but acted as his concert agent for recital tours in western Europe. In 1897 he married the pianist Vera Isakovich, and from 1898 until 1903 taught at the Moscow Conservatory. He then devoted himself entirely to composition, and in 1904 settled in Switzerland. After 1900 he was much preoccupied with mystical philosophy and his First Symphony, composed in that year, has a choral finale, to his own words, glorifying art as a form of religion. In Switzerland he completed his Third Symphony, *Le Divin Poème*, first given under A. Nikisch in Paris in 1905. The literary "program" of this work, devised by Tatiana Schloezer, with whom he had formed a relationship after abandoning his wife, was said to represent "the evolution of the human spirit from pantheism to unity with the universe ('The divine Ego')." Theosophical ideas similarly provided the basis of the orchestral *Poème de l'extase* (New York City, 1908) and *Prometheus*, which included a part for a colour organ projecting colours onto a screen during the performance, though this instrument was not used at the first performance (Moscow, 1911).

From 1906 to 1907 Scriabin toured the U.S. where he gave concerts with Safonov and M. Altschuler, and in 1908 he frequented theosophical circles in Brussels. In 1909 he was encouraged by the conductor S. Koussevitzky, who both performed and published his works, to return to Russia. About this time he was no longer thinking in terms of music alone; he was looking forward to an all-embracing "Mystery." This was planned to open with a "liturgical act" in which music, poetry, dancing, colours, and scents were to unite to induce in the worshipers a "supreme, final ecstasy." He even went so far as to proclaim himself a sort of messiah whose function it was to "sound the final chord of our race, reuniting it with the Spirit." He continued to give concerts, however, both in Russia and abroad, and in 1914 played under Sir Henry Wood in London. He wrote the poem of the "Preliminary Action" of the "Mystery" but left only sketches for the music. He died in Moscow on April 27 (new style; 14, old style), 1915.

Scriabin was a composer of the most sensitive, exquisitely polished piano music. His piano works include ten sonatas (1892–1913), an early concerto, and many preludes and other short pieces. Even his orchestral works create the impression of orchestrated piano music. An idolater of Chopin in his youth, he early developed a personal style. As his thought became more and more mystical, egocentric, and ingrown, his harmonic style—based on chords of fourths built up from the higher partials of the harmonic series—became ever less generally intelligible, and he failed to carry with him, as other harmonic innovators have done, enough admirers to establish his new idiom as a permanent, widely acceptable contribution to the language of music.

BIBLIOGRAPHY.—L. Sabaneev, *Skryabin* (1923); A. Swan, *Scriabin* (1923); G. Abraham, "Scriabin," in *Masters of Russian Music*, by M. D. Calvocoressi and G. Abraham (1936); B. de Schloezer "Scria-

bine," in *Musique russe*, vol. ii, ed. by P. Souvtchinsky (1953); B. Pasternak, "Ich Kannte Scriabin," in *Melos*, no. 26 (1959). (G. AB.)

SCRIBE, AUGUSTIN EUGÈNE (1791–1861), French playwright, an outstanding dramatic craftsman who dominated the Parisian stage for some 30 years, was born in Paris, Dec. 24, 1791. He chose as his medium the *vaudeville*, a popular form of satirical comedy with rhymed and sung couplets, which he soon began to transform by replacing its conventional characters with characters observed from contemporary society (*Une Nuit de la garde nationale*, 1815) and by introducing elements of the comedy of manners, as in *Le Charlatanisme* (1825). His first real comedy was *Le Mariage d'argent* (1827), but though without sung couplets, this play retained the dramatic structure of the *vaudeville*. Scribe maintained that his sole aim was to amuse; addressing the Académie Française, to which he was elected in 1836, he denied that the theatre's purpose was to depict society.

He had a facile pen: working alone or with collaborators, he wrote over 400 theatrical works, including a ballet—*Manon* (1830)—and opera libretti: *Fra Diavolo* (1830), *Robert le diable* (1831), *Les Huguenots* (1836), and *Le Juif errant* (1852). His best-known plays are *Le Verre d'eau* (1840), *Bertrand et Raton* (1833), and *Adrienne Lecouvreur* (1849). His plays, which appealed to middle-class tastes and prejudices, brought Scribe a fortune. He died in Paris, Feb. 20, 1861.

BIBLIOGRAPHY.—*Oeuvres complètes*, 76 vol. (1874–85); G. E. Legouvé, *Soixante ans de souvenirs* (1886–87), vol. ii; N. C. Arvin, *E. Scribe and the French Theatre, 1815–60* (1924). (D. KS.)

SCRIBES, a group of Jewish scholars frequently mentioned in the Gospels, often together with the Pharisees and sometimes together with the high priests. The word "scribe" is used in the English versions of the New Testament to translate the Greek word *grammateus* (plural *grammateis*), which in turn was used to render the Hebrew word *sofer* (plural *soferim*). In early biblical Hebrew *sofer* usually designated an important court official, but during the first two-thirds of the period of the Second Commonwealth (from about 400 B.C. to A.D. 1), when Judaea was successively a province of the Persian, Macedonian, and Roman empires, *sofer* meant an interpreter of Scripture. It was in this capacity that Ezra (about 400 B.C.) was called a scribe. In the early part of the 2nd century B.C. Jesus ben Sirach, the head of a school of wisdom in Jerusalem, celebrated the intellectual and moral preeminence of the *soferim* or scribes (Ecclus. 38:24–39:11).

During the last three centuries before the beginning of the Christian era the *soferim* were the professional interpreters and teachers of biblical law and ethics. Since the Jewish community enjoyed cultural autonomy under foreign rule, it was permitted to govern itself by its own constitution (called *politeia* in the ancient Greek sources), namely the Pentateuch. However, the Mosaic laws required interpretation and expansion to meet the needs of this later age and the work of interpretation was undertaken by the *soferim*. The two leading religious-political parties, the Pharisees and Sadducees, both had their professional scholars or scribes, but as the Pharisees enjoyed greater influence among the Jews as a whole, it was the Pharisaic scribes whose interpretation of the law was considered authoritative by later rabbinic scholars (*see* JEWISH SECTS DURING THE SECOND COMMONWEALTH). It is the Pharisaic scribes who are meant in most of the Gospel references to *grammateis*. It should be noted also that the Gospels treat this word as synonymous with *nomikoi*, "jurists," and *nomodidaskaloi*, "teachers of the law" (*i.e.*, biblical law). The chief Jewish legislative and judicial body meeting in Jerusalem from about 200 B.C. to A.D. 70, known as the great Sanhedrin (*q.v.*), was made up of scribes of the Pharisaic party as well as the scribes and priests of the Sadducean party.

Which party was in the majority throughout this period is not known, but sources indicate that on certain matters, the opinions of the Pharisaic scribes prevailed. However, there was a difference in the usage of the word "scribes" between the Jewish and Christian writings of the first two centuries A.D. The rabbinic sources of this period apply the word "scribes" (*soferim*) to the biblical interpreters and jurists of the period between Ezra and their own time, while the Pharisaic scholars or rabbis of their own period are generally called sages (*hakamim*), though they performed the same functions as the *soferim*. The latter term was usually applied in the Hebrew writings of the first two centuries A.D. to less eminent persons, such as writers of legal documents, copyists, notaries, and teachers of elementary schools. When the Gospels speak of scribes (*grammateis*) as contemporaries of Jesus, they mean the jurists and teachers who had earlier been called *soferim* but were now generally called *hakamim*.

The chief functions of the scribes throughout the greater part of the period of the Second Commonwealth were: (1) to interpret the biblical law and to provide new legislation by the construction of the written text or on the basis of the tradition of the elders (*i.e.*, the oral law); (2) to give instruction in all fields of knowledge relevant to the study of Scripture; (3) to act as legal advisers to judges and those who administered the law.

See Emil Schürer, *Geschichte des jüdischen Volkes im Zeitalter Jesu Christi*, 3rd ed., vol. ii, pp. 312–328 (Eng. trans., *History of the Jewish People in the Time of Jesus*; 1961); G. F. Moore, *Judaism*, vol. i, pp. 39–40. (RA. MS.)

SCRIBNER, the name of a family of U.S. publishers, descended from settlers in Norwalk, Conn., in the 1600s; the name, originally spelled Scrivener, was changed to its present form after 1742. Charles Scribner (1821–1871), founder of the firm, was born in New York City on Feb. 21, 1821, and was graduated from the College of New Jersey (Princeton) in 1840. He began publishing in New York as a partner of Isaac D. Baker in 1846. Upon his death, in Lucerne, Switz., on Aug. 26, 1871, the business was carried on by his three sons and in 1878 the firm was renamed Charles Scribner's Sons.

From the beginning the Scribner list was distinguished for its theological and philosophical works; in the latter part of the life of the first Charles Scribner the firm also began to publish reprints and translations of English and European works. The second Charles Scribner (1854–1930), president of the firm from 1879 to 1928, further diversified the list, and under the Scribner imprint appeared works by many of the most eminent writers of the time, including, among Americans, Henry James, George W. Cable, and Theodore Roosevelt, and among British authors, Robert Louis Stevenson, George Meredith, Rudyard Kipling, and John Galsworthy. Among the periodicals published by the firm were *Scribner's Monthly*, founded 1870, and *St. Nicholas*, founded 1873, both sold in 1881 to the Century Company (*Scribner's Monthly* being renamed *The Century Illustrated Monthly Magazine*); and *Scribner's Magazine* (1887–1939).

See Roger Burlingame, *Of Making Many Books* (1946), a centennial history of the firm; G. W. Schuyler, *Colonial New York* (1885), vol. ii, for the genealogy of the Scribner family.

SCRIPPS, EDWARD WYLLIS (1854–1926), U.S. newspaper publisher, who inaugurated the newspaper "chain" in the U.S., was born on a farm near Rushville, Ill., June 18, 1854. Educated in the district school and a private one conducted by his half sister Ellen, Scripps in 1872 joined the staff of the *Detroit Tribune*, then partly owned and managed by his half brother, James Edmund Scripps, who in 1873 started the *Detroit Evening News*, on which Edward was a reporter and finally city editor. He left his job in 1877 to tour Europe with his half brother George, and on his return began the *Cleveland Penny Press* in 1878. Buying or establishing papers, first in the Midwest and later on the West Coast, Scripps before his death expanded his chain to 34 newspapers in 15 states. (*See* NEWSPAPER: *United States: "Chains" and Consolidations*.) His papers were low-priced (originally penny papers), written for the "95 per cent," *i.e.*, the common people, politically independent, liberal, and prolabour.

In 1897, with George and Milton Alexander McRae, who were partners in the chain, he organized the Scripps-McRae Press Association, ultimately combined with another independent service as the United Press. He also organized the Newspaper Enterprise Association, the first newspaper syndicate created to serve a chain of newspapers. In 1922 he transferred his interests to his son Robert. Scripps died on board his yacht in Monrovia Bay, Liberia, on March 12, 1926.

BIBLIOGRAPHY.—Gilson Gardner, *Lusty Scripps* (1932); N. D. Cochrane, *E. W. Scripps* (1933); and *Damned Old Crank* (1951), a collection of Scripps's autobiographical writings, edited by Charles R. McCabe.

SCROFULA, the general name formerly given to tuberculosis (*q.v.*), "scrofulous," "strumous," and "tuberculous" being nearly interchangeable. The particular characteristics associated with "scrofula" have, therefore, varied at different periods when the real nature of the disease was misunderstood; but essentially what was meant was tuberculosis of the bones and lymphatic glands, especially in children, and it is in this sense that the word survives. The old English popular name was "king's evil" (*q.v.*), so called from the belief that the sovereign's touch could effect a cure.

SCROGGS, SIR WILLIAM (*c.* 1623–1683), lord chief justice of England from 1678 to 1681, presided over the trials of those accused of complicity in the "Popish Plot" of 1678 to put the Roman Catholic James, duke of York (later James II), on the throne. He was born at Deddington in Oxfordshire, allegedly the son of a butcher, but probably the child of a prosperous grazier. Educated at Oxford University and at Gray's Inn, he fought briefly for Charles I early in the Civil War. Called to the bar in 1653, he practised during the Protectorate. A man of broad wit and fond of good living, he found favour with Charles II, was knighted in 1665, and was made a judge of the Common Pleas in June 1676. He became chief justice of the King's Bench in May 1678.

Presiding over the "Popish Plot" trials, Scroggs completely trusted the revelations of Titus Oates (*q.v.*) and welcomed the verdicts of guilty on the accused Roman Catholics. Although he harried them with execration of their faith, he did not infringe the rules of evidence. In July 1679 he dismissed the flimsy case against Sir George Wakeman, Queen Catherine's physician. In subsequent trials he convicted more "plotters," but his acquittal of Wakeman (and, by implication, of the queen) led to charges that Scroggs had been bribed by the court. He was deluged with public abuse; only the House of Lords and a grateful king shielded him from impeachment by the House of Commons in January 1681. His unpopularity forced Charles to remove him from the bench in April 1681 with a handsome pension. Scroggs died in London on Oct. 25, 1683.

Scroggs has been posthumously condemned by those who accepted contemporary Whig propaganda, but his worst excesses were committed on behalf of the Whigs. Although he displayed great passion on the bench, legal historians have found nothing incorrect in his application of the existing harsh law of treason.

See F. S. Ronalds, *The Attempted Whig Revolution of 1678–1681* (1937); A. Havighurst, "The Judiciary and Politics in the Reign of Charles II," *Law Quarterly Review*, vol. lxvi (1950). (H. G. Ro.)

SCROPE (originally THOMSON), **GEORGE JULIUS POULETT** (1797–1876), English geologist and political economist, best known for his study of volcanoes, was born in London on March 10, 1797, the second son of John Poulett Thomson. He assumed the surname of Scrope on his marriage in 1821 to the daughter of William Scrope. Educated at Harrow and St. John's college, Cambridge, he visited Naples, Italy, as an undergraduate in 1816–17, where his interest in volcanoes was stimulated by the actions of Vesuvius. In 1821 he examined the extinct volcanoes of the Auvergne and collected material for his work *On the Geology and Extinct Volcanoes of Central France* which he published in 1827.

Scrope commenced his researches at a time when the doctrines of Abraham Gottlob Werner (*q.v.*) were still in the ascendant, but his studies were soon to play a part in their overthrow. His first work, *Considerations on Volcanoes* (1825), is to be regarded as the earliest systematic treatise on volcanology, being the first attempt to frame a satisfactory theory of volcanic action and to show the part volcanoes have played in the earth's history. He early appreciated the important part played by water in igneous action and he effectively disposed of the elevation theory of craters.

In his work in the Auvergne he demonstrated the volcanic origin of the basalts and the formation of the valleys of the region by river action. He was elected a fellow of the Royal society in 1826.

Soon after his marriage, Scrope settled at the family seat of Castle Combe, Wiltshire, eventually devoting his attention largely to social and political questions. He was a member of parliament from 1833 until 1868 and published a long series of pamphlets and reviews advocating free trade and social reforms, especially with regards to the poor law. Scrope died at Fairlawn, near Cobham, Surrey, on Jan. 19, 1876. (C. E. T.)

SCROPHULARIACEAE, the figwort or snapdragon family of flowering plants, containing about 250 genera and 5,000 species, found on all continents (except Antarctica) but most numerous in temperate areas. The family is especially well represented in western North America, where more than 300 species occur in California alone, and in South Africa, from which about 780 species are known. Many members of the family are cultivated for their showy blooms, which in their diversity of size, colour and form rival those of the orchids. These ornamentals, representing about 65 genera, include snapdragon (*q.v.; Antirrhinum*), foxglove (*q.v.; Digitalis*), slipperflower (*Calceolaria; q.v.*), princess tree (*Paulownia*), speedwell (*Veronica*), mask-flower (*Alonsoa*), coral plant (*Russelia*), beardtongue (*Penstemon*), *Hebe, Linaria, Angelonia* and *Torenia*. Few additional Scrophulariaceae are of economic importance. The drug digitalis (*q.v.*) is the dried leaf of the common foxglove (*Digitalis purpurea*). Some members of the family are troublesome weeds, especially butter-and-eggs (*Linaria vulgaris*), mullein (*q.v.; Verbascum*) and several speedwells (*Veronica*). Witchweeds (*q.v.; Striga*) are serious parasites of various crops, mostly grasses, in warmer parts of the old world; *S. asiatica* has become established in southeastern United States. The orange-yellow roots of *Escobedia grandiflora* are sold in Colombian markets for colouring food. *Paulownia tomentosa* (princess tree) is a timber tree in Japan. The common garden

FROM L. H. BAILEY, "MANUAL OF CULTIVATED PLANTS"; REPRODUCED BY PERMISSION OF THE MACMILLAN CO.

SOME SPECIAL CHARACTERISTICS OF THE FAMILY SCROPHULARIACEAE

snapdragon (*Antirrhinum majus*) has been much used as an experimental plant in genetics.

Most Scrophulariaceae are annual, biennial or perennial herbaceous plants, but many, especially in warmer regions, are shrubs. Among the few trees are *Halleria lucida* of eastern and southern Africa, growing 45 ft. tall and 10 in. in diameter, and *Hebe parviflora* of New Zealand, about 25 ft. tall and 2 ft. in diameter. Many members of the family are parasitic, attaching themselves to the roots of other plants by means of suckers. Some, such as Indian paintbrush (*Castilleja*), witchweed (*Striga*) and eyebright (*Euphrasia*), are partial parasites that contain chlorophyll and thus can produce their own food through photosynthesis, although drawing water and minerals from the host. Others (*Harveya, Hyobanche*), lacking chlorophyll, are total parasites. Leaves of Scrophulariaceae may be alternate, opposite or rarely whorled on the stem and are smooth-edged to deeply toothed or lobed, or, in some aquatic species, divided into threadlike segments. *Hebe* has evergreen leaves. Many Scrophulariaceae produce numerous small seeds: in a capsule of hedge hyssop (*Gratiola virginiana*) nearly 1,800 seeds may be developed, and a single monkey flower plant (*Mimulus ringens*) can bear 300,000 seeds. In contrast, *Lagotis, Tozzia* and some New Zealand *Euphrasia* produce a single seed in each cell of the fruit. Seeds of Scrophulariaceae may be winged, almost smooth, reticulate, tuberculate, ridged, etc., and are spread mainly by wind. Some genera (*e.g., Scoparia, Penstemon*), produce seeds that may be ingested by animals and expelled later, growing from the droppings; those of certain species of *Veronica* and *Melampyrum*, which develop an oil-secreting gland, are dispersed by ants that feed on the oil; and seeds of aquatic and marsh species (*e.g., Gratiola, Limosella*) may be carried in water or in mud adhering to birds and mammals.

Flowers of Scrophulariaceae are borne singly along the stem or are grouped into clusters. The often two-lipped corolla consists of five united petals, with the upper two sometimes fused to appear as one (*Veronica*). The corolla lobes are similar to strongly dissimilar. The united portion of the petals (tube) is very short (*Veronica*) to long and either flaring (*Digitalis*) or narrow (*Buchnera*), reaching a length of six inches in *Escobedia guatemalensis*, the species having the largest flowers in the family. The four or five sepals are separate to almost fully united. The stamens are usually four, but sometimes two (*Veronica*) or rarely five (*Capraria*, some *Verbascum*). The ovary is two-celled, superior and with axile placentae. Most members of the family are pollinated by bees and flies; some species, however, are adapted for pollination by hovering moths (*Harveya, Penstemon albidus*), by butterflies (*Buchnera, Sophronanthe*) or by hummingbirds (*Castilleja, Macranthera*). Nearly all Scrophulariaceae have a capsular fruit, but in *Halleria, Teedia* and *Leucocarpus* the fruit is a berry and in the South African tribe Selagineae the fruit is a schizocarp.

The largest genera of the family are *Calceolaria, Pedicularis* (louseworts), *Verbascum, Veronica* and *Penstemon*. A few fossils (*e.g., Paulownia*-like leaves) have been attributed—for the most part doubtfully—to the family.

Certain closely related families, especially the broomrape family (Orobanchaceae), are sometimes included in the Scrophulariaceae. The genus *Paulownia* is often referred to the trumpet vine family, combining as it does the arboreal habit of the Bignoniaceae with seed characters of the Scrophulariaceae.

See F. W. Pennell, "The Scrophulariaceae of Eastern Temperate North America," *Academy of Natural Sciences of Philadelphia, Monographs no. 1* (1935).

(J. W. Tr.)

SCRUBBIRD, the name applied to two species of *Atrichornis*, an Australian genus of primitive perching birds that comprises the family Atrichornithidae. The noisy scrubbird (*A. clamosa*), of western and southwestern Australia, the male of which is noisy and imitative of other birds, is brown above, each feather barred with a darker shade. The throat and belly are reddish-white and there is a black patch on the throat; the flanks are brown. This species, which inhabits the thickest scrub of brambles and dense undergrowth, is in danger of becoming extinct. The rufous scrubbird (*A. rufescens*), from the forests of New South Wales, is dark brown except for an orange-brown tinge to the underparts.

SCUDÉRY, GEORGES (1601–1667) and **MADELEINE** (1607–1701) **DE,** brother and sister, established themselves in Paris as writers and critics. Both were born at Le Havre: Georges on Aug. 22, 1601, and Madeleine on Nov. 15, 1607. Georges settled in Paris in 1630, having abandoned his military career to turn to poetry and drama. When Madeleine joined him in 1638, he had already made a name for himself as a bitter protagonist in the quarrel of *Le Cid* (*see* Corneille, Pierre). He wrote a letter to the Académie Française criticizing the new play, and his own tragicomedy, *L'Amour tyrannique* (1638), was patronized by Richelieu in opposition to Corneille. Because of her brother's literary reputation, Madeleine was at once admitted to the Rambouillet coterie, the leading Parisian salon of the day, where she met almost all the more noteworthy figures in contemporary French society and literature. (*See* Rambouillet, Cathérine de Vivonne, Marquise de.)

Before long Madeleine was producing works of her own; her lengthy heroic romances, such as *Ibrahim, ou l'illustre Bassa* (4 vol., 1641), *Artamène, ou le Grand Cyrus* (10 vol., 1649–53), *Clélie* (10 vol., 1654–60), were soon the delight of all Europe. They appeared under the name of her brother, who almost certainly collaborated in them. He may have done little more than read the proofs; on the other hand, it is likely that *Almahide, ou l'esclave reine* (8 vol., 1660–63) was written by him. The two Scudérys certainly spent most of their lives together. When, in 1644, Georges went to Marseilles for three years, as governor of the fortress Notre-Dame-de-la-Garde, his sister accompanied him. In Paris they shared the same address until 1654, when Georges' political opinions as a Frondeur caused his banishment from the capital.

Meanwhile the name of Paul Pellisson, the brilliant man of letters, was being increasingly associated with that of Mlle de Scudéry. She met him in 1652 and formed with him a close and passionate friendship terminated only by his death in 1693. At about the same time she established her own salon, the *Société du Samedi*, which soon became a well-known hive of *préciosité*.

Although Georges was an industrious and prolific writer, only *L'Amour tyrannique* and an epic, *Alaric* (1654), have survived. He was elected to the Académie in 1650 and was always a prominent figure in Paris; but it is his sister who has kept the name of Scudéry alive. Her novels are landmarks in the history of the French novel. Although they have classical or Oriental personages as heroes and heroines, their language and action reflect the fashionable ideas of the time. Several of her contemporaries can be identified in these *romans à clef*. In *Clélie*, Herminius represents Pellisson; Scaurus and Lyriane are Paul Scarron and his wife (later Mme de Maintenon); and in the description of Sappho in volume X of *Le Grand Cyrus* the author paints herself, whence her pseudonym "Sapho." It is in *Clélie* that the famous *Carte de Tendre* appears, a sort of "Map of Love's Land," in which the river of Inclination waters the villages of *Billet Doux, Petits Soins,* and so forth. This literary amusement offers an excellent example of the *esprit précieux*. The novels are full of psychological observation, and reveal the author as a mistress of dialogue. Everything she wrote was in the most decorous spirit, which contrasts rather strongly with the general arrogance of her brother.

After six years in Normandy, Georges returned to the capital, where he died on May 14, 1667. Madeleine spent her later years publishing numerous volumes of conversations. She died in Paris on June 2, 1701.

Bibliography.—H. C. Lancaster, *French Dramatic Literature in the 17th Century*, part II (1932); C. Aragonnès, *M. de Scudéry, reine du Tendre* (1934); D. McDougall, *M. de Scudéry; Her Romantic Life and Death* (1938); G. Mongrédien, *M. de Scudéry et son salon* (1946).

SCULLING: *see* Rowing.

SCULPTURE may be broadly defined as the art of representing observed or imagined objects in solid materials and in three dimensions. There are two general types: statuary, in which figures are shown in the round, and relief, in which figures project from a ground; for a further discussion of the distinction *see* Relief. Sculpture is one of the oldest and most widespread

of the arts. It is also one of the most difficult, for it requires physical labour, patience and complete control of the materials involved for a long period of time.

One approach to the subject is that of techniques and materials. For a general discussion of the methods by which pieces of sculpture are created, *see* SCULPTURE TECHNIQUES. The specific technical requirements and possibilities of the various materials are discussed in greater detail in such articles as IVORY CARVING; JADE AND OTHER HARD STONE CARVINGS; TERRA COTTA; WAX FIGURES; and WOOD CARVING. Most of these articles also give a brief historical summary of sculpture in the material under discussion.

Another approach is historical, tracing the development of sculpture in a country or in a civilization. In regard to the sculpture of the far east, for example, there are separate entries on CHINESE SCULPTURE and JAPANESE SCULPTURE; Indian sculpture is discussed in the article INDIAN ART (which also covers the art of Ceylon); the sculpture of the Indonesian archipelago is treated in INDONESIAN ARCHAEOLOGY AND ART. In the middle east, where Islam is in general unsympathetic to sculpture, little of it exists; *see* ISLAMIC ART. The sculpture of the civilizations of the ancient middle east is of archaeological as well as artistic interest and is discussed in the articles on the civilizations. *See*, for example, EGYPT: *Ancient Civilization and Culture: Arts and Crafts;* BABYLONIA AND ASSYRIA: *Art and Art Objects*. For the Minoan civilization of Crete *see* CRETE: *Archaeology: Sculpture and Modeling; see* also AEGEAN CIVILIZATION. For primitive sculpture *see* PRIMITIVE ART.

Greek sculpture forms a major part of the subject matter of the article GREEK ART; Roman sculpture is likewise discussed in ROMAN ART. This article is devoted to the sculpture of western civilization after the classical period and is divided into the following main sections:

See also biographies of sculptors discussed in the text, and articles on the art of various periods: BAROQUE ART; RENAISSANCE ART; MODERN ART; etc.; *see* also references under "Sculpture" in the Index. (X.)

I. ROMANESQUE SCULPTURE

With the dissolution of the Roman empire in the west, cultural hegemony over the Christian world passed to the Eastern empire, with its centre at the court in Byzantium (Constantinople, now Istanbul). Byzantine art, in contrast to Greek art (which as it developed became increasingly naturalistic), endeavoured to transfigure the forms of physical nature into something supernatural and transcendental, and it evolved styles of elaborate flatness, skilfully repudiating the material and the corporeal, which suggested the absolute dominion of the spiritual. As a result the art of statuary, an art of material mass, almost ceased to exist; sculpture was limited to ivory carving and goldsmith work.

In the west sculpture was generally less suspect. From the beginning the Christian Church marked the graves of apostles and martyrs with stone or bronze memorials. The character of these did not at first differ from other contemporary memorials except by the occasional use of Christian symbols. The vast number of sarcophagi in Rome, executed for the most part after the reign of Constantine I, illustrate the development of this Hellenistic art, in which classic forms and drapery, not without dignity and grace, are united intricately with the conventional decorative motives of orientalized Antioch or the genre detail of Alexandria. After the 5th century they share the technical decadence of Rome. *See* also EARLY CHRISTIAN ART.

The revival of the arts under Justinian I found a wide expression in intimate sculpture such as ivory carvings and architectural ornament. There was a sporadic appearance of monumental relief sculpture, such as that on the Anglo-Saxon crosses of England. During the Carolingian renaissance of the 9th century there was even a revival of statuary, although none of it has survived. A truly sculptural instinct distinguishes the relief scenes of the great bronze doors on the cathedral of Hildesheim in Germany (early 11th century), where the lively figures are isolated in clear, frieze-like compositions and emerge three-dimensionally, with their heads almost in the round; the vigour and virility of this exuberant masterpiece contrast strikingly with the sophistication of Byzantine art.

The bronze Easter column set up, supposedly by St. Bernward, in Hildesheim early in the 11th century, is a modest prophecy of the monumental spirit which would distinguish the sculptural decoration of the new monastic buildings rising in much of western Europe. Developed in the abbey doorways and on the pillars and capitals of cloisters, where the sculptor had to learn anew the technique of stone carving and of rendering the human figure, this spirit gradually entered once more the art of the occident.

During the 11th century more and more churches were constructed in the Romanesque style, the massive forms of which are another indication of this sculptural instinct. Romanesque sculpture culminated in France in the great semicircular relief compositions over church portals, called tympanums. The example at Moissac, which represents the Apocalyptic vision with the 24 elders, is a particularly brilliant demonstration of how devices of style can so transform the objects of nature that they seem entirely purged of terrestriality. All the forms are suspended in a predominating plane that denies physical space. Differences in scale are masterfully exploited: the tiny figures of the elders are a foil to the looming image in the centre. With great consistency, every detail has been subjected to a process of stylization that produces rhythmic patterns in the drapery, hair and feathers. The central figure is so flattened as to appear disembodied, while the two towering angels have been so attenuated that their bodies have lost all mass.

The astonishing variety which these master sculptors achieved within the confining principles of Romanesque style can be illustrated on the one hand by the tympanum of Burgundy, such as the spectral "Last Judgment" at Autun or the "Pentecost" at Vézelay, and, on the other, by the less visionary sculpture of Provence, such as that of St. Trophime in Arles or of the church in St. Gilles, which retain many of the forms and characteristics of classical antiquity. (*See* also ROMANESQUE ARCHITECTURE.)

Another sculptural form that reappeared in Europe during the latter part of the Romanesque period was sepulchral sculpture, in which a sculptured figure of the deceased was cut or molded on top of a sarcophagus or on the sepulchral slab set into the floor of an abbey or cloister. *See* SCULPTURE, SEPULCHRAL.

II. GOTHIC SCULPTURE

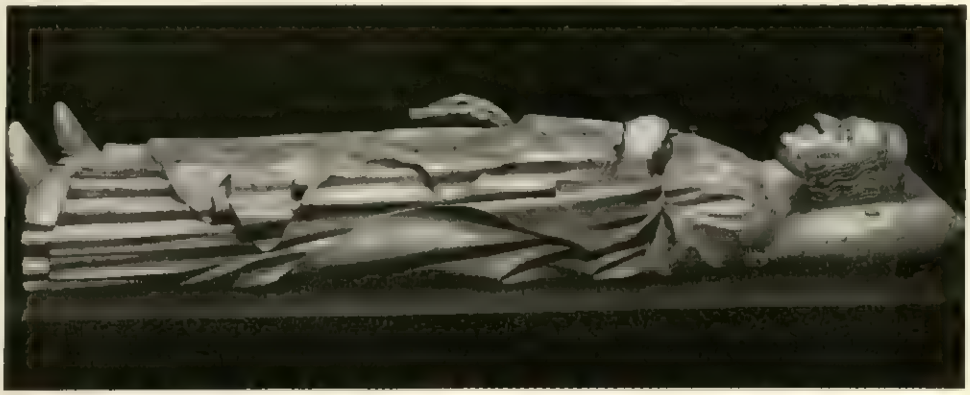
1. The Development of the Gothic Style.—The masterpieces of the rather forbidding art of the Romanesque, which is so intransigent in its rejection of earthly things, are in fact to be

The Apocalyptic vision with the twenty-four elders, tympanum of the church of St. Pierre, Moissac, France; *c.* 1135

Statue of Christ (Le Beau Dieu), south portal of Chartres cathedral, France; *c.* 1210–50

ROMANESQUE AND GOTHIC SCULPTURE

Gilded statue of the Virgin Mary (La Vierge Dorée), south transept portal of Amiens cathedral, France; 13th century



The tomb of Charles V in the abbey of St. Denis, France, by Andre Beauneveu; 14th century

Detail showing three prophets (Jeremiah, Simeon and St. John the Baptist), north portal of Chartres cathedral, France; 14th century

The Virgin Mary (Nuestra Señora la Blanca), portal of the west transept of León cathedral, Spain; mid-13th century

PHOTOGRAPHS, (BOTTOM CENTRE) BOURSAT—RAPHO GUILLUMETTE, (BOTTOM RIGHT) CLIXÉ ARXIV. "MAS," (OTHERS) GIRAUDON

SCULPTURE

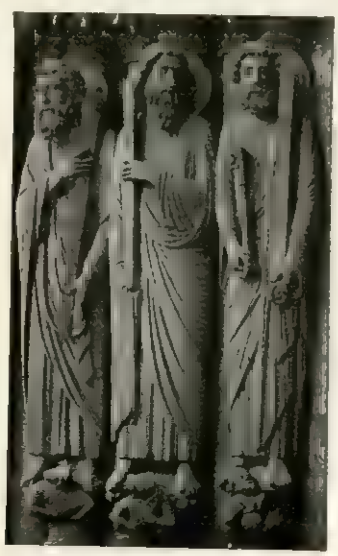
"The Last Judgment," tympanum on the west portal of St. Lazare cathedral, Autun, France; 1120–78

The marble pulpit in the baptistery of Pisa, Italy, by Niccolò Pisano; completed in 1259

Madonna and Child, in the Campo Santo, Pisa; by Giovanni Pisano

The marble tabernacle of the chapel of Or San Michele, Florence, by Andrea Orcagna (*c.* 1308–*c.* 1368)

Detail from the marble pulpit in the baptistery of Pisa by Niccolò Pisano



Detail showing statues of three saints, south portal of Chartres cathedral; 14th century

Detail of the "Well of Moses," pedestal completed *c.* 1400 by Claus Sluter. In the Dijon museum, France

ROMANESQUE AND GOTHIC SCULPTURE IN FRANCE AND ITALY

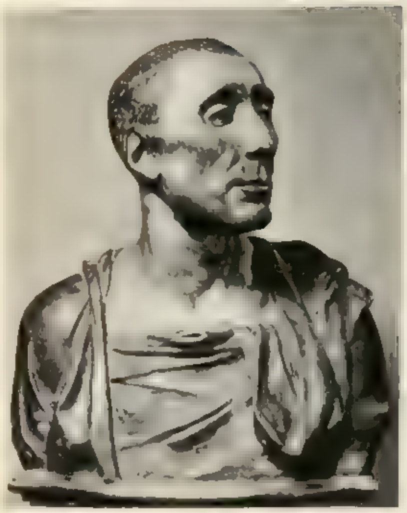
PHOTOGRAPHS, (TOP, BOTTOM RIGHT) GIRAUDON, (BOTTOM CENTRE) JEAN ROUBIER, (OTHERS) ALINARI

The bronze east doors on the baptistery of San Giovanni, Florence, Italy, by Lorenzo Ghiberti (1378–1455). The panels represent in relief scenes from the Old Testament. Each portal is surrounded by statuettes of scriptural personages, and the door frame is ornamented with a border of foliage, flowers, birds and animals in Renaissance style

"GATES OF PARADISE," FLORENCE

PHOTOGRAPH, SCALA, FLORENCE

SCULPTURE



"Niccolo da Uzzano" by Donatello (c. 1386–1466). In the Museo Nazionale, Florence

Tabernacle with "Madonna and Child" by Luca della Robbia (c. 1399–1482). In the Museo Nazionale, Florence

Panel of the "Singing [illegible] a bas-relief by Luca della Rob[illegible] cathedral at Florence. In the Museo de[illegible]omo, Florence

"David" by Donatello, the first large-scale, free-standing nude statue of the Renaissance, executed in bronze about 1430. In the Museo Nazionale, Florence

"Creation of Eve" by Jacopo della Quercia (c. 1374–1438). On the door of the church of S. Petronio, Bologna

"Mary Magdalene," wood sculpture executed c. 1455 by Donatello. In the Florence baptistery

DONATELLO AND THE RENAISSANCE

PHOTOGRAPHS, (TOP LEFT, TOP CENTRE) ALINARI, (BOTTOM LEFT, BOTTOM RIGHT) GIACOMO BROGI DI LAURATI, (BOTTOM CENTRE) POPPI

"The Nativity," five figures forming a presepio; painted terra-cotta statues from the workshop of Antonio Rossellino (1427–79). Metropolitan Museum of Art, New York city

"Virgin and Child" from the School of Auvergne, France, 12th century; statue of polychromed oak. Metropolitan Museum of Art, New York city

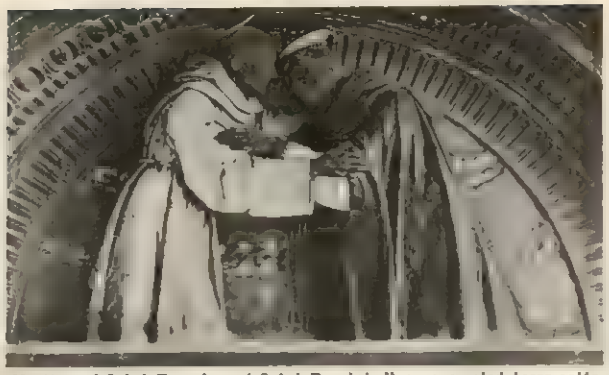
"Virgin and Child" by Andrea del Verrocchio (1435–88); polychromed and gilded terra-cotta relief executed c. 1470. Metropolitan Museum of Art, New York city

"St. Michael the Archangel" by Andrea della Robbia (1435–1525); glazed terra-cotta relief executed c. 1475. Metropolitan Museum of Art, New York city

SCULPTURES AND RELIEFS OF THE 12TH AND 15TH CENTURIES

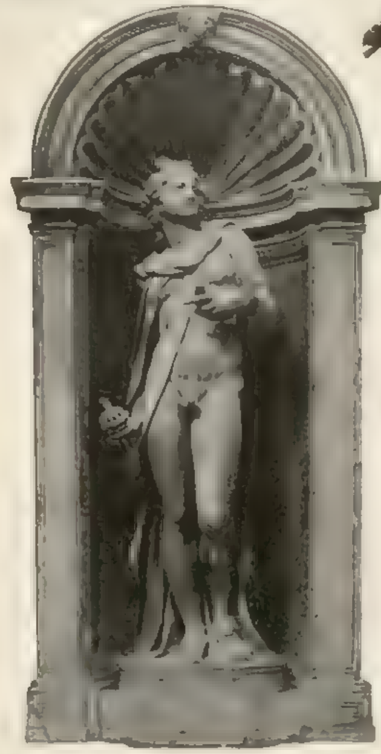
BY COURTESY OF (ALL) THE METROPOLITAN MUSEUM OF ART, (TOP LEFT) GIFT OF J. PIERPONT MORGAN, 1916, (TOP RIGHT) KENNEDY FUND, 1911, (BOTTOM LEFT) DICK FUND, 1960, (BOTTOM RIGHT) ROGERS FUND, 1909

SCULPTURE



"Meeting of Saint Francis and Saint Dominic," an enameled terra-cotta lunette by Andrea della Robbia, executed between 1490 and 1495 in the loggia of the hospital of S. Paolo, Florence

"Rape of the Sabines" by Giovanni da Bologna (1529–1608). In the Loggia dei Lanzi, Florence



"Apollo," one of four statues by Jacopo Sansovino (1486–1570) surmounting the arches of the Loggetta dei Cavalieri, Venice

"Herakles and Ant ," a bronze statuette by Anton ollaiuolo (c. 1430–98). In the rgello, Florence

"David," bronze statue by Andrea del Verrocchio, Museo Nazionale, Florence

Detail from "La Pietà" terra-cotta group by Niccolò dall'Arca (c. 1435–94), executed 1463. In the church of Sta. Maria della Vita, Bologna

"Child with Dolphin," bronze fountain figure by Andrea del Verrocchio (1435–88). In the Palazzo Vecchio, Florence

RENAISSANCE SCULPTURE

BY COURTESY OF (BOTTOM CENTRE) ENTE PROVINCIALE PER IL TURISMO, BOLOGNA; PHOTOGRAPHS, (TOP LEFT) THE MANSELL COLLECTION, (BOTTOM LEFT) F. BRUCKMANN, (ALL OTHERS) ALINARI

"Madonna and Child," glazed terra-cotta relief by Andrea della Robbia (1435–1525). Metropolitan Museum of Art, New York city

"Ange[illegible] Annunciation," statue from a group of pai[illegible]a cotta by Matteo Civitali (1436–1501). [illegible]politan Museum of Art, New York city

"Blessed Ludovica Albertoni," marble effigy by Giovanni Lorenzo Bernini (1598–1680). Church of San Francesco a Ripa, Rome

RENAISSANCE AND BAROQUE SCULPTURES

SCULPTURE

"Two Evangelists" by Jean Goujon (c. 1510?–1568?). In the Louvre

"David" by Michelangelo (1475–1564). The colossal marble statue was executed between 1501 and 1504. In the Academy of Fine Arts, Florence

"Victory" by Michelangelo. In the Palazzo Vecchio, Florence

"Diana and the Stag," a fountain designed by an unidentified artist for the chateau d'Anet. Now in the Louvre

"Madonna and Child" (1501-04) by Michelangelo. In the Church of Notre Dame, Bruges

SCULPTURE IN ITALY AND FRANCE

"Mercury" by Giovanni da Bologna (1529–1608). In the Museo Nazionale, Florence

PHOTOGRAPHS, (TOP LEFT, TOP RIGHT, BOTTOM RIGHT) ALINARI, (TOP CENTRE) GIRAUDON, (BOTTOM LEFT) FEDERICO ARBORIO MELLA, (BOTTOM CENTRE) COLLECTION ARCHIVES PHOTOGRAPHIQUES

"Deposition," marble relief by Jean Goujon (c. 1510?–68?), French. In the Louvre

"La Pietà" by Michelangelo (1475–1564). In St. Peter's, Rome

"L'Infanta Eleonora d'Aragona" by Francesco da Laurana (1425?–1502), Italian. In the Museo Nazionale, Palermo

"Madonna and Child with St. Anne" by Francesco da Sangallo (1494–1576), Italian. In Or San Michele, Florence

"Perseus With the Head of Medusa" by Benvenuto Cellini (1500–71). In the Loggia dei Lanzi, Florence

"Fontaine des Innocents" by Jean Goujon. In Paris

Unfinished marble by Michelangelo. In the Accademia di Belle Arti, Florence

RENAISSANCE AND MANNERIST SCULPTURE

PHOTOGRAPHS, (TOP, BOTTOM CENTRE) GIRAUDON, (OTHERS) ALINARI

SCULPTURE

"Milo of Crotona" by Pierre Puget (1622–94). In the Louvre

"St. Peter" by Joseph Anton Feuchtmayer (1696–1770). In the Salem monastery, Germany

"Hercules" by Pierre Puge In the Louvre

BAROQUE SCULPTURE

"Apollo and Daphne" by Giovanni Lorenzo Bernini (1598–1680). In the Museo della Villa Borghese, Rome

"Lyric Poetry" by Jean Baptiste Tubi (1635–1700) with Antoine Coysevox. In the Gardens of Versailles

Tomb of Alexander VII by Giovanni Lorenzo Bernini. In St. Peter's, Rome

PHOTOGRAPHS, (TOP LEFT) GIRAUDON, (TOP CENTRE) TONI SCHNEIDERS, (BOTTOM CENTRE) COLLECTION ARCHIVES PHOTOGRAPHIQUES, (OTHERS) ALINARI

The tomb of Richelieu by François Girardon (1628–1715). In the Church of the Sorbonne, P[illegible]

The tomb of Francis II of Brittany and his consort Marguerite de Fo[illegible] Michel Colombe (c. 1430–c. 1512). In the cathedral of Nantes, France

The tomb of Pope Clement XIV by Antonio Canova (1757–1822). In the Church of the Holy Apostles, Rome

The tomb of Marsuppini by Desiderio da Settignano (c. 1428–64). In the Sta. Croce, Florence, Italy

The tomb of Ilaria del Carretto by Jacopo della Quercia (c. 1374–1438). In the cathedral, Lucca, Italy

SCULPTURE ON TOMBS

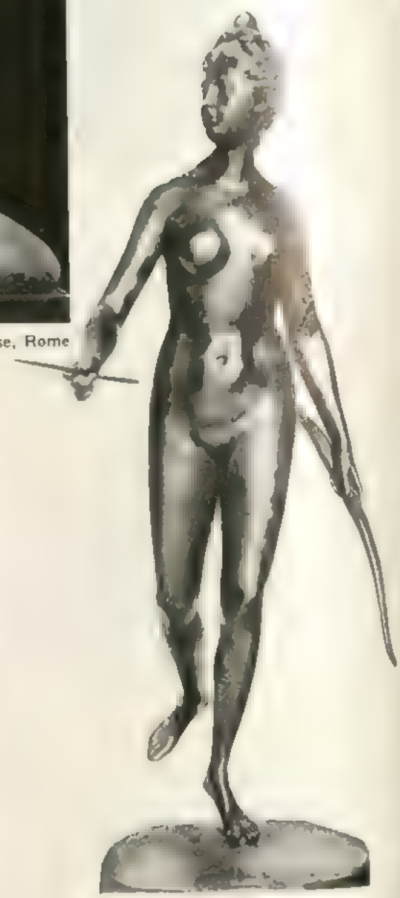
"Perseus With the Head of Medusa" by Antonio Canova (1757–1822), Italian. In the Vatican

"Jeanne d'Arc" by François Rude (1784–1855), French. In the Louvre

"Mercury" by Jean Baptiste Pigalle (1714–85), French. In the Louvre

"Pauline Bonaparte as Venus" by Antonio Canova. In the Museo della Villa Borghese, Rome

"The Bather" by Étienne Maurice Falconet (1716–91), French. In the Louvre

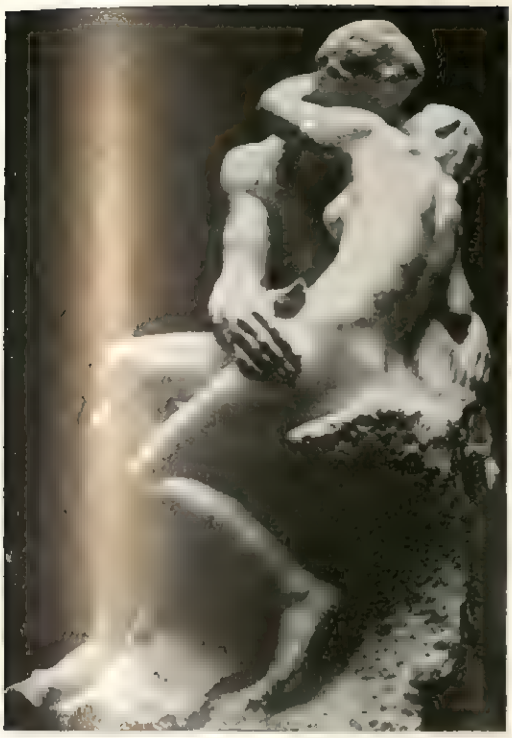


"Diana" by Jean Antoine Houdon (1741–1828), French. In the Louvre

"Nymph and Satyr" by Clodion (1738–1814), French. In the Metropolitan Museum of Art, New York city

ROCOCO AND NEOCLASSICAL SCULPTURE

BY COURTESY OF (BOTTOM CENTRE) METROPOLITAN MUSEUM OF ART, BEQUEST OF BENJAMIN ALTMAN, 1913; PHOTOGRAPHS, (TOP LEFT) ANDERSON, (CENTRE) ALINARI, (OTHERS) GIRAUDON

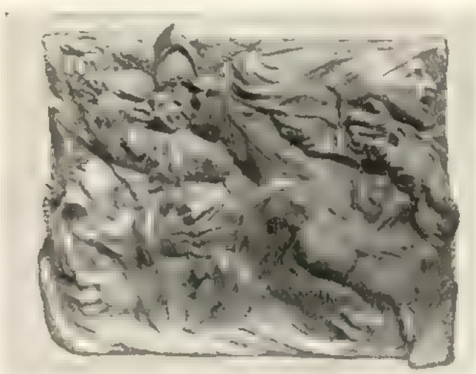


"The Kiss," a marble group by Auguste Rodin (1840–1917), French. In the Metropolitan Museum of Art, New York city

"Theseus and the Centaur Bianor," a bronze statue by Antoine Louis Barye (1796–1875), French. In the Metropolitan Museum of Art, New York city

"Athlete Struggling With Python" by Frederic Leighton (1830–96), English. In the Tate gallery, London

"Head of a Woman Wreathed With Vine Leaves," a study for "La Danse" for the Paris Opera house by Jean Baptiste Carpeaux (1827–75), French. In the Metropolitan Museum of Art, New York city



"The Carnage," bas-relief by Auguste Préault (1809–79), French. In the Chartres museum

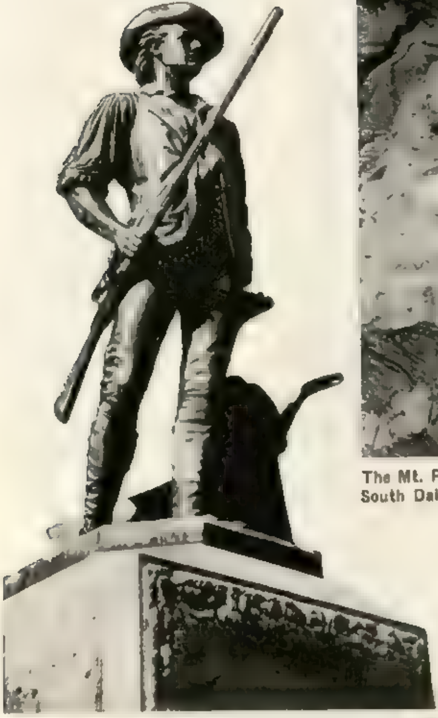
"Theseus and the Minotaur" by Antoine Louis Barye. In the Louvre

"Beside the Sea," marble statuette by Auguste Rodin. In the Metropolitan Museum of Art, New York city

THE 19TH CENTURY AND RODIN

BY COURTESY OF (TOP CENTRE, SECOND ROW CENTRE, BOTTOM RIGHT) THE METROPOLITAN MUSEUM OF ART, (TOP CENTRE) ROGERS FUND, 1910, (SECOND ROW CENTRE) FLETCHER FUND, 1925, (BOTTOM RIGHT) GIFT OF THOMAS F. RYAN, 1910, (TOP RIGHT) THE NATIONAL GALLERY, MILLBANK; PHOTOGRAPHS, (TOP LEFT, SECOND ROW RIGHT, BOTTOM LEFT) GIRAUDON

SCULPTURE



"The Minute Man" by Daniel Chester French (1850–1931), U.S. At Concord, Massachusetts

The Mt. Rushmore memorial by Gutzon Borglum (1867–1941), U.S. In South Dakota

"Marshal Ney" by François Rude (1784–1855), French. In Paris

MONUMENTAL SCULPTURE

"General Sherman" by Augustus Saint-Gaudens (1848–1907), U.S. At the entrance of Central park at 59th street and Fifth avenue, New York city

"Jeanne d'Arc" by Paul Dubois (1829–1905), French. Before Reims cathedral

"Christ of the Andes" by Mateo Alonso (1878–1955), Argentina. On the Argentina-Chile border

PHOTOGRAPHS, (TOP LEFT) DAVID CORSON—A. DEVANEY—PUBLIX, (TOP RIGHT, BOTTOM CENTRE) GIRAUDON, (OTHERS) A. DEVANEY—PUBLIX

"Column" (1923) by Naum Gabo, U.S. In the Solomon R. Guggenheim museum, New York city

"Mlle Pogany" (1928–29) by Constantin Brancusi, Rumanian. Philadelphia Museum of Art

"Standing Youth" (1913) by Wilhelm Lehmbruck, German. In the Museum of Modern Art, New York city

"Woman's Head" (1909) by Pablo Picasso, Spanish. In the Museum of Modern Art, New York city

"Bird in Space" (1919) by Constantin Brancusi. In the Museum of Modern Art, New York city

"The Horse" (1914) by Raymond Duchamp-Villon, French. In the Walker Art centre, Minneapolis

"Unique Forms of Continuity in Space" (1913) by Umberto Boccioni, Italian. In the Museum of Modern Art, New York city

AVANT-GARDE SCULPTURE AND CONSTRUCTIVISM

SCULPTURE

"Lobster Trap and Fish Tail" (1939) mobile by Alexander Calder, U.S. In the Museum of Modern Art, New York city

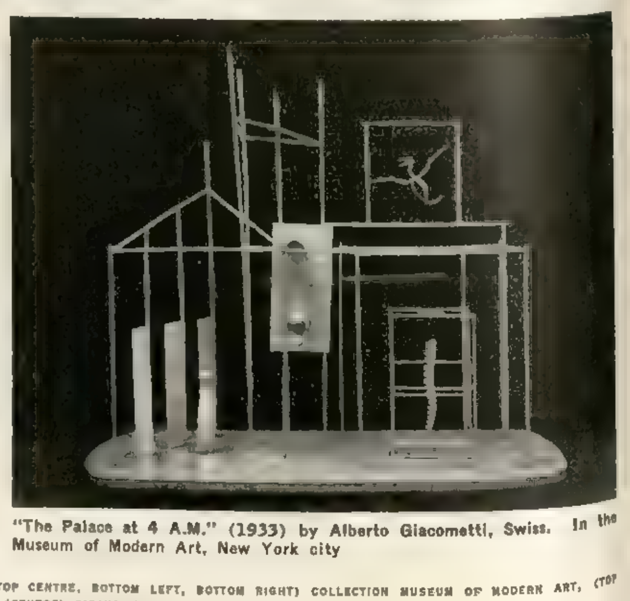
"Standing Woman" (1932) by Gaston Lachaise, U.S. In the Museum of Modern Art, New York city

"Chained Action" (c. 1906) by Aristide Maillol, French. The Metropolitan Museum of Art, New York city

SCULPTURE BETWEEN 1900 AND 1940

"Mediterranean" (c. 1901) by Maillol. In the Tuileries gardens, Paris

"Triumph of the Egg, I" (1937) by John B. Flannagan, U.S. In the Museum of Modern Art, New York city



"The Palace at 4 A.M." (1933) by Alberto Giacometti, Swiss. In the Museum of Modern Art, New York city

BY COURTESY OF (TOP RIGHT) THE METROPOLITAN MUSEUM OF ART, FLETCHER FUND, 1929; (TOP LEFT, TOP CENTRE, BOTTOM LEFT, BOTTOM RIGHT) COLLECTION MUSEUM OF MODERN ART, (TOP LEFT) MRS. SIMON GUGGENHEIM FUND, (TOP CENTRE) GIFT OF THE ADVISORY COMMITTEE; PHOTOGRAPH, (CENTRE) GIRAUDON

"Sea Quarry" (1949) by Theodore Roszak, U.S.; welded and brazed metal. In the Norton Gallery and School of Art, West Palm Beach, Fla.

"Reclining Figure" (1945–46) by Henry Moore, English; elm wood. In the Cranbrook academy, Michigan

"The Unknown Political Prisoner" (1952) by Reg Butler, English; bronze wire with stone base. In the Tate gallery, London

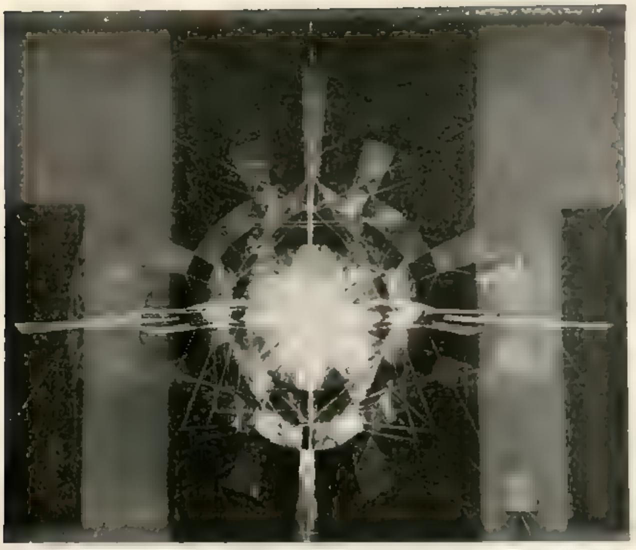
"Virgin of Assy" (1948–54) by Jacques Lipchitz, French. In the church at Assy, France

"Horse and Rider" (c. 1949) by Marino Marini, Italian; bronze. In the Walker Art centre, Minneapolis

"Tripartite Unity" (1947–48) by Max Bill, Swiss; chrome-nickel steel. In the Museu de Arte Moderna, São Paulo, Brazil

SCULPTURE AFTER WORLD WAR II

BY COURTESY OF (TOP LEFT) CRANBROOK ACADEMY OF ART GALLERIES, (TOP CENTRE) NORTON GALLERY AND SCHOOL OF ART, (BOTTOM LEFT) CHURCH OF ASSY, (BOTTOM CENTRE) WALKER ART CENTER, MINNEAPOLIS, (BOTTOM RIGHT) MUSEU DE ARTE MODERNA SÃO PAULO; PHOTOGRAPH, (TOP RIGHT) KEYSTONE



"Sun" (1953–56) by Richard Lippold, U.S.; gold wire. In the Metropolitan Museum of Art, New York city

"Social Consciousness" (1952–53) by Jacob Epstein, Engl[illegible] bronze. In Fa[illegible]mount Park, Philadelphia

"Head of a Woman" (1951) by Pablo Picasso, Spanish; bronze. In the Museum of Modern Art, New York city

CONTEMPORARY SCULPTURE

"Sentinel" (1959) by Seymour Lipton, U.S.; nickel silver on monel metal. In the Yale University Art gallery

"Family Going for a Walk" (1953) by Kenneth Armitage, English. Gimpel Fils gallery, London

"Structure" (1955) by Raymond Jacobson, U.S.; brass. Privately owned

BY COURTESY OF (TOP LEFT) THE METROPOLITAN MUSEUM OF ART, FLETCHER FUND, 1956, (TOP RIGHT) PHILADELPHIA MUSEUM OF ART, ELLEN PHILLIPS SAMUEL MEMORIAL, FAIRMOUNT PARK ART ASSOCIATION, (CENTRE) MUSEUM OF MODERN ART, BENJAMIN AND DAVID SCHARPS FUND, (BOTTOM LEFT) YALE UNIVERSITY ART GALLERY; (BOTTOM CENTRE) A. C. COOPER LTD., (BOTTOM RIGHT) MRS. LLOYD ACKERMAN

found principally on monastic buildings. They were meant for men who had withdrawn from the world. In the dynamic 12th century, the burghers of the newly powerful towns replaced the monks as the principal patrons of art, and it is to them, as lay Christians active in the world of affairs, that the art of the Gothic was increasingly addressed. The Christ of the west portal of Chartres cathedral was made soon after the radically dematerialized God of Vézelay, but by comparison seems more to partake of the substance of the physical world.

Likewise, by contrast with Romanesque art, the "Golden Virgin" of Amiens seems a credible creature of our physical world: the massive draperies and rounded upper portions impart a real bulk to her figure; she turns with a smile to a Child, who returns her affectionate gaze and whose orb seems a toy. Yet a brief comparison with any draped figure of classical antiquity will disclose how profoundly medieval this Gothic figure is, with the parts of the body completely obscured and the drapery composition not really hanging but swinging upward toward the spiritual focus of the work—in this instance, to the relation of Mother and Child, but more characteristically to the masklike faces with their great, hypnotic eyes that suggest the mystic state of the visionary. Never can the position of the legs be clearly determined under the draperies of Gothic statues; nor is there any sense of the interaction of weight and support in an organic structure, even when religious considerations demand the exposure of the body. The Gothic sculptor no longer rejects the physical world, but he selectively readmits only what he can exploit in the interests of spiritual expression.

Gothic art embodies the popular religious fervour that gained momentum all through the 12th century and culminated in the 13th, an age of wide-ranging spiritual enthusiasm and holy heroes—St. Francis of Assisi, St. Dominic, St. Thomas Aquinas and St. Louis, king of France. Huge sculpture shops were developed about the great cathedrals, Chartres, Paris, Bourges, Reims, Amiens and all the rest, and there the sculptors of all Europe were trained. The later Gothic sculpture of Bamberg in Germany, for example, traces its ancestry to the style of Reims during the previous generation; eventually the French style permeated almost the whole of Europe. Occasionally a genius of truly impressive individuality emerged, such as the master who produced the tragic reliefs on the rood screen, the Crucifixion group and the statues of benefactors for the cathedral of Naumburg in Saxony.

In the later 13th and in the 14th centuries the human and pathetic aspects of Gothic sculpture were intensified and its monumentality was often exchanged for intimacy, elegance and subtlety. Another important current, that of greater realism, emerged later in the 14th century; and in Burgundy, whence a new ascendant culture was soon to be diffused over much of Europe, this realistic tendency culminated in the work of the greatest sculptor of the end of the century, Claus Sluter. His masterpiece, the famous "Well of Moses" in Dijon, illustrates his capacity to endow the old Gothic forms with massive bulk and with trenchant realism in such details as the head, without compromising their fundamental medievalism. This style survived well into the later 15th century, producing such works as the tomb of Philippe Pot, now in the Louvre.

2. Tomb Sculpture.—The growth of naturalism in Gothic figure sculpture can be traced more continuously in tomb figures than in any other field. The tombs in the abbey of St. Denis, near Paris (1264), built by St. Louis for his ancestors and his sons, are characteristic examples of 13th-century tombs. In these the figure is modeled not in relief but in the round. Each figure, represented as in early manhood and rendered as a standing figure laid upon the ground, is graciously idealized. Each has a simple architectural framework, low in relief, and each is placed on a base embellished with pictorial reliefs. The expression is calm and benign and the draperies straight and simple with long clearly incised lines.

Characteristic of the 14th century is the tomb of Charles V at St. Denis by André Beauneveu, where the advancing realism of the century expressed itself in accurate portraiture and in more flowing and natural draperies. The figure is portrayed in the stiffness of death, and about him are grouped representations of his relatives in the costumes and attitudes of mourners. The fine tombs of the popes of Avignon—one of which is elaborated with a canopy—and the effigy of Robert d'Artois at St. Denis, clad in full armour with a lance in his hand, are other examples which illustrate the growing interest in actuality. This principle reached its widest acceptance in the 15th century. At Tournai, Belg., there was developed the type of tomb in which the figure of the deceased, no longer dead and recumbent, kneels before the Virgin or some religious object; these tombs, the production of which amounted to an industry, were exported to all parts of England, France and Germany and widely imitated. The monument to Isabelle de Bourbon in the cathedral at Antwerp (*c.* 1465) and the tomb of Louis de Male, at Lille (1455), both of which are of bronze, are characteristic examples of this art in which the elaboration of detail and the representation of action and of personality are clearly the preoccupation of the sculptor. Burgundy produced an equally realistic funerary art, of which the tomb of Philip the Bold (completed 1410), begun by Claus Sluter, is an elaborate example.

3. Italian Gothic.—During the later medieval centuries Italian sculptors turned repeatedly to the inevitable models of classical antiquity. The style of the monumental bronze doors which have survived in numerous examples also suggests the closer relations with the Byzantine east, although those in Pisa and Monreale by Bonanno of Pisa, and the famous ones in Benevento (12th century), are more vigorous and three-dimensional than Byzantine relief. The antiquarian tendency is especially pronounced in the 13th-century sculptures made at the instance of Frederick II, and particularly in those made for a triumphal arch erected at Capua, the extant fragments of which include toga-clad, laurel-crowned portrait busts and other allusions to the antique.

The artistic origins of Niccolò Pisano probably trace back to this southern school but he was active chiefly in Tuscany. His first work, the hexagonal pulpit completed in 1259 for the baptistery of Pisa, is the masterpiece of this antiquarian tendency, for there he breathed the Gothic spirit into the massive forms of antiquity and disposed them in unique relief compositions of which the teeming profusion is rigorously disciplined. In the famous herculean nude "Strength," his genius takes a form of classical antiquity and medievalizes it with a swinging Gothic compositional curve and with Gothic pathos. In his later works, the pulpit in Siena and the fountain of Perugia, he grew constantly more Gothic. His son, the prolific Giovanni Pisano, is the greatest of Italian Gothic sculptors. The range of his passionate dramatic art is seen in his masterpiece, the pulpit completed in 1301 for S. Andrea in Pistoia, with its masterfully exploited high relief and its profoundly emotive statuettes. The severer art of Arnolfo di Cambio retains more of the spirit of his master, Niccolò, as in the great canopy tombs (of Adrian V, in S. Francesco, Viterbo; of the cardinal de Braye, in S. Domenico, Orvieto), and especially in the heroic statues of the Virgin (Opera del Duomo, Florence) made for the façade, begun by Arnolfo but later destroyed, of the cathedral of Florence. Tino di Camaino amplified the sculptural enrichment of tomb composition into the splendid form of the monument for Cardinal Petroni (1318), in the cathedral of his native Siena, and endowed his figures with a reflective lyricism and a rhythmic linearity that recall the painting of that city.

Another masterpiece of the Sienese school is the extensive decoration of the cathedral of Orvieto, probably by Lorenzo Maitani; there the linear subtleties and aristocratic elegance of the school luxuriate in ornamental relief plans of great opulence, while the sharp, wiry forms of the bronze sculptures above uniquely exploit this exceptional medium by projecting boldly and piercingly from the architectural mass. In this medium of bronze, the masterpiece of the period is Andrea Pisano's famous south portal (1330–36) of the Florentine baptistery, the more severely linear figures of which are harmoniously arranged in simple, squarish compositions further clarified by gilding and by contrast with the complicated form of quatrefoil panels.

Later in the 14th century, in a spiritual reaction aggravated by the plagues which had decimated the population, there was a

tendency among both painters and sculptors to exchange the more progressive aspects of the art of their predecessors for a strangely unorthodox proliferation of forms, often less human and more formalized. This reactionary period is best represented by the great polychrome marble tabernacle (1355–59) in Or San Michele, Florence, by the leading painter and sculptor of the time, Andrea Orcagna. Other followers of the Pisano tradition carried it to northern Italy and expanded it under local and foreign influences, for example, in the virile and vigorous style seen in the rather menacing and Teutonic equestrian figures of the elaborate Scaliger tombs in Verona.

III. RENAISSANCE SCULPTURE

1. The Beginnings.—The radical innovations in Tuscan sculpture about 1400 are among the first signs of the Italian Renaissance. In the art of certain transitional figures, such as Jacopo della Quercia, the medieval suspicion of the body was dispelled and heroic nudes re-created the physical beauty of the antique in a spirit and style intelligible to the late medieval audience. One of the first monuments of the new age is Della Quercia's tomb of Ilaria del Carretto (*c.* 1406) in the cathedral of Lucca, in which the Burgundian free-standing tomb was translated into Latin, with richly sculptural figures of nude *putti* bearing garlands. The effigy partially retains the Gothic spirit, however, as in the drapery passages which conceal the lower parts of the body.

In the more progressive Florence, which for the next two centuries was to be the world centre of sculpture, much that is Gothic survived in the work of the less radical artists, such as Lorenzo Ghiberti, who won the famous competition for the north portal of the Florentine baptistery and there enriched, developed and refined the model of Andrea Pisano. The subject of the competition was the sacrifice of Isaac; Ghiberti's competition piece, in which the difficult quatrefoil form is brilliantly exploited for both narrative and aesthetic effect, represents Isaac as a bound naked youth, one of the first in a millennium to be endowed with both the physical beauty and the restrained pathos of the antique ideal. It can be compared with the less elegant but more urgently dramatic trial-piece of Filippo Brunelleschi, whose Isaac, a more boyish body studied carefully from nature, shows such acutely observed psychological effects as the way the toes have been bent and rigidly tensed with fear. Thus mankind and the physical world it inhabits are the great themes which were to be elevated to predominance in Renaissance art. Ghiberti's later east doors for the baptistery were said to be worthy of the Gates of Paradise, a sobriquet still in general use.

The sharpest break with the medieval past was made by Donatello. After several important early works, he was commissioned in 1411 to make a statue of St. Mark for Or San Michele. For this revolutionary statue Donatello restored to art a pose of ancient Greek sculpture, the weight mainly on one leg, with the hip over that leg higher and slightly advanced and the shoulder lower and slightly receding. The formula rationally and articulately expresses the interaction of weighing and supporting factors in the structure of the human body, and this statue, for the first time since classical antiquity, thus recognizes the inherent validity of man. It is, however, no mere imitation of the antique. The right arm and hand do not hang with the relaxed freedom usual in ancient sculpture but press with barely perceptible tension against the thigh, and two huge fingers toy nervously with a fold of drapery in the absent-minded, involuntary way they often do when one is under a strain. This, even more than the pathos of the furrowed brow, hints at the internal turmoil of a simple, solid and normally steady man in the grip of a tremendous experience which partakes of the divine: a conflict of external tranquility and internal tension that is wholly unclassic. This imaginative re-creation of a real personality, in all its physical, psychological and spiritual complexity, was to become the distinguishing achievement of modern western art, but in 1411 it was quite unprecedented.

At the same time, "St. Mark" is placed on an unyielding and thoroughly improbable pillow, an early foreshadowing of the almost perversely personal and ultimately inexplicable elements which constantly increase in Donatello's late work and which preclude what he must have felt as the rather complacent rationality of much antique art. The "St. George" and the "St. Louis," also made for Or San Michele, and the statues for the campanile constitute a series of constantly more profound variations on the theme of the "St. Mark." This art of characterization culminated in the astonishing achievement of the "Gattamelata," the first colossal bronze equestrian statue since antiquity, in which a mount that is effortless and masterly, in comparison to the less convincing antique prototype, is used not merely to suggest equestrian prowess but rather to enhance the majesty of a great man, the paragon of the spiritualized active life. A closer view, however, somewhat qualifies this eulogy; the head discloses many psychological facets—determination, calculation, suspicion, bitterness, introspection, even compassion—the reconciliation of many conflicting tendencies not all conventionally "good" but all marvelously human. This statue is the perfect embodiment of the Renaissance.

Donatello's bronze "David" is the first monumental nude statue since antiquity and, no less important, the first truly free-standing figure with a full range of multiple views, although the front predominates. The relief of Herod's feast in the Sienese baptistery (1423–27) is one of the earliest extant elaborated demonstrations of perspective, while other very low-relief compositions in marble (*stiacciato-rilievo*) showed how that translucent medium could be endowed with dramatic pictorial effects of atmosphere and movement. All this is only a general indication of Donatello's historical importance. In his last works, such as the pulpit reliefs in S. Lorenzo, Florence, he is the modern Christian visionary, whose individual mysticism is very unlike the institutional, collective mysticism of medieval art.

Luca della Robbia was the fourth great innovator of early Renaissance sculpture. His great marble singers' gallery for the Florentine cathedral and his many beautiful Madonna reliefs in a new medium, coloured and glazed terra cotta, restore the calm simplicity, clarity, harmony and balance of truly classic art in an unpretentious way closer to ordinary humanity than the more Olympian ideal of antiquity. His earlier work, redolent of optimism, expresses an affecting approbation of the Renaissance rediscovery of the world, a spirit that curiously recedes in the more abstract and ascetic style of his late works.

2. The 15th Century.—The innovations of these masters were refined and augmented by the less revolutionary stylists of the middle and later 15th century. Donatello's design for the tomb of the schismatic pope John XXIII in the Florentine baptistery (1425–27) exploits the great Corinthian columns of the architecture in converting a Gothic canopy tomb into a classic wall tomb composed of antique elements; this remained a model for the succeeding generations of Renaissance sculptors. Bernardo Rossellino gave it a less personal and more orthodox form in his tomb of the humanist Leonardo Bruni (*c.* 1445) in Sta. Croce, Florence. Desiderio da Settignano, a lyric and sensitive sculptor of exquisite virtuosity, made the tomb of Carlo Marsuppini (*c.* 1455) in the same church a richer and less somber elaboration of the same theme. Antonio Rossellino, Bernardo's younger brother, carried the form to the highly ornamental and pictorial extreme of the tomb of the cardinal of Portugal (1460s) in S. Miniato al Monte, Florence. The brilliant Antonio Pollaiuolo endowed his bronze nudes with an intensified animal vitality; they could remain the exhilaratingly unliteral hyperboles still permitted by the purely empirical, unsystematic anatomy of art before the reforms of Leonardo da Vinci. The work of the versatile Verrocchio ranges from a patrician charm and freshness in his lesser works to the rugged, monumental power of his great equestrian general, "Colleoni." The preoccupation with style and execution seen in the works of his contemporaries is well illustrated by his brilliant tomb of Pietro and Giovanni de' Medici (1472) in S. Lorenzo, where all reference to death is excluded from a work of pure artistic abstraction. Sepulchral art outside Florence is represented by the tomb of the doge Pietro Mocenigo (*c.* 1485) in SS. Giovanni e Paolo, Venice, by Pietro Lombardi.

In France, during the 15th century, a wholly native development

in late Gothic sculpture, called the *détente* ("relaxation"), is in some ways equivalent to the early Italian Renaissance, for the extreme realism and the intricate drapery compositions of the later Gothic were replaced by a sweet and simple idealism. At the end of the 15th century its leading exponent, Michel Colombe, actually appropriated some of the new Italian forms, as can be seen in his tomb of Francis II of Brittany in Nantes, but his refined sensitivity to feminine beauty is very French. In 15th-century Germany, however, the realism, the emotionalism and the tortuous complication of the draperies were carried by such masters as Veit Stoss to an extreme that has been termed the late Gothic baroque, although a calm lyricism akin to the *détente* tempers the carvings of Tilman Riemenschneider. The more massive and monumental style of Adam Kraft's reliefs of the stations of the cross in Nürnberg (1505) is another independent approximation of the early Renaissance. Peter Vischer, also of Nürnberg, knowingly and unslavishly synthesized the Italianate and the Germanic, but it was not until the second half of the 16th century that northern sculptors were at home in the new art imported from the south.

3. Leonardo da Vinci and Michelangelo.—The artists of the High Renaissance, after *c.* 1500, evolved a truly classic style, serene, lucid, harmonious and more monumental than the art of the 15th century. The tremendous intellect of Leonardo da Vinci so enlarged artists' understanding of physical man and the physical world that, although he was only incidentally occupied with sculpture and none of his works long survived him, all sculpture, from his own time down to the 20th-century revolutions of modern art, is indebted to him for his full realization of a complete and systematic mastery of human anatomy. This vastly enlarged the expressive scope and illusional conviction of representational figure sculpture. Now, drawing upon antique sources, a great master could fully express the High Renaissance ideal of man the beautiful, harmonious and noble as Andrea Sansovino, for example, did in the sculptures on the Sforza and Basso tombs in S. Maria del Popolo in Rome.

The potentialities of Leonardo's innovations reached their greatest fulfillment in Michelangelo. In his early Pietà in St. Peter's, Rome, a marble whose unrivaled technical perfection realized the promise of a number of the sculptor's brilliant youthful works, he imparted a mood of calm and tender elegy to the monumentality, the plasticity, the idealism and the harmonious repose seen in Andrea Sansovino's and all High Renaissance art. There is a great clarity and logic in the composition of two interlocked crescents, with draperies emphasizing the structurally and psychologically important point where the Madonna's hand supports but does not actually touch the dead body of her Son. This classic phase of Michelangelo led to the colossal marble "David" (1501–04), a splendid nude figure nearly 13½ ft. high, in which new complications began to subvert the clarity of the High Renaissance. The multiple views, in a way characteristic of classical sculpture, are compositionally subordinated to the main front view; but this is curiously at odds with the content, since one must then walk around to the right, to a very secondary view, in order to see in the face the theme of the whole work: the heroic defiance of David that blazes forth from his scornful glare. Moreover, the psychologically very particularized state of this head is in conflict, by classic standards, with its physically very generalized, wholly ideal form.

These dissonances erupted into a violent rejection of the classic style in Michelangelo's next statue, the "St. Matthew" in the Florentine academy, begun in 1506 and abandoned in a rough-hewn state that perpetuates the interrupted creative act. The traditional stance of Donatello's "St. Mark" is now wrenched into the violent pose called *contrapposto*, and in this case deliberately reversed (hip lower over the weight-bearing leg and shoulder higher), so that the device originally evolved to express repose now expresses its extreme antithesis, a terrible tension that locks the struggling figure in a painfully frustrated standstill. The motive attains to the stunning tragedy of the two nude slaves in the Louvre and of the later unfinished ones in the Florentine academy and expresses the torturing ambivalence Michelangelo felt toward the human body, which he called "the earthly prison of the soul." These works belonged to various projects for the tomb of Pope Julius II, which was originally intended as a vast monument of about 40 large statues, but, in the course of the 30 years intermittently devoted to it, was repeatedly reduced by the pope's heirs to the pitiful compromise now in S. Pietro in Vincoli, Rome.

The Medici tombs in the new sacristy of S. Lorenzo, Florence (1520–33), which were carried nearer completion, are part of an architectural ensemble by Michelangelo himself, who would also have executed the frescoes; there would have been bronze river gods below (his clay model is now in the academy) to balance the symbols of the times of day on the sarcophagi above. The Medici dukes seated above, whose idealized portraits are so generalized they represent mankind rather than individual men, are thus released by death from the prison of the body and from the bonds of time and place, and, turning toward their patron saints and the Madonna diagonally across the room, seek unity with God. Late in Michelangelo's life the terrible conflicts of these tragic works are at last resolved, as in the sublime Pietà now in the Florentine cathedral but originally intended for his own tomb by the nearly octogenarian Michelangelo, who carved his self-portrait in the Nicodemus tenderly releasing the dead Christ from his embrace. The wraithlike forms of the even more dematerialized "Pietà Rondanini" in Milan, a fragment from the very end of his nearly 90 years, testify that, as Michelangelo said in one of the poems of his old age, he had turned from ". . . painting and sculpture, to that Divine Love spreading arms wide from the Cross to embrace us."

4. Mannerism.—Even the classic calm, clarity and harmony of Michelangelo's early works, as the Vatican "Pietà," were permeated with a profound melancholy that sets them apart from a characteristically optimistic artist of the High Renaissance like Raphael. Twenty years later the spirit of that fortunate but brief period seems to have soured, and there was a deliberate reversal of all of the qualities that had been cultivated in its art. One of Michelangelo's least typical works, the "Victory" group originally intended for the tomb of Julius, is a good introduction to the new movement, which is now neutrally termed Mannerism. The great attenuation of forms, the unreal pose, the wide-eyed, fixed stare that has no object and the precariously towering verticality of the composition combine to produce a strange and disconcerting air of enigma and irrationality which is the negation of classic art and has therefore been called anticlassicism. These qualities were carried to an extreme in the typical work of the time, the "Madonna and Child With St. Anne" (1525) in Or San Michele, Florence, by Francesco da Sangallo, whose tomb sculptures, among the most impressive monuments in the style, add a sinister realism to all the other expressionistic singularities of anticlassical Mannerism. Although the greatest sculptor of the generation, Alonso Berruguete, was trained in Florence and Rome, he returned to his native Spain and there produced such masterpieces as the polychrome sculptures now in the Valladolid museum, audaciously twisted and flamelike figures of great expressive power; or the remarkable alabaster Transfiguration group placed high over the choir screen of Toledo cathedral, where the translucent material in which this turbulent composition is carved is set aglow by the light of the great rose window.

Artists of more moderate temperament avoided these extremes. Jacopo Sansovino developed the objective classic ideal of his master, Andrea, into the subjective aestheticism seen in the "Apollo" on his Loggetta in Venice. Its elegantly lifted shoulder converts what was now felt as the merely physical logic of a traditional pose into a hyperbole of grace that accords with the extreme ideal of the refined, masklike face, huge eyes and tiny mouth, with the smooth, abstract surfaces and with the sweeping, continuous curves of the contour lines. The style of this statue foreshadows the reaction, toward the middle of the 16th century, against the violence and pessimism of the current style. Benvenuto Cellini's "Perseus," and especially its fantastic base adorned with elegant and hypersensitive figures, may well be the masterpiece of this phase of Mannerism, in which the fidelity to exclusively private ideals of beauty is often carried to an esoteric extreme. This extreme is considerably modified by reference to nature and to

classical antiquity in the work of Giovanni da Bologna, who, in his earlier compositions, addressed himself to the problem of multiple views and brilliantly solved it in the famous "Sabine" group by choosing a turning action that reasonably generates a spiral composition and thus impels a continuity around the statue, the full effect of which can be perceived only in motion.

In France the Renaissance created many decorated tombs of great beauty, among which those of Cardinal Amboise in Rouen cathedral and of Francis I at St. Denis by Pierre Bontemps are perhaps the most celebrated. Jean Goujon, one of France's greatest sculptors, worked at the time; an example of his vital and gracious style is the "Fontaine des Innocents" in Paris. The French sculptor Germain Pilon moderated the extreme Mannerism of Goujon and his contemporaries with greater naturalism, and, in a way characteristic of his generation and of the new age of the Counter-Reformation, returned with impressive results to the religious themes that had little occupied the sculptors of the mid-century. His effigy of René Birague in the Louvre is a perfect balance between sympathetic naturalism and monumental restraint. The splendid "Tombs at Brou" (1505–26) near Bourg, built for Margaret of Austria by Konrad Meit of Worms, are characteristic of the ornate, lingering Gothicism of the Renaissance in Germany.

IV. BAROQUE SCULPTURE

1. The Beginnings; Bernini.—A precocious example of the 17th-century rejection of Mannerism is seen in Francesco Mochi's brilliant Farnese equestrian monuments in Piacenza (1612–29), where the forms are pierced and opened up and the momentary, unstable poses, with draperies fluttering and tails lashing, give a vivid movement that releases the figures from the Mannerist spell. Their naturalism and pictorialism are also hallmarks of the coming baroque. No field was more congenial to the spirit of baroque art than sculpture carried out on a conspicuous scale. The baroque artist achieved dramatic pictorial unity by abolishing the traditional limits separating painting, sculpture and architecture. The solid masses of sculpture and even of architecture were made to move in space by means of such motive forms as undulations; sculpture was transformed by such painter's devices as richly varied illusionistic textures, coloured materials and irregularly dappling light effects.

Giovanni Lorenzo Bernini, the greatest sculptor of the 17th and 18th centuries, amplified the style in marble, establishing the sculptural principles for those two centuries in a series of youthful works of unrivaled virtuosity, as the "Apollo and Daphne." Stone was now completely emancipated from stoniness by open form and by an astonishing illusion of flesh, hair, cloth and other textures, pictorial effects that had earlier been attempted only in painting. These qualities make what his contemporaries called his "speaking portraits" seem unprecedentedly alive; portrait sculpture for two centuries is a variation of these innovations. In the statue of St. Longinus in St. Peter's in Rome, Bernini created the characteristic formula of baroque sculpture by throwing the draperies into a violent turmoil, the complicated and broken involutions of which are not rationally explained by the figure's real bodily movement but seem paroxysmally informed by the miracle itself. The famous group of the miraculous vision of St. Theresa in the Cornaro chapel of Sta. Maria della Vittoria, Rome, is the masterpiece of baroque religious sculpture, and shows how Bernini's intellect could organize the arts of architecture, painting and sculpture in an overwhelming assault on the senses that dispels the resistance of the intellect. This ambitious plan is typical of the mature Bernini, whose spiritual and artistic aspirations now exceeded the scope of his early secular salon statues. His later works were largely religious and unprecedentedly vast in scale, as the dazzling "Cathedra Petri" which covers the whole end of St. Peter's in Rome with teeming multitudes of figures.

The tombs of Bernini are magnificent spectacles in which symbolic figures, clothed in sweeping draperies, with rhetorical gesture and expressive features, share in some emotional experience, theatrically depicted. An example is the tomb of Alexander VII in St. Peter's, Rome. The pontiff, set in a great apse, kneels on a high pedestal about which Charity, Truth, Justice and Wisdom weep disconsolately while Death, a skeleton, raises the great draperies of polychrome and gold that veil a darkened doorway. The fountain of the Triton in the Piazza Barberini, Rome, from which all clarity of profile or of shadow, all definiteness of plane, are removed, is also characteristic of Bernini's style, widely imitated throughout Europe.

Bernini's art was the basis of all baroque sculpture, but his extreme was moderated by his more conservative contemporaries, such as Alessandro Algardi (relief of "St. Leo and Attila the Hun," 1646, St. Peter's, Rome) and especially the Fleming François Duquesnoy. The latter's "St. Susanna" in the Sta. Maria di Loreto in Rome, a figure after the antique but enlivened with Berninian textures, was originally made to look toward the observer and, with a gesture, to direct his attention to the altar. The distinction between art and life which the Mannerists had cultivated is banished by this active participation of the statue in our space and our activities, another important innovation of Bernini and his followers.

2. Other Baroque Sculpture.—Duquesnoy was much admired in France, where the sculptors of Louis XIV, such as François Girardon, continued his tradition of setting correct and charming allusions to the antique in a pictorial and spatial context that is wholly baroque. Girardon's tomb of Richelieu, in the Church of the Sorbonne, Paris, is illustrative of the baroque monuments of France, calmer and more conservative than those of Italy. The dying cardinal, lying on his sarcophagus and originally gesturing in supplication toward the altar, is upheld by Religion and mourned by Science. The three figures, united by the lines of skilfully arranged draperies, are informed by a solemn and touching sentiment. The academic discipline imposed by the Sun King's ministers, especially Colbert, discouraged less tractable spirits, such as the passionate genius Pierre Puget, whose unique expressions of anguish are couched in the physical terms of highly original works like the Milo of Crotona; here the composition of a figure rigid with pain is jammed into a plane that expresses its helpless and tragic captivity, and the open form of the baroque is given the highly personal interpretation of a great gaping hole like a vortex.

Another of the sculptors of Louis XIV, Antoine Coysevox, had begun in the official "academic baroque" style, but his later works, undertaken after the death of Colbert, are witnesses of the gradual acceptance of the baroque in France, which now acquired the artistic leadership that Italy had long held over the rest of Europe. At the same time, the style was made lighter, gayer and more ornamental, in accordance with 18th-century taste, as seen in the famous "Horses of Marly" at the entrance to the Champs Élysées in Paris, by Guillaume I Coustou. Coustou's bust of his brother Nicolas has a characteristic freshness and informality whereby 18th-century artists avoided the grandeur they found pompous in the Berninian tradition.

The monument to the Great Elector in Berlin by Andreas Schlüter, a realistic and robust equestrian statue, is representative of the Berninian tradition in Germany, as is also the periwigged "Apotheosis of Prince Eugene" by Balthasar Permoser in the Barockmuseum of Vienna. The baroque in Germany attained at times astonishing vitality and elaboration of form; the Trinity column, in Vienna, is an example. In England the baroque spirit is less understood. Of the many baroque monuments in Westminster abbey, the best one, the tomb of Lady Elizabeth Nightingale, is by the Anglicized Frenchman Louis François Roubillac. The Pietà in Valladolid, by Gregorio Hernández, is characteristic of the more fervid devotional sentiment in Spain, embodied in polychrome images of vivid reality and clear formal simplicity that remain almost untouched by the style of Bernini.

3. Rococo.—This 18th-century style that reduced the baroque to exquisite refinement was the art of the aristocratic salon and boudoir. The little marble "Mercury" (1741) of Jean Baptiste Pigalle is almost wholly Berninian, except in its intimacy and deliberate unpretentiousness; even in Pigalle's most ambitious undertakings, as the tomb of the comte d'Harcourt in the Parisian cathedral, the relative scale of the figures is much reduced, the whole composition opened up, and an enacted tableau replaces the cere-

monious allegory of Bernini's tombs. At the same time, the more classical current of French sculpture continued and gained importance as the 18th century advanced. The clarified form and continuous, unbroken contours of Étienne Maurice Falconet's marble "Bather" (1757) adapt the classic tradition to a pretty and intimate rococo ideal that is the quintessence of 18th-century taste. This classicism was purified by Jean Antoine Houdon, who avoided the playful air of the rococo boudoir in his "Diana" (*c.* 1777) and his marble nude in the Metropolitan museum, New York (1782), His portraits, and especially his unrivaled busts of children, are the ultimate in the 18th-century refinement of Bernini's tradition.

V. THE 19TH CENTURY

1. Neoclassical Sculpture.—Houdon's younger contemporaries found his moderate classicism only a compromise and evolved a style so evocative of the antique that it is termed neoclassicism. This evocation does not mean imitation; Antonio Canova achieved a reflective, elegiac calm in images that allude nostalgically to the antique but that often form tableau compositions unprecedented in antiquity and still highly pictorial. The surfaces of such a masterpiece as his portrait of Pauline Bonaparte (completed in 1807) are smoothed into fluent abstractions, yet the fleshy and other pictorial textures of the baroque are not surrendered. His refined linearism, unknown in earlier sculpture, eulogizes feminine beauty in a mellifluously elaborated interplay of easy curves. His highly original composition in the Vatican "Perseus" boldly casts the figure in a swinging, continuous arc on a plane.

The tomb of Clement XIV by Canova, in the church of SS. Apostoli, Rome, commissioned in 1783 and unveiled four years later, is a good illustration of early neoclassicism. The theme is that of Bernini—the draped figure of the pontiff seated on his sarcophagus and mourned by Charity and Peace—but the dramatic action is replaced by a dreamy mournfulness, which is made impersonal by the generalized features, by classic draperies and by the definite geometry of the architectural forms. The tomb of Lord Nelson by John Flaxman, in St. Paul's, London, while remotely derived from Pigalle, shares the dignity and classic restraint of Canova, which became characteristic of all sepulchral art. The tomb of Queen Louise in Charlottenburg by Christian Rauch is one of the loveliest of these neoclassical tombs.

2. Realistic and Romantic Sculpture.—In the 19th century a new type of memorial, that of a statue placed in a public place, increasingly replaced sepulchral sculpture, and the neoclassicism that had dominated the first phases of 19th-century art was slowly modified by realism and romance. Pierre Jean David d'Angers, called upon to create in the streets of French cities many representations of famous Frenchmen, gave these not only a contemporary costume but also gestures and expressions in harmony with their characters and activities. His "Corneille" at Rouen, "General Drouot" at Nancy and "Thomas Jefferson" at Washington, D.C., are examples. His monument to the Grand Condé (1817–27) made a skilful compromise between official neoclassicism and the so-called troubadour style which the late rococo sculptors had evolved for monuments to personages of the distant past, but he imbued it with a fieriness that predicts the dashing cavalier of romanticism. François Rude richly developed this nationalization of French sculpture, as his statue of Marshal Ney in Paris demonstrates. He also did realistic subjects of everyday life, such as the Neapolitan fisherboy in the Louvre. His great relief on the Arc de Triomphe masterfully expresses movement and urgency in a composition which nevertheless maintains the formality appropriate to its architectural context; but the exciting gesticulation and the powerful, unidealized heads, especially of the screaming personification of France, were more strongly romantic in feeling.

Antoine Louis Barye's fluent modeling in clay conceived for casting in the dark bronze opposed neoclassical insistence upon white marble; but the meticulously detailed execution of his early works was surrendered in his maturity for broad modeling and monumental grandeur, as in the "Lion" of the Louvre. His increasing inclination toward sculptural rather than pictorial effects eventually led him to experiment with the stylizations of archaic Greek statuary, as in the fine group of "Theseus and the Minotaur," a work prophetic of the early 20th century. Another interesting sculptor among the romantics was the uneven Auguste Préault, whose astonishingly violent and enigmatic relief in the Chartres museum, called "The Carnage" (1834), prefigured Picasso's "Guernica."

The art of the clay modeler, as opposed to that of the stonecarver, was to have a long hegemony. In accordance with the romantic predilection for passionate and violent themes and for medieval subject matter, Jean Baptiste Carpeaux chose Dante's starving Ugolino for his major early work, but in his maturity he embodied the voluptuous gaiety of the second empire, as in his famous "The Dance" on the Paris opera house, where many of the devices of the baroque were adapted to modern taste. His spirited fountain in the Luxembourg gardens recaptured the pictorial warmth of French baroque masters. Far bolder and more fantastic modeling is seen in the statuettes the great lithographer Honoré Daumier made for himself, such as the "Ratapoil" or the relief, "The Emigrants," where the impetuosity of the modeling intensified the high emotional key of the tragedy and where the expressive distortions furnished modern sculpture with one of its most indispensable devices.

Auguste Rodin's highly naturalistic early work, "The Bronze Age" (1877), is effective because the banal studio pose of a man leaning on a staff produced an unconventional and expressive gesture when the staff was removed. From Daumier, Rodin had learned the bold modeling of surfaces that are emotive rather than literal; the statue is only a rough approximation that avoids the definitive finish of earlier sculpture and remains in a state of becoming. Eventually, Rodin even worked with mere fragments such as broken torsos, and he enormously enlarged the range of figure composition. The mass, hitherto the principal vehicle of sculptural composition, was explosively opened by these methods; instead of the interplay of solid and void of earlier sculpture, Rodin's works are fused with the surrounding space. These methods evolved in his many works, such as "Adam" (1880), "Eve" (1881) and others, originally conceived as a part of the masterpiece of modern sculpture, "The Gate of Hell," undertaken by Rodin in 1880 and never really completed. It was inevitable that the translucent nature of the marble surface should engage the attention of Rodin, and even though he always prepared the models in clay and left the execution in stone to assistants, such marbles as the "Kiss" (1885), when properly exhibited with light partly from the rear, appear to glow with the incandescence of their passionate intensity. (J. HUD.; J. HM.)

3. Late 19th- and Early 20th-Century Conservative Sculpture.—The work of Rodin is commonly regarded as marking the beginning of modern sculpture, which is discussed in a following section of this article. At the same time that the new movements were spreading from Paris throughout the western world, a strain of "conservative" sculpture continued.

Sometimes this was merely "official" sculpture. At other times it developed a considerable amount of grace and elegance. Conservative sculpture took its inspiration from nature but was a disciplined art. It represented persons and things, subject to selection and rejection, to composing and arranging, to considerations of place, material, appropriateness and so on; it was also influenced by great past examples, so that inevitably an element of eclecticism enters into this kind of work.

The conservative sculptors sought to avoid what is ugly (a matter of indifference to modern sculptors concerned with expression or "expressiveness" in various forms and not with beauty in the classical sense of the word), as well as all exaggeration of scale and affected "forcefulness." They had no desire to startle, to provide sensation or novelty. Their appeal was not to the nerves or to the tensions of modern life, and in this sense it may be said that they are "behind the times."

The general tendency for a sculptor of conservative leanings was to begin as either a neoclassicist or an imitator of the Renaissance and then in later life to make some slight concessions to the pressure of "modernism." A short summary of any contemporary or recent schools or individual artists almost always begins with

France, but precisely because France after the 1850s was in the forefront of modern movements, there is very little truly conservative sculpture to be found there. In fact it is in the United States that the conservative tendency has been most pronounced, just as the United States was the last to give up neoclassicism proper. Nevertheless, mention may be made of Paul Dubois, who was responsible for the tomb of General Lamoricière in the cathedral of Nantes (1878) and the statue of Jeanne d'Arc outside the cathedral of Reims (1896). He shows in all his works a deep feeling for the Renaissance. Louis Ernest Barrias has been admired for his commemorative statues and busts such as those of the duchesse d'Alençon (1904), the chemist Lavoisier (1898) or the young Mozart in bronze (1887). In England Sir Frederick Leighton worked in a belated neoclassic manner. In the U.S. Augustus Saint-Gaudens may be taken as the most distinguished exemplar of the conservative style. He was not perhaps as "native" as Daniel Chester French but he was typical of his time in being sensitive to French influence and international elegance, in much the same way that the architecture of the turn of the century was affected by study at the École des Beaux-Arts where Saint-Gaudens also worked. His subjects were American and many of them patriotic. Mention need be made only of the Robert Gould Shaw relief opposite the State House in Boston (1897) and the equestrian statue of Sherman (1903) at 59th street and Central park in New York. He also designed the fine new $20 gold piece of 1907 and the head on the $10 gold piece, the latter design originally projected for the figure of Nike-Eirene walking before General Sherman's horse. He is thus well known to numismatists and may be considered as an influence in saving U.S. coinage from a very nearly universal debasement of design in the 20th century. (A. K. McC.)

4. U.S. and Latin-American Sculpture to 1900.—Modern sculpture, like many other facets of modern civilization, is international in scope. Before turning to that subject, however, it will be useful to consider briefly the development of sculpture in the United States and in Latin America up to about 1900.

United States.—The United States can hardly be said to have a sculpture before the 19th century, although William Rush (1756–1833), a wood carver, has been called "the first native American to devote himself seriously and successfully to sculpture." In the middle years of the 19th century there came into prominence four sculptors: Horatio Greenough, who executed several government commissions in Washington, D.C.; Hiram Powers, known particularly for his portrait busts; Thomas Crawford, who did monumental sculpture; and William Wetmore Story, who lived and worked in Rome, where he was associated with several other prominent 19th-century Americans. Their contemporary, William Rimmer, was a teacher of Daniel Chester French; in the latter part of the 19th century John Rogers became well known for his small groups illustrating literary, historical and humorous subjects. Prominent conservative sculptors in the latter part of the century included Augustus Saint-Gaudens, discussed above; Paul Wayland Bartlett, who lived in France for many years and worked in the French romantic tradition; Frederick William MacMonnies, who also lived in France and is known for such sculpture as the Columbian fountain at the 1893 World's Columbian exposition in Chicago; and Daniel Chester French, sculptor of "The Minute Man" at Concord, Mass., the seated Lincoln in the Lincoln memorial in Washington, D.C., and other works with distinctively American subject matter. Lorado Taft is known for his monumental figures and also for his *History of American Sculpture* (1903; revised editions 1924 and 1930), the first comprehensive work on the subject.

Most of these men were well received in their time but with changing conditions in the 20th century have fallen in critical esteem. It has been suggested that of all U.S. sculptors who began their work before 1900, only two can be regarded as approaching greatness: George Grey Barnard, whose works include the colossal groups at the entrance to the Pennsylvania capitol buildings at Harrisburg, and Gutzon Borglum, known particularly for his colossal portrait sculptures such as the Mt. Rushmore memorial in South Dakota. (X.)

Latin America.—With the coming of Europeans to Central and South America, Indian symbolism and monumentality blended with Renaissance realism, baroque elegance and neoclassic refinement. Indian traits appeared in such European-introduced sculptural forms as the stone crosses that were erected in churchyards; statues, whether by European sculptors or aboriginal pupils, depicted Jesus, the Virgin Mary, saints and occasionally an earthly benefactor of the church. Materials were of wood, plant fibre pulp coated with canvas and gesso, or plaster. The statues often had real costumes and hair, glass eyes and teeth, and extremely realistic flesh, bloody, bruised and torn, with taut muscles and distended veins. Gold halos or crowns were added and costume textures were imitated by the gold-leaf-and-paint estofado technique. Many of these were undoubtedly inspired by paintings brought from Europe.

Few sculptors are known by name from the colonial period and fewer attributions are possible. At least a dozen can be identified in Mexico in the 16th century and twice that number in the 17th; the best known are José Cora of Puebla and his nephew Zacarías, and Gudiño of Querétaro. Many were both sculptors and architects, a necessity of the times. In the 18th century considerable artistic stimulus was provided by the Spanish-born neoclassicist Manuel Tolsa, first director of the academy in Mexico City, first to produce an equestrian statue in the new world (of Charles IV) and teacher of many sculptors of subsequent fame. The second most important artistic centre of the colonial era was Quito, Ecuador, which was known particularly for its decorative sculpture.

Following independence from Spain, sculptors in the new republics of Latin America, using techniques learned in Italy and France, produced hundreds of bronze and marble figures of such public heroes as Bolívar, San Martín and Morelos which adorn public squares in every city. In the late 19th and early 20th centuries, as ties with Spain were renewed, numerous monuments to Christopher Columbus were erected. The best-known single example of Latin-American monumental sculpture is undoubtedly the colossal bronze "Christ of the Andes" by Mateo Alonso of Argentina, erected in 1902 on the Argentina-Chile border to commemorate the peaceful settlement of a boundary dispute. For pre-Columbian sculpture *see* PRIMITIVE ART. (C. I. C.)

VI. MODERN SCULPTURE

1. Rodin's Immediate Successors.—Although the art of Rodin appears conservative in comparison to the painting of the time, in that he continued to use literary themes while painting did not, the new style that he evolved did much to revive sculpture's significance, and his importance to 20th-century sculpture can hardly be overestimated. His fresh search and revelation of the basic movements of modern life had a profound influence on sculptors who followed him.

Among Rodin's contemporaries another significant figure was Edgar Degas, whose sculpture, begun in the 1880s, was an intimate study of movement and light; in several respects it predicts 20th-century developments. Rodin's Italian counterpart, Medardo Rosso, lived in Paris during the 1880s; his work was known and owned by Rodin. Less gifted than Rodin, but interested in the same problems, Rosso used wax in such a way that light was suffused through sensitively modeled portraits and labile forms were created to express the flux that he felt was a condition of modern life. In Italy Russo influenced Arturo Martini and through him Giacomo Manzù, Marino Marini and Alberto Viani.

The ablest of Rodin's many pupils were Émile Antoine Bourdelle and Charles Despiau. Bourdelle's "Heracles" (1910) is an attempt to continue Rodin's active postures but the results are melodramatic and the forms are heavy and less sensitively modeled. Despiau, who was director of Rodin's shop from 1907 to 1914, also responded to the interest in classicism; his best work, "Girl From the Landes" (1904), was a balance of individual traits in the Rodin tradition, combined with graceful poses and well-rounded forms.

Two of the many other young sculptors attracted to Paris by Rodin's fame were Wilhelm Lehmbruck and Constantin Brancusi. Lehmbruck's early work has the soft modeling by touches of clay characteristic of the time, as in his "Mother and Child" (1907) and "Bust of a Woman" (1910). Brancusi's "Sleeping Muse"

(1906) and the small "Bust of a Boy With Head Inclined" (1907) reflect Rodin's later interests in the expressiveness of modeling as opposed to strenuous gesture. Pablo Picasso and Henri Matisse were also early disciples of Rodin, as was Jacob Epstein, particularly in his naturalistic and psychologically incisive portraits.

2. Avant-Garde Sculpture (1909–20).—In the second decade of the 20th century the tradition of body rendering extending from the Renaissance to Rodin was shattered and the Cubists, Brancusi and the Constructivists emerged as the most influential forces. Cubism (*q.v.*), with its compositions of imagined rather than observed forms and relationships, had a similarly marked influence.

One of the first examples of the revolutionary sculpture is Picasso's "Head of a Woman" (1909). The sculptor no longer relied upon traditional methods of sculpture or upon his sensory experience of the body; what was given to his outward senses of sight and touch was dominated by strong conceptualizing. The changed and forceful appearance of the head derives from the use of angular planar volumes joined in a new syntax independent of anatomy. In contrast to traditional portraiture, the eyes and mouth are less expressive than the forehead, cheeks, nose and hair area. Matisse's head of "Jeanette" (1910–11) also partakes of a personal reproportioning that gives a new vitality to the less mobile areas of the face. Likewise influenced by the Cubists' domination of their subject matter. Alexander Archipenko in his "Woman Combing Her Hair" (1915) explored the possibilities of concave areas of the body and replaced the solid head by its silhouette within which there is only space.

Brancusi also abandoned Rodin's rhetoric and reduced the body to its mystical inner core. His "Kiss" (1908), in which two blocklike figures share a powerful life gesture conveying the search for fulfillment, has a concentration of expression similar to primitive art but lacking its demonic power. In this and subsequent works Brancusi favored hard materials and surfaces as well as self-enclosed volumes that give to his subjects an introverted character. Brancusi was also the first artist to carve the bases of his sculptures to realize a poetic continuity, as in his bronze "Bird in Space" (1919), (the piece that became a *cause célèbre* in the 1920s when U.S. customs refused to admit it duty free as a work of art).

Raymond Duchamp-Villon began as a follower of Rodin, but his "Head of Baudelaire" (1911) contrasts with that by his predecessor in its more radical departure from the flesh; the somewhat squared-off head is molded by clear, hard volumes. His famous "Horse" (1914), a coiled springlike form bearing little resemblance to the animal itself, suggests in part the horsepower of locomotive drive shafts and conveys something of the mechanization of modern life. Duchamp-Villon may have been influenced by Umberto Boccioni, one of the major figures in the Italian Futurist movement and a sculptor who epitomized the Futurist love of force and energy deriving from the machine. In "Unique Forms of Continuity in Space" (1913) and "Head + House + Light" (1911) he carried out his theories that the sculptor should model objects as they interact with their environment, thus revealing the dynamic essence of reality.

Jacques Lipchitz came to Cubism later than Archipenko and Duchamp-Villon but after mastering its meaning he produced the best sculpture. In 1913, after several years of conservative training, he made a number of small bronzes experimenting with the compass curve and angular planes. They reveal an understanding of the Cubist reconstitution of the bodies in an impersonal quasi-geometric armature over which the artist had complete autonomy. Continuing to work in this fashion, he produced "Man With a Guitar" (1915), and "Standing Figure" (1915), in which voids are introduced, while in the early 1920s he developed freer forms more consistently based on curves.

Lehmbruck's mature style emerged in the "Kneeling Woman" (1911) and "Standing Youth" (1913), in which his gothicized, elongated bodies with their angular posturings and appearance of growth from the earth incarnate his belief in modern human heroism. In contrast to this spiritualized view is his "The Fallen" (1915–16), intended as a compassionate memorial for friends lost in the war.

3. Constructivism and Dada.—Between 1912 and 1914 there emerged an "antisculptural" movement, called Constructivism, which attacked the false seriousness and hollow moral ideals of academic art. The movement began with the relief fabrications of Vladimir Tatlin in 1913. The Constructivists and their sympathizers preferred industrially manufactured materials, such as plastics, iron and steel, to marble and bronze. Their sculptures were not formed by carving, modeling and casting but by twisting, cutting, welding, or, literally, construction: hence the name Constructivism.

Unlike traditional figural representation, the Constructivists' sculpture denied mass as a plastic element and volume as an expression of space; their compositions were often free transparent grids existing in space continuous with the spectator. In the machine, where the Futurists saw violence, the Constructivists saw beauty. Like their sculptures, it was something invented; it could be elegant, light or complex, and it demanded the ultimate in precision and calculation.

Seeking to clearly express "pure reality" with the veneer of accidental appearance stripped away, the Constructivists created imaginative objects totally devoid of sentiment or literary association; Naum Gabo's work frequently resembled mathematical models, and several Constructivist sculptures, such as those by Kazimir Malevich and Georges Vantongerloo, have the appearance of architectural models. The Constructivists created metaphors or equivalent structures of the new world of science, industry and production; many of the qualities of their sculpture appear in the furniture, architecture and typography of the Bauhaus (*q.v.*).

A second important offshoot of the Cubist collage was the fantastic object or Dadaist assemblage which, though tangent to Constructivism's iconoclasm, opposed its insistence upon rationality. Cubist collages were made from materials lying about in the studio, such as wood, cardboard, nails, wire and paper; examples are Kurt Schwitters' "Rubbish Construction" (1921) and Marcel Duchamp's "Disturbed Balance" (1918). This art generally exalted the accidental, the spontaneous and the impulsive, giving free play to associations. Its paroxysmal and negativist tenor led its subscribers into other directions but Dadaism formed the basis of the imaginative sculpture that emerged in the later 1920s.

4. Conservative Reaction (1920s).—In the 1920s modern art underwent a reaction comparable to the changes experienced by society as a whole. In the postwar search for security, permanence and order, the earlier insurgent art seemed to many to be antithetical to these ends and certain avant-garde artists radically changed their art and thought. Lipchitz' portraits of "Gertrude Stein" (1920) and "Berthe Lipchitz" (1922) return volume and features to the head, but not an intimacy of contact with the viewer. Tatlin and Aleksandr Rodchenko broke with the Constructivists around 1920. Jacob Epstein developed some of his finest naturalistic portraiture in this decade. Rudolph Belling converted from his mechanized "Head" to muscles and identity in his statue of "Max Schmeling" of 1929. Matisse's reclining nudes and the "Back" series of 1929 show less violently worked surfaces and more massive and obvious structuring.

Aristide Maillol continued refining his relaxed and uncomplicated women with their untroubled stolid surfaces; there was something hygienic about his figures that made the title "Bather" redundant. In Germany, Georg Kolbe's "Standing Man and Woman" of 1931 seems a prelude to the Nazi health cult, and the serene but vacuous figures of Arno Breker, Karl Albiker and Ernesto de Fiori were simply variations on a studio theme in praise of youth, body culture, vigilance and low blood pressure. In the United States adherents of the countermovement were William Zorach, Chaim Gross, José de Creeft, Adolph Block, Donald de Lue, Paul Manship and Wheeler Williams.

5. Sculpture of Fantasy (1920–45).—One trend of Surrealist or Fantasist sculpture of the late 1920s and the 1930s consisted of compositions made up of found objects, such as Meret Oppenheim's "Object, Fur Covered Cup" (1936). As with Dadaist fabrications, the unfamiliar conjunction of familiar objects in these assemblies was dictated by impulse and irrationality and could be

summarized by Isidore Ducasse's often-quoted statement, "Beautiful . . . as the chance meeting on a dissecting table of a sewing machine with an umbrella."

Of greater importance as art was the sculpture of a second group that included Alberto Giacometti, Jean Arp, Lipchitz, Henry Moore, Barbara Hepworth, Picasso, Julio González and Alexander Calder. Although these sculptors were sometimes in sympathy with Surrealist objectives, their aesthetic and intellectual concerns prohibited a more consistent attachment. Their art might best be called the sculpture of fantasy. Its sources were in visions, hallucinations, induced responses, reverie and memory, all realized in an intensely personal way. Giacometti's imaginatively derived "Palace at Four A.M." (1933), for example, contrasts with Bernini's "Ecstasy of St. Theresa" in that it does not interpret the artist's vision as if it had occurred in the exterior public world; Moore's series of "Forms" are partial inventions of shapes that interact on each other and space, suggesting life situations. The appeal of primitive and ancient ritual art to Moore, the element of surprise in children's toys for Calder, and the wellsprings of irrationality from which Arp and Giacometti drank were for these men the means by which wonder and the marvelous could be restored to sculpture. While their works are often violent transmutations of life, their objectives were peaceful, ". . . to inject into the vain and bestial world and its retinue, the machines, something peaceful and vegetative." ([Jean] Hans Arp, *On My Way*, Documents of Modern Art, vol. 6, p. 123, George Wittenborn, Inc., New York, 1948.)

6. Other Sculpture (1920–45).—The sculpture of Moore, Gaston Lachaise and Henri Laurens during the '20s and '30s includes mature, ripe human bodies, erogenic images showing an abnormal emphasis upon those parts of the body easily reached by the hands. These figures, reminiscent of Hindu sculpture, appear inflated with breath rather than supported by skeletal armatures. Lachaise's "Montagne" (1934–35) and Moore's several reclining nudes of the '30s are identifications with earth, growth, vital rhythm and silent power. Prior to Moore and the work of Archipenko, Boccioni and Lipchitz, space had been a negative element in figure sculpture; in Moore's "string sculptures" and Lipchitz' "transparencies" of the 1920s, it became a prime element of design.

Lipchitz' figure style of the late '20s and '30s is inseparable from his emerging optimistic humanism. His concern with subject matter began with the ecstatic "Joy of Life" (1927). Thereafter his seminal themes were of love and security and assertive passionate acts that throw off the inertia of his Cubist figures. In the "Return of the Prodigal Son" (1931), for example, strong, faceted curvilinear volumes weave a pattern of emotional and aesthetic accord between parent and child.

Among the sculptors of this period who dealt meaningfully with animal forms as well as the human figure was John B. Flannagan. His interest in what he called the "profound subterranean urges of the human spirit in the whole dynamic life process, birth, growth, decay and death" (quoted in Carl Zigrosser, *Catalog for the Exhibition of the Sculpture of John B. Flannagan*, p. 8, The Museum of Modern Art, New York, 1942) resulted in "Head of a Child" (1935), "New One" (1935), "Not Yet" (1940) and "The Triumph of the Egg" (1941).

Somewhat more mystical are Brancusi's "Beginning of the World" (1924), "Fish" (1928–30) and "The Seal" (1936). As with Flannagan, the recurrent egg form in Brancusi's art symbolizes the mystery of life. Nature in movement is the subject of Alexander Calder's mobiles, such as "Lobster Trap and Fish Tail" (1939) and others of leaves, trees and snow. In the history of sculpture there is no more direct or poetic expression of nature's rhythm.

7. Technical Developments After World War II.—"The modern artist is the counterpart in our time of the alchemist-philosopher who once toiled over furnaces, alembics and crucibles, ostensibly to make gold, but who consciously entered the most profound levels of being, philosophizing over the melting and mixing of various ingredients" (Ibram Lassaw, quoted by Lawrence Campbell in *Art News*, p. 66, The Art Foundation Press, New York, March 1954). While work in the older media persisted, it was the welding, soldering and cutting of metal that emerged after 1945 as an increasingly popular medium for sculpture. The technical and expressive potential of uncast metal sculpture was carried far beyond the earlier work of González and Picasso.

The appeal of metal is manifold. It is plentifully available from commercial supply houses; it is flexible and permanent; it allows the artist to work quickly and it is relatively cheap compared to casting. Modern metals also link sculpture to the stuffs of modern civilization and are emotionally related to that civilization. As the American sculptor David Smith has commented, "Possibly steel is so beautiful because of all the movement associated with it, its strength and functions. Yet it is also brutal, the rapist, the murderer and death-dealing giants are also its offspring" (quoted in Carola Giedion-Welcker, *Contemporary Sculpture*, Documents of Modern Art, vol. 12, p. 123, George Wittenborn, Inc., New York, 1955).

The basic tool of the metal sculptor is the oxyacetylene torch that achieves a maximum temperature of 6,500° F. (the melting point of bronze is 2,000° F.). The intensity and size of the flame can be varied by alternating torch tips. In the hands of a skilled artist the torch can cut or weld, harden and soften, colour and lighten or darken metal. Files, hammers, chisels and jigs are also used in shaping the metal, worked either hot or cold. The sculptor may first construct a metal armature which he then proceeds to conceal or expose. He builds up his form with various metals and alloys, fusing or brazing them, and may expose parts or the whole to the chemical action of acids. This type of work requires constant control and many sculptors work out and guard their own recipes.

Other sculptors such as Peter Agostini, George Spaventa, Peter Grippe, David Slivka and Jacques Lipchitz, who were interested in bringing into sculpture spontaneity, use of accident and automatism, returned to the more labile media of wax and clay, with occasional cire-perdue casting, which permit a very direct projection of the artist's feelings. The end result is often the creation of an informal image, bordering on the shapeless, in contrast to the often clean-cut shapes of metal sculpture. By the nature of the process the work is on a small scale.

Another modern phenomenon, seen particularly in Italy, France and the United States, is the revival of relief sculpture and its execution on a large scale to produce what might be called "environmental sculpture," intended to stand alone rather than in conjunction with a building. Louise Nevelson in her "Sky Cathedral" (1958), for example, employs boxes as container compartments in which she carefully disposes an assortment of forms and then paints them a uniform colour. In Europe the outstanding metal reliefs were those by Alberto Burri, Gio and Arnaldo Pomodoro, César Baldaccini, Zoltán Kemény and Manuel Rivera.

8. New Views of Nature.—Development of metal sculpture particularly in the United States led to fresh interpretations of the natural world. In the art of Richard Lippold and Ibram Lassaw, the search for essential structures took the form of qualitative analogies. Lippold's "Full Moon" (1949–50) and "Sun" (1953–56), the latter commissioned by the Metropolitan Museum of Art, New York, to hang in its room of Persian carpets, show an intuition of a basic regularity, precise order and completeness that underlies the universe. Lassaw's comparable interest in a galactocentric universe inspired his "Planets" (1952) and "The Clouds of Magellan" (1953).

In contrast to the macrocosmic concern of these two artists were the interests of sculptors such as Raymond Jacobson, whose "Structure" (1955) derived from his study of a honeycomb. Using three basic sizes, Jacobson constructed his sculpture of hollowed cubes emulating the elementariness, the general regularity but slight unpredictability and the surprising variety of the honeycomb.

Isamu Noguchi's "Night Land" is one of the first landscapes in sculpture. David Smith's "Hudson River Landscape" (1951), Theodore J. Roszak's "Recollections of the Southwest" (1948), Louise Bourgeois' "Night Garden" (1953) and Leo Amino's "Jungle" (1950) are later examples.

9. The Human Figure.—The variety of human forms in mod-

ern sculpture results in part from the intimate, personal nature of the sculptor's art and from the example of preceding artists in moving away from the imitation of the body. The thin, vertical, Etruscan idol-like figures developed by Giacometti after World War II showed his repugnance toward rounded and smooth body surfaces or strong references to the flesh. His men and women do not exist in felicitous concert with others; each form is a secret sanctum, a maximum of being wrested from a minimum of material. Reg Butler's work (such as "Woman Resting," 1951) and that of David Hare ("Figure in a Window," 1955) treat the body in terms of skeletal outlines. Butler's figures partake of nonhuman qualities and are fantasies of an unsentimental and aggressive character; the difficulties and tensions of existence are measured out in taut wire armatures and constricting malleable bronze surfaces. Kenneth Armitage and Lynn Chadwick, two other British sculptors, make the clothing a direct extension of the figure, part of a total gesture. In his "Family Going for a Walk" (1953), for example, Armitage creates a fanciful screenlike figure recalling wind-whipped clothing on a wash line. He is intrigued by flatness and takes pleasure in inducing wonder as to what is on the other side. Both Chadwick and Armitage transfer the burden of expression from human limbs and faces to the broad planes of the bulk of the sculpture. Chadwick's sculptures are often illusive hybrids suggesting alternately impotent De Chirico-like figures or animated geological forms.

Marino Marini and Luciano Minguzzi admire the amply proportioned feminine form. Minguzzi's women (*e.g.*, "Woman Jumping Rope," 1954) may exert themselves with a kind of playful abandon. Marini's (*e.g.*, "Dancer," 1949) enjoy a stately passivity, their quiescent postures permitting a contrapuntal focus on the graceful transition from the slender extremities to the large, compact, voluminous torso, with small rich surface textures.

The segmented torso, popular with Arp, Laurens and Picasso earlier, continued to be reinterpreted by Alberto Viani, Bernard Heiliger, Karl Hartung and Raoul Hague. The emphasis of these sculptors was upon more subtle sensuous joinings that created self-enclosing surfaces. Viani's work, for example, does not glorify body culture nor suggest macrocosmic affinities as does an ideally proportioned Phidian figure; his torsos are seen in a private way, as in his "Nude" (1951), with its large body and golf-ball breasts.

Perhaps the most impressive figure sculptures made in the U.S. between 1955 and 1960 were those by Seymour Lipton. Their large-scale, taut design and provocative interweaving of closed and open form restore qualities of mystery and the heroic to the human in a language and conception that seems right for an era of metals and a complex society.

10. Archaizing, Idol Making and Religious Sculpture.—After World War II several sculptors became interested in the art of early Mediterranean civilizations. The result was a conscious archaizing of the human form with the intent of recapturing qualities of Cycladic idols, early Greek and Egyptian statuary and some aspects of late Roman art.

Moore's admiration for archaic Greek sculpture produced "Draped Reclining Figure" (1952), which shows his return to the solid form and the suggestion of power and force by using drapery as a tense foil for the volumes that press against it. His "King and Queen" (1952–53) resulted from further excursions into the archaic Greek myth world.

The interest in recreating idols or totems was continued by Arp in his "Idol" (1950) and by Noguchi in his Stone Age-type sculptures for the Connecticut General Life Insurance company (Hartford, Conn.). By creating presences that elude rational definition, these artists restore to art its ancient and medieval aura of myth, mystery and magic in an age that consistently disclaims their existence.

The argument that modern sculpture is inappropriate for religious requirements is disproved by works of Lipchitz, Lassaw and Herbert Ferber. In keeping with the Jewish preference for nonfigural art, Ferber's ". . . and the bush was not consumed" (1951), commissioned by a synagogue in Millburn, N.J., comprises clusters of branches and boldly shaped weaving flames, invisibly suspended in a powerful and intimate vision that absorbs its viewers with its hypnotic rhythm. Lassaw's "Pillar of Fire," for the exterior of a synagogue in Springfield, Mass., also has a mesmerizing pattern recalling the induced images seen when gazing into fire. Lipchitz' sculpture of the "Virgin of Assy" (1948–54) was commissioned for the Catholic church at Assy, France.

Moreover, an increasing number of gifted sculptors are providing handsome liturgical objects and decorations, such as Harry Bertoia's shimmering reredos, Lipton's work for a synagogue in Tulsa, Okla., and Roszak's sculptured spire for Kresge chapel on the campus of the Massachusetts Institute of Technology, Cambridge.

11. Public and Private Memorials.—After World War II there was a flood of public memorial sculpture, and in Europe especially many of the commissions were carried out by modern sculptors. A striking war memorial in Italy is Mirko Basaldella's gate for the monument to the Roman hostages killed in the Ardeatine caves (1951). For its full effect the gate must be seen in connection with the rugged masonry wall to which it is attached. The gate is cast in metal and fashioned in a tangled, thicketlike pattern which suggests the painfully difficult passage from life to death for those who died in the caves.

Another imposing memorial is Ossip Zadkine's monument to the bombing of Rotterdam, a figure recoiling from the violence that descended from the sky. Moore's "Warrior With a Shield" defiantly raises his shield and mutilated body toward the ill-starred heavens during the battle of Britain. Epstein's public monument to "Social Consciousness" (1952–53), in Fairmount park, Philadelphia, treats the helplessness of those confronted with pressures over which they have no control. In contrast to the invulnerable champions of academic art, these sculptures image the hero in distress.

12. Other Developments.—Despite the rapid and exciting developments in both architecture and sculpture the two have seldom been meaningfully and integrally united. The architecture of Le Corbusier, Frank Lloyd Wright, Pier Luigi Nervi, Ludwig Mies van der Rohe and others occasionally shows strong sculptural qualities, but relatively rarely were their surfaces planned to receive sculpture. Free standing sculptures such as those created by Gabo, Pevsner, De Rivera, Calder and Noguchi have been used to provide intimacy and visual relief from the severity of the "cult of the cube" in architecture. The architectural firm of Skidmore, Owings and Merrill successfully used Bertoia's brilliant screens and Noguchi's sculptures and garden ideas; Roszak's "Eagle" for the American embassy in London and Moore's changeable reliefs on the London Time and Life building held out hope for further thoughtful integration of the arts.

Also of great moment is the phenomenon of the sculptor-designer who has produced important changes in furniture and industrial design. Max Bill's school in Ulm, Ger., showed great promise. Playground facilities have been revolutionized by such designs as those made by Noguchi for Creative Playthings Inc. in the U.S. and the slides, hollowed forms and climb apparatus of Egon Moeller-Nielson for parks in Stockholm, Swed. Noguchi, Moholy-Nagy, Bill, Bertoia and many other modern artists contributed to the breakdown of the distinction between the object of utility and the work of art. Not since Gothic times has sculpture shown such promise of becoming an extensive and important part of human existence.

See also references under "Sculpture" in the Index.

(A. E. El.)

Bibliography.—Sir Herbert Read, *The Art of Sculpture* (1956; 2nd ed., 1961); H. W. Janson and D. J. H. Janson, *The History of Art* (1962); Carola Giedion-Welcker, *Contemporary Sculpture: an Evolution in Volume and Space*, 3rd ed. (1961). For more comprehensive treatments *see* Sir Herbert Read and H. D. Molesworth (eds.), Acanthus "History of Sculpture Series" (1961–), published by the New York Graphic Society, and volumes in the "Pelican History of Art Series." *See* also bibliographies of articles referred to in the text.

SCULPTURE, SEPULCHRAL, a term descriptive of the figures of persons commemorated by sculpture or by linear cutting or by engraving on stone or metal slabs, either at floor level or on raised structures. The whole commemorative work may be a cenotaph or may contain or cover an interment. Among the

many ancient types of sepulchral monument the Etruscan form of the 6th and 5th centuries B.C., represented in the Flavian period by the Ulpia Epigone, is the nearest forerunner of the later Christian type in 15th-century Italy. There are good examples of Etruscan tomb monuments in the British museum but full-scale recumbent effigies of the dead were not a feature of Greek and Roman commemoration.

Medieval Effigies.—Full-scale recumbent effigies did not reappear in Europe until toward the end of the 11th century, when very interesting examples were found in Germany. There is a fine bronze grave slab of 1080 at Merseburg cathedral, and the dignified effigy of Wittekind, duke of Saxony, at Bielefeld, is late 11th century. Other interesting early examples of this period are at Quedlinburg and Magdeburg (1100–1300). There are three fine grave slabs of abbesses showing stiff and highly stylized drapery with long, hanging sleeves, at Quedlinburg. These slabs have delicate and well-carved conventional borders and their dates are between 1130 and 1150.

Among the very early ecclesiastical effigies in the cloisters at Westminster abbey is that of Abbot Crispin who died in 1117. It is of black Tournai marble and is believed to be the oldest recumbent effigy in England. Westminster abbey is extremely rich in the finest examples of most styles and periods, and the range of monumental art in and after the middle ages makes it the best single church in the world for the study of the subject.

In all countries and periods the figures representing kings, nobles, ladies, ecclesiastics, warriors, children, merchants, etc., are of the utmost value in the study of contemporary costume and armour and are only rarely archaic presentations of earlier styles; enthusiasm for antiquity revived Roman armour, togas, etc., in

(LEFT, ABOVE) JOHN FREEMAN, LTD.; (BELOW) BBC

DEATH MASKS OF (LEFT) HENRY VII AND (ABOVE) EDWARD III: (BELOW) DETAIL OF A BRONZE EFFIGY OF HENRY III IN WESTMINSTER ABBEY

the 17th and 18th centuries. Early history is that of the gradual emancipation from the flat slab shape which gave way to progressively higher relief toward the later 12th and mid-13th centuries; there are many examples in Germany and fewer surviving in Italy and France, but the much restored although still important series of royal tombs in the abbey of St. Denis, north of Paris, shows a further advance toward natural attitudes and finer treatment of drapery.

Germany was more conservative in adherence to the slab type but the English evolved splendid and vigorous figures of ecclesiastics, worked often in Purbeck marble, and also the famous mail-clad warriors sometimes in attitudes of partial activity or even of spirited action. The crossing of the legs has been wrongly supposed to be connected with the crusaders. There are fine examples of 12th- and 13th-century effigies at Ely, Salisbury, Wells and at other cathedrals, abbeys and parish churches.

In Italy, the Cosmati family made elaborate tombs with architectural canopies inlaid with mosaic and marble; beneath these, on a tomb-chest or a structure formed like a sarcophagus, there was often a recumbent presentation of the dead person coloured to show rich vestments. In 13th-century England the decoration was not achieved by mosaic but by painting and this became general by about 1300.

In France, apart from St. Denis, the medieval tomb-effigies are considerably rarer than in England. English canopied tombs, with the effigies raised on a tomb chest or rectangular structure for adding dignity to the design, are sometimes of quite elaborate architectural form; good examples are Archbishop de Gray at York and Bishop Giles of Bridport at Salisbury.

The 14th century was a period of great charm in sculptural developments and the richer tombs became soaring, pinnacled structures forming canopies over finely worked effigies in stone, wood or fine alabaster; magnificent examples are at Canterbury, York, Gloucester and, above all, at Westminster. Most medieval effigies have appropriate animals at their feet.

The opening of the century gave the superb royal effigies in gilt bronze in the Saints' chapel at Westminster; their contrasts with the Italian Renaissance figures of Henry VII and his queen in the same church and in the same material are important in the study of the different art forms. Another magnificent later example in bronze is Richard Beauchamp, 1454, at Warwick dated between the early Westminster bronzes and the Henry VII tomb by Torrigiano.

There are good examples of warriors in Exeter (1320) and Bristol (1300) and of ladies at Alnwick (1330) and at Westminster. Brasses abound and there is an especially fine group at Cobham, Kent. An English sculptor was employed for a papal tomb at Avignon. Very beautiful examples in stone are at Ewelme and in the Holland monument which was moved after World War II to St. Peter ad Vincula in the Tower of London.

The later effigies in the period are recumbent with the hands together in prayer, but there are a few interesting exceptions as at Tewkesbury where the effigy of Lord Edward le Despenser (1375) kneels with the hands in prayer under a high canopy on the roof of a chantry chapel. The removal of late coats of paint and varnish by R. P. Howgrave-Graham shortly before and after 1940 revealed a completely natural figure with fine lifelike face colouring and the eyes directed toward the altar. The attitude is unique at such an early date in England. In Bakewell, there is a very unusual wall monument showing only the upright busts of a knight and his lady in a canopied niche.

Weepers, sometimes representing relatives, must be included as effigies and were in some cases portraits, as seems probable in the gilt bronze statuettes on the tomb of Edward III at Westminster. These weepers surround monuments in standing attitude but are not necessarily shown mourning. About 1400 was produced the highly ornate tomb of Philip the Bold, duke of Burgundy, now in the Dijon museum, where intensely dramatic professional weepers were introduced for the first time by Claus Sluter.

The tomb of Philippe Pot, seneschal of Burgundy, in the Louvre, is in painted stone and brings into monumental sculpture the element of motion; the armoured effigy lies in the quiescent atti-

MEDIEVAL EFFIGIES

(Left) Bronze grave slab of Rudolf von Schwaben, in Merseburg cathedral, 1080; (above) Richard Beauchamp, in Warwick church, 1454 (a full-length recumbent figure); (centre bottom) kneeling figure of Edward le Despenser, in Tewkesbury abbey, 1375; (right) weepers from the tomb of Philip the Bold, now in the Dijon museum, about 1411 (eighteen inches high)

(LEFT) BILDARCHIV FOTO MARBURG; (TOP) COMMON GROUND, LTD., (CENTRE BOTTOM) NATIONAL BUILDINGS RECORD; (RIGHT) R. REMY

tude of prayer on a slab which is supported on the shoulders of bearers hooded and gowned as mourners and seeming to march in a funeral procession. They carry shields of arms. The date of this is about 1480.

The Renaissance.—In 15th-century Italy the architectural format of the Cosmati was transformed into a vehicle for relief sculpture of subtle and gentle beauty by such men as Agostino di Duccio; yet the same century, so curiously marked in much of its art by morbid insistence on death and bodily decay, emphasizes this with dramatic grimness, especially in northern Europe. The dead person appears in the full pomp of active life on a hollow architectural structure with openings making visible inside a cadaver, that is, a lean and emaciated figure or skeleton with decayed and shrunken flesh attached and with reptiles and worms infesting it. The contrast is enhanced by the dimness of light and the placing of the cadaver almost at floor level. The cadaver monuments are almost wholly confined to ecclesiastical dignitaries.

Classical grandeur and elaboration in monuments was developed fully in Italy long before anything comparable to it appeared elsewhere. Rich and costly marble was easily available and was of the utmost value in Renaissance exuberance. Symbolism and allegory appeared early in Italy in the Renaissance period, and the monument of Francis II, duke of Brittany, and his duchess (1502–07) in Nantes cathedral is a fine French example. The effigies on it are delicate in treatment and appear to be portraits. Very fine statues at the four corners symbolize Justice, Force, Prudence and Temperance. In Italy, the monuments of the great attain to extraordinary magnificence and the effigy tends to be a glorification of the man, sometimes without much of the earlier religious solemnity and devotion. He may even be seated on a life-sized horse raised high on, and within, elaborate erections of marble.

At the end of the 16th century and early in the 17th century, the fantastic and opulent splendour of Elizabethan and Jacobean monumental art was based on classical motives and the effigies were brightly painted as to faces and hands, armour and costumes being also highly coloured. The great and wealthy as well as ordinary merchants and citizens in England and France may kneel on their tombs: man and wife on opposite sides of a lectern below which the family of boys and girls pray, ranged in two rows of graduated size according to age, the boys on one side and the girls on the other. Each of these figures is a true monumental effigy and in some cases may be a portrait. Some of the grandest examples are in Westminster abbey where the tombs in the aisles of Henry VII have been brought back to their original almost gaudy splendour.

The whole concept of the treatment of monumental sculpture was changed by the powerful incidence of Michelangelo in the Medici chapel in Florence where the two figures sit in niches above and symbolic figures recline on broken pediments. While old forms continued, the new concepts gained ground in northern Europe. A bust of the deceased in a niche was common and as time went on the figures would sit, like Bernini's popes, recline on one elbow, swoon, stand or float heavenward encouraged and supported by lachrymose female figures or mourned by "well-to-do cherubs" weeping small bunches of marble tears.

A new symbolism surrounds the effigy which is often a finely sculptured and realistic portrait but may be reduced to a mere low-relief medallion cherished by a tearful female figure. Sculptural symbolism emphasizes death, loss and grief and is seldom religious in feeling, though epitaphs may be full of genuine piety.

The Neoclassical Revival.—In the neoclassical revival, represented in Italy by Antonio Canova (1757–1822), in England by John Flaxman (1755–1826) and in Denmark by Bertel Thorwaldsen (1770–1844), there was a chastened sobriety and increased reserve. (The 19th century, however, while giving examples of dignified and reverential treatment with careful portraiture, was often distinguished by vapidity and poor sentiment.) The depth to which the sculptured effigy could sink in the 18th century is seen in the Westminster monument of Sir Cloudesley Shovel; but Westminster contains a wealth of fine sculpture by famous 17th- and 18th-century artists who are fully dealt with by K. A. Esdaile (see *Bibliography*).

Portraiture in Monumental Effigies.—The earliest occurrence of portraiture in monumental effigies has been a controversial subject, and opinion has tended toward the belief that all tomb effigies were merely conventional until the 14th century was well advanced. It may be assumed that early portraiture was confined to the great and wealthy and that the numerous effigies turned out commercially from sculptural workshops were made to pattern, with minor variations as in the monuments on which they were placed. Notable disputed instances of early date are the superbly designed English effigies in cast bronze of Henry III who died in 1272 and of Eleanor of Castile (1291), both made by William Torel, a London goldsmith. The king's face is full of characterization which harmonizes with his known personality and, allowing for some stylization, it seems to show very clear elements of portraiture. Against this is the argument that it was made long after his death, but this is invalidated by the discoveries made about 1950. These discoveries were made during the restoration by Howgrave-Graham of the earlier funeral effigies that lay in state for royal obsequies in England and France and had always been accepted as the unimportant conventional handwork of craftsmen.

It was found that the faces of these effigies were of two kinds, two being cast from molds taken from the actual face very soon after death. The others were carved in wood from such death masks and all were finished and painted to simulate full life. At Westminster, these effigies were kept by the monks and so were available to the makers of the monuments and would naturally be used as models. The French effigies were preserved in a similar way by the monks at St. Denis, where the monarchs were buried. The view of Sir W. St. John Hope that a funeral effigy of Henry III was made would seem to suggest that it was the basis for Torel's bronze and that the tomb effigy is a somewhat stylized portrait in which his drooping eyelid was naturally ignored.

Portraiture seems likely in the effigy of Edward II in Gloucester cathedral, and the probability of portraiture as a frequent aim, at least of the great and the wealthy, is enforced by the skill and the interest in physiognomy found in portrait-corbels of 15th-century craftsmen in Lincoln and Durham cathedrals, in Westminster abbey and elsewhere.

In early monumental brasses, where the effigy is purely linear, portraiture is rare but it seems more usual in the incised stone and marble slabs that were made in northern France and are remarkably plentiful in the cathedral at Châlons-sur-Marne. (*See* SARCOPHAGUS; TOMB; SCULPTURE.)

Death Masks.—A death mask is a wax or plaster cast of a mold taken from a dead face. Death masks are true portraits, only requiring the correction of postmortem changes in the eyeballs. From ancient Egyptian times they have been made as aids to portrait-sculptors, and latterly as mementos of the beloved dead.

In medieval France and England actual death masks were used for the royal funeral effigies that lay in state, but the Westminster abbey series is unique, as the revolutionaries destroyed those in France. Some at Westminster served as models for wood carvers, but R. P. Howgrave-Graham's discoveries show that two are actual masks applied to the effigies.

The mask of Henry VII is probably the finest in existence, while that of Edward III is the earliest European example; it records the facial distortion due to his fatal stroke. From the 13th century, death masks provided models for the sculptors of tomb effigies.

BIBLIOGRAPHY.—A. P. C. Michel, *Histoire de l'art depuis les premiers temps chrétiens jusqu'à nos jours,* 8 vol. (1905–29); G. S. Davies, *Renascence. The Sculptured Tombs of the 15th Century in Rome, etc.* (1910); John W. Pope-Hennessy, *Italian Gothic Sculpture* (1955); G. Higgins, "On the Work of Florentine Sculptors in England in the Early Part of the 16th Century" (specially valuable for Torrigiano and Benedetto da Rovezzano), *Archaeological Journal* (Sept. 1894); G. Ferrari, *La tomba nell' Arte italiana* (1917); A. Venturi, *A Short History of Italian Art,* trans. by E. Hutton (1926); W. Bode, *Florentine Sculptors of the Renaissance* (1902; Eng. trans. 1928); A. F. Blunt, *Art and Architecture in France, 1500–1700* (1953); A. Gardner, *Medieval Sculpture in France* (1931); M. Aubert, *La Sculpture française du moyen âge et de la renaissance* (1926); P. Vitry and G. H. Brière, *Documents de sculpture française* (1910); F. Ingersoll-Smouse, *La Sculpture funéraire en France au 18ième siècle* (1911); A. Gardner, *English Medieval Sculpture,* with valuable bibliography (1951); E. S. Prior and A. Gardner, *An Account of Mediaeval Figure-Sculpture in England* (1912); Joan Evans, *English Art 1307–1451* (1949); F. H. Crossley, *English Church Monuments, A.D. 1150–1550* (1921); A. C. Fryer, *Wooden Monumental Effigies in England and Wales,* rev. ed. (1924); L. Stone, *Sculpture in Britain: the Middle Ages* (1955); M. D. Whinney and O. Millar, "English Art, 1625–1714," *Oxford History of English Art* (1957); K. A. Esdaile, *English Monumental Sculpture Since the Renaissance* (1927); R. Gunnis, *Dictionary of British Sculptors, 1660–1851* (1953); R. Kautzsch, *Der Mainzer Dom und seine Denkmaler* (1925); P. M. Halm, *Studien zur süddeutschen Plastik* (1926); Karl Gröber, *Alte Kunst in Bayern* (1924); E. L. Fischel, *Mittelrheinische Plastik des 14 Jahrhunderts* (1923); H. von Beenken, *Romanische Skulptur in Deutschland* (1924).

(R. P. H.-G.)

SCULPTURE TECHNIQUES. The two basic methods of creating sculpture are the direct method—that is, by direct carving—and modeling, in which the piece is modeled in clay, or similar material, and is then reproduced in plaster, bronze or some other material.

This article briefly describes modeling, casting in plaster and bronze, stone carving, wood carving and certain modern techniques. These are the techniques used with the more common materials of sculpture. Clay modeling as described here is that done preliminary to casting; for sculpture in clay itself *see* TERRA COTTA. *See* also METALWORK, DECORATIVE: *Techniques of Metalworking.*

MODELING AND CASTING

Modeled Sculpture.—Generally, all sculpture that is to be cast in bronze or other metals or materials is first modeled in moist clay, plasteline or wax. Clean moist clay is preferred by the majority of sculptors.

In the time-honoured technique for modeling a small massive unit of sculpture, the sculptor kneads his moist clay until it has a smooth, consistent texture. He rolls chunks of clay into a number of loaf shapes and piles them up near his modeling stand, which is equipped with a turntable. He begins by bending and squeezing a few pieces of clay to form the main geometric shapes of his planned sculpture. He then breaks off smaller pieces of clay from his remaining clay pieces and builds up his sculpture shape, turning his clay model as he works and studying it from all angles so that the whole unit develops consistently. Detailed surface shapes and textures are developed until the clay model is finished.

FROM LOUIS SLOBODKIN, "SCULPTURE PRINCIPLES AND PRACTICE," © 1948, WORLD PUBLISHING CO.

ARMATURE FOR A STANDING MALE FIGURE WITH SOME CLAY ATTACHED

In larger sculpture an inner structure called an armature is built to support the clay, which otherwise would collapse from its own weight. A few blocks of wood nailed together, a heavy lead pipe or a galvanized iron pipe secured to a baseboard will serve as the armature for a life-sized head or similar sculpture unit. Larger pieces of sculpture are supported by more complicated armatures constructed of lead pipe, iron rods or pipes and wood. A combination of these materials is used in the huge armatures for monumental sculpture.

Clay models are sprinkled with water to keep the clay pliable while the models are being developed, and they are carefully wrapped in moist cloths between sessions of work. Plasteline and wax remain pliable without water and are the favoured materials for fragile, perforated sculpture units and for preparatory sketches for large sculpture. Plasteline

used for large sculpture does not need as much supporting armature as moist clay does.

The sculptor's most important modeling tool is his own thumb. Polished hardwood blades (carved from boxwood and similar fine-grained woods) are also used to apply clay. Bent wire tools doweled to wooden handles are used to help pull the surface of the clay together and for cutting away excess clay or gouging into the surfaces. While the main mass of clay shape is being organized the sculptor may use a crude block of wood or a paddle to pat and block his clay into shape.

When the model is completed it is cast in plaster, either by a professional plasterer or by the sculptor himself. (Ls. S.)

Plaster Casting.—The three methods of plaster casting are the waste mold, the gelatin mold and the piece mold. The waste mold is the simplest and is used where only one cast of the model is desired. (In the process the mold is destroyed.) Both the gelatin mold and the piece mold can be reused; the piece mold is the most complicated to make but gives an exceedingly accurate cast.

In all plaster casting the model is first covered with a coat of shellac, so that the mold does not stick to the clay.

FROM ARNOLD AUERBACH, "MODELLED SCULPTURE AND PLASTER CASTING"; ELEK BROOKS, LONDON

PLASTER CAST BY ARNOLD AUERBACH USING THE WASTE MOLD PROCESS: (TOP LEFT) CLAY MODEL WITH DIVISION LINE MARKED WITH METAL; (TOP CENTRE) COMPLETED MOLD; (TOP RIGHT) BACK HALF REMOVED; (BOTTOM LEFT) FRONT MOLD OPENED AND CLEANED; (BOTTOM, CENTRE AND RIGHT) PROGRESSIVE STAGES IN CHIPPING THE MOLD FROM THE CAST

Waste Mold.—Division lines are first made on the clay model with small pieces of brass, or a thin clay wall, so that the mold will be marked for cutting when it is finished. This line of division is across the top of the head and behind the ears. The model is then covered with a coating of coloured plaster, made by adding bluing or dry colour to the water with which the plaster is mixed. (Later, when the molder is chipping the plaster mold from the final cast, this coating warns him that he is near the cast and must chip carefully to prevent cutting into it.) Over the first coating of coloured plaster (about $\frac{1}{8}$ in. thick), a thicker coating of white plaster is applied. If the piece is large or has parts that would break easily, this coating is reinforced with other materials. When this coating has set, the mold is cut along the marked division line and is opened. The clay of the model is dug out carefully in order not to damage the mold. Then the mold is washed, soaped with green soap (melted in hot water) and allowed to cool. All the soap is carefully removed from the mold and a light coating of olive oil is applied to insure the proper separation of the mold from the final cast.

Various methods are used for the final casting. In one method used for a head, for example, the sections of the mold are put together and held in place by scrim dipped in plaster or tied with wire. The mold is then turned upside down and plaster of the proper consistency is poured in. If the model is small it may be entirely filled, thus forming a solid cast. Larger pieces may be cast hollow, the plaster being evenly distributed by continually turning the mold. A thickness of $\frac{3}{4}$ in. is sufficient to make a strong cast, especially if it is reinforced on the inside with burlap dipped in plaster or with iron pipes. When the plaster cast has set properly, the outer wrappings are removed and the plaster mold is chipped away with a mallet and chisel. The final cast may be retouched by wetting the cast and applying to it plaster of a thinner consistency than the casting plaster.

Gelatin Mold.—This procedure is much more complicated than waste molding: proper temperature of the gelatin and correct timing in opening the mold at every stage of the operation are essential. In addition, the sections of the cast must be keyed to each other with great precision.

A layer of clay filler, about $\frac{3}{4}$ in. thick, is first placed over the model, with a clay wall marking the line of division. This layer will subsequently be replaced by gelatin, so it must be even and without weak places.

Thereafter, the procedure, briefly described, is as follows. An outer shell of plaster is made over the filler, and when it has set the shell is opened and the clay filler removed. The sections of the outer shell are replaced in their original position around the model and are bound together. Casting gelatin is poured in through a hole in the top, filling the space between the model and the outer shell. When the gelatin has cooled, the shell is opened and the clay of the model is cleaned from the gelatin mold. The gelatin mold is then greased, and, with the shell to hold it, is put together again; then the plaster, mixed to the proper consistency, is poured into it. When the plaster has set, the mold is opened and lifted away from the cast. After the mold is cleaned and greased it can be reused for casting.

Piece Mold.—As the name implies, a piece mold is made in small pieces which are then reassembled for the final casting. The process is very complicated and can be done only by persons who are skilled in working with plaster. The caster begins by determining how many pieces are required for the model he is working on, and then he makes them, one at a time, using a wall of clay or plasteline to mark off the area of each piece. The pieces, which are made of plaster, are then reassembled and an outer shell is

made to hold them in place while the final plaster cast is being made. After each casting the pieces are cleaned and oiled in preparation for the next casting. Of the three methods of plaster casting, this is the most accurate. Its other advantages are that the mold may be reused and will last indefinitely, where gelatin molds deteriorate in a short time.

Finishing and Colouring.—The plaster cast may be either an intermediate stage or the final stage of the piece of sculpture. If the cast represents an intermediate stage, it is later cast in bronze, or some other material, or reproduced in stone, as described in following sections of this article. If the plaster cast is the final stage, it is finished to resemble bronze. Of the various methods of doing this, the most common is the oil-paint method. One coat of shellac is applied as a prime. After this has dried, a coat of brown oil paint is applied and allowed to dry. If a greenish blue tint is desired, a very thin mixture of light-green oil paint is applied to the brown coat. When this is almost dry it is wiped off lightly with a rag, allowing the green colour to remain in the deep parts. For a metallic effect, a little gold or bronze powder mixed with dryer may be applied to the high parts. When the piece is thoroughly dry it is rubbed with wax, which subdues any unpleasant shiny effect. Similar procedures may be used to apply any other desired colour.

Bronze Casting.—Next to stone, the most popular medium with sculptors throughout the ages has been bronze, an alloy made up principally of copper, with tin, zinc and lead in varying proportions. (The most common formula is 85% to 90% copper and from 10% to 15% tin, zinc and other nonferrous metals.) The popularity of bronze is due to its permanence and the freedom of expression it gives the artist. Its tensile strength is high, which makes possible a movement that would be impossible in stone; it is comparatively light and therefore adaptable to many uses; and it affords a wide variety of surface finishes.

Metal casting is a highly technical operation that requires much special equipment and is done in foundries by skilled craftsmen who usually each specialize in one phase of the operation. There are two methods, the *cire-perdue*, or lost-wax, method and the sand process. *Cire-perdue*, the older method, is considered especially good for reproducing faithfully delicate and intricate detail, but the results obtained from either method depend to a very large extent upon the skill of the craftsman.

"Cire-Perdue" Casting.—After the foundry receives the piece of sculpture (usually in the form of a plaster cast), the first step is the making of a gelatin mold. For smaller sculptures, this mold is made of the whole piece; for larger sculptures, the model is divided into sections and a mold is made of each section. The gelatin mold, which bears the details of the model in reverse, is then cleaned and made ready for the wax cast.

For the wax cast, the inside of the gelatin mold is painted with molten wax which is then reinforced with sheets of warm wax pressed against it. The thickness of the wax must be controlled so that it does not exceed the desired thickness of the final cast in bronze. (The bronze in a one-foot-high statuette, for example, is about ⅛ in. thick; for larger pieces the thickness increases but seldom exceeds ½ in.) The gelatin mold is then taken off, leaving a hollow wax replica of the sculpture, which the sculptor may retouch if he wishes. At this stage, gates and vents in the form of wax rods are attached to the figure in such a way as to carry off the melting wax during the baking of the final mold.

The final mold is made of a heat-resisting semiliquid silica compound poured into the wax mold to form the core and built up around the outside to form a jacket. This entire mass is put into a brick fire chamber and baked for two or three days until the wax has melted away. The mold is then removed from the oven and is packed in sand and earth in a floor pit. The molten bronze is poured from crucibles into a large opening in the mold and runs through the gates, filling the space left empty by the melting of the (lost) wax cast. When the bronze has cooled, the silica jacket is broken away, the core is tapped out and the bronze is dipped in acid for cleaning. The piece is then touched up with hand tools, by a craftsman or by the sculptor himself. At this stage the seams and rods of crude bronze formed by the gates and vents are removed and any defects in casting are dealt with. The piece is then ready to receive its patina, as described below.

Sand Process.—This is similar to the piece-mold process of plas-

FROM JULES STRUPPECK, "THE CREATION OF SCULPTURE": HOLT, RINEHART AND WINSTON, INC.

METAL FIGURE CAST BY JULES STRUPPECK USING THE CIRE-PERDUE METHOD: (TOP LEFT) THE MODEL AND ITS MOLD, SEPARATED AND CLEANED. (TOP RIGHT) WAX MODEL WITH RODS ATTACHED. (BOTTOM LEFT) THE FINAL MOLD; (BOTTOM CENTRE) THE CAST AS IT COMES DIRECTLY FROM THE MOLD. AFTER THE RODS OF CRUDE METAL ARE REMOVED, THE FIGURE IS REWORKED AS NECESSARY AND A PATINA IS USUALLY APPLIED. (BOTTOM RIGHT) FINISHED FIGURE

ter casting. The mold is made by gently hammering against the sculptor's original model a very fine damp French sand, a composition of clay, silica and alumina, which when it dries becomes quite hard. The mold is made in two halves—usually the front and the back of the piece. For each half, the mold is divided into several pieces, the size and shape of which are governed by the shape of the model. These pieces, which are usually about two inches in thickness, are removed from around the plaster model and are reassembled in a containing iron frame called a flask. The flasks for each half of the sand mold will fit together exactly, and when assembled are held together tightly by clamps and bolts.

Inside each sand mold there is built up a core of sand, so shaped as to leave space for the bronze between the core and the mold (or the core may be made exactly the size of the mold and then shaved down). The core is suspended inside the mold with pins going through the core and the mold. Grooves are made for the bronze to enter, and gates and vents are added for escaping gases. All of this—mold and core—is baked for several hours until it is hard. Molten bronze is then poured in, filling the space between core and mold. When the bronze has cooled, the sand mold is removed and the finishing proceeds as in *cire-perdue* casting.

The sand process is frequently used for very large sculptures; these are cast in sections which are then welded together.

Patina.—Bronze will naturally become dark and beautiful, but because this requires a long period of time most sculptors colour their casts. The copper and copper compounds in the bronze make possible many variations of green, red, black, brown and blue. One method for producing a patina is to heat the bronze with a blow torch and while it is heated apply acids with a brush. The bronze is washed with water at intervals to keep the application of acids thin. For outdoor sculpture a method of applying the acids on a cold metal is preferable.

During the Renaissance and earlier, the gilding of bronze was popular, gold leaf being applied to the surface with an adhesive medium. The same method is used in silver gilding.

Another gilding process, which was used by Lorenzo Ghiberti, Benvenuto Cellini and others, is fire gilding: an amalgam of gold and mercury is applied to the surface of the bronze with a brush and the metal is then heated until the mercury evaporates, leaving the gold fused with the bronze surface. Paduan and Florentine bronzes of the Renaissance were sometimes painted with lacquer which gave a rich brown colour. Polychromatic effects may be achieved by the inlay of other metals and alloys, such as silver, copper, niello and electrum. (J. A. C.)

STONE CARVING

Direct Carving.—The sculptor begins by marking on all sides of his block of stone (with greaseless chalk or opaque water colour) a rough outline of his planned sculpture composition.

If he is working in marble, limestone, sandstone, alabaster or other soft stone, he then proceeds as follows. With a pointed chisel and his stone mallet he knocks off the big chunks of stone that are not included in his composition. Then with his pointed chisel held at an acute angle to the stone (almost parallel to the surface) he roughs out the approximate shape of his sculptural idea, with constant attention to the grain or stratification of the stone or marble. (A misdirected chisel may split a brittle marble or ruin a stone.) He carves always around the stone, crisscrossing the direction of his chisel point as he carves. When he has roughed the stone down to about an inch from the final surface, he carves with his tooth chisels. He perfects the shape as well as he can with his tooth-edged chisels and at last he carves with various-sized flat straight-edged chisels. All through his carving the sculptor keeps a sharp edge on his chisels. The stone is finished with straight-edged chisels, rasps, files and abrasives and is perhaps polished with pumice.

In working with granite, which is composed of hard crystals fused together by great heat and pressure, the sculptor cannot cut across the surface the way he carves the softer, stratified stones. The crystals must be broken by pounding directly into the block. A granite block is roughed out with a heavy pointed chisel (called a bull point); then the granite crystals are crushed further with a pick (a hand hammer with a pickax head) and then with a bush hammer (a heavy square-headed hammer with a waffled striking face). The granite sculpture is finished with chunky flat-tipped chisels and mallet. Finally, it may be polished with carborundum stones and other abrasives. (Ls. S.)

BY COURTESY OF MALVINA HOFFMAN

THE POINTING PROCESS FOR STONE CARVING: (TOP) POINTING MACHINE HUNG ON ORIGINAL MODEL; (CENTRE) POINTING MACHINE TRANSFERRED TO MARBLE; DRILLING IN PROGRESS; (BOTTOM) USING TOOTH CHISEL AND METAL HAMMER TO REMOVE SURFACE BETWEEN POINTS

Pointing.—In stone carving done by the pointing process, a model or plaster cast is made and is then reproduced in stone with the aid of a measuring device called a pointing machine. As a technique of creative sculpture it has limitations, but it has been much used, especially in the 19th century.

With the pointing machine, three fixed points are established on the model: one at the uppermost extremity (the top of the head for a bust or upright figure) and two on the base, at opposite horizontal extremities. Mounds of plaster protrude about one inch from the finished surface of the model at these points. Metal sockets are imbedded in the mounds to secure the three fixed pins of the pointing machine: the one on top to receive the downward pointing pin and the ones at the bottom to receive pins extending from the lower horizontal bar. Three similar sockets to coincide with those on the model are placed in the block of stone. The three sockets will hold the pointing machine fast, so that the movable pointer may be manipulated to locate a point anywhere on the model.

When a desired point is located on the model (the first points are usually those which protrude the farthest, such as the nose) the movable pointer is fixed securely with screws. The pointing machine is then attached to the block of stone and a hole is drilled into the stone deep enough to allow the end of the pointer to slide forward to the same position that it did on the model. Other points are established in the same way until the stone is drilled with many holes, the bottom of each corresponding to a point on the model. The block may then be cut away until the bottoms of the holes are reached and the statue is thereby roughed out. The first points are established some distance away from what is to be the final surface. As the carving progresses the points are increased in number until an accurate reproduction of the model is attained in stone. The final points on the stone are usually left about $\frac{1}{32}$ in. higher than those on the model in order to enable the sculptor to put the finishing touches on the stone himself. (J. A. C.)

WOOD CARVING

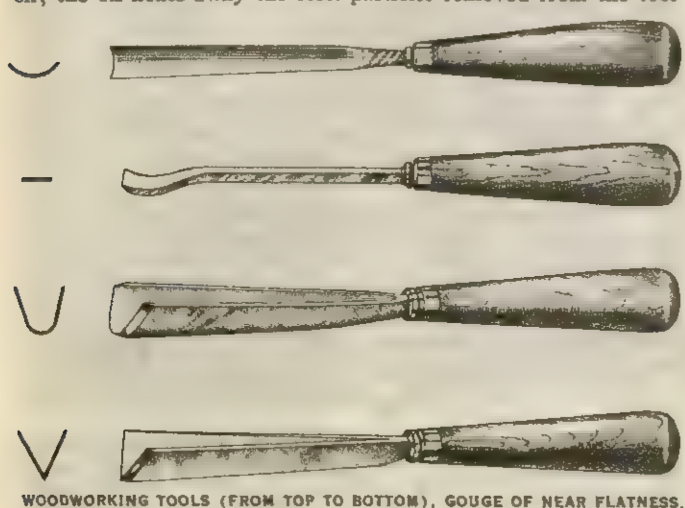
A suitable wood and several tools are needed for transforming the artist's conception into a wood carving. It is of the utmost importance that the idea or design selected be capable of execution in the block of wood chosen, for while it is possible to force an idea

upon a piece of wood, it is much better to use its natural beauty and structure as an aid in designing the carving. Thus, if a figure or animal is to be portrayed in wood, the design should be somewhat abstract; academic or realistic results are not easily attained except in relief carving because of the limitations of the material.

The problems that arise in carving are those of related forms and planes. An experienced artist may solve these as he works, letting the idea develop from the shape and grain of the material, but a less experienced sculptor will do better to first work out a full-size sketch in soap or plaster. After the elementary problems of placement have been solved, the carver proceeds to transfer the idea to the wood, and in so doing makes any changes that are needed to improve the idea. In most cases it is possible to eliminate details that do not contribute to the finished effect. The use of many details in carving proclaims either the amateur or the skilled carver with little imagination.

The wood itself should be of dense, even grain, such as mahogany or walnut, and must be seasoned; green or unseasoned wood will crack and split as the carving progresses, owing to evaporation of moisture. The moisture content varies with the wood and with the humidity. Wood will also absorb moisture and expand or swell, thus changing the contours of a piece of sculpture; for this reason air-dried wood, which absorbs less moisture, is preferred to kiln-dried wood. The colour and prominence of the grain is also to be considered. A wood with a distinct grain is sometimes more suitable for a carving with simple forms that utilize the markings as part of the design; in a more intricate carving the design would be lost in the confusion of grain and detail. Burls, roots and crotches are utilized for their unusual markings and necessitate a very carefully conceived design to produce a fine work. Knots and checks (cracks caused by shrinkage) can be used or ignored, unless they interfere with the design. Cracks can be filled with wood putty or stick shellac which comes in colors to match the wood. Wood with an alternate hard and soft grain is to be avoided; the chisel will cut the hard grain and compress the soft, causing small holes that are very hard to remove or conceal. Contrary to what might be thought, a soft wood is not easier to carve; razor-sharp tools are required to cut it, whereas a hard wood will cut easily. In carving, the best procedure is to cut across the grain rather than with it; this method makes for a consistent control over the cutting action of the chisel or gouge and will enable the carver to achieve a better and more even texture on the finished wood.

As for tools, a great number are manufactured, but the beginner needs a minimum of three—a gouge of near flatness, a deep gouge or fluter and a flat chisel, preferably a skew chisel. A sharpening stone will also be required to keep the tool edges in condition. Tools are usually supplied ground sharp but not honed—which means that the final cutting edge must be put on the tools by the carver. This is done by moving the tool back and forth over a fine stone called a hone, which is kept well oiled with a light machine oil; the oil floats away the steel particles removed from the tool



WOODWORKING TOOLS (FROM TOP TO BOTTOM). GOUGE OF NEAR FLATNESS, FRONT BENT CHISEL, TWO DEEP GOUGES

edge and thus preserves the cutting action of the stone. The dirty oil must be removed with a cloth, and new oil added. The tool should be held at the angle to which it has been ground on the bevel side and rubbed along the length of the stone (with a slightly rotary motion in the case of the gouges). The rough edge on the inside of the tool may be removed by the use of a slip stone, which comes in various shapes to fit the different tools. Tools should be of the best quality, and once sharpened will not need more than a few minutes care for each carving. Also needed is a mallet, which should be made of wood, plastic or rawhide to prevent damage to the ends of the tools when they are struck.

After the wood has been selected, a study of the grain will be helpful in placing the design. By looking at the end of the block one can see the circular ends of the grain, and by visualizing the effect of cutting into a series of cylindrical planes one can determine the result of the markings. This may seem complicated, but by continued observation the carver can learn to predict the finished result with great accuracy. The next step is to outline the design on the block with chalk, making allowance for the roundness. If a band saw is used to cut along the lines, the result will be a flat surface where it should be round. To prevent this, the carver should use a deep gouge to cut into the block along the lines of the design, never in a perpendicular to the surface but always at an angle. This will allow the forms to grow and often appear larger than the original flat surfaces. Tearing or splitting of the wood is an indication of dull tools. In case of a knot, the carver should work around it from all directions to prevent pieces from loosening or coming out. After the forms have been roughed out with the deep gouge, the shallow gouge is used to finish the carving. The skew chisel is used to get into corners where two or three planes meet.

If the carver prefers to leave the tool marks, they should be cut of even width and should present an over-all pattern that does not detract from the forms of the carving. If a smooth finish is desired, a shoemaker's rasp and riffler rasps are necessary to obtain a fine even surface. Rifflers come in various shapes to enable the carver to reach into small crevices, but a coarse garnet paper can be folded and used as well. After the rasping and filing, the carving is sanded with various grades (textures) of garnet paper. The last paper used should be a very fine one which will impart a glasslike finish to the carving. Sanding is always done with the grain to avoid scratch marks. The final step is to fill the cracks with stick shellac of matching colour and then give the entire carving a coat of clear or coloured wax. The wax should be allowed to dry in for ten days and then should be polished with a soft cloth. Many coats of wax can be added; in case of water spotting, the wax can be removed with solvent and the carving rewaxed. (J. A. Go.)

MODERN TECHNIQUES

In the 20th century, experimentation with materials—for special artistic effects and also in search of materials less expensive than the traditional ones—occasioned experimentation with techniques as well. Wrought iron has been used by Julio González, among others, using traditional ironworking techniques to produce sculpture. "Cast stone," used by Wilhelm Lehmbruck and others, is made of ground stone, sand and portland cement, in varying proportions, and is cast solid in a piece mold.

Many modern pieces are made of several materials. Alberto Giacometti's "The Palace at 4 A.M.," for example, is constructed in wood, glass, wire and string. Mobiles are constructed of various materials, such as wire, iron, sheet steel, steel rods, sheet aluminum and the like, which are so balanced as to move gently in space; the form is said to have been originated by Alexander Calder. Another modern technique is welding. (*See* SCULPTURE: *Technical Developments After World War II.*) (X.)

Sculpting in Plaster.—Working directly in plaster of Paris is another method used by some modern sculptors. The technique has a great deal to recommend it. The speed at which the plaster sets gives the work a great sense of immediacy and enables the sculptor to translate and achieve his original idea quickly; the material can be carved as well as modeled in the final stages; and,

finally, if the piece is to be cast in bronze, the process of plaster casting is eliminated.

In this method, a strong armature of iron must be constructed, preferably by welding, and should be painted with Brunswick black or an equivalent as a protection against rust, which would eventually rot and discolour the plaster. In roughing out the general bulk of the figure, wood-wool soaked in plaster may be used; this gets the sculpture up to size very quickly and, as the wood-wool is porous, makes the final result much lighter. At this stage any available padding, such as odd scraps of wood, can be used. In the beginning small quantities of plaster should be mixed because it sets quickly (in about five minutes unless size has been added). When the freshly mixed plaster begins to thicken it can be applied to the armature with large metal spatulas or builders' plastering tools. (Wooden tools are not suitable.) After a thickness of plaster has been built up over the armature, larger quantities may be mixed, as the new plaster adheres more readily to the plaster already there.

Before work is resumed on a plaster figure that has dried, it must be soaked very thoroughly with water; otherwise the old plaster will draw the moisture out of the new and cause it to crumble. (El. F.)

Bibliography.—Malvina Hoffman, *Sculpture Inside and Out* (1939); Louis Slobodkin, *Sculpture: Principles and Practice* (1949); Albert Toft, *Modelling and Sculpture*, rev. ed. (1950); William Zorach, *Zorach Explains Sculpture* (1947).

SCUNTHORPE, a market town and municipal borough (1936) in the Brigg parliamentary division of the Parts of Lindsey, Lincolnshire, Eng., 28 mi. N.W. of Lincoln by road. Pop. (1961) 67,324. Area 12.3 sq.mi. Parts of the parish church date back to the Middle Ages, notably the 12th-century pillars and the 13th-century porch. In 1871 the population was only 616, but with the discovery of ironstone in the area toward the end of the 19th century, the foundations of the iron and steel industry were laid. Subsidiary industries are the road-making and building materials industry, tar and benzole distillation, light engineering, and the manufacture of clothing and footwear.

SCURVY (Scorbutus) is a vitamin-deficiency disease characterized by debility, blood changes, spongy gums, and hemorrhages in the tissues of the body. In former times scurvy was common and often fatal among sailors. It has also occurred among soldiers on campaign, in beleaguered cities, among communities in times of scarcity, and in prisons and other public institutions. It was early recognized that scurvy and diet were connected, but it was not known until 1932 that the cause is deficiency of vitamin C (ascorbic acid; *see* Vitamins). This explains the occurrence of scurvy when fresh vegetables or fruit—rich in vitamin C—are unobtainable and its disappearance when they are administered.

The symptoms come on gradually with failure of strength, and mental depression. Then follow sallow complexion, sunken eyes, tender gums, and muscular pains. These symptoms may continue for weeks, gradually worsening. Teeth fall out and hemorrhages, often massive, penetrate muscles and other tissues. Disorders of vision, particularly night-blindness, may occur. The last stages of scurvy are marked by profound exhaustion, fainting, and complications such as diarrhea and pulmonary or kidney troubles, any of which may bring about death. Even in desperate cases, however, recovery may be anticipated when the deficient vitamin is supplied, by injection or orally (vitamin pills or amended diet).

The regulated administration of lemon juice in the British Navy, begun in 1795, had the effect of virtually extinguishing scurvy in the service. Lime juice—lower in ascorbic acid—was substituted in 1865, perhaps for reasons of economy, and British sailors were widely nicknamed "Limeys." Similar regulations introduced by the British Board of Trade in 1865 benefited the mercantile marine.

SCUTAGE (Escuage) (French: écuage, from Lat. *scutum*, a shield) means "shield money" and was in feudal law the pecuniary commutation of the military service due by the holder of a knight's fee (*see* Knight Service). A lord might accept from his vassal a sum of money (or something else of value, often a horse) in lieu of service on some expedition. The system was advantageous to both sides and grew rapidly with the expansion of money economy in Europe in the 12th and 13th centuries.

Scutage was taken in various countries, including France and Germany, but was most highly developed in England, where it is first mentioned in 1100. It seems to have been levied at first on ecclesiastical tenants in chief, who had difficulty in finding their full quota of knights for the king's army. It soon became a general tax on knight's fees and fractions thereof. The amount varied, but was usually one mark (13*s.* 4*d.*) or one pound silver on the knight's fee. Henry II (1154–89) once took two marks (1159) and King John (1199–1216) sometimes three. This latter rate became standard after 1246. Though the crown could demand scutage, tenants could not refuse to perform military service if it was required. But from the time of Richard I (1189–99) special "fines" (payments larger than the routine scutage) were accepted from tenants in chief in lieu of service on a particular campaign. As a result of the frequent and heavy scutages exacted by King John, clause 12 of Magna Carta (1215) forbade the levy of scutage without the consent of a great council. During the 13th century scutages and fines continued, the latter becoming more general. Scutage, collected from mesne tenants who had not attended a campaign, was divided between the king and those tenants in chief who had served in person. But by the 14th century scutage was becoming obsolete. It had also been of legal importance as the test (according to the lawyer Henry de Bracton) of tenure by military service. *See* also Feudalism.

Bibliography.—J. F. Baldwin, *The Scutage and Knight Service in England* (1897); J. H. Round, *Feudal England* (1895); J. H. Ramsay, *A History of the Revenues of the Kings of England, 1066–1399*, 2 vol. (1925); H. M. Chew, *The English Ecclesiastical Tenants-in-Chief and Knight Service* (1932); S. K. Mitchell, *Taxation in Medieval England*, ed. by S. Painter (1951); C. W. Hollister, *The Military Organization of Norman England* (1965). (R. C. v. C.)

SCUTARI: *see* Shkoder.

SCUTARI, the name formerly used for Uskudar (*q.v.*), a town of Turkey on the Asiatic shore of the Bosporus, opposite Istanbul. The name Scutari came into use toward the end of the Byzantine Empire. Scutari was the base (1854–56) of the British Army during the Crimean War and the site of the military hospital made famous by the work of Florence Nightingale (*q.v.*).

SCYLAX of Caryanda (in Caria) (6th century B.C.), ancient Greek explorer who was a pioneer in geography and the first Western observer to give an account of India. It is known from Herodotus that he was sent by Darius (*c.* 515 B.C.) to explore the course of the Indus and that he returned by sea after two and a half years to the Isthmus of Heroonpolis (Suez). References in ancient authorities seem to show that he left a record of his voyages, but none of the references is even nearly contemporary and it is difficult to say in what manner knowledge of Scylax was transmitted. The few fragments suggest a work in the form of a *Periplus* (mariner's coastal guide), with some geographical and ethnographical description and, in the account of India, some element of the fabulous. The extant *Periplus* which has come down under his name cannot, however, be the work of Scylax of Caryanda. It is a compilation of matter of various dates completed about 350 B.C. Though dry and practical, it is of considerable interest in form and language as the earliest surviving example of this branch of technical literature.

The Suda Lexicon attributes to Scylax a work on the Carian Heraclides of Mylasa, who distinguished himself in the revolt against the Persians (498 B.C.) and possibly in the second Persian War (480 B.C.). That the explorer and geographer was a pioneer also in historical-biographical literature is improbable.

Bibliography.—Fragments ed. by F. Jacoby, *Die Fragmente der griechischen Historiker*, part IIIC, no. 709 (1958); *Periplus* ed. by C. Müller, *Geographi Graeci minores*, vol. i, pp. 15–96 (1885). *See* also F. Gisinger in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft*, 2nd series, vol. 2, col. 619–646 (1927); J. Oliver Thomson, *History of Ancient Geography* (1948). (Ar. Ca.)

SCYLITZES, JOHN (fl. 2nd half of the 11th century), Byzantine historian, was the author of a *Synopsis historiarum* dealing with the years 811–1057. Known as the "Thracesion," *i.e.*, apparently originating from the Thracesion theme in Asia

Minor, he was a high officeholder at the Byzantine court, being described in manuscripts as *curopalates* and *drungarios tes vigles* ("commander of the imperial guard"). According to George Cedrinus (*q.v.*) he was also a *protobestiarios* ("grand chamberlain"). He was known as a contemporary of Michael Psellus (*q.v.*) and may have written a legal work for Alexius I in 1092.

In his history he drew on the work of Theophanes Continuatus for the early part, and then largely relied on sources no longer extant and on contemporary oral tradition. In some manuscripts the work is extended to cover the years 1057–79; this section is of little independent value and is known as *Scylitzes Continuatus*. There is no edition of the Greek text of this important historian, who is known through the almost literal reproduction of his work by George Cedrinus.

See bibliography and list of manuscripts in G. Moravcsik, *Byzantinoturcica*, 2nd ed., vol. i, pp. 335–340 (1958); G. Ostrogorsky, *History of the Byzantine State*, p. 188 (1956). (J. M. Hy.)

SCYLLA AND CHARYBDIS, in Greek mythology, the two immortal and irresistible monsters who beset the narrow waters traversed by Odysseus in his wanderings (later localized in the Strait of Messina, Scylla being on the Italian and Charybdis on the Sicilian shore).

Scylla, daughter of Phorcys and Cratais, is a supernatural creature, with 12 feet and 6 heads on long snaky necks, each head having a triple row of sharklike teeth, while her loins are girt with the heads of baying dogs. From her lair in a cave she devours whatever ventures within reach, and takes toll of six of Odysseus' companions. She is sometimes said to have been originally human in appearance, but transformed out of jealousy through the witchcraft of Circe or Amphitrite into her present shape. She is apt to be confused with the quite distinct Scylla who betrayed her father, King Nisus of Megara, out of love for Minos.

Charybdis, lurking under the fig tree a bowshot away on the opposite shore, drinks down and belches forth the waters thrice a day and is fatal to shipping. The shipwrecked Odysseus barely escapes her clutches by clinging to a tree until the improvised raft which she has swallowed floats to the surface again after many hours.

Charybdis is an imperfectly personified whirlpool. Scylla was often rationalized in antiquity as a rock or reef. Both give poetic expression to the dangers confronting Greek mariners when they first ventured into the perilous and uncharted waters of the western Mediterranean. The proverbial use of Scylla and Charybdis for the hazards attending any deviation from the true course is as old as St. Augustine. (D. E. W. W.)

SCYPHOZOA: *see* COELENTERATA; JELLYFISH.

SCYTHIANS, a people who during the 8th–7th centuries B.C. moved from central Asia to southern Russia, where they founded an empire that survived until they were gradually overcome and supplanted by the Sarmatians (*q.v.*) during the 4th century B.C.–2nd century A.D.

Until the 20th century most of the information concerning the Scythians, as the Greeks called them, or the Sakas as the Persians termed them, was based on the knowledge that Herodotus acquired when he visited Olbia, a Greek city founded *c.* 645 B.C. at the confluence of the Bug and Ingul rivers. He incorporated all that he had learned about them into the fourth book of his history. In modern times the work carried out in Siberia by Russian scholars and archaeologists produced so much new information that the word "Scythian" is often extended to include the nomadic kinsmen of the Scythians who remained in central Asia when the latter moved into Europe.

In its exact context the term "Scythian" should, however, be applied only to the tribes that settled in what is now southern Russia. The excavations begun in 1929 at Pazyryk in the Altai by M. P. Gryaznov and resumed in 1947 by S. I. Rudenko proved that the culture, art, and way of life of the tribesmen who continued to live in western Siberia and the Altai when the bulk of the Scythians marched into Iran and southern Russia remained closely akin to that of the migrants, the entire community being of Iranian origin and sharing a common language. Herodotus attempted to distinguish between the various tribes that settled in what later became Russia, and he was able to name among others the Callipidae, the Alazones, the Aroteres, the Neuri, the Androphagi, and the Melanchlaeni, defining the areas that each inhabited. These western Scythians carried on a lively trade with those who had remained in central Asia.

History.—In the 9th century B.C. the Scythian and kindred tribes were probably concentrated somewhat to the east of the Altai, but it was not until the Chinese emperor Hsüan Wang (827–781 B.C.) decided to send an armed force to curb the fierce Hsiung-nu, who had begun to make a practice of raiding China's western boundaries, that the Scythian nomads became restless. When the Hsiung-nu were forced back from the Chinese frontier and, in retreating, dislodged the Massagetae, who occupied the grazing grounds to the north of the Oxus (Amu-Darya) River, and when the latter in their turn assaulted their immediate neighbours, the Scythians, a wide-scale nomadic migration was set in motion. There is reason for thinking that the struggle for grazing land was rendered more acute by the beginning of a severe period of drought and that this factor may well have decided the Scythians to move westward in search of satisfactory pasture instead of remaining to fight for their traditional rights.

The Scythians were accomplished horsemen, having been among the earliest people to master the art of riding. Their mobility gave them a great advantage over their neighbours, so that when they eventually advanced westward across the Oxus they moved so fast that both Herodotus and contemporary Persian sources refer to the remarkable suddenness of their appearance on Iran's northeastern border. Their advance brought the Scythians into fierce conflict with the Cimmerians (*q.v.*), who had for centuries enjoyed possession of the Caucasus and the plain lying to the north of the Black Sea. But the Cimmerians still fought on foot and the Scythian cavalry quickly gained the upper hand. Some Cimmerians retreated through the Daryal (Dariel) Pass and were pursued across the Volga, where the Scythians were able to destroy and supplant them.

Meanwhile another Scythian force chased the rest of the Cimmerian Army across Urartu (Armenia), while a third entered the Derbent defile and reached Lake Urmia some time during the reign of King Sargon of Assyria (722–705 B.C.), linking up there with the second contingent to continue the fight against the Cimmerians. Thus strengthened, the Scythian troops were able to force the Cimmerians into a steady retreat which lasted for about 30 years, ending only when both combatants had reached the borders of Assyria. Then the Scythians formed an alliance with King Esarhaddon of Assyria (reigned 681–669 B.C.), but they soon abandoned this and concentrated on destroying the Cimmerians, giving the latter no respite till they had forced them back across Phrygia into Lydia, where they were finally wiped out.

This astonishing series of victories brought great fame to the Scythians. Their chieftain Partatua (Bartatua) and his son Madyes were quick to take advantage of it by setting themselves up as rulers of Urartu and all the lands stretching to the Halys (Kizil Irmak) River, establishing their capital at Saqqez. They invaded Syria and Judaea *c.* 625 B.C. Later they reached the borders of Egypt, but Psamtik I (663–610 B.C.) wisely decided to check any further advance by purchasing peace terms from them.

Meanwhile the Medes had become masters of Persia. They considered that the Scythians' increasing might presented a real threat to their own security, and they decided to concentrate their efforts in launching a decisive attack against the tribesmen. Their better-disciplined troops eventually contrived to push the Scythians northward back through Urartu, whence they had first appeared. Although the nomads thereby lost the control they had wielded during the previous 28 years over most of Asia Minor, their lands still stretched from the Persian border, through the Kuban into most of southern Russia. When forced by the Medes to retreat, some of them likewise settled between the Caspian and the Aral seas, where they intermingled with their Dahae kinsmen to produce the people who, some three centuries later, were to become known as the Parthians, while others penetrated into India and established kingdoms there (*see* SAKA).

The Scythians who settled in the Kuban and southern Russia quickly attained a position of importance. Many of them became extremely prosperous. Graves of the 7th–6th centuries B.C. situated in the Kuban abound in objects of gold and other precious materials, and although a form of patriarchal rule remained in force, it is clear that a class of wealthy chieftains or nobles was beginning to come into being there. In the 6th–5th century B.C. the richest burials were in southern Russia and the Crimea. These are associated with a relatively small number of Scythians who, as the Royal Scyths, established themselves as rulers of the area. Isolated groups of their tribesmen penetrated as far as what became Hungary and East Prussia. The kingdom of the Royal Scyths developed into a community which was to enjoy considerable economic power till the 1st century or so B.C. Its political importance was established when it was able to resist Darius on his invasion of Scythian territory about 513 B.C.; the Scythians resorted to a scorched-earth policy, which enabled them to withstand the Persian onslaught and obliged Darius to beat a hasty retreat in order to preserve his army. His consequent withdrawal from the plain lying to the northwest of the Black Sea left the Scythians in control of it; some of the Greek cities of the Pontus even had to pay the Scythians an annual tribute in order to preserve their independence. Scythian power remained paramount there till the 4th century B.C. when the Sarmatians appeared on the Don. Thenceforth the new invaders began steadily to increase their pressure on the Scythians till, in the 2nd century B.C., they managed to confine them within the Crimean area, gradually supplanting them as rulers of the plain until, in the 2nd century A.D., they finally succeeded in destroying the last remnants of this once powerful community.

The Royal Scyths.—Though the patriarchal order and the nomadic way of life remained general, the Royal Scyths were ruled by a sovereign whose powers were transmitted to his son. He was surrounded by a group of wealthy nobles who were virtually his courtiers. The dynasty claimed descent from Targitaus, the founder of the house of Phalatae; but Targitaus was probably a legendary figure, and it is more likely that Colaxis was the first of the royal line. He was succeeded by Spargapeithes, who handed the crown to his son Lycus, who in turn passed it to his son Gnurus at about the time when Partatua and his son Madyes had established themselves as co-rulers in Transcaucasia (*c.* 630 B.C.). In 589 Saulius reigned over the Royal Scyths. His brother Anacharsis (whose existence is doubted by certain authorities) went to Greece as his ambassador and became so much respected there that the Greeks spoke of him as "the Scythian eloquence." The throne then passed to Idanthyrsus who, with his brothers Taxacis and Scopasia, expelled Darius. Ariapeithes, the son of Idanthyrsus, came to the throne next, marrying in turn a Greek woman, a Scythian, and a daughter of a Thracian chieftain. He was succeeded by his son Scyles, who was murdered shortly before the time of Herodotus by his brother Ostomasades because of his fondness for Greek manners, dress, and religious observances. The murderer succeeded his victim on the throne. His successor, Arianthus, is remarkable for having organized a census of his people. The next known ruler, Aristagorus, reigned *c.* 495 B.C. A descendant, Ateas, was killed in 339 B.C. at the age of 90 while fighting against Philip II of Macedonia. King Agarus' reign likewise ended with his death in battle (310 B.C.). Nothing is known of the kings who followed him till Scylurus established himself in 110 B.C. at Neapolis (in the Crimea) and struck Scythia's first coins at Olbia. He was succeeded by his son Palakus, the last Scythian sovereign whose name has been preserved in history.

Hippocrates described the Scythians as inclined to laziness, fatness, and gaiety, likening their appearance to that of eunuchs and commenting on their sexual impotence; but his statement is corroborated neither by the events in their history nor by the appearance of those Scythians whose features have been preserved on contemporary Greek metalworks—for example, on the electrum kumiss jug from the Chertomlyk burial.

Administration Under the Royal Scyths.—For administrative purposes the Royal Scyths divided their kingdom into four districts, each of which was controlled by a governor provided with a paid military bodyguard. The governor dispensed justice and levied taxes from the settlers and tribute from certain of the Greek Pontic cities. The tribal way of life remained in force; communities were governed by their elders or chieftains, these dignitaries being periodically summoned to attend a general assembly held in the presence either of the governor or of the sovereign. These assemblies differed from the gatherings held by the sovereign in the spring of each year in order to inspect and feast his army. In times of war the country was divided for purposes of recruitment into three sections, the enrolled men being later grouped into units, each one of which was commanded by its own officer.

The Army.—The army consisted entirely of freemen, *i.e.*, Scythian tribesmen; they were fed and clothed but paid no wage, though those of them who could produce the head of a soldier killed in battle were entitled to share in the day's booty. According to Herodotus, the Scythians scalped their victims, fashioning their skulls into cups which they wore attached to their belts, using them to pledge an oath in a mixture of blood and water. Such cups mounted in delicately worked gold have been found in several tombs.

Many Royal Scyths wore bronze helmets and chain-mail jerkins of the Greek type lined with red felt. Their shields were generally round and made of leather or iron, often decorated with a central gold ornament in the form of an animal, but other tribesmen carried square or rectangular shields made of wood or leather. All used a double curved bow, shooting over the horse's left shoulder; arrows had trefoil-shaped heads made, according to date, of bronze, iron, or bone. Arrows and bow were carried in a gorytus (bow case) slung from the left side of the belt. Their swords were generally of the Persian type, with a heart-shaped or triangular crosspiece intricately ornamented. According to date, the blades were of bronze or iron; in southern Russia, the sheaths were often encased in gold worked into embossed designs and offset with paste or ivory inlay and gems. Their knives were of various shapes and lengths, some being curved in the Chinese manner. They wore the dagger attached to the left leg by straps, and many carried spears or standards surmounted by bronze sculptures depicting real or imaginary beasts.

Horses.—Every Scythian owned at least one gelding to serve as a riding horse, but the wealthy possessed a great many mounts; most Scythians also owned oxen or rough ponies, which served as beasts of burden. The finest riding horses were of the Fergana breed, but the majority were Mongolian ponies. The Scythians devoted much time and attention to their horses and ornamented all their trappings. Bridles were provided with metal cheekpieces in the shape of animals, and the leather straps were adorned with embossed, cutout, or appliqué designs, which also often represented animals. Saddles consisted of two felt cushions mounted on wooden frames bound with yellow or red and sometimes embellished with gold plaques; the felt saddle cloths were adorned, as were the seats of the saddles, with appliqué designs. Metal stirrups were not known to the Scythians, but there is reason to believe that they rested their feet in felt or leather supports. Women lived in subservience and semiseclusion, traveling with their children in covered wagons with solid wheels and a central shaft along which mules or oxen could be yoked in pairs.

Possessions.—The frozen tombs of Pazyryk (*see* below, *Tombs*) contained many well-preserved articles of clothing, all of which were profusely trimmed with complicated embroidered and appliqué designs; the clothes of the wealthy in southern Russia were covered with tiny gold-embossed plaques, sewn to the garments. At Pazyryk felt appliqué wall hangings were found, some displaying religious scenes featuring the Great Goddess or anthropomorphic beasts, others with geometric or animal motifs. Even the embalmed body of a man was covered with tattooed designs of real and mythical beasts executed with great spirit and delicacy. Felt rugs were found, as well as a knotted, woolen pile carpet of Persian origin, from the 5th century B.C., displaying figures of riders, elks, and stars. Felt cushions and mattresses, wooden tables with carved or turned legs and detachable, slightly hollowed, traylike tops, and wooden blocks serving as stools or head rests were customary.

FROM T. TALBOT RICE, "THE SCYTHIANS"; REPRODUCED BY PERMISSION OF THAMES AND HUDSON, LTD.

SCYTHIAN SETTLEMENTS AND BURIAL GROUNDS. INSET MAP SHOWS AREA OF HIGHEST CONCENTRATION IN AND NEAR THE CRIMEAN PENINSULA (TAURIC CHERSONESE), SOUTH RUSSIA

The tombs of the Eurasian plain as a whole produced a mass of tools and domestic utensils, many of them made of precious materials. Rich jewelry and arms of great value and beauty are frequently found in the burials belonging to the western section. In addition, each burial throughout the entire area contained a cast bronze cauldron of the distinctive Scythian shape. These cauldrons vary in size from quite small examples to others weighing as much as 75 lb.; an overwhelming majority have a solid base, shaped like a truncated cone, round which the fire was heaped. The upper section is a semispherical bowl provided not with a loop handle but with handles (shaped like animals) fixed to the rim opposite each other. The finest cauldron, found at Chertomlyk in the Dnieper district, has six handles. At Pazyryk, small cauldrons filled with stones and hempseeds were found standing beneath leather or felt tentlets with three or six supports. Herodotus referred to what he termed a Scythian purification rite which, he noted, consisted in inhaling the fumes of hempseeds thrown onto hot stones; the passage was well-nigh incomprehensible till archaeologists discovered that a smoking outfit of this sort had been provided for each person buried at Pazyryk, making it clear that hemp fumes were inhaled for pleasure and not, as Herodotus assumed, as part of a religious observance.

Occupations and Beliefs.—The Scythians were keen huntsmen and fishermen; they were skilled at curing hides; they excelled at working metals; and the settlers were good agriculturalists. Though they had neither an alphabet nor, until later times, a coinage, they carried on a lively trade not only with the inhabitants of central Asia but also with the Greeks of the Pontic cities, often exchanging surplus goods and furs for Greek luxuries such as fine ceramic wares.

The Scythians worshiped the elements but they were not a devout people and never felt the need for temples. Their deepest feelings were centred on the Great Goddess, Tabiti-Hestia, the patroness of fire and beasts, and she alone of all their deities figures in art. They also worshiped Papeus-Zeus, god of the air; Apia-Gaea, goddess of the earth; Oetosyrus-Apollo, god of the sun; Artimpaasa-Aphrodite, goddess of the moon; Thamunasades-Poseidon, god of water; and, in time of war, possibly also Ares and Heracles. They do not appear to have had any priests, but they had a class of magicians and soothsayers, the Enaries, who spoke in high-pitched voices and dressed as women. The Scythians believed that these feminine characteristics had been inflicted upon them by the Great Goddess as a punishment for desecrating her shrine at Ascalon; they were probably eunuchs and may well have been the men Hippocrates had in mind when describing the appearance of the Scythians.

Burial Customs.—The Scythians venerated the graves of their ancestors, sparing neither wealth nor labour in providing vast tombs richly furnished and equipped. Their funerary rite was elaborate and as costly in lives as in material. In the Altai, interment took place only in the spring and autumn weeks, and embalming was therefore essential. The Scythians of southern Russia also resorted to it, adopting the process Herodotus ascribed to them. Once the corpse had been prepared, it was customarily placed on a cart. Then the mourners (in the case of a chieftain, the entire tribe; for a lesser person, the family and friends), having cropped their hair, wailing and tearing their clothes, would conduct the dead man during 40 days through the lands that had once belonged to him or, if he had none, would escort him on visits to his relatives and friends.

On the 40th day the body was carried to the tomb and, while one of the dead man's wives, his principal servants such as his cupbearer and groom, and a substantial number of his riding horses were being put to death, his body was lowered into the main burial chamber; the dead man's newly killed companions and servants were then laid in the graves which had been prepared for them. The bodies of the horses were also placed within the mound, the tombs of the Scythian and kindred nomads all taking the form of horse burials.

All the dead wore their finest jewelry and clothes and the best of their arms and accoutrements, and they also took with them the objects considered necessary for their future life, only the best of their possessions being chosen for the tomb. The horses were harnessed in their finest trappings; at Pazyryk some were even provided with elaborate masks made of felt and leather touched up with gold, some in the shape of a stag's head, complete with antlers, some surmounted with figures of birds. The number of horses varied with the period and the region, in some cases running into hundreds.

Finally, according to Herodotus, the Scythians marked the first anniversary of a chieftain's death by choosing 50 men from his bodyguard, all of them freeborn tribesmen, and putting them to death, together with their mounts; the men were then seated on their horses and the bodies impaled together on stakes set in a circle round the tomb, so that the faithful dead might forevermore guard it from desecration. It is unfortunate that this method of protection, so costly in human and animal lives, failed to achieve its purpose; almost every Scythian grave opened by archaeologists has been found to have been rifled in antiquity.

Tombs.—Tombs of the Scythian and kindred tribes are found from the Altai to eastern Europe. In the Altai the method of construction sometimes led to the formation of a layer of ice above the grave and to the preservation of the tomb contents as efficiently as in a modern deepfreeze. As a result, many of the materials that perished in the western burials have been preserved in the Asiatic tombs. The frozen burials of Pazyryk date from the 5th to the 3rd century B.C. There the size of the mounds is measured by their circumference rather than by their height, for all are low, being topped only with a covering of boulders. It was this boulder cover that caused the formation of a layer of ice when the rainwater that seeped through it to the earth filling beneath became frozen. The richest Scythian tombs are situated in the Soviet Union, primarily in the Kuban region and in the kingdom of the Royal Scyths (*i.e.*, on the Dnieper and in the Panticapaeum

region in the eastern Crimea), though poorer burials are found over a much larger area. The tombs of the chieftains are extremely elaborate constructions, the larger having as many as five chambers, each measuring some 9 ft. by 15 ft. and 7 ft. in height; they often lie as much as 40 ft. below virgin soil, while the mounds erected on them vary from 400 to 1,200 ft. in diameter.

In southern Russia, at a chieftain's death a sloping trench was first dug; at its further extremity a shaft was sunk into virgin soil to form the main burial chamber; subsidiary graves, as many as five in number, were prepared for the chieftain's companions. Next the trench had to be transformed into a corridor and then the graves had to be made to resemble the rooms or tents in which the dead had spent their terrestrial existence. This was accomplished by smoothing the sides of the excavated areas and facing them, according to the region, with tree trunks, matting, birch bark, thatch, or even fabrics. In the Crimea painted decorations were often preferred, while at Pazyryk wall hangings were customary. Hooks were fixed into the walls and spare clothing hung on them; shelves were built round the sides to hold provisions. The ceilings of the subsidiary graves were treated in much the same manner as the walls, but four posts were often sunk into the main grave to support either a canopy placed above the corpse or a roof; floors were often covered with gravel or rushes. The dead were orientated toward the east, laid on mattresses, often set on biers; in the Altai, however, the bodies were laid in coffins made of hollowed-out tree trunks.

In southern Russia the chieftain's body occupied the main chamber; his wife's was placed in the next best; and his retainers were laid in the others, the grooms always being given preferential treatment. No special graves were provided for the horses, their bodies being laid round the tomb in as orderly a manner as possible; but when a large number of horses were sacrificed the corpses were often piled one upon the other in a heap. At Pazyryk, on the other hand, the horses were ordinarily buried in a separate but adjoining chamber decorated in much the same manner as the section of the tomb reserved for the chieftain and his companions.

BY COURTESY OF THE HERMITAGE, LENINGRAD

ARTIFACTS OF SCYTHIAN CIVILIZATION

(Top) Gold stag plaque, probably central ornament from a shield; 6th century B.C. Found at Kostromskaya Stanitsa. (Bottom left) Gold comb with ornament showing battle between Scythians; 5th–4th century B.C. Found at Solokha. (Bottom right) Horse's burial mask of felt, leather, gilded hair, and copper in shape of stag's head; 5th century B.C. Found at Pazyryk

In Hungary the burials of Zöldhalompuszta and Tápiószentmárton are among the most important because of the magnificent gold plaques in the shape of stags found there; in eastern Germany that of Vettersfeld is of particular interest on account of the gold figure of a fish of remarkable design discovered in it, together with interesting items of military equipment.

Art.—Nothing that has so far been found in the Altai has proved of the same intrinsic value as the objects discovered in the western burials, but they show that similar work was produced by the artists of the entire Eurasian area, whether they worked in wood, leather, bone, appliqué felts, and coarser metals such as bronze and iron or whether they used more precious materials such as silver, gold, or electrum. The artists and craftsmen of the Scythian age were fascinated by the animals they saw or imagined, and though numerous objects were decorated with geometric forms, many of them symbolic or magical, the art of the period is essentially an animal art. Combat scenes between two or more animals are numerous, as are single animal figures. Many real or mythical beasts are represented, the majority of the types having roots in deep antiquity, but the Scythians fashioned them in a manner that was new and characteristically their own. As is to be expected with nomads who were constantly on the move, the objects they produced are generally small in size, but many are made of precious materials and practically all are of superb workmanship.

The gold figures of semirecumbent stags, measuring some 12 in. in length, are outstanding; they were probably used as the central ornaments for the round shields carried by many Royal Scythian fighters. In pre-Scythian times the stag was used as a totem by the tribesmen of Siberia, but in the Scythian period the emblem had lost much of its religious significance. Perhaps the loveliest of the gold stags is the 6th-century-B.C. example from the burial of Kostromskaya Stanitsa in the Kuban, but versions of the 5th century B.C. from Tápiószentmárton in Hungary and of the 4th century B.C. from Kul Oba in the Crimea are scarcely less beautiful. Certain scholars ascribe the former to Thracian and the latter to Ionian craftsmen, yet all conform wholly to Eurasian traditions, and much more concrete evidence is necessary before the former of these attributions can be accepted as final. In the three the stag is shown in a recumbent position, with its legs tucked beneath its body, but with its head raised and its muscles taut so that it gives an impression of rapid motion. The stag also figures prominently in the art of the Pazyryk people, whether as leather or felt cutouts or as carved wooden terminals.

The Scythians' artistic idiom is one of great compression as well as of synthesis; contrasting positions of the body are combined with astonishing skill to depict every possible aspect of the animal when visualized during all its diverse activities, zoomorphic junctures of fabulous intricacy helping to evoke the animal kingdom in its abundant multiformity. Thus, though the art is basically representational in character, it is at the same time impressionist in spirit, while it also often verges on the abstract in conception; yet, however complex its elements, they are fused in the finished work into a single entity of compelling force and beauty.

BIBLIOGRAPHY.—*Ancient:* Herodotus, iv, 1–142 (Eng. trans. by G. Rawlinson); Strabo, vii, ch. 3–4; xi, ch. 1, 3, 7; Pliny, iv, 79 *et seq.;* Ptolemy, *Sarmatia;* Diodorus Siculus, ii, 43 *et seq.;* Justin, i, ch. 1, 8; ii, 1–5.

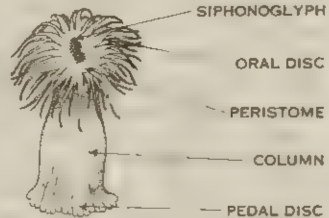
Modern: E. H. Minns, *Scythians and Greeks in South Russia* (1913); G. I. Borovka, *Scythian Art* (1928); V. Pârvan, *Dacia* (1928); M. I. Rostovtsev, *Skythien und der Bosporus* (1931); B. N. Grakov, *Skify* (1947); S. I. Rudenko, *Der zweite Kurgan von Pazyryk* (1952); T. Talbot Rice, *The Scythians* (1957); T. A. H. Potratz, *Die Scythen in Südrussland* (1963). (T. T. R.)

SCYTHOPOLIS: *see* BEISAN.

SEA ANEMONE, a sedentary and usually solitary marine animal whose flowerlike appearance is striking; the name anemone is taken from the Greek word for windflowers. Sea anemones are polyplike coelenterates belonging to the class Anthozoa (*q.v.*). The

radial arrangement of the usually colourful tentacles around the disc that bears the mouth at its centre is reminiscent of the petals of a flower.

Anemones vary in size from polyps only a fraction of an inch in both length and diameter to giants measuring about three feet in diameter. These carnivorous animals are common along all of the sea shores of the world, abundant in bays and estuaries, and occur even in the greatest depths of the oceans; they are never found in fresh waters. Most anemones adhere tenaciously by their bases to hard substrates such as rocks, other animals, pilings, wharves, and ship bottoms. They seldom move, although a slow creeping motion is possible. Some anemones do not possess a base, but have instead a specialized aboral region, the physa, which makes it possible for them to burrow into soft muddy or sandy bottoms, where they will be found with just their tentacles and oral area exposed. A few anemones with a base, such as *Stomphia* and *Boloceroides*, can release their grip upon the substrate and by movements of the column or tentacles can swim short distances. *Minyas* has a base that secretes a porous, chitinous mass that serves as a float in carrying the anemone about at the surface of the water. Anemones are devoid of a skeleton but they may secrete a horny cuticular covering or, by using various adhesive structures on the body wall, cover themselves with sand grains, bits of shell, or other foreign objects.

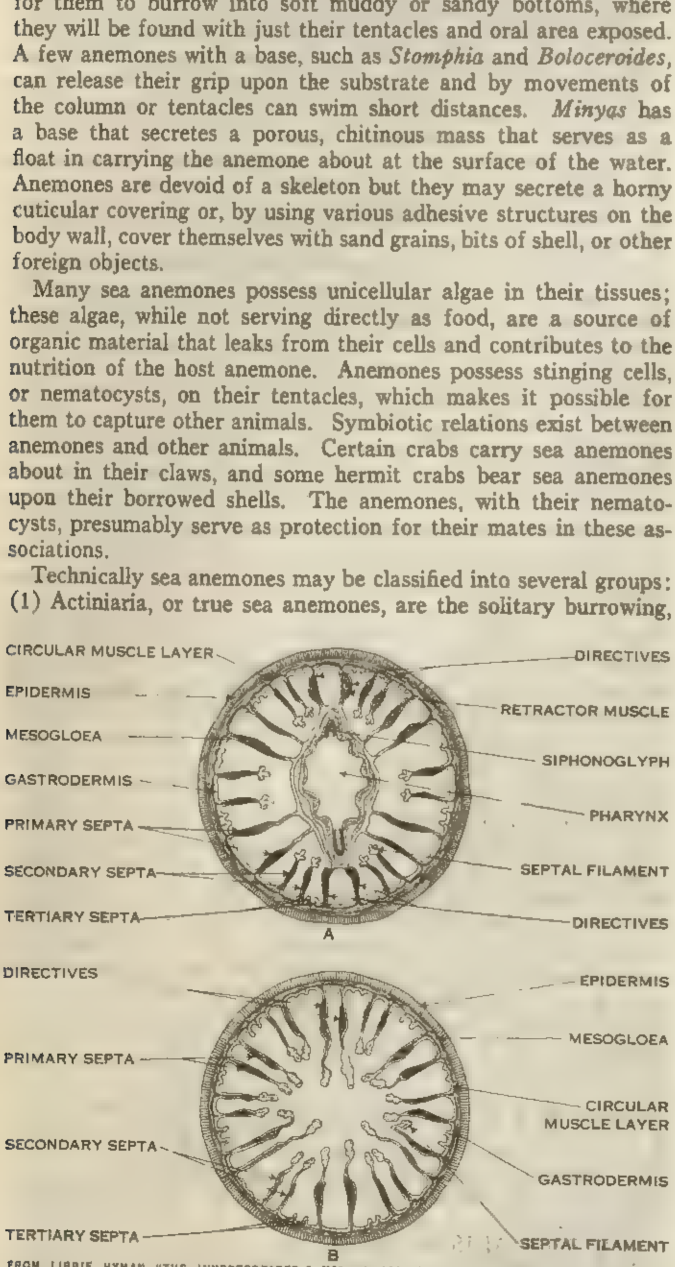


FROM LIBBIE HYMAN "THE INVERTEBRATES," VOL. I (1940); REPRODUCED BY PERMISSION OF MCGRAW-HILL BOOK COMPANY, INC

FIG. 1.—GENERAL FEATURES OF A SEA ANEMONE

Many sea anemones possess unicellular algae in their tissues; these algae, while not serving directly as food, are a source of organic material that leaks from their cells and contributes to the nutrition of the host anemone. Anemones possess stinging cells, or nematocysts, on their tentacles, which makes it possible for them to capture other animals. Symbiotic relations exist between anemones and other animals. Certain crabs carry sea anemones about in their claws, and some hermit crabs bear sea anemones upon their borrowed shells. The anemones, with their nematocysts, presumably serve as protection for their mates in these associations.

Technically sea anemones may be classified into several groups: (1) Actiniaria, or true sea anemones, are the solitary burrowing, attached, or free floating animals; (2) Ptychodactiaria, or deep water anemones, are found in the Arctic and Antarctic; (3) Corallimorpharia are solitary or tropical forms intermediate between corals (Madreporaria) and sea anemones (Actiniaria); (4) Zoanthidea are solitary or colonial anemones, many species of which occur on the stalks of sponges, sea fans, and sea feathers; (5) Ceriantharia are solitary types that live in tubes formed of mucus and extruded nematocysts. These animals usually occur in soft bottoms in subtidal locations.



FROM LIBBIE HYMAN "THE INVERTEBRATES," VOL. I (1940); REPRODUCED BY PERMISSION OF MCGRAW-HILL BOOK COMPANY, INC.

FIG. 2.—CROSS SECTIONS OF A SEA ANEMONE (A) THROUGH THE PHARYNX AND (B) BELOW IT

BIBLIOGRAPHY.—O. Carlgren, "A Survey of the Ptychodactiaria, Corallimorpharia and Actiniaria," *Kungl. Svenska Vet. Hand.* Series 4, vol. 1, pp. 1–129 (1949); L. H. Hyman, *The Invertebrates:* vol. 1, *Protozoa through Ctenophora* (1940); T. A. Stephenson, *The British Sea Anemones:* 2 vol. Ray Society, London (1928, 1935).

(C. HA.)

SEABORG, GLENN THEODORE (1912–), U.S. physical chemist and co-winner with Edwin M. McMillan of the 1951 Nobel prize in chemistry, is known principally for his work on the synthetic transuranium elements. He was born in Ishpeming, Mich., on April 19, 1912, and was educated at the University of California (A.B., Los Angeles, 1934; Ph.D., Berkeley, 1937). He was successively research associate, instructor and assistant professor (1937–45), becoming professor of chemistry and associate director of the Radiation laboratory in 1945. He served as chancellor, 1958–61.

With his co-workers, he added (1940–55) eight new elements encompassing atomic numbers 94–101, of which plutonium (94) is the best known because of its use as a nuclear explosive and for nuclear power. During World War II, which he spent as a section chief at The University of Chicago metallurgical laboratory, the first industrial production of plutonium was undertaken in the newly devised uranium reactors, and Seaborg had the primary responsibility for isolating the plutonium from the reaction products. The other new elements were: americium (95), curium (96), berkelium (97), californium (98), einsteinium (99), fermium (100), mendelevium (101). Following the tentative identification of nobelium (102) in Sweden in 1957, he predicted, at the Geneva conference on atomic energy (1958), the possible discovery of six more synthetic elements. In Jan. 1961 President Kennedy named him chairman of the Atomic Energy commission. (I. P.)

SEABURY, SAMUEL (1729–1796), American clergyman, first bishop of the Protestant Episcopal Church in the United States, was born in Groton, Conn., on Nov. 30, 1729. After graduating from Yale in 1748, he served for four years as catechist in Huntington, N.Y. In 1752 he went to Great Britain and, after studying medicine for a year at Edinburgh, was ordained deacon and priest successively on Dec. 21 and 23, 1753. Returning to America, he served as rector in New Brunswick, N.J. (1754–56); Jamaica, N.Y. (1757–66); and Westchester, N.Y. (1766–75). He took a loyalist stand in the Revolution and wrote some pamphlets on that side under the pseudonym of A. W[estchester]. Farmer. These led to his being arrested by the revolutionary party and held prisoner in Connecticut for several weeks. After his release, he returned to Westchester for a while but later sought refuge with the British forces in New York, where he served for a time as chaplain of a loyalist regiment. He was there at the end of the war when the clergy of Connecticut, meeting in Woodbury on March 25, 1783, made him their second choice for bishop. As Jeremiah Leeming, the first choice, declined, Seabury was sent to England to seek consecration. A special act of parliament was required to permit the consecration of American bishops, and various circumstances connected with Seabury's election made the archbishops hesitate to seek the power in his case. He then went to Scotland, where he secured consecration on Nov. 14, 1784, from bishops of the nonjuring Episcopal Church in that country. In the meantime, Episcopalians in the middle and southern states had begun a movement that led to the adoption of a constitution for the church in 1786 and the consecration of two bishops in England in 1787. Between that event and the next meeting of the General convention in 1789, the leaders of the church were concerned with the problem of uniting the two episcopal lines. The difficulties arose from Seabury's political views and from uncertainty as to the canonical status of the Episcopal Church in Scot-

land. A conciliatory spirit finally prevailed on both sides, and the whole Protestant Episcopal Church was united in the second session of the General convention of 1789. A third bishop was consecrated in England, and all four bishops joined in the consecration of Thomas John Claggett as bishop of Maryland on Sept. 17, 1792, the first consecration on American soil. Though disappointed in some features of the church's constitution and of the Book of Common Prayer as revised in 1789, Seabury secured their acceptance in Connecticut, where he continued to serve as bishop and as rector of the church in New London until his death on Feb. 25, 1796.

See also PROTESTANT EPISCOPAL CHURCH.

BIBLIOGRAPHY.—E. E. Beardsley, *Life and Correspondence of the Rt. Rev. Samuel Seabury* (1881); W. J. Seabury, *Memoir of Bishop Seabury* (1908); E. L. Pennington, *From Canterbury to Connecticut* (1941). (W. W. Ms.)

SEA COW, an aquatic mammal of the order Sirenia (*q.v.*) but in a restricted sense the recently extinct Steller's sea cow (*Hydrodamalis stelleri*). This gigantic relative of the manatee and dugong (*qq.v.*) formerly inhabited Bering and Copper islands in the Bering Strait near Kamchatka. It was discovered in 1741 and described by Georg W. Steller, who accompanied Vitus Bering in his voyage of exploration. It was killed in large numbers by Russian sealers and fur hunters, who found the animals easy to hunt; by 1768, less than 30 years after its discovery, the sea cow apparently was exterminated.

The sea cow reached a length of about 24 ft., with a relatively small head and broad, horizontal, forked tail fluke. The thin skin was dark brown in colour, sometimes streaked or spotted with white. The jaws were toothless but provided with ridged horny plates. The flippers were short and blunt, lacking the terminal joints of the digits. When discovered, these sirenians were numerous in bays, where they browsed upon the abundant seaweed. Bones of the sea cow are still found there from time to time. (J. E. HL.)

SEA CUCUMBER, popular name, after their shape, for warty-skinned marine animals of the class Holothurioidea, the most aberrant and diversified class of the Echinodermata (*q.v.*).

RUSS KINNE-PHOTO RESEARCHERS INC.

SEA CUCUMBER (CUCUMARIA FRONDOSA) PHOTOGRAPHED OFF THE COAST OF MAINE

The body is usually elongate, often sausage-shaped, with a spacious body cavity, or coelom, and well-developed muscles that enable the holothurians to move in wormlike fashion. In size, sea cucumbers range from about an inch (order Elasipoda) to several feet (order Apoda) in length. They are usually dull in colour, being shades of gray, drab, brown, and black; however, some are brightly coloured. The skeleton is usually reduced to minute spicules, important for classification; rarely, larger plates are present, and a few forms presumably have lost all spicules. The water vascular system, characteristic of all echinoderms, is well developed in some orders and almost nonexistent in others. Food is collected by means of tentacles, modified tube feet that surround the mouth. In the bizarre primitive deepwater forms, the Elasipoda, the tentacles are hardly more than enlarged tube feet by means of which the animals scoop up the ooze of the sea bottom. In more advanced groups, the tentacles may become shovellike organs, as in the many reef forms of the tropics, or they become richly branched and are used as a net in which plankton particles are strained and afterward transferred to the mouth, where the tentacle is licked clean. In the streamlined members of the order Molpadonia, which tunnel their way through mud, the tentacles are small and used as digging tools, while at the same time food is taken in. The wormlike Apoda, which burrow, use their tentacles in the same manner, while forms that live on reefs have large feathery or plumose tentacles that sweep up food particles. The Holothurioidea comprise about 500 species, ranging from extreme shallow water to the greatest depths of the ocean. A few are utilized as food (*see* BÊCHE-DE-MER). (EL. D.)

SEAFORD, an urban district and seaside resort in the Lewes parliamentary division of East Sussex, Eng., 4 mi. SE from Newhaven by road. Pop. (1961) 11,021. The town is sheltered by the South downs behind and the high chalk cliff of Seaford head to the east. In former days the Ouse entered the English Channel there, and the natural harbour so formed accounts for the origin of Seaford, probably in Roman times, though it is not mentioned in Domesday. It became a corporate limb of the Cinque Port of Hastings, and was doubtless of considerable importance until about the end of the 14th century, when its rapid decline began as a result of the silting up of the harbour and the increase in the size of ships. In the 16th century the town was finally deserted by the Ouse, which now runs into the sea at Newhaven, and no revival of its prosperity occurred until the early 19th century, when it began to be frequented as a seaside town. Seaford is noted for the number of schools established there. The beach is pebbly though there are sands at low tide.

SEAHAM, a seaport and urban district in the Houghton-le-Spring parliamentary division of Durham, Eng., 5 mi. SSE of Sunderland by road. Pop. (1961) 25,889. Seaham is mentioned in a document of 930, but modern Seaham dates only from 1828, when the 3rd marquess of Londonderry built a harbour for the export of coal from the Rainton pits. Later the three Seaham pits were sunk.

The district was first known as Seaham Harbour and comprised Seaham and Dawdon, but in 1937 New Seaham was incorporated and the name changed to Seaham.

SEA HORSE, a small marine fish whose horselike appearance in the head and arched neck has long provided a motif in art. Sea horses belong to the genus *Hippocampus*. Like other members of the pipefish family (Syngnathidae), they have the body enclosed in bony rings, the tiny mouth at the end of a tubular bony snout, the gills lobate, the pelvic fins lacking, the anal fin minute, and the dorsal fin serving as the organ of locomotion by means of wavelike rippling. They differ from typical pipefishes in the heavier build, in carrying the head at right angles to the body, in lacking the caudal fin, and in the more extreme development of the tail end of the body into a prehensile appendage. As in the pipefishes, the female deposits the eggs in a pouch under the tail of the male, who then broods the young until they hatch. About 50 species, ranging from an inch or two to nearly a foot in length, inhabit warm seas. Sea horses are sold as dried curios, and are sometimes kept in aquariums.

LILO HESS FROM THREE LIONS

MALE SEA HORSE (HIPPOCAMPUS) WITH YOUNG JUST EMERGING FROM SWOLLEN POUCH

See also FISH. (C. L. Hs.)

SEA ISLANDS. These low-lying islands form a chain off the coast of South Carolina, Georgia and Florida between the Santee and St. Johns rivers. The more important islands are Folly, Edisto, Hunting, Parris, Port Royal and St. Helena off South Carolina; Tybee, Wassaw, Ossabaw, St. Catherines, Sapelo, St. Simons, Jekyll and Cumberland off Georgia; and Amelia off Florida. A number of wildlife refuges and parks have been established on the islands and some are popular resorts. The first Europeans who were to claim the Sea Islands and partly to occupy some of them were the Spaniards, who established Roman Catholic missions on most.

By the end of the 17th century the Spaniards had been driven off the islands north of Florida and early in the 18th century the English laid claim to them, making them a part of the colony of Carolina. An aura of romance and speculation enveloped the islands for more than two and a half centuries, beginning in 1717 when they were granted to a Sir Robert Montgomery, who made them a part of his margravate of Azilia on the mainland, and called them the Golden Islands in a promotional booklet he issued in 1720. Sir Robert's scheme met with failure, and when Georgia became a colony in 1732 and settlement began the next year, the islands were among the first parts of the colony to be occupied. The battle of Bloody Marsh fought on St. Simons in 1742 saved the colony from destruction by the Spaniards. These islands were early looked upon as private little kingdoms. St. Catherines was awarded to Mary Musgrove, a half-breed Indian princess, in payment of a debt claimed against James Oglethorpe, the principal founder of the colony; and before the end of the colonial period the island came into the possession of Button Gwinnett, one of Georgia's three signers of the Declaration of Independence. Rice and cotton plantations were established, especially on the South Carolina islands of St. Helena and Port Royal and the fine, long-stapled Sea Island cotton was developed. After the American Civil War, abandoned plantations were confiscated and the land was given to freed slaves; the population of the Sea Islands of South Carolina still is predominately Negro. (*See* also PORT ROYAL; GULLAH.)

In the 20th century, in part because of boll weevil infestation of the cotton, a more diversified agriculture developed, including corn, peanuts, potatoes, poultry and truck gardening, and Sea Island shrimp, crab and oysters were harvested in quantity. A U.S. marine corps training installation is located on Parris Island.

Of the Islands off Georgia, almost all of Sapelo in the antebellum period became the domain of Thomas Spalding, prominent Georgia slaveholder, planter and legislator; in the second half of the 20th century it remained in private hands. In the last half of the 19th century Jekyll was made an exclusive winter playground for members of the Jekyll Island club; and the Carnegie family secured most of Cumberland about the same time for the same purpose.

The state of Georgia bought Jekyll from the club and since 1947 it has been a state park. Fort Frederica national monument, established 1945, is on St. Simons which with adjoining Sea Island also has resort facilities. Jekyll, St. Simons and Sea Island are near Brunswick (*q.v.*). The Florida Sea Island, Amelia, was first settled by Oglethorpe in 1735. It became part of East Florida, Spanish, in 1783 and was ceded to the United States with the rest of Florida in 1821. The northern part of the island is occupied by a state park.

See E. M. Coulter (ed.), *Georgia's Disputed Ruins* (1937); E. M. Coulter, *Thomas Spalding of Sapelo* (1940); B. Vanstory, *Georgia's Land of the Golden Isles* (1956). (E. M. Co.)

SEAL, an aquatic mammal of the family Phocidae, suborder Pinnipedia, order Carnivora (*see* CARNIVORE). The phocids are the true seals; the eared seals, comprising the fur seals and sea lions (including the trained seals of the circus), belong to the family Otariidae (*see* SEA LION). The body is torpedo-shaped. The limbs are short, with the fingers of the forelimbs joined to form a flipper bearing separate claws, and the hind limb with the toes joined by a web. In swimming, the body is propelled by side-to-side strokes of the paddlelike hind flippers, each flipper acting alternately in the power stroke. On land, progress is laborious, the body being hitched forward by the forelimbs and pushed along by the pelvic region alternately in a looping movement.

The seal differs from the sea lion chiefly in its mode of swimming, its gait on land and its lack of external ears. There are about 18 kinds of seal, those of temperate and polar seas being more numerous than those of the tropics; one species is confined to fresh water. Seals range in size from the little fresh-water seal (*Pusa sibirica*) of Lake Baikal in south Siberia, about 3 ft. long, to the enormous sea elephant or elephant seal (*q.v.*) of the subantarctic regions, the bulls of which reach a length of 16 to 18 ft.

Nearly all seals are gregarious, at least when breeding; some kinds assemble in enormous herds on sea beaches or floating sea ice. In species such as the elephant seal and gray seal (*Halichoerus grypus*) the males take possession of harems of cows and drive rival bulls away from their territory. One young at a time is normally born to each adult cow and suckled for a comparatively short while; the pups gain weight very rapidly on the diet of seal milk, which contains about 50% fat. During the period of suckling the cow remains ashore and does not feed. The newborn pups are clothed in a coat of soft, silky fur, which is molted about the time of weaning and replaced by coarser hair. In some species, such as the common or harbour seal (*Phoca vitulina*) and the sea leopard or leopard seal (*Hydrurga leptonyx*), the first molt is precocious and may take place partly or completely before birth. The fluffy birth coat is generally white but sometimes, as in the elephant seal, it is black. The cows are again impregnated soon after the birth of the pups. The gestation period is thus about 11 months or a little more; some species (possibly all) exhibit delayed implantation, in which the growth of the embryo is arrested at a very early stage and renewed some months later, so that active gestation may actually occupy only about seven months.

Seals appear to be long-lived, with a potential span of 30 or 40 years, but their average life in the wild is considerably shorter because of heavy mortality from accident and misadventure among the younger animals. Some kinds of seal, such as the common seal, inhabit coastal waters, but others, like the harp seal (*Pagophilus groenlandicus*) of the Arctic, the Weddell seal (*Leptonychotes weddelli*) and the crabeater seal (*Lobodon carcinophagus*) of the Antarctic, live in the open sea (*i.e.*, are pelagic) and generally leave the water only to lie on floating sea ice. The coastal species are more or less sedentary, but the pelagic ones make extensive, regular migrations. Seals produce a variety of sounds from the wailing bleat of the pup through snarls and barks to the windy roars of the elephant seal. The food of many kinds of seals consists mainly of fish; some species consume large quantities of squid, other mollusks, and crustaceans. The crabeater seal lives on planktonic crustaceans that it strains from sea water through the sieve formed by the serrated cusps of its side teeth. The leopard seal of the antarctic feeds largely on penguins and other sea birds.

Seals have a thick layer of blubber under their skin and some of the gregarious species are killed in large numbers by man for the sake of the oil that can be rendered from it. Their hides are sometimes used for leather, and the skins of newborn phocids in their fluffy natal coat are prized in the fur trade. *See* SEALING; *see* also references under "Seal" in the Index. (L. H. M.)

See V. B. Scheffer, *Seals, Sea Lions and Walruses: a Review of the Pinnipedia* (1958); Carleton Ray, "Locomotion in Pinnepeds," *Nat. Hist., N.Y.*, vol. lxxii, no. 3 (March 1963).

SEA LAW: *see* MARITIME LAW.

SEA LILY, the popular name for marine animals whose adult sedentary stage, stalked and flowerlike, suggests a lily; these forms belong to the Crinoidea, a class of echinoderms that includes also the stalkless feather stars. Sea lilies occur in deeper water, often in large numbers. Only about 80 species are present today, most under 2 ft. in height, while in earlier geological periods thousands of species, some of giant size (up to 70 ft. high), existed. The remains of extinct sea lilies preserve well and are important as guide fossils.

A typical sea lily has a tall, slender stem, composed of small, five-lobed calcareous discs and a small cup-shaped body, rigid with plates, with the mouth and anus placed on the upper, soft-skinned side. From the edge of the cup arise, in the simplest forms, five articulated arms with alternating rows of shorter branches, the pinnulae. In most forms the arms fork near the base so that a multiple of the original five arms is developed. The animals feed on microscopic particles carried to the mouth along ciliated grooves present on the upper side of the arms and pinnulae. The water vascular system is well developed; the podia, which fringe the arms are delicate cylindrical types that, besides serving sensory and respiratory functions, assist in collecting food. (El. D.)

SEALING is the taking of certain genera of seals for their fur, hides, blubber or oil. Most commercial sealing is conducted on land or on ice at the breeding grounds, or rookeries, particularly those in the north Pacific and north Atlantic. For many years the taking of seals was accompanied by great waste with the result that the seal herds were threatened with extinction, but sealing in most areas is now regulated by international conventions or other agreements that aim at conserving the breeding stock. Pelagic sealing, *i.e.*, the taking of seals in the open sea while the animals are migrating to the breeding grounds, was once practised indiscriminately in the north Pacific but is now prohibited because it was especially wasteful. As a general rule, regulated sealing is conducted so as to harvest only the superfluous males.

Commercial sealing is now conducted mainly in the Pribilof Islands (*q.v.*) of the north Pacific, on pack ice in the northeast and northwest Atlantic, and in rookeries of the South Pacific. The most important seals in the sealing industry are the fur seal in the north Pacific, the harp (Greenland) seal in the north Atlantic, and the elephant seal in the South Pacific. The ringed (jar) seal is also important. (*See* Seal.)

The North Atlantic.—Seals were first taken in the subarctic region of the north Atlantic by whalers when whaling began there early in the 17th century. Later, in the most prosperous days of north Atlantic whaling, the whalers neglected the seals but by the latter part of the 19th century they included them again in their catch. Sealing as a separate occupation had meanwhile developed in the north Atlantic during the 18th century. The chief species taken were the Greenland, harp or saddleback seal (*Pagophilus groenlandicus*), hooded or bladdernose seal (*Cystophora cristata*) and bearded or squareflipper seal (*Erignathus barbatus*).

The greatest annual kills in the north Atlantic took place off Newfoundland and in the Gulf of St. Lawrence between 1820 and 1860, when more than 500,000 seals were taken in several years. The largest annual kill on record is that of 1831, when 687,000 seals (mostly harp seals) were taken by about 300 ships and 10,000 men. After 1910, between 250,000 and 300,000 seals were taken annually, except during World War II, when most of the sealing ships were inactive.

The ringed, or jar, seal (*Pusa hispida*), the commonest and most widely found seal of the arctic, is one of the most important animals to the Eskimo. In some places it was and still is the backbone of the native economy. In the Canadian arctic between 30,000 and 50,000 seals are killed annually and the skins of about 10,000 of these animals are traded to the Hudson's Bay company each year. The skin of the one-year-old animal is particularly prized by fur dealers and is called the silver jar. The gray seal (*Halichoerus grypus*) and the common or harbour seal (*Phoca vitulina*) are not the subject of organized commercial exploitation. Both the gray seal and common seal, however, are killed by fishermen, in some places for a government bounty, because they are the primary hosts for certain parasites of commercial fish and because they are a nuisance around certain fishing operations.

Sealing in the north Atlantic is done in the spring on floating pack ice from the White sea to the Gulf of St. Lawrence. The chief species taken is the harp (Greenland) seal. There are three distinct breeding stocks, all of which are exploited commercially. An eastern stock breeds in the White sea (now a Soviet fishery exclusively), a central stock in the region of Jan Mayen island (U.S.S.R. and Norway) and a third off Labrador and Newfoundland and in the Gulf of St. Lawrence (Norway, Canada, U.S.S.R., U.S., Denmark and France). The species is strongly migratory; it occupies the pack ice area during the breeding season in the spring and then moves north to open water in the arctic in summer and autumn.

The white-coated pups, which are born on the ice, form the basis of the international Atlantic commercial seal fishery. Before World War II, almost 90% of the catch was of pups. Sealing in this area was primarily a Newfoundland industry. After 1949 the centre of the industry shifted to Nova Scotia.

In another post-World War II development, a Norwegian fleet of about 16 sealing vessels began to fish off Newfoundland and in the Gulf of St. Lawrence when their Soviet concessions in the White sea expired. About the same number of Canadian vessels were added to the northwest Atlantic sealing fleet at the same time. After these ships appeared, many older seals began to be taken until they finally constituted about 30% of the catch.

The pups are taken on the ice beginning early in March. Bedlamers (immature harp seals one year old or older) and adults are shot on the ice while molting during April and May. The pelts with their adherent blubber, called sculps, are removed on the ice and dragged or winched to the sealing ship; there they are stored in iced or refrigerated holds. The annual catch of harp seals in the 15 years following the end of World War II, including landsmen's net catches in Canada and Greenland, averaged about 225,000 pups and 85,000 immatures and adults, with about 90% of the catch shared equally by Canada and Norway. Sealers in both east and west areas of the north Atlantic used airplanes after World War II to locate herds.

Oil, fur and leather are the three major commodities produced from seals in the north Atlantic. The pup is the basis of the fur stock, though the short hair of the immatures is also used. Neither the pup nor the immature furnishes true fur and the value of their pelts is less than that of the true fur undercoat of the fur seal. The white coat of the pup consists of fetal hair that is replaced, beginning about 10 days after birth, by a short, spotted hair coat that grows in under the white coat. There were no formal laws in the early 1960s governing the killing of seals in international waters of the western Atlantic and the herd was being seriously reduced. In 1961 Canada proposed to the International Commission for Northwest Atlantic Fisheries (ICNAF) that the conservation of harp and hooded seals be placed under that organization. This was accepted by the commission and ratified by all member countries. Sealing in the eastern Atlantic was regulated by a Soviet-Norwegian treaty on arctic marine mammals.

The North Pacific.—The fur seals, which, with the sea lions, belong to the family of eared seals (Otariidae), differ from other seals in that they possess a permanent undercoating of short, soft fur that constitutes the sealskin of the costumer (*see* Sea Lion). The Pacific fur seal (*Callorhinus ursinus*) began to be exploited after V. J. Bering made his voyage to the north Pacific in 1741. The migrations of the fur seal cover the area from the latitude of southern Japan on the west and from southern California in the east to the great rookeries on the Komandorski (Commander) and Pribilof islands on the north. At their heights in the 19th century the Pribilof herd included about 2,500,000 seals and the Komandorski group totaled more than 1,000,000. A herd on Robben (Tyuleniy) Island in the Sea of Okhotsk was of minor importance. These herds were greatly reduced in the last half of the 19th century and by 1897 did not exceed a combined total of 600,000 animals. As a result of international treaties and control by the U.S. bureau of fisheries, the Pribilof herd increased from 125,000 in 1911 to about 1,000,000 in the second half of the 20th century.

Pelagic sealing in the north Pacific began in the early 19th century. At first the catch averaged 75,000 annually but after about 1868 it increased rapidly. Beginning in 1879, sailing vessels carrying canoes were employed to attack the migrating seals, which were thus harvested just before the breeding females gave birth to their pups; in some of the catches, half of the animals were pregnant females. At one time the pelagic sealing vessels exceeded 100 in number and some of them carried as many as 25

canoe crews. It has been estimated that by 1902 about 1,000,000 seals had been taken at sea.

The greatest catch, however, was always made on shore. From the Pribilof and Komandorski herds nearly 2,500,000 seals were taken on land between 1868 and 1897.

From a conservation standpoint it is fortunate that the habits of the seals on land permit their taking with the least possible depletion of the breeding stock. The young males, or bachelors, leave the rest of the herd to rest and sleep on beaches that are adjacent to but apart from the breeding grounds. There they are surrounded at night by the sealing crews, rounded up in droves of from 1,000 to 3,000 and driven inland to the killing grounds. The large droves are broken up into groups of from 20 to 50; the males three years of age or in that size class are knocked down with clubs, while those that are too large or too small are allowed to escape. The skins are removed, salted in kenches and, when cured, exported.

Many laws and treaties have been enacted to protect the north Pacific seals while they are at sea and thus to ensure a continued sealing industry. In 1911 an international convention was signed by Great Britain (for Canada), Japan, Russia and the United States to prevent pelagic sealing by their nationals. Land sealing also was regulated. At that time the herd on the Pribilofs numbered about 125,000 animals. The annual take there now runs between 50,000 and 70,000 skins, and the herd numbers well over 1,000,000 animals. The convention of 1911 was terminated in 1941 by the withdrawal of Japan. A provisional agreement between Canada and the United States ensured continued pelagic protection off North America. Then in 1957, Canada, Japan, the U.S.S.R. and the U.S. signed the Interim Convention on Conservation of North Pacific Fur Seals. This pact was to last six years, during which time commercial pelagic sealing continued to be prohibited, and an extensive pelagic and land research program on fur seals was undertaken by the four member countries to form the basis for negotiating a new convention. The North Pacific Fur Seal commission was set up by the four member countries to coordinate the work. At the end of the six-year period the time for research was extended. (*See* also ALASKA: *Diplomacy Involving Alaska;* WILDLIFE CONSERVATION: *Conservation of Sea Mammals and Fish: Seals.*)

The South Atlantic and Pacific.—The southern fur seal (*Arctocephalus australis*) was once taken at the Galapagos Islands, Tierra del Fuego and Lobos Islands. This or other species also were taken at South Africa, Australia, New Zealand and many points near the Antarctic circle. It was taken before 1793 and was once abundant in South Georgia and other dependencies of the Falkland Islands.

Great numbers of sea leopard or leopard seal (*Hydrurga leptonyx*), sea elephants or elephant seals (*Mirounga angustirostris* and *M. leonina*), Weddell seal (*Leptonychotes weddelli*) and other species were seen by early explorers and voyagers, but at first the skins of the fur seals alone seem to have been taken. One of the earliest recorded landings for seals is that of the Argentine ship "Juan Nepomucena," which brought in 13,000 skins in 1820. In that and the two following years more than 90 vessels, divided about equally between Great Britain and the United States, worked the southern grounds. In the first season, catches of 18,000 were not unusual for individual ships and five British ships took a combined total of 95,000 seals. Seal oil and blubber, particularly from the elephant seals, began to be taken. James Weddell estimated that in the two seasons of 1820–21 and 1821–22, 1,200,000 fur seals were taken from South Georgia and 320,000 from the South Shetlands, along with 940 tons of elephant seal oil. By 1892, because of the diminution of the herds, sealing vessels returned to their home ports from South America with mixed cargoes; and though in the early 1890s a Scottish whaling expedition to the Ross sea took 20,000 sealskins with four ships, by the end of the 19th century the fur seal had almost completely disappeared from the Falkland Island dependencies and other seals, sea elephants in particular, had greatly diminished in number.

From 1881 sealing in these territories has been regulated; closed seasons were introduced and sealing is now permitted only by licences, which may specify both the kind and number of seals taken. The capture of fur seals is prohibited.

Most sealing in the southern region is conducted in pursuit of the elephant seal; this sealing has responded satisfactorily to regulations. The elephant seal is taken by whalers, but pups may not be taken nor, as far as practicable, are female seals to be harvested. There is also a closed season and areas along certain stretches of coast are closed. The absence of segregation of young males on the rookeries is a hindrance to the observance of the regulations. Elephant seals are of great size, the females reaching 11½ ft. and the male sometimes 22 ft. in length. Mating takes place within the harem immediately after the young are born early in October, and the young, which are born singly, are usually weaned in November; these circumstances have determined the closed season. Shortly after mid-20th century, the number of seals taken at South Georgia was about 6,000 per annum.

See ELEPHANT SEAL.

BIBLIOGRAPHY.—V. B. Scheffer, *Seals, Sea Lions and Walruses* (1958); K. W. Kenyon *et al.*, "A Population Study of the Alaska Fur-seal Herd," U.S. Fish and Wildlife Service, Special Scientific Report, *Wildlife no. 12* (1954); H. D. Fisher, "Utilization of Atlantic Harp Seal Populations," *Trans., 20th N. Amer. Wildl. Conf.* (1955); B. Rasmussen, "Exploitation and Protection of the East Greenland Seal Herds," *Norwegian Whaling Gazette*, no. 2 (1957); B. Roberts, "The Protection of Fur Seals in the North Pacific Ocean," *The Polar Record*, no. 30 (1945); R. M. Laws, "The Elephant Seal Industry at South Georgia," *The Polar Record*, no. 46 (1953); F. H. Fay, "History and Present Status of the Pacific Walrus Population," *Trans., 22nd N. Amer. Wildl. Conf.* (1957); A. W. Mansfield, "The Walrus in the Canadian Arctic," Fisheries Research Board of Canada, *Arctic Unit Circular no. 2* (1959), "Seals of Arctic and Eastern Canada," Fisheries Research Board of Canada, *Bulletin no. 137* (1963); Bureau of Fisheries, *Alaska Fishery and Fur Seal Industries*, annual reports of the Alaska division, U.S. Fish and Wildlife Service; David Sergeant, "Harp Seals and the Sealing Industry," *Canadian Audubon* (March–April 1963). (H. D. F.)

SEALING WAX was once widely used to seal letters and attach impressions of seals to official documents. In medieval times it consisted first of a mixture of beeswax, Venice turpentine, and colouring matter, usually vermilion, but when lac from the East Indies was introduced into Europe by the Venetians it displaced the beeswax. Sealing wax was prepared by melting the rosin in a copper or earthenware pot and adding the colouring matter slowly while stirring. The molten mixture was poured into stick-shaped molds. In later times, sealing waxes containing admixtures of chalk (calcium carbonate), magnesium carbonate, and barite white (barium sulfate) and employing mineral pigments were developed. In very inferior waxes, ordinary rosin was substituted for lac.

Since the advent of the gummed envelope and various methods of affixing official seals, sealing wax has had only a few applications as a sealing material. It has been used with some success in decorative handwork. (E. L. Y.)

SEA LION, a name given generally to all members of the eared seal family (Otariidae), also called otaries; in a more restricted sense it is applied to the larger forms, such as the northern or Steller's sea lion (*Eumetopias jubata*), and to species of *Zalophus*, to which belongs the California sea lion (*Z. californianus*). Otaries, along with true seals (Phocidae) and walruses (Odobenidae) constitute a suborder of Carnivora (*see* CARNIVORE: *Seals and Seal Allies* [*Pinnipedia*]).

There are about 15 species of otaries, frequenting the North and South Pacific ocean, the South Atlantic and the southern oceans but entirely absent from the North Atlantic. One of the southern species extends up the west coast of South America as far as the Galápagos Islands. Some species of otary undertake regular and extensive annual migrations; those breeding on the islands of the North Pacific ocean journey as far as California every year. One of the smaller species, the California sea lion, is commonly kept in captivity and is the kind generally seen in circuses, trained to balance objects on its snout.

All otaries have a thick layer of blubber beneath the hairy skin and, though fur seals are principally valued for their pelts, the so-called "hair seals" are hunted for the oil obtainable from their blubber.

Otaries differ from true seals or phocids in having external ears

and a longer and more distinct neck, and in the formation of their limbs. The limbs, as in the true seals, are flippers with the digits joined, but the toes of otaries bear separate straplike extensions beyond the last joint. The locomotion of otaries is quite different from that of phocids. Their forelimbs are longer than those of phocids and their hindlimbs are long enough to be rotated forward under the body when the animal is on land. The body can thus be lifted clear of the ground; otaries can lope along at a clumsy gallop for a limited distance at considerable speed. In swimming, the body is propelled by powerful strokes of the forelimbs, whose effective area is increased by a fold of skin joining the trailing edge to the body wall; the hindlimbs, unlike those of the true seals, are of secondary importance in swimming.

Adult males and females differ greatly in size. The mature bulls in some species are more than twice as large as the cows and have an enormous development of the neck muscles and of the mane on the head and neck.

The Otariidae also includes the fur seals. These have a thick, soft fur beneath their coarse outer hair; the pelts of the young bulls and of the females form the seal fur of commerce. The wild herds are scientifically managed on their breeding grounds so that an annual crop of pelts is harvested without risking extermination of the animals—a possibility in former times, when indiscriminate slaughter was practised.

Various kinds of fur seals are found in both northern and southern subarctic and temperate seas. The northern fur seal (*Callorhinus ursinus*) ranges along the North American coast from Alaska to the Mexican border and breeds in the Pribilof Islands. The southern fur seals (*Arctocephalus* or *Arctophoca* species) range from sub-Antarctica to southern South America, Africa, Australia and New Zealand.

The voice of an otary is harsher than that of a phocid, and ranges from a sharp bark or honk to a growling roar. Otaries are strictly gregarious in the breeding season; some species gather on remote islands in herds numbering millions. They have a well-developed harem system, with the master bulls fiercely defending their territories and the younger, bachelor bulls gathering in segregated herds.

Otaries feed on squid and fishes. During their time ashore the bulls fast for several months; but the cows leave the beaches at intervals, presumably to feed, after their pups are born. The gestation period is about 10 or 11 months; but, as with true seals, there is probably a delay in implantation of the embryo.

(L. H. M.)

SEALS. The word "seal" (Latin *sigillum;* Old French *scel*) denotes the result of the impact of a hard engraved surface, the die or matrix, upon a softer material such as clay or wax, producing a device in relief, the impression, whereby from remote antiquity, as shown by cylinder seals from Babylonian sites, from Egypt, and from Crete, and later by the intaglio gems of classical Greece and Rome, the authenticity of documents was established in the manner of the modern written signature. The term "seal" is apt to be ambiguous, and in medieval times was sometimes used indifferently for the matrix or the impression; where there is doubt, it should be applied to the latter. For seals in antiquity *see* BABYLONIA AND ASSYRIA: *Art and Art Objects: Glyptic;* AEGEAN CIVILIZATION; CRETE: *Archaeology: Seals;* GEM.

EUROPEAN SEALS

After the fall of the Western Roman Empire in 476 the practice of sealing declined, and except that the popes, following a custom that originated with the Eastern emperors, used bullae (or lead seals) from the middle of the 7th century, and some very debased examples were used by the Visigoth and Merovingian kings, it was not revived in Western Europe until Carolingian times under Pepin the Short (d. 768). The seals then employed were either antique gems or contemporary derivatives: Charlemagne (d. 814) had a gem engraved with Jupiter Sarapis; later the German king and Holy Roman emperor Otto I the Great (d. 973) used a portrait bust of himself. In England the antique gem appeared on the counterseal of Durham Cathedral in the 10th century. Rock crystal was specially favoured, a conspicuous example being the crystal intaglio of Lothair II, king of Lorraine (d. 869), now set in a reliquary cross and preserved in the cathedral treasury at Aachen. Engraved gems were generally set in a metalwork mount on which the legend was engraved.

The earliest example of a "seal of majesty" with the monarch crowned and enthroned appears to be that of the emperor Otto III (d. 1002). In England there is no evidence of this before Edward the Confessor, after whom there is a continuous series down to modern times.

Before the end of the 11th century seals were used only by sovereigns, the higher clergy, and nobles, but by the end of the century sealing had become more general, and by the 13th century seals were used by craftsmen and tradesmen. While one seal would suffice for these, persons holding high rank would require more. The sovereign in addition to his great seal would have a privy or secret seal, a signet, generally a ring, and deputed seals for courts and officials, and various official seals of commerce. Bishops, like corporate bodies, had normally a seal of dignity, another *ad causas* (for current business), a private seal, and a signet. The lesser seals of persons who also possessed seals of office are known as "small seals."

The study of seals (sigillography) is important not only in the fields of genealogy and diplomatics but also in art history, particularly since most seals can be fairly closely dated. In Europe the evolution of style can be followed in miniature by this means. The late 12th century marks the beginning of the great period of seal engraving, which reached its zenith in the 13th and 14th centuries. The Renaissance brought a taste for opulent and intricate detail in the devices even in the heraldic shields, where perhaps it is shown at its best. About this time signet rings, often set with engraved gems, had their vogue. For the 17th century and later nonannular fob seals and desk seals are characteristic, surviving until modern times.

The Matrix.—For the use of engraved gems as seals, *see* GEM. The commonest materials were bronze or latten, but gold matrices are recorded as having been used by royal personages and many silver examples have survived in all categories. Other materials were lead, pewter, ivory, and occasionally jet. (In the case of bullae, where the impression itself was usually of lead, the matrix might be of steel or iron.)

Where seal and counterseal were of the same size, one matrix would be provided with pierced lugs, into which pegs on the other would be inserted, the impression being secured by a press. More usually, however, the matrix had one face, and was applied by means of a handle, a flange at the back, a ridge with a loop for attachment, or a hexagonal cone surmounted by a pierced trefoil with or without a loop at the top.

Sizes and shapes varied, the largest seals being generally, though not invariably, official, either secular or religious. Of the many possible shapes the most usual were the circular and the oval or pointed oval, the latter particularly appropriate for representing standing figures.

Engravers.—Research has revealed the names of many English seal engravers, and there are records from the 12th century; apparently the earliest extant example is the silver corporation seal of Exeter, signed by a certain Luke, perhaps about 1200. For official seals there is a fairly continuous record, containing such makers of royal seals as Walter de Ripa (first great seal of Henry III), Nicholas Hilliard (1547–1619), goldsmith and miniaturist (second great seal of Elizabeth I), Thomas Simon, the finest medalist and engraver of his time (seals of Cromwell and Charles II), and members of the Wyon family (active from George IV to Victoria). J. Roman (see *Bibliography,* below) quotes many names of French seal engravers, and some names of craftsmen from other European countries are given in W. Ewald, *Siegelkunde* (1914).

Seal Impressions.—Seal impressions are most commonly of wax—either of natural colour, or red, black, green, or brown. The majority of metal bullae are of lead, like the numerous surviving papal examples, but some sovereigns had bullae struck in silver or gold either as a mark of their own dignity or to confer special honour on the recipient of a charter.

BY COURTESY OF (LEFT AND RIGHT) THE PUBLIC RECORD OFFICE, LONDON; (TOP CENTRE) ARCHIVES CANTONALES VAUDOISES, LAUSANNE; (BOTTOM CENTRE) THE TRUSTEES OF THE BRITISH MUSEUM

(LEFT AND RIGHT) SECOND GREAT SEAL OF ELIZABETH I OF ENGLAND, DESIGNED AND ENGRAVED BY NICHOLAS HILLIARD AND DERICKE ANTHONY, 1584–86: OBVERSE, THE QUEEN ENTHRONED, REVERSE, THE QUEEN ON HORSEBACK; (TOP CENTRE), "SEAL OF MAJESTY" OF THE GERMAN OTTO III (980–1002), DEPICTING HIM CROWNED AND ENTHRONED; (BOTTOM CENTRE) OBVERSE OF THE FIRST SEAL OF THE ENGLISH KING EDWARD THE CONFESSOR, WHO REIGNED 1042–66

Except in the case of metal bullae, where suspension has always been the rule, seals were applied directly to documents (*en placard*) in the earlier periods. In England suspended seals came into being in the 11th century; in France impression on the deed continued till the beginning of the 12th and intermittently until its revival toward the end of the 13th. Suspension was achieved by cutting a tongue along the bottom of the document, to which a single or double seal could be affixed, or by a slit in the bottom of the document through which a parchment strip could pass, the loose ends being joined by the seal. From about the end of the 12th century silk and woolen cords were used by the higher ranks for important documents. About the end of the medieval period attachment *en placard* again became general.

The Counterseal.—The larger of the suspended seals normally bore a counterseal on the back for greater security, either the reverse of a double seal, or a secret seal, a signet, or an *ad causas* seal.

Devices and Legends.—The face of the majority of seals shows a central device with a surrounding inscription or legend commonly indicating the person or institution entitled to the seal, although on small private seals it may be a secular motto or a devotional sentiment. Latin appears at all periods, but English and French are found from the 13th century, sometimes along with Latin.

The word *sigillum* precedes the owner's name and description in the genitive case, but for royal seals the nominative without *sigillum* is used, as also on the earlier medieval bishops' seals of dignity. Abbreviations, owing to limited space, are common, and signs by way of punctuation usually separate the words. The types of lettering are, in very roughly chronological order, Roman capitals, Lombardic, black letter, and in post-medieval times Roman capitals again.

Types of Seals

Secular.—*Royal Seals.*—The earliest and most complete series is that of the French kings, where evolution can be followed from Carolingian times, the grand series of round seals with an enthroned figure of the king beginning with Henry I (d. 1060). In England the first royal seal which ranks as a "great seal" is that of Edward the Confessor, impressions of which are still extant. The seal was furnished with a counterseal, the design being nearly identical with that of the obverse. William the Conqueror, as duke of Normandy, used an equestrian seal, representing him mounted and armed for battle. After the conquest of England, he added a seal of majesty, copied from the seal of Henry I of France, as a counterseal. In subsequent reigns the order of the two seals was reversed, the seal of majesty becoming the obverse, and the reverse being the equestrian seal, a pattern that has continued to modern times. In the later Middle Ages royal seals became increasingly elaborate and a climax of magnificence was reached in the seal of Henry V of England (d. 1422).

Other Categories.—Interesting and important officers' seals are those of the admirals, extant examples being of value in showing the evolution of the ship in medieval and later periods. Commerce and industry are illustrated by seals used for customs, seals relating to the delivery of wool and hides, and cloth-subsidy seals. Unofficial commercial seals are exemplified by the seals of the livery companies and companies trading abroad. On many seals of private traders the device or the legend may indicate the occupation. Local seals comprise those of countries, towns, and their officials, various devices being found, sometimes the place itself. Castles are common, being a feature of English sheriffs' seals; some seals show bridges, while others may have a religious subject or merely a shield of arms. An exceptional class is the English series for labourers' passes under the Statute of Cambridge, 1388.

Universities show the chancellor and masters in convocation, while the colleges have patron saints or founders, a religious emblem as a shield of arms; the faculties and nations generally had seals, a common design being a doctor or master lecturing.

Religious Seals.—*Episcopal.*—The characteristic medieval seals of dignity from about the 11th century show the bishop blessing, robed and mitred and holding a crozier. As time went on, changes in vestments appeared and elaboration in the form of canopies, flanking figures, and tabernacle-work, giving place later to scriptural subjects or often to groups of saints, with the bishop in adoration at the base, with or without a shield of arms.

Monastic Seals.—Much the same evolution is seen in monastic seals, where the device is often appropriate to or illustrative of the institution. Armorial bearings are a common feature, whether subordinate to the principal subject or, later, the sole device. An exceptional group are the greater seals of the bishops of Durham as lords palatine with the enthroned figure on one face and the bishop as an armed knight on the other.

Some fine examples are found, in which the device is a building, probably that indicated by the legend; the splendid seals of Boxgrove Priory, Sussex, and Haltemprice Priory in the East Riding of Yorkshire are conspicuous examples.

Papal Seals.—Papal bullae form a class of their own. On the obverse are the heads of SS. Peter and Paul, and on the reverse the pope's name. Archaic from the beginning, they have so continued, in spite of attempts by Paul II to alter and by Julius II to improve the design.

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

REVERSE (ABOVE) AND OBVERSE (BELOW) OF THE SEAL OF WILLIAM I (THE CONQUEROR) WHO REIGNED 1066–87: REVERSE SHOWS THE KING ENTHRONED; OBVERSE SHOWS HIM FULLY MOUNTED AND ARMED

Personal Seals.—With private and personal seals a division falls naturally between heraldic and nonheraldic. The earliest type was the equestrian, showing the mounted owner fully armed, which began in the 11th century. From the second quarter of the 12th century, these seals show armorials and are valuable illustration of the changes in arms and armour; their best period is the late 13th and early 14th centuries. The purely armorial seal dates from the late 12th century; on it the shield is the sole or principal device, as in many of the categories of official commercial seals. Private seals probably appear about the end of the 13th century, usually small and circular; the succeeding centuries show increased elaboration of the means of filling the field, until in the 15th and 16th centuries the full armorial achievement is developed.

BY COURTESY OF THE BIBLIOTHEQUE NATIONALE, PARIS

SEAL OF THE UNIVERSITY OF PARIS, 1292: VIRGIN AND CHILD ENTHRONED (TOP CENTRE) SURROUNDED BY A BISHOP AT LEFT, STUDENTS READING BELOW, AND SAINT AT RIGHT

The seals of persons not entitled to arms show a great variety of devices, conspicuous among which are merchants' marks. Some bear their owner's name round fleurs-de-lis crescents, stars, and various conventional designs; others have legends describing the seal or expressing sentiments. Animal figures are also found on some seals, with relevant legends. (A. B. To.)

Small Seals.—Small seals supplemented the great seal. They were not duplicates of it nor, although in an emergency they could be used in its stead, were they intended to be substitutes. For instance, a late medieval bishop, with a seal *ad causas* for the administration of his diocese, would use a signet (sometimes called a seal *ad arma*) for personal correspondence and secular business; he might occasionally use this signet as a counterseal. Although originally designed for the private purposes of their owners, some small seals acquired an official character. Most medieval sovereigns had at least two small seals, which were placed in the custody of certain officials. Except in England, they were invariably controlled by chancery, as for example, in the papacy and in France. (*See* CHANCELLOR: *France*.) The English small seals' freedom from such direction allowed them to evolve autonomous offices which gave rise to some of the chief ministries of state.

Like some other developments in royal administration, the first small seal known to have been used in England dates from the reign of King John. This was the privy seal kept by the clerks of the king's chamber (*q.v.*) for use in business relating to the king's domestic economy. When the wardrobe assumed responsibility for such matters in the reign of Henry III, the privy seal was transferred to this department, and under Edward I and Edward II the controller of the wardrobe had charge of it. Although chancery at first enrolled writs of privy seal, the wardrobe itself recorded these writs in registers in the reigns of Edward I and his son. Many drafts of privy seal letters, originally kept in files, remain from this period.

As a result of the action of the ordainers in 1311, the privy seal was removed from the wardrobe and entrusted to its own keeper; the succession of keepers of the privy seal can be traced from 1312. The keeper became one of the foremost officers of state, ranking after the chancellor and treasurer. Except for the tenure of Nicholas Carew (1371–77), the seal was kept by clerks who usually received bishoprics in recognition of this service. The title lord privy seal first appeared when the office was held by Richard Foxe (1487–1516). Since 1530 the title has been held by laymen, who until recent years were usually peers.

The records of the privy seal office continued to be kept in files in the 15th century. Despite misfortunes, many have survived, some being published under the title of *Proceedings and Ordinances of the Privy Council of England* (ed. by Sir N. H. Nicolas, 1834–37). This mistaken description arose from the frequent notices on these records of activity by the king's council. As the council did not possess a seal or keep its own registers of proceedings until the 16th century, the privy seal office acted as the council's secretariat. It remained, however, a major instrument of the medieval king's personal government. Thousands of writs of privy seal which were sent into chancery to authorize the chancellor to issue letters under the great seal survive to this day, and many others are to be found in the records of the exchequer and other departments which required such writs before expenditure could be incurred. From drafts and formularies it can be observed that the privy seal was used for royal letters sent to foreign monarchs and to officers and subjects in England as well as those overseas. One important function of the privy seal office was the preparation of contracts for military service to the crown; the privy seal was attached to the part of the indenture kept by the king's retainer.

A number of royal seals smaller than the privy seal appeared during the 14th century. Edward II and Edward III had each in their early years a secret seal kept by their chamber clerks. Edward III also had small seals known as the *signum*, the novel signet and a griffin signet (the last for the administration of chamber lands). These seals were all short-lived, and a second small seal did not attain a significant position in the royal administration until the reign of Richard II. The rise of the signet and of the office of its keeper, the secretary, was due to the loss of royal control over the privy seal, which had attracted the attention of the baronial opposition. Although royal control was not lost permanently, the established position and multifarious activities of the privy seal office now demanded that it should have a fixed headquarters: the keeper rented an inn in London and the seal "went out of court." From time to time, however, the keeper was required to attend the sovereign on his journeys, and the privy seal long retained a great advantage over the signet in that it was well known and accepted without question. From 1418 to 1421, when Henry V was in France, there were two privy seal matrices, one with the king, the other with the council in England.

The uncertain standing of the signet was revealed in the reign of Henry VI. The king did not have a signet, or a secretary, in

BY COURTESY OF (TOP LEFT AND TOP RIGHT) THE TRUSTEES OF THE BRITISH MUSEUM; (CENTRE, BOTTOM LEFT AND BOTTOM RIGHT) THE SOCIETY OF ANTIQUARIES, LONDON

RELIGIOUS SEALS

(Top left, top right) Obverse and reverse of the seal of the Benedictine Priory of St. Mary and St. Blaise, Boxgrove, Sussex, 13th century: obverse shows the facade of the priory incorporating a scene of the Annunciation and portraits of Christ and St. Blaise, reverse shows the Virgin enthroned; (centre) seal of Alexander Neville, archbishop of York, 1374–88, shows the archbishop seated, surrounded by saints and angels; (bottom left and bottom right) obverse and reverse of the seal of Merton Priory, Surrey, 1241: obverse shows the Virgin enthroned, reverse depicts the patron saint, St. Augustine, robed and blessing

BY COURTESY OF THE PUBLIC RECORD OFFICE, LONDON

SMALL SEALS

(Top left) Signet of Richard II, 1396, with a shield of arms: arms of St. Edward impaling old France and England quarterly with a crown above; (top right) privy seal of Edward I, 1304, the earliest extant impression of an English seal, with a shield of the arms of England; (bottom left) privy seal of the Court of Wards and Liveries, 1535: ornamental shields of arms of France and England quarterly ensigned by a crown and supported by a lion and griffin, all upheld by two naked children; (bottom right) secret seal of John, Count of Mortain (later King John of England), 1189–94, showing him mounted, armed and bearing a shield

England until he came of age in 1436, and in his periods of incapacity they ceased to be employed. Under Edward IV, however, the signet office was so securely established that it survived during the brief reign of his young son, Edward V. The earliest surviving register of the signet office dates from this time (1483). The monarch now tended to prefer to employ the signet to make the royal will known. The secretary supplanted the keeper of the privy seal as the king's attendant secretarial officer, and also took over his leading part in the management of relations with foreign powers. The activities of the privy seal office became increasingly formal, consisting largely of the issue of letters ordered by warrants under the signet.

Until 1533 the office of secretary was held by a clerk whose personal standing was not particularly prominent, but Thomas Cromwell, the first layman to hold the office, achieved such power that after his fall in 1540 Henry VIII appointed two secretaries. Cromwell was the first secretary of state in the modern sense, and from his time the office has followed its own line of development. The secretary detached himself from the signet office, which also became a purely formal department, issuing letters when required to do so by warrants under the royal sign manual. Up to the 19th century, the signet and privy seal offices were merely links in the bureaucratic chain whereby grants by the sovereign were ultimately made effective by letters patent under the great seal in chancery. The signet office was abolished in 1851 and the privy seal office in 1884. The lord keeper was retained and is now a member of the cabinet while there are seven secretaries of state who still each receive three seals from the sovereign at the time of their appointment and use the second of these (the lesser signet) for all government business requiring the sanction of an official seal. (R. L. S.)

SEAL OF THE UNITED STATES

When the United States came into existence the use of seals to authenticate important state documents was the prevailing practice, and the matter of obtaining a suitable seal was one of the first to which attention was given; on July 4, 1776, within a few hours after agreeing to the Declaration of Independence, the Continental Congress named a committee of three, Benjamin Franklin, John Adams, and Thomas Jefferson, "to bring in a device for a seal for the United States of America." There were, however, many delays; the report of this committee was tabled and two further committees were appointed before a decision was reached, on June 20, 1782.

The design in the report then accepted incorporated the suggestions of a number of people, but those primarily responsible were William Barton of Philadelphia, who had some knowledge of heraldry, and Charles Thomson, also from Philadelphia, the secretary of the Congress.

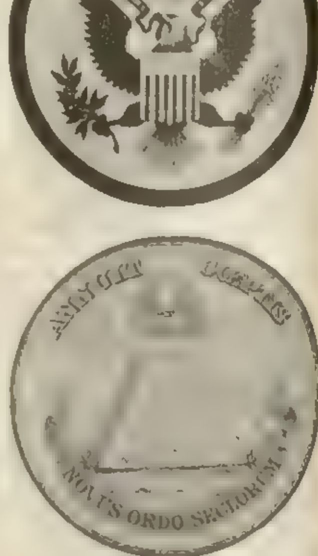
The description or blazon, which reads as follows, remains part of the law of the land today:

ARMS. Paleways of thirteen pieces, argent and gules; a chief, azure; the escutcheon on the breast of the American eagle displayed proper, holding in his dexter talon an olive branch, and in his sinister a bundle of thirteen arrows, all proper, and in his beak a scroll, inscribed with the motto, *"E Pluribus Unum."*

For the CREST. Over the head of the eagle, which appears above the escutcheon, a glory, or, breaking through a cloud, proper, and surrounding thirteen stars, forming a constellation, argent, on an azure field.

REVERSE. A pyramid unfinished. In the zenith, an eye in a triangle, surrounded with a glory proper. Over the eye these words, *"Annuit Cœptis."* On the base of the pyramid the numerical letters MDCCLXXVI. And underneath the following motto, *"Novus Ordo Seclorum."*

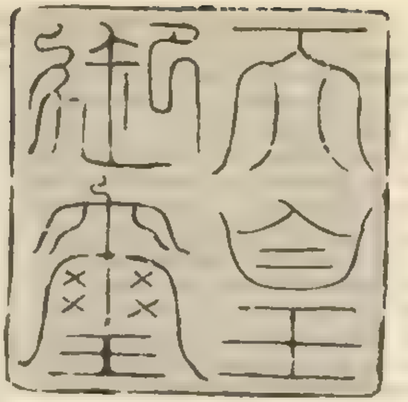
Soon after the action of Congress the obverse of the seal was cut in brass and was used as early as Sept. 16, 1782, on a full power issued to General Washington to arrange with the British for exchange of prisoners of war. By the act of Congress of Sept. 15, 1789, which changed the Department of Foreign Affairs to the Department of State, the seal of 1782 was declared to be the seal of the United States and the secretary of state was made its custodian.



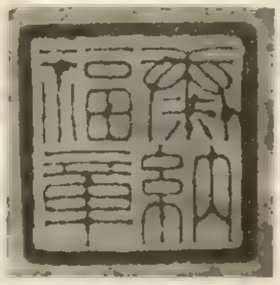
OBVERSE (ABOVE) AND REVERSE (BELOW) OF THE GREAT SEAL OF THE UNITED STATES

In the course of its history six (or possibly seven) dies of the seal of the United States have been cut and used officially. The first, executed in brass, was employed as early as Sept. 16, 1782, and as late as April 24, 1841. This seal measures about $2\frac{1}{4}$ in. (6 cm.) in diameter. Its distinguishing characteristics are a border resembling a chain of flowers, six-pointed stars and the arrows touching the border. Intended for impression on wax, it had only one face, cut in intaglio. Almost invariably it was impressed on a circular paper wafer, a thin layer of red wax being introduced between the wafer and the document for the double purpose of attaching the wafer and bringing out the device in relief. The second die was cut in 1825. About $4\frac{1}{2}$ in. (11 cm.) in diameter, it depicts the eagle realistically rather than heraldically. This die did not supersede the first, but was employed concurrently with it, being reserved for preparing pendant seals.

The third die, which superseded the first, was used from April 1841 until November 1877. Of approximately the same size as the seal of 1782, it differs in the style of its execution. The device has the appearance of being crowded toward the top; the stars, which for the first time are five-pointed, are minute; and the sheaf of arrows departs from the law in that it includes not thirteen, but six. During the first twenty years or more of its service, this die, like that of 1782, was impressed on a wafer over wax; thereafter glue or paste replaced the wax for attaching the wafer to the document, and there are indications that a crude counter-die may have been provided. The die that superseded the third was used from November 1877 until April 1885. Measuring about 2¼ in. (6 cm.) in diameter, it was executed in close imitation of the seal of 1841; and like that seal it departs from the law in having only six arrows in the sinister talon. It is readily distinguished from the earlier seal by the larger size of its stars; and it was provided with a counter-die.





(Both items, above left) Ivory seal with lion handle, inscribed with the name Chang En-jung: China, 19th century; (top centre) emperor's seal, Japan; (circular seal at centre) seal of Toyotomi Hideyoshi, 16th-century Japanese general; (right and lower right) soapstone seal with lion handle, inscribed with the name T'ai Na-fu: China, 19th century

BY COURTESY OF (LEFT, ABOVE AND BELOW) YAMANAKA AND COMPANY; (CENTRE, TOP AND BOTTOM LEFT) "SEKAI DAIJITEN"; (RIGHT AND CENTRE BOTTOM RIGHT) MRS. ALFRED E. COHN

Criticism of the faulty design of the seal then in use led to an act of Congress of July 7, 1884, appropriating $1,000 to "enable the Secretary of State to obtain dies of the obverse and reverse of the seal of the United States, and the appliances necessary for making impressions from and for the preservation of the same." The design of the obverse was an enlargement of the seal of 1782 with modifications aimed at artistic improvement and stricter adherence to the original resolution creating the seal. Although the act of 1884 provided also for cutting the reverse, it was decided to leave this provision unexecuted; and the reverse of the seal, "spiritless, prosaic, heavy, and inappropriate," has remained uncut and unused to this day. Its diameter of 3 in. (8 cm.) distinguishes the die of 1885 from all previous dies. Provided with a counter-die, it was usually impressed over a paper wafer pasted to the document; but there are examples of its impression directly upon a document, a practice authorized by act of Congress of May 31, 1854. About 1888 the present style of wafer, with invected edge, replaced the serrated form previously used.

In 1903 a new die was engraved in hardened steel, following exactly the design of the seal of 1885. This die was first used on Jan. 27, 1904, and continues in current service. Measuring 3 in. (8 cm.) in diameter, it may be distinguished from the seal of 1885 by its greater depth and by minute differences in the rays of the "glory." In the 1885 seal all the rays are solid lines; in the 1904 seal every other ray is a dotted line. Like the earlier seal, it is provided with a counter-die; and it is usually impressed over a paper wafer pasted to the document, although examples are to be found without the wafer.

Legally, the seal has two designations, "the Great Seal" and "the Seal of the United States," both of which appear in acts of Congress and in a decision of the Supreme Court and both of which are in general use.

It has a limited use which is strictly guarded by law. With the expanding functions of the government, the extent of its use has been curtailed from time to time by act of Congress or executive order. For instance, where formerly the seal was affixed to all civil (not military or naval) commissions signed by the president, now persons appointed by the president to serve under Cabinet officers other than the secretary of state are commissioned under the seals of the respective departments. Except for some proclamations and the commissions of some civil officers, the seal is now used only in connection with international affairs.

Apart from the seal, and as the emblem or coat of arms of the nation, the device of the obverse is employed officially in innumerable ways, and sometimes in more or less modified form, for purposes of decoration or identification. It appears on medals, on stationery, on publications, on currency, on flags, in paintings, and as architectural adornments; it forms part of the seal of the Department of State and, with differences, part of the seal of the president; and it is displayed in colour over the entrance of embassies and other foreign service installations.

The symbolism of the obverse of the Great Seal is conventional and well known: the American bald eagle; the motto *"E Pluribus Unum"* ("One from many"); the olive branch of peace and the arrows of war; and symbols of the 13 colonies. That of the reverse is less familiar. The design incorporates the date of the founding of the nation, 1776; an unfinished pyramid, suggesting the firm and durable building of the new nation, not complete, however, and having room for other states; a single eye surrounded by the sun's rays, suggesting the eye of providence surrounded by the light of the universe; and two mottoes, *"Annuit Coeptis"* ("He [*i.e.,* God] has favoured our undertakings") and *"Novus Ordo Seclorum"* ("A new order of the ages"), both adapted from Virgil.

The states also have and use seals; for illustrations of these *see* UNITED STATES (OF AMERICA): *Administration and Social Conditions: State Governments.* (R. S. PA.; X.)

CHINESE AND JAPANESE SEALS

The use of seals for impressing personal names, ranks, and titles of office has existed in China since earliest times, and was introduced into Korea and Japan as soon as they came under the influence of Chinese culture. In China bronze seals have survived from the 5th century B.C., and even when allowance is made for the Chinese habit of attributing later practices to ancient, even legendary, rulers, the mention of seals in literature suggests that their use began earlier. The first strictly datable mention relates to the year 544 B.C., when a messenger carried to Duke Hsiang of the

state of Lu a "sealed document" (*hsi shu*). The Han Chiu I, a work on the organization of government by Wei Hung (1st century A.D.), records some regulations for official seals: those of the feudal princes were to be known as *hsi* and those of other nobles as *yin*, and both were to be of gold; the seals of ministers or generals, also of gold, were to be called *chang* and those of officials with a stipend of 2,000 piculs of grain or more to be known as *chang* and made of silver. There is no sign, however, that this distinction of terminology was ever closely observed, although it became customary to reserve the word *hsi* for imperial seals. In T'ang times *hsi* was replaced by *pao* in this sense. *Hsi* were traditionally large and square, made of jade, ivory, or precious metal and richly carved. The most famous imperial seal (now known only from the histories) was that of Shih Huang Ti, the first emperor of the house of Ch'in and unifier of the empire. It was made of fine jade with a knob in the form of a one-horned dragon, and bore the inscription:—"The Emperor glorious and long-lived, the recipient of Heaven's command." It is said to have been handed down the line of the succeeding emperors of the house of Han. A bronze seal of historical interest was found in 1784 in Fukuoka Prefecture, Kyushu, Jap. It reads "King of the Han Wei-nu country." The Later Han history records the arrival of Japanese tribute-bearers at the Chinese court possessing seals which had been accorded by the Han emperor Kuang-wu (A.D. 25–57), of which this is presumed to be an example.

The use of seals by private persons as a signature and a mark of possession on books, documents, paintings, and calligraphies survives to the present time. As a signature the seal generally bears the surname and personal name followed by the word *yin* (seal impression) set in square or oblong shape, and reading top to bottom and right to left, although this direction varies. Every Chinese who might be called on to sign his name in the past possessed and often carried such a seal. The seal (the word is used for both the die and the imprint) might represent the characters of the name in solid line or reserved against a background of the ink. The latter is always red. Even from the Han period the custom was to use a special form of characters—the *chuan* or seal script—which shows considerable deviations from the normal hand.

But besides signature seals, persons of literary and artistic pretensions had made for themselves ornamental seals, sometimes round, oval, or gourd-shaped, in which they recorded the name of their study or library, their own literary pseudonyms and sobriquets, words of commendation for the paintings, etc., on which the seals were impressed. Artists frequently had scores of such seals, to the inconvenience of art historians and the convenience of fakers. For although a signature seal might pass as a legal identification, subject to verification, seals could be counterfeited as readily as brush-written signatures forged, and some authorities would maintain that the authenticity of a painting can never rest merely on the evidence of seals.

The carving of inscriptions on seals, even in jade, became a refined hobby; and new seals might be commissioned or presented on any auspicious occasion. One famous example, familiar to students of painting, is the seal adopted by the great collector Emperor Ch'ien-lung on his birthday, which was impressed on many of the paintings in the imperial collection (now in the National Palace Museum, Peking).

See also references under "Seals" in the Index. (W. Wn.)

Bibliography.—*British Isles:* W. R. Anson, *Law and Custom of the Constitution,* 5th ed. (1922); T. F. Tout, *Chapters in the Administrative History of Mediaeval England,* 6 vol. (1932); W. de G. Birch, *Catalogue of Seals . . . in the British Museum,* 6 vol. (1887–1900), *History of Scottish Seals* (1905); A. B. Tonnochy, *Catalogue of British Seal-Dies in the British Museum* (1952); *Guide to the Seals in the Public Record Office* (1954); H. S. Kingsford, *Seals,* no. 30 in "Helps for Students of History Series" (1920); A. Wyon, *The Great Seals of England* (1887); G. Pedrick, *Monastic Seals of the 13th Century* (1902), *Borough Seals of the Gothic Period* (1904); Sir H. C. Maxwell Lyte, *Historical Notes on the Great Seal of England* (1926); H. Laing, *Descriptive Catalogue of . . . Ancient Scottish Seals,* 2 vol. (1850, 1886); E. C. R. Armstrong, *Irish Seal-matrices and Seals* (1913).

Continental Europe: L. Perrichet, *La Grande Chancellerie de France des origines à 1328* (1912); F. H. O. Morel, *La Grande Chancellerie royale, 1328–1400* (1900); L. Douet d'Arcq, *Collection de sceaux,* 3 vol. (1863–68); G. Demay, *Inventaire des sceaux de la Flandre* (1873), *Inventaire des sceaux de l'Artois et de la Picardie* (1877), *Inventaire des sceaux de la Normandie* (1881), *Inventaire des sceaux de la Bourgogne* (1912); J. Roman, *Manuel de Sigillographie française* (1912); O. Posse, *Die Siegel der deutschen Kaiser und Könige von 751 bis 1806,* 5 vol. (1909–13), *Die Siegel des Adels der Wettiner Lande bis zum Jahre 1500,* 4 vol. (1903–11); M. Tosi, *Bullaria e Bullatores della Cancellaria pontificia,* in *Gli Archivi Italiani* (1917); F. de Sagarra, *Sigillografia catalana* (1916). (A. B. To.)

Seal of the United States: Continental Congress Papers in the Library of Congress; Archives of the Department of State in the Department and in The National Archives; *Journals of the Continental Congress, 1774–1789* (1904–37); *Statutes at Large of the United States;* G. Hunt, *History of the Seal of the United States* (Department of State, 1909). (R. S. Pa.; X.)

Chinese and Japanese Seals: V. Contag and C. C. Wang, *Maler- und Sammler-Stempel aus der Ming- und Ch'ing-Zeit* (1940); R. H. van Gulik, *Chinese Pictorial Art as Viewed by the Connoisseur* (1958). *See* also essays, monographs, memorials, etc., by writers of the Sung, Yüan, and Ming dynasties reprinted in the K'ang Hsi Encyclopaedia. (W. Wn.)

SEAMANSHIP, the art of handling a ship or boat under any and all conditions of weather, tide, current, or other influence affecting its movement or safety. The term is also applied to the skill attributed to a good seaman. It should not be confused with navigation nor with pilotage (*qq.v.*), but it is a companion art to these; for without good seamanship in the actual maneuvering of the ship, good practical results cannot be achieved by the navigator or pilot.

In elementary form, seamanship is involved in the handling of even the smallest floating craft. In sailing ships, seamanship is largely concerned with rigging (*q.v.*), the making and shortening of sail (*q.v.*), and the correct manipulation of sails and rudder (*q.v.*) for maneuvering the ship. In power-driven ships, ability to maneuver the vessel by means of screw propeller and rudder replaces skill in handling sails. However, in the handling of all vessels it is necessary for the seaman to know the effect on the ship of wind, tide, and current; the behaviour of a vessel in a seaway; the various methods of anchoring in a harbour or mooring to a dock; the safety precautions required aboard ship; and the nautical rules of the road (*see* Rules of the Road at Sea).

Seamanship also involves a multitude of other services incidental to maritime work such as the proper loading of cargoes; shifting of weights; making the ship snug for heavy weather; handling the ship's boats; care of passengers; attention to the ship's fittings; and knotting and care of ropes.

Good seamanship is acquired by practice and natural aptitude. It cannot be taught by precept or the study of books alone. (M. H. I.)

SEAMI MOTOKIYO: *see* Zeami (Seami) Motokiyo.

SEA PEOPLES, a modern collective name for groups of aggressive seafarers who appeared in the eastern Mediterranean toward the end of the Bronze Age, especially in the 13th century B.C. The Sea Peoples are held responsible for the destruction of old powers such as the Hittite Empire. Their activities were but one aspect of major migrations of peoples which violently interrupted the old order elsewhere; *e.g.,* in Greece where the Mycenaean civilization was destroyed. The precise extent and the origin of the upheavals were still uncertain in the 1960s. The difficulty lies in the abrupt halt of historical sources (near eastern ancient archives) as a result of the invasions and destructions. Principal but one-sided evidence for the Sea Peoples comes from Egyptian texts and illustrations recording the successful defense of Egypt against the new enemies. Additional information can be gathered from Hittite sources, from archaeological data, and by comparison with later historical material (*see* Hittites). Recently, the archives of the north Syrian city of Ugarit (modern Ras Shamra) and new finds of tablets from the Hittite capital (Boğasköy) have added dramatic details concerning coastal raids and the final naval battles around Cyprus shortly after 1200 B.C.

The Egyptians waged two wars in which the Sea Peoples were involved. The Egyptian king Merneptah (*c.* 1236–1223 B.C.) in the fifth year of his rule defeated a Libyan attack from the west. The Libyans were supported by peoples called "northerners from all lands" whose names can be transcribed as Ekwesh, Teresh, Luka, Sherden, and Shekelesh. These five allies must have crossed the Mediterranean by ship to join the Libyans, and thus were

naval powers of some, if modest, importance. The second encounter of Egyptians and Sea Peoples took place under the rule of Ramses III (*c.* 1195–1164 B.C.) who faced major invasions in his fifth and his eighth year after the Hittite Empire and the Syrian coast had been devastated by the onslaught of "foreign countries" who started the troubles "in their islands." The attacks came by sea and by land, and the listing on this occasion is: Peleset, Tjeker, Shekelesh, Denyen, and Weshesh (*see* EGYPT: *History: New Empire: 20th Dynasty*).

Tentative identifications of the Sea Peoples are as follows:

Ekwesh: a group of Bronze Age Greeks (Achaeans [*q.v.*]; Ahhiyawa in Hittite texts).

Teresh: Tyrrhenians (Tursenoi) known to later Greeks as sailors and pirates from Asia Minor, ancestors of the Italian Etruscans (*see* ETRUSCANS: *Origins*).

Luka: a coastal people of western Asia Minor, also known from Hittite sources. They are known to have raided the coast of Cyprus (Alashiya) in the 14th century B.C. The name survives in classical Lycia on the southwest coast of Anatolia.

Sherden: Sardinians to judge by their name and appearance. The Sherden acted as mercenaries of the Egyptians in the battle of Kadesh (*c.* 1299 B.C.) where Ramses II fought the Hittite king Muwatallis. In this land battle the Hittites had the Luka among their allies. In Egyptian illustrations the Sherden wear helmets with horns and carry swords and round shields, equipment similar to that of bronze warrior figurines found on Sardinia.

Shekelesh: The name suggests that this tribe is identical with the Sicilian tribe called Siculi (*q.v.*).

Peleset: generally admitted to refer to the Philistines (*q.v.*) who settled in Palestine after their defeat in Egypt. In the reliefs of Ramses III's funeral temple at Medinet Habu, the Philistine (Purasati, Pulesati) sailors are shown wearing stiff crowns (often called feather crowns, but perhaps a special haircut combined with diadem and chinstrap). The same headgear is apparent in terracotta images on sarcophagi from Beth-shan (Beisan) and on seal impressions from Enkomi on Cyprus; in both cases, contact with Philistines or Sea Peoples is clear. The Philistines after their arrival in Palestine made a decorated kind of pottery which is of Aegean derivation; their connections with Crete are archaeologically probable.

Further identifications (Tjeker-Teukroi; Denyen-Danuna or Danaoi) are much more uncertain.

The problems connected with the Sea Peoples concern their provenance and the chronology of their migrations. It is not known if any of them came from the western Mediterranean before the attacks on Egypt (*e.g.*, the Sardinians and Sicilians, who certainly lived in the west later). The extent to which the Aegean islands were responsible for the unrest is uncertain. The Trojan War can be seen as a prelude to the Sea Peoples' raids in the eastern Mediterranean, and both Ekwesh and Philistines have Aegean connections. The Luka and Teresh can be interpreted as Anatolian peoples; for the others, there hardly seems room along the known shores of the eastern Mediterranean. The coming of iron into the ancient world may be connected with the Sea Peoples, in which case part of their strength may have derived from eastern Asia Minor, perhaps the Black Sea coast.

See J. B. Pritchard (ed.), *Ancient Near Eastern Texts Relating to the Old Testament*, pp. 262–263 (1950); A. C. Vaughan, *Those Mysterious Etruscans* (1964). (M. J. ME.)

SEAPLANE, an airplane capable of taking off from and landing on water; depending upon its design, it is also sometimes referred to as a flying boat, float plane or hydroplane. The first practical seaplanes were built and flown in the United States by Glenn H. Curtiss in 1911 and 1912. For the next 30 years they played an important role in the development of aviation.

The Curtiss inventions led swiftly to the famous British F-boats of World War I, which originated such naval missions as over-ocean reconnaissance and patrol, aerial antisubmarine warfare and mine laying, and air-sea rescue. After the war, commercial versions of the same seaplanes set the range and endurance records of the time, and inaugurated the first international airlines from the U.S. to Cuba and Canada. In 1919, the U.S. navy's water-based NC-4 conquered the North Atlantic. In 1924 four single-engine landplanes of the U.S. army were converted to water operation and made the first flight around the world.

In the 1930s, the largest and fastest aircraft in the world were seaplanes. Their utility and versatility were dramatized by a Russian flight from Moscow to New York via Siberia, by Gen. Italo Balbo's mass flights from Rome to Rio de Janeiro and Chicago, Ill., by Col. Charles A. Lindbergh's flights over the Pacific, and by Adm. Richard E. Byrd's extensive mapping of Antarctica—all with marine aircraft. Seaplanes were used by the German airline, Deutsche Lufthansa, for transcontinental and ship-to-shore airmail service, and by the U.S. Pan-American Airways to initiate regular transpacific and transatlantic service. The "China clipper" became one of the most famous types flown across the Pacific ocean. Racing seaplanes held the absolute speed record from 1931 to 1939, and many of their advanced features were later incorporated in the famous British fighter planes of the early 1940s.

After the outbreak of World War II, the military and commercial significance of seaplanes gradually diminished, partly because of the construction of land bases and aircraft carriers. Following World War II, the development of water-based aircraft continued on a small scale in Great Britain and the U.S., but major advances in aeronautics of this era were made with land-based equipment.

Characteristics and Types.—A seaplane must have sufficient buoyancy to float on water and must also have some means for supporting its weight while moving along the water surface at speeds up to flying speeds. It must be able to take off and land with a margin of stability and control on the part of the pilot; its structure must be strong enough to withstand the shock of landing; and its water resistance must be low enough to permit reasonably short take-off runs.

Ways of meeting these requirements were provided by Curtiss in the form of float seaplanes which are essentially landplanes with buoyant floats or pontoons substituted for the landing wheels, and flying boats in which the main floats and fuselage are combined in a single boatlike body. Single float seaplanes and single hull flying boats require side floats or wing-tip floats to keep them upright. Twin float seaplanes do not require the auxiliary floats, nor do twin hull flying boats and single hull boats with stub wings or sponsons located at the water line.

The addition of a retractable wheel landing gear to a float seaplane or flying boat, also first accomplished by Curtiss, creates the amphibian aircraft capable of operating from land runways or water. A post-World War II development was the pantobase or all base airplane incorporating devices for operating from water or from a variety of unprepared surfaces such as snow, ice, mud and sod.

Performance.—When buoyant floats are added to an airplane in place of landing wheels, the plane's weight and aerodynamic drag are generally increased and its flight performance is impaired. These penalties also exist to a degree in smaller propeller-driven flying boats since the hull must be larger than a landplane fuselage to provide water clearance for propellers, wings and tail surfaces; but in larger sizes the apparent weight penalty may disappear because of a disproportionate increase in the weight and complexity of the landplane gear. The amphibian and pantobase airplanes naturally have additional weight penalties arising from their many different functions.

Jet propulsion promises the seaplane exemption from the size required for propeller clearance. It offers for the first time true size and performance parities with landplanes having similar wings and power plants. Nevertheless, flight performance penalties are crucial items in seaplane technology; traditionally they have been the subject of much scientific investigation on the part of aircraft manufacturers and aeronautical research laboratories.

Seaplane Research and Development.—With continual advances in power and flight speeds, aeronautics has been favoured by a concentration of scientific research effort never before known in the history of transport. In this effort, all known mathematical tools have been brought to bear, and the use of scale models for systematic experiments on the problems of flight has become a

sophisticated art. The seaplane, along with other aircraft, has benefited from this scientific climate, and its particular drawbacks have been the subject of considerable aerodynamic and hydrodynamic research, notably by such government laboratories as those of the National Advisory Committee for Aeronautics (N.A.C.A.) in the U.S., the Royal Aircraft Establishment (R.A.E.) in Great Britain, the Deutsche Versuchsanstalt für Luftfahrt (D.V.L.) in Germany, and the Guidonia aeronautical centre in Italy. These laboratories have provided staffs, wind tunnels and seaplane towing tank facilities for theoretical treatments of the fundamental phenomena involved, experimental verifications of the theories and evaluations of their applications to practical aircraft.

Model investigations of seaplane components and configurations were originally carried out in ship model testing establishments. In the 1920s and 1930s specialized towing tanks were built by D.V.L., N.A.C.A., R.A.E. and Guidonia with higher testing speeds for simulating seaplane operating conditions up to the minimum flight speeds of aircraft. These tanks were later augmented in the western world by high-speed facilities at the experimental towing tank of the Stevens Institute of Technology in the U.S., and by several private facilities operated by seaplane manufacturers in the U.S. and Great Britain.

As a result of hydrodynamic research of the post-World War II era, particularly in the U.S., where the department of the navy offered encouragement, important scientific progress was made in theories of planing surfaces, water impact loads and submerged lifting elements (hydrofoils). Better design methods were devised for achieving optimum proportions and shapes of flying boat hulls, improved take-off and landing qualities, and extended rough water capabilities in the face of increasingly severe military performance requirements. Because of limited actual procurement and operation of seaplanes, however, methods for water handling and beaching seaplanes at their terminals and at austere forward bases remained in a relatively primitive stage of development.

Hydro-skis.—Jet propulsion brought with it an improved design concept for water-based aircraft wherein the buoyant float system or oversized flying boat hull is dispensed with and fully retractable planing elements are provided for taking off and landing in much the same manner as with retractable wheels. With this concept, the seaplane floats at rest on the optimum fuselage for flight. The planing elements or hydro-skis are extended for water operation and retracted flush with the fuselage while in the air. Hydro-skis have proved to be compatible with advanced transonic and supersonic airframe and power plant arrangements. Because of their relatively small size and high loadings, they markedly reduce the loads and motions induced by wave impacts.

Land-Water Operations.—As a further extension of the hydro-ski principle, small landplanes have been operated off beaches and ramps on hydro-skis with no attempt to provide flotation at all. In this method of operation, pioneered by Canadian bush pilots with snow skis, the aircraft accelerates on land to the relatively low minimum planing speed of the skis; it then enters the water to complete its take-off run. On landing, the procedure is reversed and the airplane comes to rest conveniently out of the water. This novel means of operation requires runways or cleared areas only a few hundred feet long instead of the thousands of feet needed for conventional landplane operation.

Vertical Take-Off and Landing.—With aircraft designed to rise from and return to the earth vertically, the usual hydrodynamic requirements of water basing virtually disappear. Helicopters have been successfully converted to water operation simply by installing inflated rubber float systems for buoyancy and stability while afloat or by making the fuselage watertight and adding side floats for stability. A further development has been a helicopter fuselage shaped somewhat like a rudimentary flying boat hull for improved water taxiing characteristics.

Notable Seaplanes.—The largest seaplane built was the U.S. "Hughes Hercules" cargo transport flying boat, completed in the latter stages of World War II. This aircraft had a wing span of 310 ft., a gross weight of 400,000 lb. and was powered by eight 3,000-h.p. engines. Because of temporary wartime restrictions on aluminum, it was constructed entirely of molded plywood. The war ended before the plane actually went into service, but it demonstrated that seaplanes are not bound by the same size limitations as landplanes.

BY COURTESY OF U.S. NAVY

U.S. NAVY MARTIN P5M-2 "MARLIN"

Another notable seaplane was the "Mars" flying boat built by the Glenn L. Martin Co. In one short flight in 1949 a "Mars" carried more than 300 persons. A four-engine military transport, it had a wing span of 200 ft. and made regular flights from California to Hawaii.

The British continued the production of large propeller-driven flying boats in the post-World War II era. Their development culminated in the Saunders-Roe SR/45 "Princess" transport for luxury passenger service between London and New York. This aircraft had a pressurized hull for high-altitude operation, a gross weight of 300,000 lb. and was powered by ten turboprop engines of 3,000 h.p. each. It had a pay load capacity approaching 50,000 lb., a range of more than 4,000 nautical miles and a cruising speed of 300 knots at an average altitude of 30,000 ft. It was test flown in 1952, but was not placed in service because of withdrawal of government support. U.S. procurement in the same time period was entirely of naval seaplanes; there was no parallel commercial activity. The U.S. navy continued production of replacements for the World War II Martin PBM "Mariners" until 1949 and in the 1950s built more than 200 Martin P5M "Marlins", all-weather flying boats for service with the fleet. The "Marlins" were 72,000-lb. aircraft, driven by two 3,400-h.p. reciprocating engines and carried a large payload of electronic submarine detection equipment.

The first turboprop seaplanes in the United States were the Convair P5Y and R3Y "Tradewind" flying boats intended for naval patrol and logistic transport missions. The latter had a maximum gross weight of more than 170,000 lb. and a cruising speed exceeding 300 knots at an altitude of 25,000 ft. One version incorporated a bow-loading arrangement for movement of troops and equipment on and off beaches; another, multiple in-flight refueling equipment for carrier-based fighters.

The first turbojet seaplane was the British Saunders-Roe Sr/1A experimental water-based fighter, built and flown in 1947. It employed a conventional straight-wing flying boat arrangement with a single jet intake in the nose of the hull. The U.S. navy unveiled the XF2Y-1 "Sea Dart" developed by Convair in the early 1950s. The "Sea Dart" had a delta-wing transonic configuration with air intakes located above and behind the pilot's canopy. It represented the first literal application of the retractable hydro-ski principle. It was operated successfully with a twin ski system mounted on shock absorbing struts, and with a smaller single ski system mounted on fixed struts.

Conclusion.—Airplanes having thrust-weight ratios of less than one and capable of utilizing natural waterways as their runways have in the past been an effective answer to man's desire to move quickly at will about the earth. Many of their intrinsic advantages—superior economy, safety, flexibility of operation—un-

fortunately have been lost sight of by commercial and military interests. In the second half of the 20th century, they continue to offer to airline planners a practical means of alleviating growing problems of airport cost, congestion, noise and safety. Proponents of the seaplane contend that, for military use, they can provide invulnerable bases, dispersal of weapons systems, relief from undependable land-base arrangements with grasping or unstable governments and the safest test beds for gaining experience with nuclear propulsion away from inhabited areas.

BIBLIOGRAPHY.—H. C. Richardson, *Flying Boats* (1942); R. S. Fogg *et al.*, *Seaplane Flying and Operations* (1949); *Jane's All the World's Aircraft* (annual); D. B. Mac Diarmid, "Seaplanes in the U.S. Coast Guard," *S.A.E. Jour.*, vol. 59, no. 8, pp. 56–57 (Aug. 1951); H. Knowles, "Future of Flying Boat," *Royal Aero. Soc. Jour.*, vol. 56, no. 497, pp. 322–354 (May 1952); F. Bowhill, "Flying Boat and Its Place in Aviation," *Inst. Transport Jour.*, vol. 24, no. 9, pp. 321–330 (March 1952); M. J. Brennan, "Design and Development Aspects of Princess," *Aircraft Eng.*, vol. 24, no. 285, pp. 326–334 (Nov. 1952); E. G. Stout, "High-Speed Water-Based Aircraft," *S.A.E. Jour.*, vol. 61, no. 3, pp. 17–19 (March 1953); J. M. Bruce, "Felixstowe Flying Boats," *Flight*, vol. 68, no. 2445, 2447, 2448; vol. 69, no. 2452; G. Loening, "Economics of Large Aircraft," *S.A.E. Jour.*, vol. 64, no. 4, pp. 48–55 (April 1956); A. D. Struble, Jr., "Navy Nuclear Seaplane," *S.A.E. Jour.*, vol. 65, no. 10, pp. 64–66 (Sept. 1957); W. Green, *Floatplanes* (1962). (J. B. PA.)

SEA POWER. In its classical sense this term refers to all the means by which a nation extends its military power onto the seas. It is important both in peacetime and in wartime, for by its mere existence sea power affects a nation's influence in peacetime.

Sea power is sometimes confused with the possession of a large merchant marine, but a nation may have a considerable merchant marine without possessing notable sea power, as do Norway and Greece and as did the United States before 1860. Conversely, a nation may possess great military power upon the seas without boasting much of a merchant marine, as did the United States for the two decades preceding World War I. On the other hand, Great Britain was for centuries the world's leading naval power in the military sense and also possessed the world's largest merchant marine.

The main purpose of sea power has always been to protect friendly shipping from enemy attack and to destroy or hinder the enemy's shipping, whether that shipping was used for commercial or military purposes. In the time-honoured phrase, it has aimed at "using the sea for oneself and denying its use to the enemy." When one belligerent or the other has virtual control of surface shipping in portions of the seas, the situation has been traditionally known as "command of the sea." Attaining such command in the areas of chief importance has usually been the first aim of the naval antagonists. In a great naval war each side may command different areas of the sea, and each may try to extend its own area of command and to dispute the command in those areas claimed by the enemy. Large areas may be so generally in dispute as not to reflect a condition of command at all.

The strategic functions of sea power during wartime fall into five major categories:

1. Sea power protects the movement over water of one's own military forces and their supplies to coastal areas where they may be landed and used against enemy forces. This may mean landing an army on a hostile shore or on a friendly shore for operations in nearby territories.

2. It also protects from enemy attack the friendly shipping that carries the commodities of ordinary seaborne trade, including industrial goods and raw materials that support the nation's war economy.

3. It prevents the enemy from using the sea to transport his own military forces. It may thus constitute the primary means of defending one's homeland or overseas territories against enemy invasion.

4. Sea power may exert military and economic pressure on the enemy by preventing him from importing by sea commodities necessary for his prosecution of the war. It also prevents him from exporting products and thus obtaining funds to pay for military commodities received from neutrals to whom he does have access. It may exert pressure upon neutrals to prevent them from trading with the enemy. This use of sea power is known as blockade (*q.v.*) and has usually been exercised according to specific procedures prescribed by international law.

5. Naval forces have also been used to bombard land objectives from the sea. In the first half of the 20th century this function of sea power grew enormously in importance. The development of the aircraft carrier added a new dimension to this bombardment capability, as did the missile-firing nuclear submarine. In the early 1960s the nuclear submarine was on the way to becoming the single most important instrument of sea power; it was scarcely distinguishable in function from strategic air power in a general nuclear war.

The Instruments of Sea Power.—Sea power has never meant merely warships. It has always meant the sum total of the weapons, institutions, installations, and geographical circumstances that enable a nation to exercise military power at sea. The airplane used in the control of seaborne transportation functions as an instrument of sea power even when it operates from a land base, as was true of the British Coastal Command in World War II. It is also true that aircraft operating from carriers represent the extension of sea power even when they are attacking targets deep inland. Except for the great increase in bombardment of shore or inland targets from the sea, the functions of sea power were the same in World War II as they were in the 16th century, when warships specifically designed for fighting (as distinguished from armed merchantmen) first appeared.

Early Warships.—The coming of the first modern naval vessels designed and operated as warships coincided in time with the abandonment of oar propulsion and with the mounting of "great artillery" on ships. The broadside arrangement of guns was not compatible with the use of oars and the oars themselves were made unnecessary by developments in the art of sailing. The standard fighting ship in the English Navy became the galleon, a ship with two or three decks carrying its main batteries in the broadside and lighter pieces fore and aft. Armed merchant ships of 300 tons burden and over, called "great ships," supplemented the galleons. Such were the ships that won the great victory in 1588 over the Spanish Armada (*q.v.*), which contained large, slow ships with relatively few short-range guns. The Spanish ships were supposed to close with the enemy so that the soldiers with which they were crowded could board the enemy vessels. Thus the English victory was the victory of a new and revolutionary conception of naval battle: the English ships refused to let the Spanish ships get close enough for boarding and pounded them with guns of superior range and power.

It soon became evident that the ship that was powerful enough to be the mainstay of the fighting fleet and to meet on more or less equal terms the best ships the enemy could bring into battle was too large and too expensive—and also too heavy and too slow—to serve those functions of sea power that required numerous armed ships of good speed. For example, the pursuit and capture of enemy merchant vessels required the kind of ship that early became known as a cruiser. Thus during the 17th and 18th centuries the warship evolved into two major types. The main fighting ship of the fleet was the "ship of the line" or "line-of-battle ship," so called because it was meant to "fight in the line." Such ships were two-deckers or three-deckers (rarely also four-deckers) with heavy broadside armament as well as heavy timbers in their walls to keep out enemy shot. The other main type comprised the lighter and faster cruisers, of which the largest was the frigate, a vessel with a single deck of guns of lesser calibre than those of the line-of-battle ship. Like the frigate (but smaller) was the corvette, and below it came the sloop of war, usually employed as a dispatch vessel. The designations "frigate" and "corvette," after having been out of use for 100 years, were restored during and after World War II to denote warships that were intended mostly for escort work and were intermediate in size between cruisers and destroyers.

Modern combat ships fall into three major categories: (1) ships that fight chiefly by means of the aircraft launched from their decks; (2) those that fight primarily with guns or with rocket-propelled missiles; and (3) those that fight mainly with under-

water weapons such as mines, torpedoes, and depth charges. Among the latter is a further sharp distinction between surface and submarine craft. The fact that submarines also launch intermediate-range missiles indicates the extent of overlap among the three categories.

Aircraft Carrier.—The aircraft carrier (*q.v.*) appeared at the very close of World War I, the first being the British "Argus." Early in World War II carriers of the Royal Navy played an important part, as in the Mediterranean Battle of Cape Matapan (March 1941) and in the dramatic and successful chase of the German battleship "Bismarck" (May 1941) in the North Atlantic. In the Pacific campaigns the carrier won for itself the dominant place in the battle fleet, taking the position of primacy that the battleship had enjoyed for three centuries. The carrier was a floating air base that combined the mobility and sea-keeping power of the large warship with the advantages deriving from the use of aircraft. In the later stages of World War II, the U.S. Navy was able to concentrate such numbers of carriers in its task forces as to enable them to overwhelm with their aircraft the local air defenses of the archipelagoes against which they moved. Carriers furnished the major ground-support aviation to the United Nations during the early phases of the Korean War (1950-53) before air bases were established ashore, and they continued throughout the war to provide a valuable supplement to land-based aviation. Carrier-based planes also played a major role in Vietnam in the mid-1960s.

The addition of atomic weapons to its magazine enabled the carrier to bid for an important part not only in limited war on the Korean model but also in strategic bombing in any general war, though partisans of land-based (or submarine) bombardment power maintained that these same weapons had made the large carrier much too vulnerable to be worth its huge cost. The navy reply stressed the argument that the mobility of the carrier gave it a kind of protection not available to air or missile bases. With nuclear propulsion the mobility of the carrier was greatly increased.

The large carriers in the U.S. Navy are called attack aircraft carriers; others, designated support aircraft carriers, are considerably smaller, originally being built on converted cruiser hulls. During World War II a large number of escort carriers, nicknamed "jeep" carriers, were built on converted merchant ship hulls. They were invaluable for providing air support to the antisubmarine patrols in mid-ocean.

Battleship.—The battleship (*q.v.*), a direct descendant of the old line-of-battle ship, represented the gunpower of a fleet in its most impressive form. It was designed to stand up, along with its mates, to the most powerful gun-fighting ships in an enemy fleet. The largest of the Allied types in World War II reached a size of approximately 58,000 tons displacement, much of it devoted to armour; it carried 14-in. or 16-in. guns in U.S. and British types and up to huge 18.1-in. guns in two Japanese ships; it also carried an array of as many as 150 lesser guns for antiaircraft purposes; and in the case of the "Iowa" class it boasted a speed equal to or surpassing that of most cruisers. This prodigious piece of fighting machinery became obsolescent as its guns were far outranged by the bomb- and torpedo-carrying aircraft launched from carriers. Any hope that it might still serve a useful function disappeared in face of the development of missiles with nuclear warheads that could be mounted on a destroyer and could easily outrange and totally destroy any type of gun-firing ship. In the early 1960s neither the Royal Navy nor the U.S. Navy retained any battleships in commission, though the U.S. Navy had several in the reserve fleet.

Cruiser.—The cruiser (*q.v.*) is a gun-firing or rocket-firing ship, smaller than the battleship and designed to cost less to build and operate so that larger numbers can be made available for patrol purposes. In size and armament, cruisers have ranged from types that were almost indistinguishable from battleships to those very close to the size of a destroyer. The battle cruiser of World War I had the size and gun calibre of a battleship but sacrificed weight of armour for greater speed. The U.S. Navy used the term "large cruiser" for three vessels of the "Alaska" class built in World War II, displacing about 27,000 tons each and carrying 12-inch guns. Similar vessels in foreign navies, like the German "Gneisenau" and "Scharnhorst" and the French "Dunkerque" and "Strasbourg," were called battleships. Definitely in the cruiser category, however, are fighting ships that range from about 5,000 tons standard displacement to 17,000 tons (in the U.S. "Salem" class).

The well-known distinction between light and heavy cruisers does not refer to the size of the ship but to the size of its guns. The Washington Naval Limitation Treaty of 1922 specified that vessels carrying guns of from 6.1 in. to 8 in. (the limits then set for cruisers) should be termed heavy cruisers. Ships with smaller guns should be called light cruisers.

Destroyer.—In the category of vessels whose weapons are mainly of the underwater type are the destroyer and the submarine (*qq.v.*). The modern destroyer is a vessel of less than 3,500 tons displacement, carrying in the U.S. Navy guns of the same size (dual-purpose 5-in.) as some of the smaller light cruisers. However, it carries fewer such guns, usually five and only rarely as many as eight, and sacrifices armour entirely. Some of its guns have been replaced with guided missiles.

The destroyer takes its name from the fact that it was originally designed to destroy with its guns the torpedo boats that menaced the battle fleet (hence, "torpedo-boat destroyer"). Later the destroyer acquired both the torpedoes and the function of the boats it had been designed to destroy. In World War I the destroyer with its depth charges became the chief means of fighting submarines, and thus an essential escort vessel. For the latter purpose it was supplemented in World War II by the destroyer escort, a slower, smaller, and cheaper vessel, much more easily constructed and manned, built in large numbers to cope with the German U-boats.

Submarine.—The submarine had its prototype in a one-man vessel designed and built during the American Revolution by David Bushnell, and actually tried, though without success, against British warships. The submarine became an important naval vessel when on the eve of World War I it adopted the diesel engine, the periscope, and the gyrocompass. Its tactical advantage lay in its ability to gain concealment and protection by slipping under the waves. It was the only type of warship that could operate independently for extended periods in seas otherwise dominated by the enemy. Its torpedoes enabled it to attack vessels of all sizes, including battleships and aircraft carriers, but its greatest triumphs were achieved against merchant ships. The tremendous tonnages of Allied shipping sunk by the German U-boats in 1917 and again in 1942–43, and the comparable accomplishments of U.S. submarines in the Pacific during the whole of World War II, showed the potentialities of a submarine offensive against an opponent lacking the specific means of defense.

The versatility of the submarine was demonstrated in World War I when submarines not only fired torpedoes and laid mines but came to the surface on occasion and used their deck guns against unarmed merchantmen. The experimental British M-class of three vessels actually mounted 12-in. guns. The large French submarine "Surcouf" used in the early part of World War II carried on its deck two 8-in. guns and a hangar for a seaplane. Submarines also proved useful for patrolling and for special assignments such as landing a few men on a coast to carry out a secret mission. This versatility was vastly increased with the development in the United States of the nuclear-powered submarine capable of discharging while submerged ballistic missiles with nuclear warheads and with a range of over 1,500 mi.

The overwhelming defeat ultimately suffered by the German U-boats in both world wars was due to the characteristic weaknesses of undersea craft of the time, especially their low submerged speed. But developments after World War II promised to wipe out some of these disadvantages. The introduction of nuclear propulsion gave the submarine almost unlimited range and enabled it to travel at high speeds while submerged. Because it does not waste power in producing wave action, a submarine can have a greater speed submerged than afloat for any given propulsion thrust. The submarine also increased the depth to which it could dive, modern maximum depths being secret but certainly more than

800 ft. In addition, the use of missiles with nuclear warheads gave it a strategic bombardment capability.

Other Types.—Intermediate between combatant and noncombatant types are the vessels that lay and sweep up mines. Ships of almost all sizes and types have served as minelayers, including submarines and cruisers, and their strategic importance in both world wars was enormous. A small and slow but indispensable and especially valiant ship has been the minesweeper. The development of new techniques for dealing with ever more ingenious types of mines has been one of the great sagas of modern naval warfare (*see* MINE, NAVAL).

Ships that are not used directly in combat but enable combat vessels to function are tenders of various kinds, repair ships, store ships, oilers, fleet tugs, transports, cargo ships, hospital ships, and various other vessels that go to make up what is known as the "fleet train." In the latter stages of World War II in the Pacific, the technique of using a fleet train was perfected to the point of enabling battle squadrons or "task forces" to achieve a range and sea-keeping capability unknown at least since the days of sail.

Bases.—Comparable to the fleet train but immobile and more elaborate in its facilities is the naval base or naval air station. The importance of bases to sea power has been reflected in centuries of struggle to acquire and retain them. Britain's ability after the Dutch Wars of the 17th century to contain the whole continent of Europe with sea power stemmed not alone from its continuing though sometimes tenuous superiority in ships and seamanship but also from the possession of a particularly advantageous system of bases, most of which had been acquired through the exercise of sea power. The first and greatest British naval base was the United Kingdom itself, with its many fine harbours, splendidly situated for the control of sea-lanes from northern and western Europe to other countries. Bases at Gibraltar, Malta, and Alexandria in the Mediterranean enabled the British to control the southern exits from Europe.

The entire Pacific campaign of World War II was mostly a contest for bases that would enable American naval and air power to exercise their dominance closer and closer to the heart of Japan. The great fleet train could supplement the bases but could not substitute for them. It was the acquisition of the Mariana Islands, Iwo Jima, and Okinawa that finally enabled Allied air and naval power to hurl the blows that brought Japan to her knees. On the other hand, mobile fighting power could acquire through capture the bases necessary to its function, but no number of land bases could substitute for a gross shortage of floating fighting power.

The Role of Sea Power.—The usually great and sometimes decisive importance of sea power in war has been generally appreciated in modern times by maritime nations. But the incisive summing up of the role of sea power was the work of Alfred Thayer Mahan, a captain (later retired as a rear admiral) in the U.S. Navy. In 1890, toward the end of his active service career, he published *The Influence of Sea Power upon History, 1660–1783,* which described and analyzed the part of sea power in the great struggles of the period indicated in the title. This book gained extraordinary attention and won for its author not only enduring international fame but also the assurance of a large audience for each of the many succeeding volumes on naval history and strategy that he published. The German kaiser was one of the several national leaders who were captivated by Mahan's books, which stirred up at least equal interest also in Great Britain. Thus Mahan's labours provided ideological impetus to the great Anglo-German naval race that preceded World War I.

One who reads Mahan's sober and scholarly volumes today may find it difficult to understand the excitement they caused and the undoubtedly great influence they had on the naval policies of the major powers of his time, including the United States and Japan. Even after his death, Mahan's views continued for a considerable time to enjoy respect if not veneration. His emphasis on the role of sea power in war, which could not fail to result in some exaggeration, was intensified by others who found his message attractive. Where Mahan himself had used the cautious word "influence," others spoke of sea power as if it inevitably determined the course of every major war. World War I was later interpreted by many writers as having been essentially decided by the naval blockade of Germany, some going so far as to assert that the gigantic struggles ashore were just so much useless bloodletting.

Actually, the Allied blockade would not have been so important militarily if the battles ashore had not used up resources which the blockade prevented the Germans, but not the Allies, from replenishing. Also, had the initial German war plan (the Schlieffen Plan) succeeded, as it almost did, the war might have been decided ashore very quickly—as the Franco-German War had for all practical purposes been decided in a few weeks in 1870 despite clear French naval superiority. It was the stalemating of the battle lines in Europe following the first Battle of the Marne that permitted Allied sea power to exercise its throttling effect. In any case, the Allied naval contribution to victory was actually much greater than is symbolized by the word "blockade," since British and later American participation in the war hinged upon the successful use of sea power to transfer men and commodities overseas. Thus, one of the truly decisive campaigns of the war was the battle in the Atlantic against the German U-boat—a battle that was to take place again during World War II.

Following World War II, on the other hand, popular opinion—and also much professional opinion—regarding the experience of that war was distorted in the opposite direction. Germany's lesser vulnerability to the effects of blockade as compared with her situation in the previous war, and the new and dramatic role played by tactical and strategic air forces caused many observers to withhold due credit from Allied sea power despite its enormous and vital contribution to victory. World War II was in fact the conflict in which sea power reached the apogee of its influence on history. The first great air war and the war that saw the most titanic battles of all time on land was also by far the greatest of naval wars. It could hardly have been otherwise in a war that was truly global, where the pooling of resources of the great allies depended upon their ability to travel the highways of the seas. The sometimes threatened but always maintained maritime communications, especially between the United Kingdom and the United States, and also between them and the Soviet Union, provided a sharp and telling contrast with the isolation from each other suffered by three major Axis powers. Even Germany and Italy, though contiguous, were prevented by Allied sea power from using those sea communications that in peacetime had carried the bulk of their commerce with each other.

Similarly, the critical role played by United Nations and especially U.S. sea power in the Korean War lacked dramatization because of the absence of naval opposition—except from mines laid defensively in large numbers around the ports of North Korea, especially Wŏnsan. But overwhelming UN naval superiority was the key factor that made it possible to resist Communist aggression in Korea. Aircraft carriers also provided important aviation support to UN ground forces, especially in the critical early phases of the war.

Naval forces also made possible several amphibious operations, including the spectacular "envelopment from the sea" at Inch'on, where U.S. forces were landed on the flank of and behind Communist lines, causing a complete collapse of the enemy position and a precipitate retreat (*see* KOREAN WAR).

Nuclear Weapons and the Future of Sea Power.—The development of nuclear weapons necessitated an entire reevaluation of the role of sea power. Even if it is assumed that a nation under bombardment by numerous thermonuclear weapons could continue to maintain a navy, which might itself be a target for such weapons, a general war of that kind would almost certainly move too fast to permit sea power to exercise its traditional function. But against this proposition two important facts have to be assessed. First, with the new kinds of forces being made available to it, especially in the form of missile-launching submarines and other missile-launching vessels, the navy of the future may play an important and possibly a dominant role even in general war, though strictly of a bombardment nature. In other words, the function that was once decidedly subsidiary to other functions involving control of sea communications could become the only important one in general war. Second, if the war of the future is a limited

war on the Korean model, it will call upon naval power to play much the same part it has played in many wars of the past.

See also AIR POWER; FLEET, NAVAL.

BIBLIOGRAPHY.—A. T. Mahan, *The Influence of Sea Power upon History, 1660–1783* (1890), *Influence of Sea Power upon the French Revolution and Empire, 1793–1812*, 2 vol. (1892), *The Life of Nelson* (1897), *Sea Power in Its Relations to the War of 1812* (1905), *Naval Strategy* (1911); J. S. Corbett, *Some Principles of Maritime Strategy* (1911); Bernard Brodie, *Guide to Naval Strategy*, 3rd ed. (1944), *Sea Power in the Machine Age*, 2nd ed. (1943); A. F. Wescott (ed.), *American Sea Power Since 1775* (1947); Harold and Margaret Sprout, *The Rise of American Naval Power, 1776–1918* (1939), *Toward a New Order of Sea Power* (1940); Arthur Bryant, *Years of Endurance, 1793–1802* (1944), and *Years of Victory, 1802–1812* (1945); H. W. Richmond, *The Navy as an Instrument of Policy, 1558–1727* (1953), *Statesmen and Sea Power* (1946); S. W. Roskill, *The Strategy of Sea Power* (1962). For descriptions of warships in the world's navies, the standard reference book is the British annual, *Jane's Fighting Ships*. (B. BR.)

SEA SERPENT. Enormous serpents, both terrestrial and marine, are subjects around which have arisen such an array of legends and stories that it is almost impossible to disentangle fiction from fact. So far as terrestrial snakes are concerned it seems fairly safe to assume that there are none in the few remaining unexplored parts of the world that greatly exceed in size those that are already known. In the depths of the sea, possibly, may still be gigantic creatures of which there is no knowledge. Such a creature, however little related to the true serpents, would still qualify very well for the popular idea "sea serpent."

Up to the present no animal has been captured that has not, on examination by competent persons, proved to belong to a previously well-known group. A large number of the well-authenticated stories of monstrous marine creatures seem to be explicable as incorrect observations (because of abnormal visual conditions or ignorance) of animals already quite well known.

Many possible explanations have been put forward to account for sea serpents. A number of porpoises swimming one behind the other and rising regularly to take air might produce the appearance of a very large serpentlike creature progressing by a series of vertical undulations. A flight of seafowl and a brood of ducks have been mistaken for a large snake swimming at the surface of the water. Large masses of seaweed half awash have, on more than one occasion, been believed to be some gigantic animal. Basking sharks (*Cetorhinus maximus*), which have a habit of swimming in pairs one behind the other with the dorsal fin and the upper lobe of the tail just above the surface, produce the effect of a body 60 ft. or more long; even a simple partially decomposed specimen which was cast ashore was reported in all good faith as a sea serpent.

In the same category as basking sharks may be mentioned tunnies (*Thunnus thynnus*), porbeagles (*Lamna cornubica*) and chimaeras (*Chimaera monstrosa*), which at various times have been incorrectly recorded by observers unfamiliar with them. Ribbonfishes or oarfishes (*Regalecus*), which attain a length of 20–30 ft. and are snakelike in shape, have been suggested as the possible explanation of some so-called sea serpents, particularly of those reported from the Mediterranean, where these animals are most common. Nemertines, which may reach a length of 30–45 ft., have also been suggested as a possible explanation of some records. Sea lions when breaking surface for breath might, if seen from an unfamiliar viewpoint or in a fading light, be mistaken for much larger, snakelike animals.

Giant squids (*Architeuthis* species) are undoubtedly the foundation on which many accounts are based; these animals, which may attain a total length of 50 ft., are sufficiently uncommon to be unfamiliar to the majority of persons. They do occasionally frequent those regions from which many accounts of sea serpents have come—Scandinavia, Denmark, the British Isles and the eastern coasts of North America. One of these animals swimming at the surface with the two enormously elongate arms trailing along through the water would produce almost exactly the picture that many of the strangely consistent independent accounts require; a general cylindrical shape with a flattened head (=posterior end of the squid's body), appendages on the head and neck (=lateral fins and edge of mantle), colour dark, lighter beneath, progression steady and uniform, body straight but capable of being bent and spouting water (=water ejected from siphon). Further, sperm whales are known to kill and devour *Architeuthis* and similar cephalopods, and one of the most graphic accounts of the sea serpents speaks of it as in conflict with a whale around which it had thrown two coils and which it ultimately dragged below the surface; actually, it seems quite probable that the whale was eating a giant squid whose tentacles, thrown round the whale in the struggle, were mistaken for the coils of a snake, and that the whale, far from being dragged under, merely sounded with its prey in its mouth. A. C. Oudemans in his *The Great Sea Serpent: an Historical and Critical Treatise* (1893) gives a full account of all these possibilities.

A stranded animal carcass on the coast of Lower California was plausibly reported as a sea serpent. The remains were those of a large beaked whale. The body had been eviscerated (by sharks presumably) and the skin remaining was twisted up so that between the huge head and the large body there appeared to be a long and plesiosaurlike neck.

The discovery of an enormous eel larva in the Pacific, for which no adult is known, suggests that there may be an undiscovered gigantic kind of eel. This would of course perfectly satisfy the popular concept of a sea serpent. (H. W. P.; K. P. S.)

SEASICKNESS: *see* MOTION SICKNESS.

SEASONS, divisions of the year according to consistent changes in character of weather in the course of the annual cycle. The seasons named in European languages are associated with a yearly cycle in the life of plants, particularly of cultivated plants. Winter is the season of dormancy, spring that of sowing (Lat. *satio[n]*, whence Fr. *saison*, Eng. season) and germination; summer is the period of growth and maturity and autumn the time of harvest. This fourfold division of the year can seldom be recognized in the annual cycle of weather itself.

Outside the tropics, the essential characteristic of this cycle is

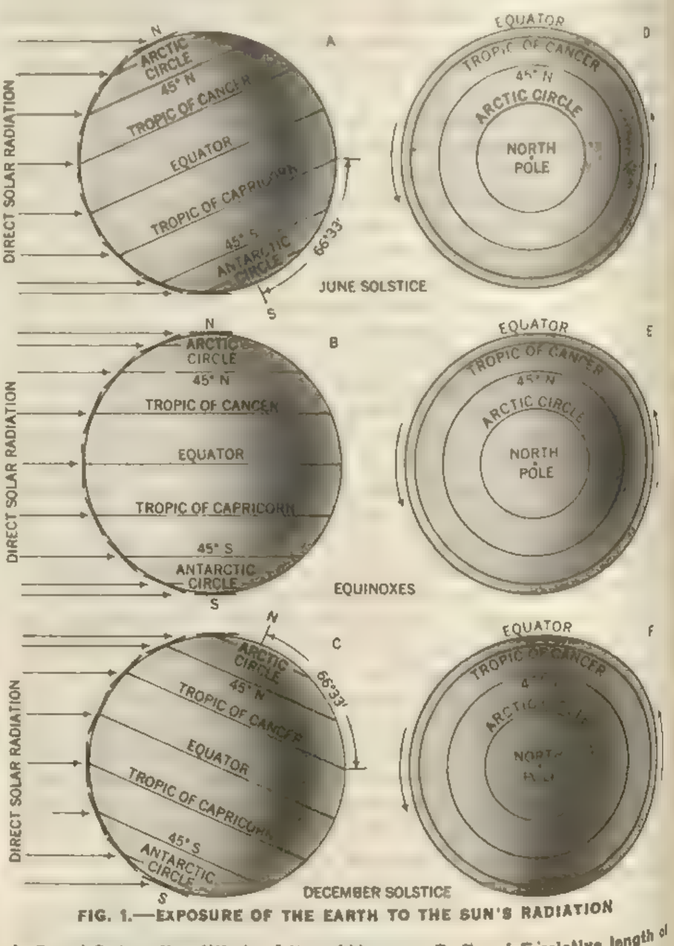


FIG. 1.—EXPOSURE OF THE EARTH TO THE SUN'S RADIATION

A, B and C show the attitude of the midday sun; D, E and F relative length of daily periods of sunshine and darkness

a pendulation of temperature between a single maximum and a single minimum. Other differences among the seasons are subordinate to the cycle of temperature. Thus, only the extreme seasons have a distinctive character; if the husbandman's yearly round had not left so profound an impress on European speech, spring and autumn would be considered merely transitional periods, hardly worthy of names co-ordinate in rank with those of summer and winter.

In the tabulation and discussion of climatologic observations made in middle and high latitudes, it is customary to combine the calendar months into seasons as follows:

Northern hemisphere:	*Winter*	*Spring*	*Summer*	*Autumn*
	December	March	June	September
	January	April	July	October
	February	May	August	November
Southern hemisphere:	*Summer*	*Autumn*	*Winter*	*Spring*

This division is only a convenience; in the appropriate northern latitudes March and November may be in fact winter months, May and September summer months. In relation to crops, the year is often divided into two parts, separated by the dates of the earliest and latest killing frosts in autumn and spring respectively.

Solar Radiation.—The seasonal changes of temperature are dependent upon the annual variation in the angle of the sun's rays and in the length of daylight, which determine the amount of solar radiation absorbed daily by the surface of the earth and the atmosphere (*see* CLIMATE AND CLIMATOLOGY: *Climatic Controls; Solar Radiation*). Assuming a cloudless sky and constant transparency of the atmosphere, the amount of solar radiation received at a point on the earth's surface at any moment is approximately proportional to the sine of the angular height of the sun above the horizon (the solar altitude). The amount received during any day evidently increases with (1) the length of the daily period of sunshine and (2) the sun's altitude at the successive moments of the day. The contrast between summer and winter is the result of seasonal variation in these two factors. Both have their maximum at the summer solstice and their minimum at the winter solstice. They result, in turn, from the changes in exposure of the earth to the solar radiation during its annual revolution about the sun. As the earth moves in its orbit, its axis maintains a nearly constant orientation in space, inclined about 66°33′ to the plane of the orbit. Figure 1, A, B and C, show the attitude (position) of the earth's surface with respect to the sun's rays at the solstices and equinoxes. The lines tangent to the meridian at the left of the diagrams represent the plane of the horizon at selected latitudes. The smaller of the two angles made by one of these lines and the shaft of the arrow that represents the solar rays is the sun's altitude at noon on the day and in the latitude indicated.

Diagrams D, E and F of fig. 1 illustrate the seasonal variation in length of the daily period of sunshine in the northern hemisphere. The great circle that separates the illuminated and shadowed hemispheres swings in an annual period back and forth across the segments of the earth's surface enclosed by the polar circles. In the daily rotation of the earth, any point in the hemisphere (except a segment near the pole at times other than the equinoxes) is carried alternately through sunshine and shadow. The part of the daily cycle during which the sun is above the horizon at any point is represented by the fraction of its parallel of latitude that lies within the illuminated hemisphere, and the part during which the sun is below the horizon by the fraction of its parallel that lies within the shadowed hemisphere.

The graphs of fig. 2 show the annual variation, by weeks, of the average daily amount of radiation from sun and sky actually received on a unit horizontal surface (the average "insolation") at four stations in North America. Outside the tropics, the maximum and minimum of the annual march of insolation usually fall near the solstices. The amount of heat produced by insolation depends on the fraction that is absorbed. Among types of surface common on the earth, a snow cover reflects most and absorbs least of the incident radiation. Hence, the annual march of absorbed radiation on land surfaces that are covered with snow

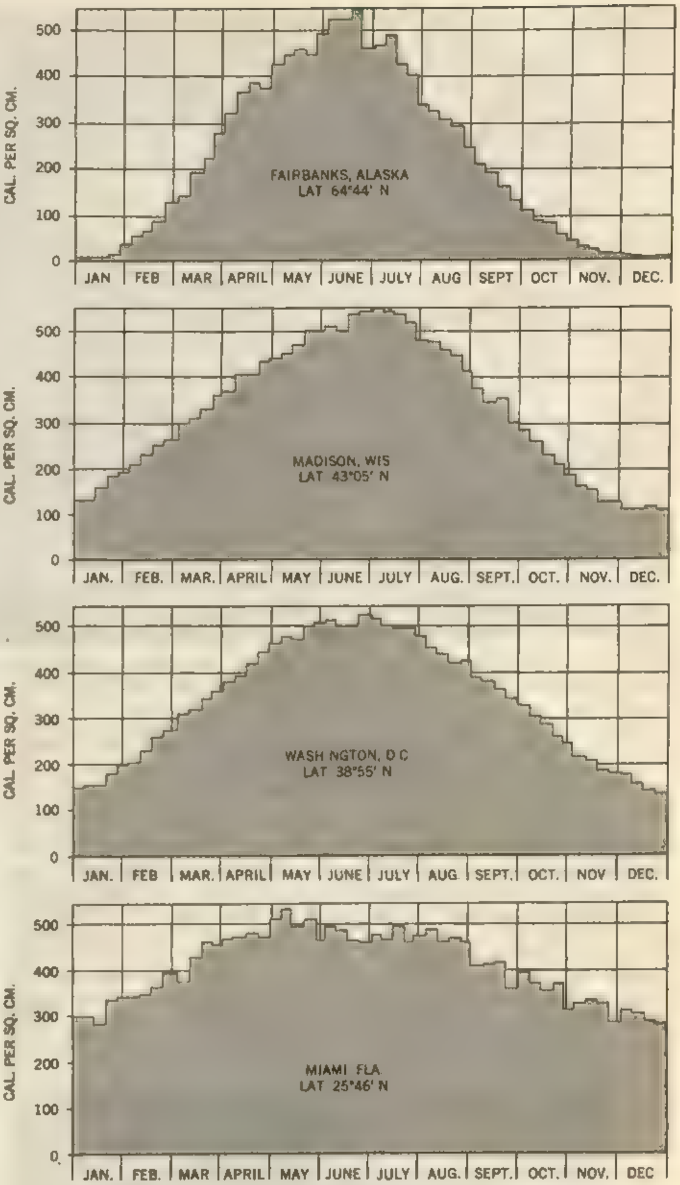


FIG. 2.—RADIATION FROM SUN AND SKY

Weekly means of daily amounts of radiation received on unit horizontal surface at four stations in the United States. Data from U.S. Weather Bureau

in winter and free of snow in summer has a larger amplitude than is suggested by the total incident radiation plotted in fig. 2, since a larger fraction of the total is absorbed in summer than in winter.

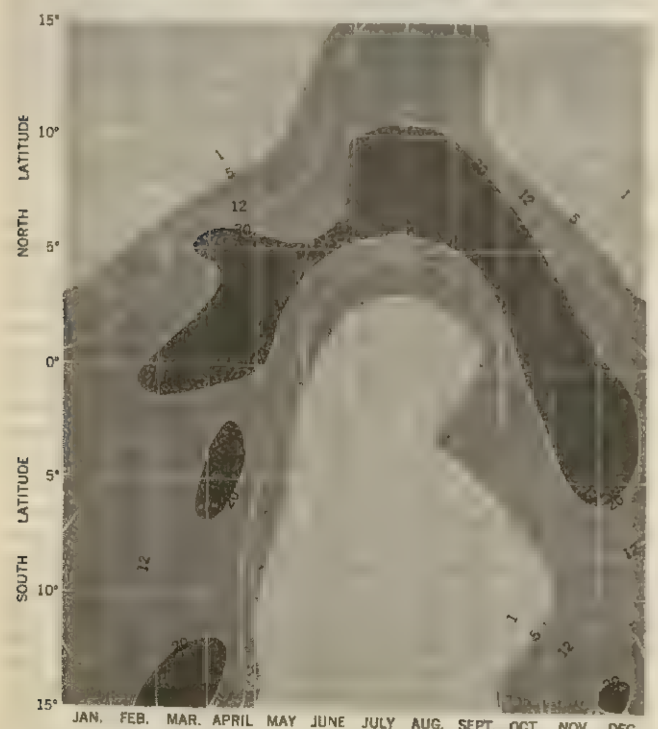
Annual Temperature March in High and Middle Latitudes.—The temperature of the air near the earth's surface, whether land or water, follows closely the temperature of the surface itself. How large a rise in temperature of the surface and the air close to it is produced by the absorption of a given amount of radiation depends on the disposition of the heat so produced. If heat is removed rapidly from the surface as radiation is absorbed, whether into the air, by radiation into space, or downward into the substratum, the temperature at and near the surface rises less than it does if the heat is removed more slowly. The processes that remove heat most rapidly from an absorbing surface are evaporation of water, the blowing of cool air over the surface and, in water bodies, the mixing, effected by the wind, of water at the surface with deeper water. All these processes reduce the rate of warming, and prevent the temperature of the surface from rising to as high a maximum as is attained by dry surfaces not effectively cooled by wind.

A surface of land or water attains its maximum temperature at the time when it becomes warm enough to lose heat to the atmosphere and space as rapidly as it is gaining heat by absorption

of radiation. This equilibrium cannot be reached until after the maximum of the march of absorbed radiation; that is, in most places, after the summer solstice. It is attained when the curve that represents the rate of absorption of radiation, declining from its maximum, intersects the curve of loss of heat from the surface. The cooler the surface is, in relation to the radiation absorbed, the longer is the time required for the curve of absorbed radiation to decline to the level of the curve of loss of heat. With a given march of absorbed radiation, therefore, the maximum of the annual march of temperature occurs later at cooler surfaces than at warmer ones.

A corresponding effect on the minimum of the march of temperature is produced by the return to the surface in winter of the heat transferred downward into the substratum in the preceding summer. With a given annual march of absorbed radiation, the amount of heat accumulated in the oceans and large lakes in summer and returned to their surfaces in winter is many times the turnover of heat below a land surface. Again with a given annual march of absorbed radiation, the larger the turnover of heat the higher is the temperature of the surface at its annual minimum, and the longer is the minimum delayed after the winter solstice. Thus, the characteristic differences in annual march of air temperature between the oceans and the lands that are embodied in the terms "maritime" and "continental" arise from the difference in annual heat turnover below their respective surfaces. Over the land, the maximum and minimum of the annual march of air temperature occur about one month after the solstices. Over the oceans, they are retarded about two months. As a result of the circulation of the atmosphere, the margins of the continents share the maritime regime of temperature.

Seasons in Low Latitudes.—From middle latitudes toward the equator, the amplitude of the annual march of insolation diminishes, and the annual cycles of insolation and temperature become increasingly susceptible to distortion by seasonal variation in cloudiness and rain. In fig. 2, there is a suggestion of flattening at the peak of the march of insolation at Washington, D.C., the consequence of summer cloudiness; and at Miami, Fla., the peak is broadly flattened. In tropical latitudes, seasonal differences in weather depend more on the shifting of the great wind belts of the earth than on the annual march of insolation. The alternation of rain and drought becomes more important than the



ADAPTED FROM KOPPEN, "DIE KLIMATE DER ERDE: GRUNDRISS DER KLIMAKUNDE"

FIG. 3.—MEAN MONTHLY PRECIPITATION, IN CENTIMETRES, ALONG THE MERIDIAN OF 15° E. LONGITUDE WITHIN 15° OF THE EQUATOR

change in temperature through the year. Seasonal alternation of rain and drought itself produces seasonal change in temperature, rainy seasons in the tropics being almost without exception cooler than dry seasons, regardless of the time of year in which the seasons occur.

The zone of convergence between the trade winds of the two hemispheres is a rainy belt, and rainy weather accompanies it as it swings northward and southward in its annual cycle. Where the belt is narrow, its pendulation may bring two rainy and two dry (or less rainy) seasons per year to places near the equator. Fig. 3 shows the distribution of precipitation through the year in the western part of equatorial Africa, along the meridian of 15° E longitude. There the meteorologic processes are more nearly symmetrical with respect to the equator than in any other land area in comparable latitudes. Between the equator and the parallel of 5° N latitude a fairly symmetrical sequence of two rainy and two dry seasons in the course of the year may be identified. In most tropical areas, however, the alternation of seasons includes no more than one rainy and one dry season. Relief of the land and exposure to winds from different directions introduce exceedingly wide variations in length and intensity of these seasons.

In India a strongly marked seasonal alternation of rainfall and drought extends northward into latitudes in which there are distinct temperature seasons, summer being rainy, winter dry. The result of this superposition of seasons based on temperature and on raininess is a sequence of three seasons of different lengths: a cool season in the months December–February, a hot season from March to mid-June and a rainy season from mid-June to November. The warmest season thus precedes the rainy period, in which cloudiness strongly reduces insolation. A similar superposition of temperature and rainfall seasons is observed, though in less extreme form, in the outer parts of the tropical belt in Africa and South America. In those continents the highest temperatures of the year are sometimes recorded before the rainy season, sometimes after it.

See also METEOROLOGY; WEATHER FORECASTING.

BIBLIOGRAPHY.—J. von Hann, *Handbuch der Klimatologie*, vol. 1, 4th ed., rev. by K. Knoch (1932); W. G. Kendrew, *Climates of the Continents*, 5th ed. (1961); W. Köppen and R. Geiger (eds.), *Handbuch der Klimatologie*, vol. 1 ff. (1930–40). (J. B. L.)

SEATON VALLEY, an urban district (1935) in the Blyth parliamentary division of Northumberland, Eng., 8 mi. NNE of Newcastle. Pop. (1961) 26,095. Area 19.5 sq.mi. The district lies in the centre of the south Northumberland coalfield and has ten mines employing the greater part of the population. Other industries produce heat-resisting materials, bricks, fire-lighters, and briquettes; agriculture ranks second to coal in importance.

Seaton Valley comprises ten mining villages; the biggest are Cramlington, Shiremoor, and Seaton Delaval. The latter, the administrative centre, takes its name from the De la Val (Delaval) family, which dates to Norman times. Seaton Delaval Hall (1720), just outside the district, is one of Sir John Vanbrugh's great achievements; nearby are the mausoleum (1776) and Sterling Castle (a ruin), both built by Delavals. Backworth Hall (1780) is a miners' welfare centre.

SEATTLE, the chief city of Washington, U.S., situated on a neck of land between Elliott bay (Puget sound) and Lake Washington (freshwater); a port of entry, headquarters of the Washington customs district; the seat of King county and the largest city of the Pacific northwest. It is 125 nautical miles (232 km.) from the Pacific ocean; 110 mi. S. of the Canadian border. Seattle's population was 557,087 in 1960. The standard metropolitan statistical area includes King county and, from 1959, Snohomish county; the population in 1960 was 1,107,213. The 1960 census showed nine other cities in the metropolitan area with a population over 5,000: Renton, Auburn, Kirkland, Kent and Bellevue in King county, and Everett, Edmonds, Lynnwood and Marysville in Snohomish county. (For comparative population figures *see* table in WASHINGTON: *Population*.)

Central Seattle is situated on a series of hills, some reaching 500 ft. above the harbour, in surroundings of great natural beauty.

WIDE WORLD

THE SEATTLE WORLD'S FAIRGROUNDS WITH MT. RAINIER IN THE BACKGROUND. AT RIGHT THE 600-FT. SPACE NEEDLE

Ninety miles to the south in the Cascade range is Mt. Rainier (*see* MOUNT RAINIER NATIONAL PARK). To the west, beyond Puget sound, rise the jagged Olympic mountains, which shelter the city from the heavy winter rains and winds of the coast. Beyond Lake Washington to the east are the high Cascades, which protect Seattle from the winter cold of the mid-continent. Prevailing winds (average velocity 8.9 m.p.h.) are from the southwest in winter and the northwest in summer. Average annual precipitation is 33.44 in. (84.9 cm.), two-thirds of it falling between October and March.

The salt-water harbour measures five miles across from West point on the north to Alki point on the south, and includes the east, west and Duwamish waterways, extending inland on the south side of the bay; Smith cove, on the north side; and Shilshole bay, the western outlet of the Lake Washington ship canal, north of West point. The canal (8 mi. long, minimum depth 28.5 ft.) connects Puget sound with Lake Washington, passing through Lake Union. The locks near the west end of the canal, which overcome the difference of 26 ft. between water levels, accommodate ships 760 ft. long. Elliott bay, where shipping is concentrated, has an entrance width of 2.5 mi. and is very deep and free from natural obstructions. Along the water front the hills have been graded down to give a comparatively level area for the business district. This is built up with several large hotels, public buildings (including the modern public library, 1960), and high business buildings, most of them erected either after World War I or in the post-Korean war building boom. A concrete pontoon bridge across Lake Washington, the floating portion of which is 6,561 ft. long, was completed in 1940. A second Lake Washington bridge, northwest of the first, was completed in 1963.

History.—The first white settlement was made at Alki point in Nov. 1851 by a group of middle westerners, most of whom moved around on Elliott bay the following February. The new town was named for a friendly Indian chief. In 1853 a plat was filed and Seattle became the seat of King county, Oregon territory. Later that year the region became part of the newly proclaimed Washington territory. Lumbering was the chief activity, the steam sawmill of Henry Yesler giving Seattle its early importance. By 1855 the population was 300. In Jan. 1856 it was attacked by neighbouring Indians and successfully defended by the U.S. sloop of war "Decatur." The city was incorporated in 1869. Early growth was slow, the 1870 census showing a population of 1,107, which rose to only 3,533 in 1880; in this period the chief exports were lumber, fish and coal (from the mines to the east and south of Lake Washington). After 1880 the rate of growth increased rapidly. In 1883 a spur line connected the city with Tacoma, main terminus of the Northern Pacific, the first transcontinental railroad to reach Puget sound. For some years thereafter (a period of intense rivalry between Seattle and Tacoma) this rail service was unsatisfactory. Not until 1893, when the Great Northern reached the city, did Seattle become a full-fledged rail terminus. In 1889, the year Washington attained statehood, most of the business district of Seattle burned but this did not hamper the city's growth. The population in 1890 was 42,837. The arrival of the first Japanese steamer in 1896 marked the beginning of considerable foreign trade.

Of greater significance was the Yukon mining strike, which touched off the great gold rush beginning in 1897. Almost overnight Seattle became an important commercial centre, the outfitting point for prospectors and the port to which they shipped their gold.

Subsequent Alaskan gold strikes, notably at Nome and Fairbanks, gave more impetus to Seattle's growth. Led by City Engineer Reginald H. Thomson, Seattle prepared for the future by grading the downtown hills (a project lasting for decades), by securing an abundant supply of water from Cedar river (in the Cascades, 26 mi. S.E.) and by building the first municipally owned hydroelectric plant in the country (1905). In 1909–10 the Alaska-Yukon-Pacific exposition was held at Seattle. Between 1905 and 1910 ten cities and towns were annexed and the city's population grew to 237,194. By 1910 Seattle had been reached by two other major railroads. The opening of the Panama canal in 1914 gave a new stimulus to the city's commerce, as did the completion of the Lake Washington ship canal two years later. During World War I Seattle built more ships than any other port of the U.S. The average number of wage earners in the city's manufacturing establishments rose in five years (1914–19) from 11,523 to 40,843; the value of the output, from $64,475,000 to $274,431,000. During the same period unions grew rapidly but the labour movement suffered a major setback with the failure of a general strike in 1919 that lasted for five days. The 1920 census showed a population of 315,312. The decade of the 1920s was one of moderate growth; that of the 1930s was one of stagnation and depression. World War II brought a great boom, with shipyards and the aircraft industry playing important roles.

The period after World War II was characterized by a resumption of the rapid growth of population, with an increasing diversity of industry.

Between April and Oct. 1962 the "Century 21" exposition (Seattle World's fair) was held on a 74-ac. tract of land one mile from the central shopping district with which it was connected by a high-speed monorail train. In addition to its own cultural and commercial values, the exposition left a legacy in the form of renovated and new permanent civic buildings.

Government.—The city is governed under a charter amended in 1946 to provide for a park superintendent and an administrative assistant to the mayor. Municipal elections are nonpartisan. The mayor and city councilmen are elected at large for four-year terms; beginning in 1962, council candidates had to file for designated positions on the ballot. The city owns its street transit, water supply and hydroelectric generating and distributing systems.

Harbour development and administration of public terminal facilities are in the hands of a public corporation called the Port of Seattle, created in 1911 by the people of Seattle and King county and administered by a three-man board of elected commissioners. Also under its jurisdiction is the Seattle-Tacoma International airport.

"Metro."—The Municipality of Metropolitan Seattle ("Metro") was established by popular vote in 1958, after the passage of an enabling act by the state legislature. Of the six functions authorized by the legislature (water supply, planning, rapid transit, parks-recreation, garbage disposal and sewage disposal), only the most pressing task of sewage disposal was permitted in the implementing act.

The Metro area, as initially established, covered 230 sq.mi. in northern King county, including ten incorporated cities and towns (Seattle, Renton, Bellevue, Medina, Clyde Hill, Hunt's Point, Beaux Art, Houghton, Kirkland and Bothell) plus unincorporated areas. Governed by a council consisting of representatives of the governmental bodies concerned, Metro quickly embarked upon the building of a sewage disposal system.

Commerce, Industry and Transportation.—Seattle is the leading commercial, industrial and financial centre of the Pacific northwest. Its geographical position in relation to the orient and Alaska makes it a natural receiving and distributing point for trans-Pacific and Alaskan traffic and the Panama canal gives access to Atlantic and Gulf markets, while the products of its tributary territory and its own manufactures supply staple articles for outgoing freight. A large part of the mail moving across the Pacific is handled in Seattle. In the early 1960s leading imports were limestone, lumber, gypsum rock, bananas and newsprint; the chief exports were coal, wheat, barley, tallow and wheat flour. Seattle is the principal outfitting point for the fisheries of the North Pacific and the chief supply point and wholesale market for logging camps and the agricultural northwest. Cheap electric power combined with abundant raw materials of certain kinds and distance from the older industrial centres has stimulated industry. The city has over 1,000 manufacturing establishments.

In addition to highway and railroad transportation facilities, Seattle is served by airlines with flights to Canada, Alaska, Japan and other points in the far east as well as to other points in the U.S. It is a port of call for steamers with sailings to Alaska and California, across the Pacific and through the Panama canal to eastern U.S. and European ports. It is the centre of the system of ferries owned by the state.

Education and Cultural Activities.—The public-school system includes elementary, special, junior high and high schools and a technical school. There are also Roman Catholic parochial schools and other denominational and nondenominational private schools. The University of Washington (*see* WASHINGTON: *Education*) is in Seattle. Other institutions of higher learning in the city are Seattle university (Roman Catholic, established 1891) and Seattle Pacific college (Free Methodist, chartered 1891). An educational television station broadcasts from the University of Washington. The public library system includes, in addition to the main library, several branches and bookmobiles, a municipal reference library and a library for the blind. Other cultural features of the city include the Seattle Art museum, the Frye Art museum, the Museum of History and Industry and the Seattle Symphony orchestra.

Parks and Recreation.—Seattle's public park system of over 3,500 ac. includes parks, community playfields, neighbourhood playgrounds, golf courses and beaches. The area also offers many opportunities for water sports and hunting and fishing.

See also references under "Seattle" in the Index.

BIBLIOGRAPHY.—Edwin J. Cohn, Jr., *Industry in the Pacific Northwest and the Location Theory* (1954); Murray Morgan, *Skid Road: an Informal Portrait of Seattle* (1951); Calvin F. Schmid, *Social Trends in Seattle* (1944). (R. E. BU.)

SEA URCHIN, popular name for spiny marine animals, members of the Echinoidea, a class of the Echinodermata (*q.v.*). The name *Echinus*, applied to a common genus, refers to the similarity between some of the Mediterranean urchins and the European hedgehog, called *echinos* by the ancient Greeks, which when rolled up forms a spiny ball of similar size. The majority of the sea urchins, the Regularia, conform to this type, whereas the remaining, the Irregularia, cake urchins, sand dollars (*q.v.*), and heart urchins, deviate in many respects.

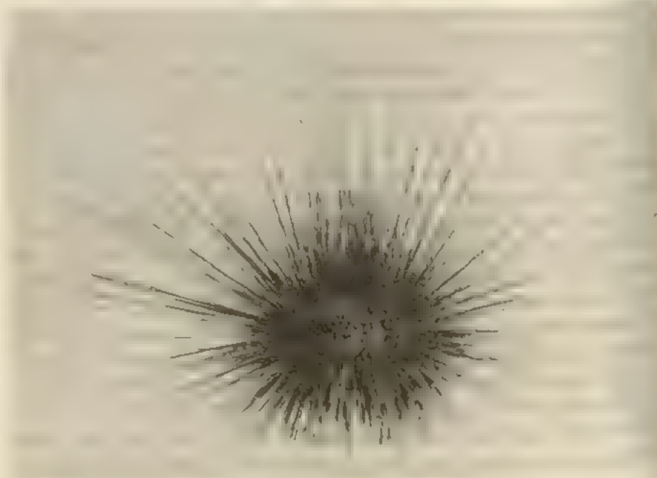
The regular sea urchins have a more or less spherical shell, or test, of closely fitting calcareous plates, arranged in five double rows, the radials, perforated by doubled pores, each pair admitting one tube foot, and five double rows without such pores, called the interradials. On the under side of the test is a large hole covered by skin; the mouth and the elaborate dental apparatus are placed in the centre of the hole. On the opposite pole is a smaller opening, the vent or anus. (*See* fig. 1.) The plates of the test carry smooth knobs, each articulating with a calcareous spine that can be moved by the action of basally placed muscles. The

BY COURTESY OF THE AMERICAN MUSEUM OF NATURAL HISTORY

FIG. 1.—TEST OF REPRESENTATIVE SEA URCHIN

Dorsal view (left), with spines removed, and (right) ventral view showing mouth opening and five teeth

size, form, and number of the spines determine the external appearance. In one primitive form, *Cidaris*, the larger primary spines, few in number and shaped like a policeman's night stick, are surrounded by a circle of smaller secondary spines. In *Diadema*, well known from most coral reefs, the spines are very long and slender and are effective as protective lances. (*See* fig. 2.) The long, thick spines of the slate pencil urchins (*Heterocentrotus*) were actually used as writing pencils by the early missionaries in the South Sea islands. A close relative has spines shaped like short flat discs, which fit together to form a mosaic of hexagonal plates. From muddy bottom in deep water is known one sea urchin with long slender spines, gently arched and with the tip expanded into a small hoof; by means of these modified spines the sea urchin can move over the soft bottom with little difficulty. A characteristic of most sea urchins is the pedicellariae—minute, three-pronged pincers, often provided with poison glands, borne



BY COURTESY OF THE AMERICAN MUSEUM OF NATURAL HISTORY

FIG. 2.—SEA URCHIN (DIADEMA ANTILLARUM)

From the West Indies and Florida

on slender stalks. They remove dirt, pass food particles along, and serve as effective weapons. From the five radial canals of the water vascular system, which characterizes all echinoderms, arise the podia or tube feet, which pass out through the pore pairs of the test and end in a terminal sucking disc.

The sea urchin's masticatory apparatus ("the Lantern of Aristotle") is a complex structure consisting of five ribbon-shaped teeth—rootless, like the incisors of a rodent—each of which is encased in a jaw, a triangular pyramid. Each jaw has the apex directed downward and is united by short muscles to its neighbours. From raised points along the edge of the large opening in the test extend five pairs of bandlike muscles, which attach to the lower part of the jaws. Five similar pairs of muscles, placed between them, attach to the upper ends, or bases, of the pyramids. By contracting or extending these muscles, the sea urchin can adjust its dental apparatus to facilitate chopping up larger pieces of algae, scraping off the minute algal growth on the rocks, grinding up smaller animals, and, in certain species, boring a hole in rock wherein the animal may spend its life. The majority of the regular sea urchins are vegetarians, but some live chiefly on smaller bottom animals and others subsist on the decaying debris of the muddy ocean floor.

Sea urchins are free-living forms, creeping on the sea bottom. They are able to move in all horizontal directions, at right angles to the axis that runs from the mouth to the anus. The Irregularia, however, are burrowers that have acquired a secondary bilaterality and can move in one direction only. In these forms, the anus has shifted down in one of the interradii and the radius opposite becomes the fore end.

In sea cookies, or cake urchins, and in sand dollars, which live in shallow water exposed to wave action, the test is thick, flattened and reinforced by internal pillars. The masticatory apparatus is preserved but modified. The mouth opening in the test is small and food is swept in along ciliated grooves and then sifted and ground up by the teeth.

In heart urchins, which live at greater depth on or in mud, the test is fragile and the mouth has moved forward and become a transverse slit; the complex dental apparatus has disappeared. Many of these forms live in burrows lined with mucus secreted by glands on the spines. The food is collected by means of long tube feet that reach up to the surface of the mud through a hole in the burrow and sweep up what organic particles there may be on the muddy surface.

There are about 700 species of sea urchins now in the seas, a fraction of what existed in former times. Because of their easily preserved skeletons, these forms constitute one of the most important groups of fossils used by students of paleontology. *See* also references under "Sea Urchin" in the Index.

See L. H. Hyman, *The Invertebrates*, vol. 4, *Echinodermata* (1955); R. T. Jackson, *Phylogeny of the Echini* (1912). (EL. D.)

SEBASTIAN, SAINT, a martyr believed to have suffered in Rome during the persecution of the emperor Diocletian. His feast day on Jan. 20 occurs in the early Roman martyrology of the chronographer of A.D. 354, as well as in 5th-century Hieronymian martyrology. According to his legend he was a captain of a Roman cohort who converted many soldiers and was condemned by Diocletian to be shot to death by archers. When they had left him for dead, a Christian widow, Irene, nursed him in her house. After recovering he presented himself before the emperor who immediately ordered him to be beaten to death. His body, thrown into a sewer, was found by another pious woman, whom the saint told in a dream to bury it near the catacombs. His relics are believed to be in the basilica of S. Sebastiano on the Appian Way, which attracted many pilgrims in the Middle Ages. He is frequently represented in art as a beautiful youth wounded by arrows.

See H. Thurston and D. Attwater (eds.), *Butler's Lives of the Saints*, vol. i, pp. 128–130 (1956). (H. C. G.)

SEBASTIAN (1554–1578), king of Portugal from 1557, was born at Lisbon on Jan. 20, 1554, the posthumous son of the heir to the throne John (d. Jan. 2, 1554). He succeeded his grandfather John III in 1557. During his minority the country was governed by the regencies of the widowed queen Catherine (1557–62), a sister of Charles V, and of John III's brother Cardinal Henry (1562–68), who succeeded Sebastian as king of Portugal. The period of the regencies was disturbed by troubles in Portugal's overseas dominions in Morocco, India, and Brazil.

Sebastian assumed personal control in 1568. Fanatically religious from an early age, he also devoted himself to hunting and military exercises, and his sole ambition was to lead a crusade against Morocco. He reversed John III's policy of gradual withdrawal from Morocco and embarked on new conquests. After a preliminary reconnaissance at Ceuta in 1574 he organized a large expeditionary force despite the advice of Philip II of Spain and Cardinal Henry. The army landed at Arzila in July 1578 and at the Battle of the Three Kings near Alcazarquivir on Aug. 4, 1578, sustained a disastrous defeat in which Sebastian was killed. His body was buried at Alcazarquivir, but rumours spread in Portugal that he was still alive. After Spain's conquest of Portugal (1580), Philip II recovered the body in 1582 on payment of a large ransom, but the movement known as Sebastianism grew, its adherents believing that Sebastian would return as the saviour of Portugal. Four pretenders to the throne impersonated him between 1584 and 1598. (DA. A. P.)

SEBASTIANO DEL PIOMBO (SEBASTIANO LUCIANI) (*c.* 1485–1547), Italian painter of the Venetian school who became the friend and protégé of Michelangelo, was born in Venice. According to G. Vasari, who knew him, he was 62 when he died in Rome in 1547, and it is therefore assumed that he was born in 1485. The same authority tells us that he learned "the first rudiments from Giovanni Bellini, at that time an old man. And afterward, when Giorgione had established in that city (Venice) the methods of the modern manner, with its superior harmony and its brilliancy of colouring, Sebastiano left Giovanni and placed himself under Giorgione, with whom he stayed so long that in great measure he acquired his manner." The truth of this statement can be seen in such works as the "Madonna and Child with SS. Catherine and John the Baptist" (Accademia, Venice), the organ doors of S. Bartolomeo a Rialto, Venice, and the altarpiece of S. Giovanni Crisostomo, Venice, recorded by Vasari as having been confused with the work of Giorgione himself. After Giorgione's death in 1510 Sebastiano is said to have finished some of his pictures, and there is a group of paintings whose authorship is still undetermined between them, including such works as the "Adulteress" (Glasgow) and the "Judgment of Solomon" (Bankes collection, Kingston Lacy, Dorset). The "Salome" in the National gallery, London, is dated 1510 and is therefore a key work of this period; Vasari himself once attributed it to Giorgione before realizing his mistake.

Sebastiano perhaps might now have inherited Giorgione's reputation and commissions, but in the spring of 1511 he was invited to Rome by the Sienese banker Agostino Chigi, who had just built the Villa Farnesina by the Tiber and was having it decorated. Sebastiano left Venice and returned only once (1528/29), spending the rest of his life in Rome. His decorations in the Farnesina met with success and he came into contact with Raphael and his pupils, who were also working there, and became his admirer. Later, perhaps quite soon after 1511, Sebastiano transferred his allegiance to Michelangelo and it seems that Michelangelo saw that if he were to correct Sebastiano's draftsmanship, Sebastiano's gifts as a painter were hardly to be equalled in Rome. The first result of this form of collaboration was the "Pietà" (Museo, Viterbo), for which Vasari says Michelangelo supplied "the invention and cartoon," the actual execution being by Sebastiano. The same thing happened with the "Flagellation" painted in oil on the wall of S. Pietro in Montorio, Rome. Michelangelo supplied a drawing for this as early as 1516, although the painting was executed only in 1520–24.

A direct rivalry between two groups came to a head when Cardinal Giulio de' Medici commissioned the "Transfiguration" from Raphael and the "Raising of Lazarus" from Sebastiano. The "Lazarus" (now in the National gallery) was commissioned in 1516 and finished in 1519; Raphael's "Transfiguration" was still unfinished at his death on April 6, 1520. For the next 11 years Sebastiano was reckoned the best painter in Rome, particularly of

portraits, which he produced with less delay than was usual with his other works.

In 1531 Cardinal de' Medici, now Pope Clement VII, gave him the well-paid office of the "piombo" (the supervision of the lead seals attached to papal bulls) and Fra Sebastiano, as he now became, worked less and less. One of his late works (1531 or later) is the "Lady as St. Agatha" in the National gallery. He also invented a method of painting on stone and a technique for painting in oil on plaster walls, and he tried to persuade the pope to make Michelangelo use oil instead of fresco for the "Last Judgment" in the Sistine chapel. This led to a coolness between them which lasted almost to Sebastiano's death, which was on June 21, 1547.

BIBLIOGRAPHY.—G. Vasari, *Lives of the Painters . . .*, English ed. by C. du G. de Vere, vol. vi (1913); L. Dussler, *Sebastiano del Piombo* (1942); R. Pallucchini, *Sebastian Viniziano* (1944); J. Wilde, *Catalogue of Italian Drawings . . . in the British Museum. Michelangelo . . .* (1953) on the relationship between Michelangelo and Sebastiano. (P. J. My.)

SEBORRHEA: see SKIN, DISEASES OF.

SECCHI, (PIETRO) ANGELO (1818–1878), Italian Jesuit and astrophysicist, noted for his pioneer investigations in solar phenomena and in stellar spectroscopy and spectral classification, was born at Reggio Emilia, Italy, on June 18, 1818. He entered the Society of Jesus in 1833, and became lecturer in physics and mathematics at the Collegio Romano, Rome, in 1839. Driven into exile in 1848 by the proscription of the Jesuits, he went to Stonyhurst college, Eng., and then to Georgetown university, Washington, D.C., where he pursued his mathematical and scientific work, returning to Rome as professor of astronomy and director of the observatory at the Collegio Romano in 1850. Here he erected a new observatory, equipped with a Merz refractor, over the church of Sant' Ignazio, where he carried out his researches into stellar spectroscopy, terrestrial magnetism and meteorology. His meteorograph was demonstrated at the Paris exhibition of 1867.

Secchi's main achievement was his division of stellar spectra into four groups: (1) stars like Sirius, with strong hydrogen lines; (2) stars similar to the sun, with numerous fine lines; (3) stars of the α Herculis type, with nebulous bands degraded toward the red; and (4) carbon stars, with violet-degraded bands. These divisions were subsequently developed into the Harvard classification, based on a simple temperature sequence, that forms the basis of modern systems. Secchi was also active in solar observation; he proved that prominences are appendages of the sun and discovered many features of their behaviour; he also photographed the corona at the eclipse of 1860. He was, in addition, an active observer of double stars. He died at Rome on Feb. 26, 1878.

Secchi wrote *Sulla mesura della base trigonometrica eseguita sulla Via Appia nel 1854–5* (1855); *Catalogo delle stelle di cui si è determinato lo spettro luminoso* (1867); *Sugli spettri prismatici delle stelle fisse* (1868); *Le Soleil* (1870); *Le stelle* (1877).

See Agnes M. Clerke, *A Popular History of Astronomy during the Nineteenth Century* (1902). (B. E. J. P.)

SECONDARY EDUCATION, a stage in education beyond the elementary (U.S.) or primary (British) level, usually beginning at age 11 or 12 and generally constituting either terminal education or preparation for college—in the United States, high school education. Secondary education constitutes a serious and difficult problem in virtually every country of the world. The difficulties arise from a number of causes: changing demands of modern culture; economic transformation in process everywhere; the ideal of equality of educational opportunity for all; and recruitment and training of young persons of talent to meet the needs and the higher intellectual standards of modern societies. To find a solution for these difficulties involves not only a reorganization of the educational systems of most countries but the abandonment of the monopoly of the traditional concept of secondary education.

This article discusses the background of these difficulties and the emergence and development of systems based on the U.S. common school or single-track plan and the British and European specialized school or multitrack plan, and analyses efforts to provide equality of educational opportunity for all.

For the history of secondary education in various countries *see* EDUCATION, HISTORY OF. For discussion of the organization of secondary education in various national systems of education *see* SCHOOL ADMINISTRATION. *See* also *Education* sections of articles on individual countries.

BACKGROUND OF THE PROBLEM

Classical Curriculum.—This concept can be traced to the grammar and rhetoric schools of Greece and Rome with emphasis on literary training. Expanded into the seven liberal arts the concept was directed to a different aim in the middle ages. The rediscovery of Greek and Latin authors established the predominance of the classical studies in the grammar schools in European countries and then in the American colonies (*see* CLASSICAL EDUCATION). As commercial relations and travel expanded, demands for a type of secondary education to meet new needs were met by the establishment of schools in the 17th and 18th centuries. These schools, however, did not enjoy the prestige of the grammar schools, particularly because they were refused recognition by the universities. Justification for the retention of classical studies, which became meaningless because more attention was paid to form than to content, was that even if they were useless, they trained the mind and the mind so trained could without difficulty attack new situations and problems. The classical curriculum enjoyed a monopoly as the only type of secondary education to the middle of the 19th century. Criticisms were met by an internal reform (neo-humanism) giving more attention to content and meaning, which originated in Germany in the 18th century.

From the early years of the 19th century the traditional classical schools began to enjoy special prestige not only because their graduates were admitted to the universities, but also because they were granted special privileges (*Berechtigungen* in Germany and *sanctions* in France), and in England, after the reforms of Thomas Arnold at Rugby, social recognition as character training institutions. Changing economic and cultural conditions, however, stimulated protests against the monopoly of the classical schools and a demand either for the introduction of modern subjects (foreign languages, sciences, mathematics) or for new types of secondary schools for the teaching of these subjects. By the end of the 19th century different types of curricula (classical, semiclassical and modern) in separate or in the same schools, received equal recognition and privileges. In England, however, publicly maintained secondary schools were not established until after 1902, and the private endowed schools, the best of which were known as "public schools," continued to provide an education which was mainly classical with other subjects grudgingly admitted. The older English universities made Latin and Greek compulsory requirements for admission well into the 20th century. In none of the countries was secondary education for girls provided at public expense until the 19th century, and it was late in that century that opportunities for their admission to universities were provided.

Secondary education constituted a separate type distinct from the elementary schools in all countries except the United States. An insignificant number of pupils passed on from the elementary to the secondary schools, preparation for which was given in different schools (*Vorschulen* in Germany, *classes préparatoires* in France and preparatory schools in England). These schools were fee paying and since their curricula were designed to prepare pupils directly for admission to the secondary schools, the pupils had an advantage over those who wished to transfer from the elementary to the secondary schools. Further, the secondary schools, since they charged tuition, were selective and drew their pupils from the wealthy classes, who could afford to pay the fees and the cost of maintaining their children while at school. Poor pupils of ability might obtain scholarships in competitive examinations, but their attendance usually required some financial sacrifice on the part of parents.

U.S. Common School.—The United States was the first country to provide equality of educational opportunity and to depart from the traditional dual type of organization by adopting the common school system in which elementary and secondary educa-

tion were articulated to facilitate continuous progression through the levels of the system. Some doubt still persisted in the 19th century as to whether the high school grades, in which other than elementary school subjects could be taught, came under the provisions for the establishment of common schools. In 1874, in a court decision on the Kalamazoo (Mich.) case, the provision of secondary education at public expense was declared to be constitutional. Another feature of the U.S. high school was the gradual expansion of the curriculum which included the subjects of the colonial Latin grammar school and the practical subjects of the private academies which emerged during the period of the American Revolution. A new type was the English Classical school established in Boston in 1821. The core of the high school curriculum consisted of subjects required for entrance to college whether by examination or by accreditation, first adopted by the University of Michigan in 1871. The proper content of a secondary education was just as much debated in the United States as in other countries and was defined at the end of the 19th century by the reports of the Committee of Ten (1893) and of the Committee on College Entrance Requirements (1899). It was generally considered that what was good for college entrance was also good for a liberal education.

Expansion of secondary education coincided with increasing urbanization and industrial expansion. Both factors made the provision of high schools possible and higher family incomes enabled parents to keep their children in schools beyond the elementary grades. Two other influences contributed to the spread of opportunities for secondary education for all. The first of these was the investigations conducted early in the 20th century of the doctrine of formal discipline and transfer of training in which it was found that neither developed spontaneously from the study of traditional subjects but could be cultivated under certain conditions. The conclusion was immediately interpreted to justify the provision of any subject that a pupil might request or be interested in. The second influence followed studies of elimination (the number of pupils who left school after a year or two) and of mortality (the number who stayed the course but failed to pass the final examinations). Both investigations resulted in throwing the doors of the high schools wide open for the admission of any subject that pupils, parents or the public demanded. The high school consequently became nonselective and the expansion upward of the universal school was encouraged by leaders who showed a greater interest in keeping pupils in school than in the quality of subjects taught.

The outstanding feature of this system is that pupils pass on normally from the primary to the postprimary school without having to pass entrance examinations or paying tuition. The high school has become the school for all youth, who are not distributed into separate schools but find the courses best adapted to their needs and abilities in the same comprehensive high school with the advice of guidance officials or parents. The weaknesses of the system, which were not recognized until after World War II, are that the high school tended to cater to the average pupils and to ignore those above and below the average; that all subjects were considered to be of equal value educationally under the quantitative system of points and units; and that pupils were not required to stay long enough with any subject to acquire a real mastery. The only requirements that were specifically defined are those necessary for entrance to college.

Dual Plan.—The organization of secondary education in the United States stands out in direct contrast to the educational systems of other countries except the few that adopted the American plan. The characteristic organization outside the United States has been on the dual plan; elementary and secondary education have been distinct types and only a minority of the elementary school pupils passed on to the secondary schools, generally only if they were bright and could win scholarships through a competitive examination. The secondary schools drew their clientele from private preparatory schools. Further the courses in the secondary schools were definite and were limited to the study of the mother tongue and national literature, foreign languages and literature (ancient or modern), mathematics and science, history and geography, some art, music and physical training. The affiliations of the complete secondary schools were with the universities rather than with the elementary schools; pupils who did not complete the full course generally entered white collar rather than manual or technical occupations.

The development of commerce and industry during the 19th century produced a demand for some form of intermediate education, less academic and more practical than the secondary and not as selective. The demand was met in Germany by the creation of *Mittelschulen*, in France of *écoles primaires supérieures* and in England of higher elementary schools and later in some localities of central schools. The curricula of these schools were somewhat more advanced than those of the elementary schools; pupils remained longer and were prepared for employment in business or industry as white collar workers but generally in a lower grade than pupils who came from secondary schools. Through the educational systems a sort of hierarchy was created and the more advanced secondary education was sought frequently for the status and privileges that it conferred. The cleavage between secondary and elementary or postelementary education was maintained by their administration in separate departments of the central authorities. The certificates conferred by different types of schools also contributed to the cleavage, since those of the secondary schools enjoyed higher social and occupational values.

Reform Movements.—This situation continued into the 20th century and gave rise to the reform movements proposed after World War II. In Germany secondary education for boys was provided in nine-year courses in *Gymnasien* where the chief subjects were Latin and Greek; the *Realgymnasien* with Latin and modern subjects; and *Oberrealschulen* with modern subjects. The final examination (*Abiturientenprüfung*) led to a certificate accepted for entrance to a university. Incomplete six-year courses were offered, with the same emphases as in the corresponding complete courses, in *Progymnasien*, *Realprogymnasien* and *Realschulen*. Girls received their secondary education in *Lyzeen* and *Oberlyzeen*. Secondary education in France was provided in *lycées* (state-maintained schools) and *collèges* (municipally maintained); classical, semiclassical and modern courses were available in each school, with special courses in the final or seventh year in philosophy or mathematics. The leaving examination, held in two parts at the ends of the sixth and the seventh years, led to the *baccalauréat*, a prerequisite for university entrance.

Secondary education in England was not provided at public expense until the Education act of 1902 set up local education authorities (L.E.A.s) and permitted them, if of a certain size, to establish and maintain secondary schools. The provision of public secondary education followed Matthew Arnold's plea to "organize your secondary schools" and the report of a commission, with James Bryce as chairman, appointed (1894) to consider "the best methods of establishing a well-organized system of education in England." Undoubtedly the growing competition of Germany exercised some influence in the same direction. Until passage of the act of 1902, secondary education was provided in ancient endowed schools, which included the "great public schools" and many others, day and boarding, which began to be called public schools and a large number of private schools. Many new public schools were established after Thomas Arnold's reform at Rugby. All schools, public and private, charged fees; the majority were boarding schools and their curricula were predominantly classical with instruction in modern languages (including English) and science permitted but not particularly favoured. Their affiliations were with the universities of Oxford and Cambridge, which in the latter part of the 19th century conducted school examinations. Schools were not supervised by any government authority, unless they had endowments to administer. Able pupils from elementary schools could win scholarships in competitive examinations to enter secondary schools at about age 11; normally such schools were day schools since boarding school fees were generally beyond the means of boys coming from elementary schools.

Girls received their secondary education at home or in private schools or schools provided by a voluntary organization such as the Girls' Public Day School trust (1872). Those who completed

their secondary education and passed the necessary examination could enter the few women's colleges in Oxford or Cambridge but could not (until the 20th century) proceed to degrees; others continued their education in one of the provincial universities on an equal footing with men.

The secondary schools established under the Education act were organized on the model of the existing secondary schools. Pupils were admitted by entrance examinations and tuition fees were charged until 1907, when the government regulations required the schools to provide "free places" up to 25% of the enrollment for pupils coming from elementary schools; the free places were subsequently replaced by "special places," the percentage of such places could be from 25% to 50%, and a means test was introduced. The curriculum was designed to provide a general or liberal education, and included English language and literature, at least one foreign language, geography, history, mathematics, science, organized games, physical exercises, music and drawing. Until 1917 the charge was made that the secondary school pupils had to take too many external examinations. In 1917 the number of external examining bodies, normally associated with the universities, was reduced to eight, and under the government's regulations pupils could sit for only two examinations—the first or school certificate, usually taken after four years of secondary education, and the higher school certificate at the end of the course. The certificates were accepted for entrance to the universities.

Patterns of Organization.—The four systems of secondary education—the German, French, English and American—have either influenced the organization of secondary schools directly or are found in other countries. Thus the German type exercised an influence on the types of secondary schools in the Scandinavian countries; French *lycées* were reproduced not only in French-speaking countries but in all the east European and middle east countries, where the French ministry of education often conducted examinations for the *baccalauréat*. English type public schools, boarding schools to meet the needs of a sparse population, were reproduced in Australia and New Zealand, until with the rise of urban centres day secondary or high schools were established similar in aim and curriculum to the local education authorities' schools in England with an emphasis primarily on an academic curriculum leading to university entrance. In New Zealand secondary education in rural areas is provided in district high schools attached to the primary schools; in urban areas two-year intermediate schools have been established leading to the four-year high school. Multipurpose schools are provided in Australia in sparsely populated areas, but in urban centres there exist academic, commercial, technical and home science schools.

In Canada the organization of secondary education varies from province to province; Quebec has replicas of the French and English systems to meet the needs of the different language and religious groups. In the other provinces the majority of pupils attend the publicly maintained schools, which may offer four, five or six year courses, and may be academic only or comprehensive as in the United States. In South Africa there is a common pattern in the fact that secondary education is continuous with primary, extends from standard VI to standard X with examinations at the end of standard VIII in two provinces and a final examination at the end of standard X, accepted for university entrance.

None of the patterns mentioned is found in Latin-American school systems. They have, however, solved one of the serious problems in having an end-on system—the pupils pass on to the secondary schools directly from the primary schools. Secondary education, given in *colegios*, *liceos*, *ginásras* (Brazil) or *institutos de segunda enseñanza* (Cuba), is five or six years in length. Generally the course is divided into two parts or cycles—the first cycle, usually three years in length, is general and common to all, and the second cycle provides for specialization for two or three years in humanities or sciences, depending on the requirements of the faculty that is to be entered in the university. Success in final examinations leads to the *bachillerato* in humanities or sciences. In Chile, a number of experimental schools have been established with courses divided into three parts—a common course of liberal education for all, an exploratory course to discover the pupils' special abilities and interests and a specialized course in humanities or sciences for entrance to the universities.

SECONDARY EDUCATION FOR ALL

World War I marked the beginning of a turning point in the history of secondary education in many countries outside the United States. The basic aim in the new movement was to establish a common system of education founded on the principle of equality of educational opportunity for all. Under various names (*Einheitsschule* in Germany, *école unique* in France and secondary education for all in England) the movement involved the abolition of the traditional dual system (elementary schools for one class of pupils and secondary for a selected group). It also implied a common foundation in education for all at least up to the age of 11 or 12 years. For the time being the term secondary education was not redefined. In Germany, the first step was to abolish the privileged status of the *Vorschulen* and in France the *classes préparatoires* and the same primary education was to be provided for all irrespective of the school that they attended. To enable pupils who did not transfer from the basic school (*Grundschule*) at the normal age of 10, a new type of secondary school (*Aufbauschule*) was established in Germany. In France fees for secondary education began to be abolished in 1931.

In England, following the report on *The Education of the Adolescent* (1926) of the board of education's consultive committee (Sir Henry Hadow, chairman), it was decided that more educational opportunities should be provided and that all pupils should be given some form of postprimary education from the ages of 11 to 15. The age of transfer from the primary to the postprimary stage was thus established at about 11; pupils who did not pass on to a secondary school received a more advanced education in "senior tops" than was previously available in the last three years of the traditional elementary schools. The senior schools, however, continued to be administered by the same department in the central board of education as the primary schools. The raising of the age of compulsory school attendance, recommended by the Hadow committee, was not actually achieved until 1947.

In the United States the problem of the articulation of primary and secondary education had been attacked earlier and for two reasons. It was thought: (1) that the transition from the eighth grade of the elementary school to the first year of the high school was too abrupt; and (2) it was considered desirable to introduce pupils to secondary school subjects slowly and to explore their abilities and aptitudes. One solution was to establish junior high schools to which pupils passed at the age of 12; another was to organize the last two grades of the elementary school as an intermediate section to serve as a gradual introduction to the high schools. The educational organization might be on the 8–4, 6–6 or 6–3–3 basis. In the United States, as in other countries, the end of World War I represented a new stage in educational development. Enrollment in the high schools began to rise rapidly and resulted in serious discussions not only of the aims of secondary education but also of the appropriate subjects to be taught. It had long been obvious that the traditional academic subjects were not suited either to the abilities or the interests of many adolescents. The issue was discussed by numerous committees of which the most influential was the Commission on the Reorganization of Secondary Education which in its 1918 report recommended the following objectives: health, command of fundamental processes, worthy home membership, vocation, civic education, worthy use of leisure and ethical character. These "cardinal principles" did not solve the curriculum problem. The practices varied throughout the country; in some schools various courses were offered; in others programs were tailor made for each pupil according to his abilities and interests. The only requirements that remained specific were subjects prescribed for college entrance with some leeway for electives. Entrance to college could be obtained under the widespread system of accreditation or in the case of private colleges by taking the examinations of the college entrance examination board established in 1899. Outside these requirements for pupils headed for college, the greatest flexibility prevailed and the number of "offerings" in the high

schools increased until a high school certificate meant only that its possessor had spent a certain number of years in the institution or had accumulated 14 to 16 units of miscellaneous subjects.

U.S. Appraisal.—The increasing enrollment in U.S. high schools, attended by about 70% youth between 14 and 18, brought with it a large number of problems, particularly with reference to the content of secondary education. There was at no time any suggestion that a return should be made to selective practices or to separate schools. Criticisms began to be made of the quality of high school achievement. It was stated (*e.g.*, Harvard report *General Education in a Free Society*, 1945) that "the tendency is always to strike a somewhat colourless mean." The Educational Policies commission of the National Education association, on the other hand, based its report *Education for All American Youth* (1944) on the theory that many pupils were not getting an education suited to their abilities, interests and needs and stressed common learnings and vocational preparation for all. At the same time, another group found that of the pupils enrolled in high schools 20% could profit from an academic course, 20% from a technical course and 60% needed quite a different type of education which came to be known as "life-adjustment training" (see *Vitalizing Secondary Education*, U.S. Office of Education Bulletin, no. 3, 1951).

The quality of education became a matter of public concern during and after World War II. Selective service boards had to reject many draftees for illiteracy and physical defects and the armed forces had the task of training many others to the level of functional literacy. The deficiencies of secondary and higher education were revealed when an adequate number of personnel could not be found to fill positions that required a mastery of mathematics, science and foreign languages. Little attention was paid to these revelations until the successful launching of the first sputnik by the U.S.S.R. in 1957 caused a panic throughout the U.S. because of the apparent superiority of the U.S.S.R. in science and technology. Studies were made of the Soviet educational system and groups of educators went to Russia. The general conclusion was that the Soviet system would not be suitable for U.S. education, which would have to be improved internally. Efforts were made in different parts of the country to insist on higher standards of achievement, to require more homework and to set up higher requirements for high school graduation. One aspect that was neglected was the need of teachers with better mastery of special subjects. With the aid of funds from foundations and government agencies, experiments were initiated to improve the courses in mathematics and physics and to provide additional training in summer schools for teachers of these subjects. The emphasis on more science and mathematics was widespread, while inadequate attention was paid to the humanities and social sciences, necessary for a balanced educational program.

Attention was focused on the issue of quality, standards and curriculum by the publication in 1959 of *The American High School Today* by James B. Conant and *Education and Freedom* by Adm. H. Rickover. Conant recommended that all pupils in the comprehensive high schools, irrespective of their special programs, for graduation should be required to study four years of English, three or four years of social studies and at least one year of mathematics and science. For pupils able to profit from an academic course, estimated at 15% to 20%, he advocated four years of foreign languages, four of English, four of mathematics, three of science and three of social studies. Since the small high school is normally unable to provide such a full course, he suggested the establishment of a number of central high schools. Rickover's thesis was that "only massive upgrading of the scholastic standards of our schools will guarantee the future prosperity and freedom of the republic." He urged a program similar to that recommended by Conant and the establishment of special high schools to demonstrate the new program.

It has been generally recognized that the changes taking place after the end of World War II demand better trained manpower and leaders in every aspect of national life—political, social, cultural, economic and scientific and technological—and that a general or basic liberal education should be available for all adolescents. The danger that threatened was a certain imbalance arising from the emphasis on science and technology. The most serious difficulty in implementing the desired reconstruction of secondary education lay in the inadequate supply of well-qualified teachers, caused in part by better paying jobs in other occupations. This situation stimulated plans for the wider use of educational television, programmed instruction and other devices.

Secondary Education in the U.S.S.R.—The educational systems in the Union of Soviet Socialist Republics are organized on the principle of the common or unitary plan with four years of primary school (entered at the age of 7) and three years of junior and three years of senior secondary schools. The incomplete systems consist of seven years, the complete of ten years of schooling, but the rural areas generally have only the four-year primary schools. Secondary education is free and generally coeducational. The curriculum was academic and included: Russian or native language and literature, foreign languages, mathematics, sciences, history, geography, constitution of the U.S.S.R., drawing, singing and physical training, with many hours of homework and stiff examinations, especially at the end of the course. In 1958 the system became subject to criticisms of which Premier Nikita Khrushchev was a leading spokesman. He attacked the emphasis on academic education as unrelated to life and production needs of the country, and particularly because it developed in the pupils a contempt for physical labour. Students who could not gain admission to the universities were unprepared to enter industrial or agricultural occupations. He recommended the academic course for pupils of recognized superior ability, especially in mathematics and science, music or fine arts. The majority should, after seven or eight years of schooling, continue their education in factory vocational schools or work during the day and attend evening schools. Education, following the principle stated by Lenin, should be linked to socially useful labour and production. Ensuing changes added an 11th year to the ten-year system, with polytechnical subjects at all levels and work experience in the last two years. In the mid-1960s the schools returned to the ten-year system with a more orthodox academic program while retaining emphasis on scientific and technical training to produce skilled workers.

Reorganization.—*England.*—The demand for educational reconstruction to provide greater equality of opportunity, raised during World War I, was continued in most of the western European countries and began to be met after World War II. (The postwar changes in the German Federal Republic are given under EDUCATION, HISTORY OF.) The groundwork was laid in England by the White Paper, *Educational Reconstruction*, issued in 1943 by the board of education; and in France by the report of the Langevin commission (*La Réforme de l'Enseignement*, 1946). The recommendations of the White Paper and the report of the Spens committee, *Secondary Education* (1938), were incorporated in the Education act, 1944. Local education authorities were required to provide secondary education for boys and girls above the age of 11, suited to "their different ages, abilities and aptitudes." Three types of schools were proposed—grammar, technical and modern, and combinations of these in multilateral schools. Grammar schools give the traditional academic education; modern schools give a combination of general and practical education, which may include a foreign language; technical secondary schools, still few in number, combine general and technical subjects. Fees have been abolished in publicly maintained schools. Academic or grammar school courses are also provided in direct-grant schools, which receive grants from the government on condition that free places are provided for pupils coming from primary schools, and in independent or private schools, including the public schools. In 1952, the examination system was changed and the general certificate examination was introduced to be taken at about 16 to encourage pupils to stay in school to that age; papers may be taken at ordinary, advanced or scholarship standards. Since the subjects to be taken are not prescribed, modern school pupils began to sit for the examination.

The two chief problems of secondary education in the early 1960s in England centred on (1) the examination by which pupils were allocated to one of the three schools, sometimes referred to

as educational types, and on (2) the separation of pupils into different schools. It was argued that 11 plus could be too early an age to discover the abilities and aptitudes of pupils and that the separation of pupils into different schools was socially divisive. The ministry of education permitted the establishment of comprehensive high schools—about 45, most of them in London and Coventry. The Labour party was strongly in favour of the comprehensive schools (see *Learning to Live: Labor's Policy for Education*, n.d.). The conservative policy was to improve the buildings and equipment of all public secondary schools, to reduce the size of classes and to retain the grammar schools for the ablest children (*see* Ministry of Education, *Secondary Education for All, A New Drive*, 1958).

France.—The system of education in France is being gradually reorganized by decrees; there had been no legislative enactment for reconstruction. The age of compulsory attendance was raised to 14 in 1936. The majority of pupils continue in elementary schools for eight years, of which the last two are practically oriented. At age 12 pupils may take an entrance examination to some form of secondary education: *cours complémentaires*, attached to the primary school and offering a modern or prevocational course; *collèges modernes* (formerly *écoles primaires supérieures*), which provide a general and practical course; *collèges techniques* with academic and technical subjects leading after six years to a new *baccalauréat technique* and to more advanced technical education; and the traditional *lycées* or *collèges*, the former state and the latter municipal institutions. Courses of the *lycées* and *collèges* have been increased, particularly in the second cycle (fifth and sixth years), and provide for a number of different combinations of academic subjects; the first part of the *baccalauréat* is taken at the end of the second cycle. The seventh or last year is devoted to specialization in philosophy–letters, experimental sciences or philosophy–sciences, mathematics or technical and economic sciences. At the end of this year the second part of the *baccalauréat* is taken. Students are no longer admitted to universities on the *baccalauréat* alone but must spend a preliminary propaedeutic or preparatory year in a university and pass another examination before being finally admitted.

Transition.—Secondary education is everywhere passing through a transition stage as a result of 20th-century changes in culture and civilization. The problems that emerge involve a reconsideration of the concept of a liberal education, development of a better understanding of contemporary issues, national and international, and training for adaptability to social and economic changes. In the 19th century the industrial revolution hastened the movement for universal elementary education to produce, among other ends, literate workers. The 20th-century industrial revolution, based on scientific and technological advances, demands more education and is producing occupational redistribution, with prospects of further changes from automation. Redistribution of wealth and the shift of social classes have made the prolongation of education possible. Intensification of nationalism and the consequent economic competition have stressed the importance of well-trained manpower, so that education is being prolonged both in democratic and in communist states. The keynote everywhere is equality of educational opportunity, which brings the problems of allocation according to abilities and aptitudes and of determining the nature of general education for all and the time to begin specialization. *See* ELEMENTARY EDUCATION; COLLEGE; UNIVERSITY; WOMEN'S EDUCATION, HISTORY OF; ADULT EDUCATION; *see* also references under "Secondary Education" in the Index.

BIBLIOGRAPHY.—James B. Conant, *Education and Liberty* (1953); J. F. Cramer and G. S. Browne, *Contemporary Education* (1956); E. Hylla and W. L. Wrinkle, *Die Schulen in West-Europa* (1953); I. L. Kandel, *History Of Secondary Education* (1933), *The New Era in Education* (1955) and *American Education in the Twentieth Century* (1957); V. Mallinson, *Comparative Education* (1958); UNESCO, *World Survey of Education*, vol. i (1955), vol. iii (1961). (I. L. K.)

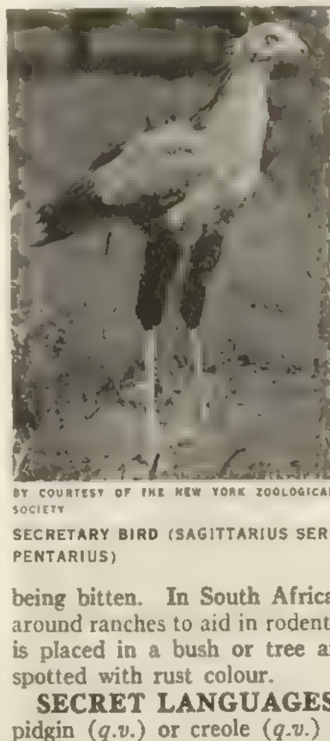
SECONDARY ERA: *see* MESOZOIC ERA.

SECOND BAKU (VTOROYE BAKU) is the popular name for a group of very rich oil fields of the U.S.S.R., lying between the middle Volga and the Urals. These fields have surpassed the old Baku (*q.v.*) oil field as the major source of petroleum in the Soviet Union. The Second Baku lies in the Tatar and Bashkir autonomous republics and the Kuibyshev, Orenburg and Perm *oblasts*. The presence of oil there has been known since the 18th century, but only during the first five-year plans were sufficiently deep bores put down. The first well was at Verkhne-Chusovskiye Gorodki near Perm in 1929, followed by Ishimbai (1932) and Syzran (1935). Since then, production has risen very rapidly from about 6,000 tons in 1929 to about 75,000,000 in 1958.

The Second Baku comprises five major groups of oil fields: (1) the Perm area, with Krasnokamsk as the largest centre; (2) on the Volga near Kuibyshev with Syzran as its centre; (3) the Tuimazy area, with its new oil town of Oktyabrski (since 1955 a new and rich field has been developed close by at Shkapovo); (4) Buguruslan in Orenburg *oblast*; (5) the Belaya valley around Ishimbai and Sterlitamak (*q.v.*). Oil occurs in rocks of Silurian to Permian age, and in general at considerable depth. At Tuimazy, wells reach 9,600 ft.

Natural gas is also found in quantity in the Second Baku, especially in the Buguruslan area. Refining is carried on at a number of centres in each field, notably at Krasnokamsk, Ufa, Sterlitamak and Syzran. (R. A. F.)

SECRETARY BIRD, an African bird with long legs, standing nearly four feet high. Only one species is known definitely, *Sagittarius serpentarius*, constituting the family Sagittariidae. One of the most powerful birds of prey (Falconiformes), it is found locally south of the Sahara. From the back of the head and nape hangs an erectile tuft of long black feathers which reminded early naturalists of a secretary with quill pens in his hair. Around the eyes is orange skin; the head, neck, and back are bluish-gray, the lower surface black; tail quills are banded with black and tipped with white; the beak is hooked.



BY COURTESY OF THE NEW YORK ZOOLOGICAL SOCIETY

SECRETARY BIRD (SAGITTARIUS SERPENTARIUS)

The secretary bird feeds on insects and reptiles and can kill the most venomous snakes, striking them repeatedly with its taloned feet. The long legs together with the bird's habit of leaping back after each stroke preserve it from being bitten. In South Africa it is sometimes tamed and kept around ranches to aid in rodent and snake control The huge nest is placed in a bush or tree and in it are laid two white eggs, spotted with rust colour.

SECRET LANGUAGES. Quite the reverse of a jargon, pidgin (*q.v.*) or creole (*q.v.*) language, each of which serves to broaden communication between groups, a secret language is used by a special group to preserve its identity and to exclude outsiders. Thus the term refers rather to the social function of a speech form than to any property of its structure. The field as a whole is poorly studied, but there are many isolated cases on record. Occasionally, as with Romany (Gypsy) languages and with the Todas in India, a quite distinct language serves the function of secrecy. An imperfectly remembered language, preserved under unusual conditions, may serve ritualistic purposes, as with the west African Lucumí in Cuba.

Most so-called secret languages are grammatically the same as the language of the surrounding community, but consist in systematic substitutions for individual words. In this case it is difficult to distinguish clearly between secret languages in a strict sense and argot, slang (*q.v.*), which uses ephemeral substitutions, and caste, class and technical occupational dialects. Such lexical substitution is recorded among the Sema Nagas and the Langos and among outlaws, thieves, organized criminals, itinerant peddlers and entertainers and religious groups in many parts of the

world. Similar substitutions are used informally by lovers and children. Systematic insertion or substitution of sounds in the normal language of the community is familiar to us as "pig Latin"; analogous practices occur in other western cultures and are elaborately used by Tagalog adolescents in the Philippines.

The principal criterion for application of the term secret language seems to be a concerted effort on the part of the users to exclude others and prevent their acquiring it. By contrast, tabu words are known to a large part of the community, but are by convention not spoken.

BIBLIOGRAPHY.—R. Iversen, *Secret Languages in Norway*, 3 vol. (1944–50); O. Jespersen, *Language* (1922); L. Bloomfield, *Language* (1933); H. C. Conklin, "Tagalog Speech Disguise," *Language*, 32:136–139 (1956); "Linguistic Play in its Cultural Context," *Language*, 35:631–636 (1959). (E. P. H.)

SECRET SERVICE, U.S., a bureau of the treasury department, charged by law with the following duties:

1. Protection of the person of the president of the United States and members of his immediate family, the president-elect and the vice-president.

2. Detection and arrest of any person committing any offense against the laws of the United States relating to coins, paper money, checks and securities of the United States and of foreign governments.

3. Detection and arrest of any person violating certain laws relating to the Federal Deposit Insurance corporation, federal land banks, joint-stock land banks and national farm loan associations.

Origin and Early Duties.—Before the American Civil War, the paper currency of the United States consisted of notes issued by banks operating under federal or state charters. As there was no federal paper currency at that time congress left the detection and prosecution of counterfeiters to the states, local authorities and bank associations. The first paper money issued by the U.S. government consisted of noninterest-bearing treasury notes authorized under acts of congress passed in 1861. An 1862 act authorized the issue of "United States notes," popularly referred to as "greenbacks," and the National Bank act of 1863 provided for the issuance of national bank notes (*see* BANKING: *The United States*).

During the Civil War funds were appropriated by congress to suppress the widespread counterfeiting of the currency. The solicitor of the treasury supervised the expenditure of the money appropriated for this purpose; most of the money was used for the payment of rewards to individuals for furnishing information leading to the arrest and conviction of counterfeiters. As this method of operation did not prove to be effective it became necessary to make a definite and centralized effort to suppress counterfeiting and to restore and maintain confidence in the federal currency. Accordingly, in July 1865, Hugh McCulloch, then secretary of the treasury, created the U.S. secret service and appointed William P. Wood as its first chief. Wood had carried out many investigative assignments during the Civil War for the secretary of war, Edwin M. Stanton and was experienced in hunting down counterfeiters. His force at the outset numbered about 30 men. The headquarters office was in Washington, D.C., and field offices were soon established in other cities.

Although the secret service was organized primarily to suppress counterfeiting, other departments of the federal government soon requested the use of secret service agents to investigate violations of many other laws. Since there was then no other general federal law-enforcement agency, it became common practice for the treasury department to loan secret service agents to other government departments to do important investigative work. From 1905 to 1908, by direction of Pres. Theodore Roosevelt, the secret service also investigated extensive land frauds, including those resulting from passage of the Homestead act (1862) which opened lands to settlement. Through dummy entrymen, many cattle barons in the west obtained rich grazing lands intended for homesteaders, and other unscrupulous landowners obtained lands rich in coal, oil and timber by fraudulently claiming them for agricultural purposes. In one such investigation near Durango, Colo., a secret service agent was shot and killed while investigating the suspected theft of coal from government property. The secret service exposed numerous land frauds, recovered millions of acres for the government and prosecuted many influential businessmen and public officials. Following this exposé, congress restricted the work of the secret service to its normal treasury department duties, and several secret service agents were permanently transferred to the department of justice and other government departments to form separate investigative units.

With the outbreak of World War I in 1914, Pres. Woodrow Wilson lifted previous restrictions and directed that the secret service be employed to uncover violations of neutrality. Numerous cases involving espionage and counterespionage were successfully completed by the secret service during the war and its agents worked in close co-operation with the military intelligence services. In the mid-1920s the secret service also investigated the Teapot Dome oil reserve scandals.

Protection of the President.—In 1901 Pres. William McKinley was shot and killed at Buffalo, N.Y., the third president of the United States to meet death at the hands of an assassin. Immediately thereafter, agents of the secret service were assigned to guard Pres. Theodore Roosevelt. In 1913 secret service protection was extended to the person of the president-elect, and in 1917 to the members of the immediate family of the president. Authority to guard the vice-president at his request was given the secret service by act of congress in 1951. During the early decades of the 20th century the secret service also provided protection for foreign rulers who visited the United States until this responsibility was transferred to the security division of the department of state. In carrying out their duty to guard the president secret service agents remain near him at all times and accompany him on all his travels. They went to Panama, for example, with Pres. Theodore Roosevelt, the first president to set foot on foreign soil while in office. Since then, secret service agents have guarded presidents during many historic trips abroad.

While the entire force of the secret service participates to some extent in the protective function, selected special agents known as the White House detail are assigned to stay with the president at all times. Protection of the president involves much more than keeping bodyguards around him; it also entails what is sometimes called "preventive protection." When the president travels, an agent of the White House detail goes in advance to the cities to be visited and joins agents from the secret service field offices nearest to those cities. These men confer with city and state police and other local authorities, determine the safest routes of travel, arrange for temporary residence if the president is to remain overnight and take other steps to provide maximum protection.

When the president rides in an open car through crowded streets, secret service agents walk or ride alongside and keep close watch on the bystanders. These protective measures are supplemented by investigations of persons who threaten the president in letters and other communications. All such communications are carefully analyzed by a group of specialists in what is called the protective research section of the secret service. In spite of these measures, on Nov. 22, 1963, Pres. John F. Kennedy was assassinated while riding in an open car in Dallas, Tex., the first president to meet that fate since the secret service was assigned responsibility for guarding the president in 1901. [The Warren commission report of Sept. 24, 1964, criticized the service's procedures and its inadequate liaison with the F.B.I. and other agencies. In November 1965 the top echelon of the service was reorganized to meet the Warren commission criticisms.]

White House Police Force.—On May 14, 1930, the White House police force, which had been organized in 1922 to protect the executive mansion, executive office and White House grounds, was placed under the supervision and control of the chief of the secret service. Members of the force receive training in the use of revolvers, submachine guns, riot guns, gas grenades, and in the techniques of disarming an opponent. They also receive training in fire fighting and rescue work.

In Nov. 1950 members of the White House police force and secret service agents foiled an attempt by two Puerto Ricans to assassinate Pres. Harry S. Truman, who was then residing at Blair House while the White House was being repaired. One of the

attackers was killed and the other wounded; one White House policeman, Leslie Coffelt, was killed in the struggle and two others were seriously wounded.

Treasury Guard Force.—On July 1, 1937, by direction of the secretary of the treasury, the supervision and control of the treasury guard force was delegated to the chief of the secret service. The treasury guard force protects the main treasury building and annex and a huge quantity of securities in the custody of the treasury department. Members of this force receive training in handling firearms and disarming opponents and in fire protection, alarm systems, first aid and other specialized training.

Other Activities.—The secret service performs a number of services aimed at prevention of counterfeiting and forgery of currency, checks, bonds and other obligations of the U.S. government. These measures include lectures by special agents, exhibits, pamphlets, circulars and an educational film. The secret service has published two informative and educational booklets, *The United States Secret Service, What It Is, What It Does* and *Know Your Money*. *See* also COUNTERFEIT MONEY.

See W. S. Bowen and H. E. Neal, *The United States Secret Service* (1960). (U. E. B.; X.)

SECRET SOCIETIES, any of a large range of membership organizations or associations having secret initiation or other rituals, oaths, grips (handclasps) or other signs of recognition. Elements of secrecy may vary from a mere password to elaborate rituals with a private language and peculiar ceremonials, costumes and symbols. The term may be applied to such widely divergent groups as U.S. college fraternities, the Ku Klux Klan and international Freemasonry as well as to similar phenomena in primitive cultures. For discussion of the latter *see* SECRET SOCIETIES, PRIMITIVE.

Among the earliest secret societies of which historical evidence exists were the oriental mystery cults and the religious mysteries of ancient Egypt, Greece and Rome, which had secret rites, initiations and revelations of still more ancient wisdom (*see* MYSTERY; *see* also MITHRAISM; ORPHEUS; OSIRIS; PYTHAGORAS AND PYTHAGOREANISM). Whereas the mysteries employed secrecy to guard religious truths, other groups have been forced to adopt secrecy to escape or survive suppression and persecution, as in the case of the early Christians in pagan Rome and, in their turn, of various heretical groups in the middle ages (*see* EARLY CHRISTIAN CHURCH; INQUISITION; MANICHAEISM). Medieval guilds (*q.v.*) resorted to solemn initiatory oaths and other elements of secrecy primarily for economic self-protection. Throughout history revolutionary, subversive and conspiratorial groups have organized secretly, as in the case of the Sons of Liberty. The repression of liberal, nationalist and republican movements in Europe in the 19th century, for example, produced an underground network of revolutionary secret societies (*see* EUROPE: *History: The 19th Century, 1815–1914; see* also CARBONARI; CAMORRA). Other examples may be found in the Fenian Irish Republican Brotherhood (*see* FENIANS, AMERICAN; IRELAND: *History*) and the Decembrists or Union of Salvation in Russia (*see* DEKABRISTS). The very existence of secret societies has prompted antagonisms and fostered accusations of immorality, subversion and heresy. Such accusations were made against the Roman mysteries and early in the 14th century were used to justify the ruthless suppression of the Knights Templar (*see* TEMPLARS). The early 19th-century Antimasonic movement in the United States offers another interesting example of opposition to secret societies (*see* ANTIMASONIC PARTY). Many modern secret societies were formed primarily for social and benevolent purposes and to carry out charitable and educational programs: these have been especially numerous in the United States and in the later 19th century attracted large numbers of immigrants who sought companionship and guidance among people who spoke their native language and followed their customs. In many communities such societies in the 20th century have continued to provide the principal means of members' social and civic activities. (*See* FRATERNAL ORGANIZATION; FRATERNITY AND SORORITY.)

With all their diversity of type and origin, secret societies have certain characteristics of structure and function in common and some of their ceremonials reveal surprising similarities. Historic and other details of the more important groups are covered in separate articles under their own names.

Structure and Function.—Secret societies are made up, *ipso facto*, of persons presumably oriented toward similar ends, and these ends usually manifest the characteristic differentiating secret societies from all others—that is to say, the ends are secret. Moreover, admission to membership almost always involves the explicit obligation to preserve such secrecy, and penalties for its violation are likewise explicitly stated. The explicitness involved may sometimes apply only to the members of the society, for secrecy may be so complete that even the existence of some societies is not revealed to outsiders; revolutionary, heretical and similarly subversive secret societies are cases in point. More frequent is partial secrecy: the existence of the society is publicly acknowledged or even proclaimed, as by the Ku Klux Klan (*q.v.*) in the U.S. after the Civil War and again in the 1920s and the 1950s and 1960s; at least some of the ends are made generally known; parts of the society's ceremonial are performed openly; and public co-operation with other groups having fundamentally differing ends may occasionally be undertaken.

But, obviously, secret societies would lose their reason for existence if secrecy were ever entirely abandoned. Many fraternal organizations, for example, maintained the secrecy of their rituals into the second half of the 20th century although, as in the case of college fraternities and sororities, these survived largely as formalities. In most instances, the core of the binding secrecy is to be found in the society's ceremonial. The essential part of this is rarely if ever legitimately known to those who are not initiates, particularly where the really significant ends are concerned. In order to ensure full and exact knowledge of these ends on the part of the initiates, the ceremonial stresses painstakingly accurate repetition and close guardianship. It is often designed to provide a strong emotional appeal, impressing the members with the gravity of the ceremonial occasion and the authenticity of the knowledge thereby revealed. In many secret societies the ceremonial is cast in dramatic form and contains episodes taken from holy books, revered legends, episodes thought to be of crucial historical importance, etc. Oftentimes members play parts enabling portrayal of the origin of the society, and in this portrayal the candidate for initiation usually has a key role. For instance, he may undergo a symbolic journey fraught with obstacles and temptations and at the end thereof receive the "truth" or esoteric wisdom viewed as the society's characteristic possession (*see Ceremonials*, below). In this process physical objects such as keys, pillars, swords, books, globes or staves may be endowed with symbolic meaning, so that their display on later occasions helps to reinstate, psychologically speaking, the awesomeness of the initiatory ceremony.

Many secret societies operate through a system of degrees of progressively higher rank in which secrets are revealed step by step. Initiation is therefore hierarchical; members at the higher levels are more fully aware of the ends pursued by the society than are those at the lower. Consequently, secrets of recognition are graded. That is to say, although there is ordinarily a grip, password, ceremonialized greeting in question and answer form, esoteric phrase, or secret jargon serving many of the purposes of a special language that distinguishes even the lowest initiate from nonmembers, the society has secrets within secrets. Those more fully initiated make every effort, by the use of special names, ordeals or revelations, to set themselves apart, on the one hand, and on the other to stimulate the lower ranks to the effort necessary to reach the exalted degrees.

The sedulous preservation of higher secrets serves several other purposes. For instance, beginning initiates are thereby impressed with the necessity for silence. Not only is this the case, but the art of remaining silent without giving offense to fellow members at lower levels is imparted by direct example. This is especially important when "final truth" and the real ends of the society are known only to those in the more advanced degrees, and even more so when, as in a few societies, the supreme leaders remain unknown to the rank and file membership. An essential technique

in all of this is that secrets remain unwritten, so far as possible; they must therefore be transmitted verbally in a sort of master-pupil situation. Frequently the transmission takes place under striking ceremonial conditions, reinforced by oaths of allegiance coupled with detailed specifications of dire punishment for traitors. In many modern secret societies such punishment seldom if ever occurs, but there have been instances of rigidly enforced discipline, especially in societies of subversive type—and some modern secret societies, in their early stages, were regarded as subversive.

The effects of secrecy on personality are many, but among them may be listed the growth of a sense of fusion, of a "mystic tie," induced by the sharing of secrets under the appropriate ceremonial circumstances. Further, the appropriately initiated person may effectively acquire norms or standards that extend or even substitute for the norms of the larger society of which he is apparently an integral part. Some secret societies, indeed, lay claim to the total personality of the fully devoted member, but this claim is virtually impossible of fulfillment.

Even though not fulfilled, however, the claim to the total personality means that sharp distinctions may be drawn between members and nonmembers, or in-groups and out-groups; some secret societies, even when their ends are not overtly subversive, may therefore operate in ways such that they tend to split larger societies. This being the case, supporters of various institutions within the larger society may become quite antagonistic to secret societies in general, resulting in accusations of overtly traitorous, heretical, immoral or similarly unworthy ends and their accompanying ceremonials. Political antagonism to secret societies has of course been much in evidence under totalitarian regimes—Nazi, Fascist and Communist—where all groups not controlled by the state are suppressed to the greatest possible extent.

Where secret societies effectively co-ordinate with the larger societies of which they are parts, as is often the case, the co-ordination is frequently linked with class affiliation. Secret societies recruited from the upper classes are more prone to support the existing social order than radically to challenge it; at most they aim at the "moral regeneration" of the larger society. Secret societies drawn from the ranks of the disaffected, however, are seldom free from subversive intentions and may become drastically revolutionary, heretical or even criminal, as the Mafia and the Molly Maguires (*qq.v.*).

An astonishing number of secret societies, when thoroughly investigated, can be shown to have ceremonials testifying to common origins or, at the very least, remote historical connections. At the same time, some secret societies bearing the same name and practicing the same ceremonial, in all essentials, show striking variations from one country to another (*see* ROSICRUCIANISM). (HD. BR.; J. K. RH.)

Ceremonials.—The following account of the ceremonials of the ancient Chinese Hung or Triad society by Freemason historian J. S. M. Ward indicates some of the similarities in the ceremonials of different groups.

The Hung society of China was founded, or perhaps reorganized, in A.D. 386 by the Buddhist patriarch Eon or Hwui-Yin, to spread the cult of Amitabha Buddha. Contemporary with the ancient mysteries and itself a great mystery rite, it survived over 1,500 years. The Hung rituals as they evolved showed a blending of Taoist-Buddhist ideas having curious analogies with the Egyptian *Book of the Dead*, and with certain "higher degrees" in western Freemasonry. The ceremony symbolized the journey of the soul through the Underworld and Paradise to the Holy City of the Gods, called the City of Willows, and interwoven with this was an allegory of the experiences of the mystic in his quest for union with the Supreme Being. As regards its analogies with Masonry, practically every important incident is found in certain higher degrees in England and America, while most of the hand signs are known to many Freemasons.

The ceremony comprised four sections. First the traditional history was given to the candidates in the anteroom before they entered the lodge. It was a moving story, wherein a body of monks who had helped the emperor were requited by him with the foulest treachery, all being murdered save five, who became the founders of the order. There were three villains, and for political purposes one was a Manchu emperor, either Khang Hsi, or, in some versions, his son, but originally the story was allegorical.

After this the candidates were "prepared" in the anteroom. The most notable incidents were (1) ceremonial washing and changing into white robes to symbolize not only mourning but that they themselves were dead; (2) the right arm, shoulder and breast, and also the left knee, were made bare; and (3) grass slippers were substituted for ordinary boots. Meanwhile the master opened and consecrated the lodge and invested his officers.

The third section dealt with the actual admission of the candidates, who had to pass through three gates inside the lodge and take the oath of blood brotherhood by mingling their blood with that of all members present in a cup of wine, from which each person present drank. (Women as well as men were eligible.)

The last section consisted of a catechism; the master asked a series of questions, which the conductor answered for the candidates. These revealed that they had been on a long and mysterious journey, first by land and then by boat, till they reached the City of Willows. Throughout the whole of this part of the ceremony great stress was laid on numbers, which had a definite mystical significance. The triangle also played an important part in the ritual, hence the name "Triad" society. The brotherhood had many aliases, the most famous being "The Society of Heaven and Earth." The significance of the ceremony was revealed by the opening questions:

> Master: Whence come you?
> Vanguard: From the East.
> Master: At what time?
> Vanguard: At sunrise, when the East was light.

See FREEMASONRY; *see* also references under "Secret Societies" in the Index.

See J. S. M. Ward and W. G. Stirling, ***The Hung Society***, 3 vol. (1925–26). (J. S. M. W.)

SECRET SOCIETIES, PRIMITIVE, a term loosely applied to a medley of associations (in nonliterate societies) which have little in common beyond an element of secrecy. It can be applied to the secret cabals of witches and warlocks, *e.g.*, the Mbatsav of the Tiv (*q.v.*) in Nigeria, which exist only in the minds of the more impressionable members of the community, or to such secret organizations of criminals or revolutionaries as the famed Leopard society of Sierra Leone or the Mau Mau of Kenya (*q.v.*) where the secrecy relates primarily to the membership of the association which is known only to the superior officers of the society. The most common reference, however, is to those societies or clubs in which membership conveys enhanced social status and the secrecy relates only to portions of the rituals. Secret societies of this sort form part of a wider category of associations which R. H. Lowie has called sodalities and which includes feasting clubs, dramatic and recreational societies, religious fraternities, professional unions of doctors, diviners and other specialists, initiation schools, associations of titled chiefs like the "palace societies" of Benin (*q.v.*) or the Ozo societies of the Ibo (*q.v.*) of Nigeria, associations of warriors like the military societies of the Plains Indians of North America and the head-hunting societies of west Africa or of southeastern Asia, and mutual aid societies which range from the Chinese tongs of Malaya to the tribal and patriotic unions of Nigeria and the native brotherhoods of North America.

The theories put forward by H. Webster, H. Schurtz and L. Frobenius about the origin and development of secret societies and other sodalities from associations based on the solidarity of persons of the same sex and age, and by W. H. R. Rivers and F. W. Butt-Thompson about their diffusion in Melanesia and west Africa were no longer accepted by anthropologists in the 1960s. No satisfactory classification or definitive terminology has been formulated because many of the varied functions these associations perform are complementary and interchangeable. Some ethnographers apply the term secret society to any associations which manifest some element of secrecy, while others prefer to

avoid it altogether; however, when the term is retained, usually it is limited to a more specific type of association whose primary function is the performance of rituals, the mysteries of which are known only to initiates. The society is represented to the uninitiated as an organization which can control a particular supernatural force or being, in most cases a nature spirit or spirits or the ancestral ghosts of the community. The well-being of the community is believed to be dependent on the propitiation of this supernatural power. The society thus is able to discipline the community in its name. Meetings, sometimes in secret places, may not be attended by the uninitiated under severe penalties, though they are allowed to hear the mysterious noises which may emanate from the participants. The society also provides periodic public spectacles of pageantry and dancing in which the spirits to which it ministers are represented by some of its members, usually with their identity concealed by masks or other disguises.

Secret societies of this type and the masks and other sculpture associated with them reached a very high degree of sophistication among the Northwest Coast American Indians and in parts of west and central Africa and of Melanesia (*see* MASK). The world-wide distribution of these societies cannot be explained in terms of diffusion from any particular centre; it is uneven and erratic, and even in such regions as west Africa where these societies are very prominent there are areas (notably in Ghana) where they are conspicuous by their absence.

Secret societies may be exclusive or inclusive in their membership. In some cases they are limited to a few ritual specialists and wealthy men, in others (*e.g.*, among the Kwakiutl of British Columbia) they extend to all adult and near-adult men and women who can claim to belong to the tribe by right of descent or marriage. In general, however, membership is confined to the adult men of the community (*e.g.*, the Ekpo society of the Ibibio) and although among the Mende and adjacent tribes of Sierra Leone and Liberia the women have their own secret societies (*e.g.*, the Bundu society), women and children are usually regarded as the uninitiated and profane section of the community. As boys approach manhood they are initiated into the sacred mysteries the men control.

In some communities the men are organized as a single society with a number of subdivisions or grades, according to the size of the community. The grading may be based purely on age, as among the Murngin of Australia where a man as he matures progresses through a succession of grades each with its own special mysteries into which he is initiated (*see* AGE SET). Grading also can be based on technical efficiency as in the Sekiapu society of the Kalabari of Nigeria, which performs a seven-year cycle of masked plays and grades members as those who are learning the plays, those who are performing the current cycle and those who have completed their cycle. Where wealth is unevenly distributed in a community, it may replace age as the essential criterion, and the gifts and feasts required for admission into the superior grades are sufficient to exclude all but the richest members of the society.

Where there is a single or dominant secret society it frequently acquires political functions, particularly in communities which lack strong chiefs and other centralized forms of government. Here the society forms a powerful focus of unity as it brings together in a single, self-disciplining association all the men of importance in the community without regard to their clan or lineage affiliations. It provides them with a forum for political discussion and negotiation of other business in a private, uncontroversial atmosphere freed from the tensions of interclan and intersectional rivalries which characterize the village councils and other public meetings. Thus such secret societies as the Ogboni of the Yoruba (*q.v.*) of Nigeria or the Harihu of the Elema tribes of the Papuan gulf, though they may not be represented in the formal government councils and moots of their community, are the real power behind them; they enforce the decisions of these councils. Because it can enforce its authority such a secret society also tends to become a judicial tribunal for the settlement of disputes and for the arrest, trial and punishment of criminals. So effective were some of these societies in their political and judicial functions that in some cases, notably in the west African port of Old Calabar in the 19th century, the Ekpe secret society had taken over all the functions of government. The two kings were the head and deputy head of the society and with other members of its highest grade (*nyamkpe*) formed the legislative and executive council of the community as well as its supreme court. The members of the second highest grade of Okpoho (Brass) Ekpe enforced obedience to their laws and decisions using as their agents and executioners junior members of the society disguised as the attendant spirits of Ekpe, the forest demon, to whom the society ministered.

While some communities have single, all-purpose societies of this type, others distribute political, ritual, educational and other public duties between a number of different societies, some of them secret as here defined. In some cases (*e.g.*, among some Yoruba, Ibibio, Ibo or Yakö communities of Nigeria) there is one, sometimes more than one, inclusive secret society into which almost all the men of the community are initiated and which is responsible for the masked plays and other public pageantry and dancing; there are also a number of more exclusive societies (each with special social duties) which are joined by the more socially ambitious men as they rise in the social scale; and there is one dominant and exclusive society with membership confined to the leading men of the community. In other cases no society is supreme and political; ritual and other authority is divided among a number of societies co-ordinated in a complex ceremonial organization. Among the American Indian tribes of the northwest coast the principal objective of these ceremonies is the validation of social status, in particular that of chiefs; among the Pueblo Indians (*q.v.*) of the southwestern U.S. it is the fertility of crops and the control of cosmic forces on which this fertility depends. Among the Santo Domingo Pueblo the ritual business of the community is distributed among 14 societies, 10 of them secret in the wider sense in that each has its own secret rituals relating to particular activities (one being concerned with war, one with hunting magic and the others with curing specific diseases) and 4 of them secret in the narrower sense that they are responsible for the kachina (ancestor) cult and the masked dances associated with it. All members of the Pueblo are grouped into two of these secret societies (Turquoise and Squash), each with its own kiva (underground lodge) and kachina cult into which only the men are initiated. Each kiva is associated with a more exclusive secret society named respectively Koshaire and Quirana after the particular ghosts or spirits to which it ministers. All these four societies are involved in private and public rituals designed to produce rain and to control the movements of the sun while the kachina rituals of the secret societies are intended to secure the intervention of the tribal ancestors for the same purpose. For a discussion of similar associations in less primitive cultures *see* SECRET SOCIETIES.

BIBLIOGRAPHY.—R. C. Abraham, *The Tiv People* (1940); K. J. Beatty, *Human Leopards* (1915); F. Boas, *The Social Organization and the Secret Societies of the Kwakiutl Indians* (1895); F. W. Butt-Thompson, *West African Secret Societies* (1929); R. H. Codrington, *The Melanesians* (1891); P. Drucker, *The Native Brotherhoods* (1958); F. R. Eggan, *Social Organization of the Western Pueblos* (1950); C. D. Forde (ed.), *Efik Traders of Old Calabar* (1957); K. L. Little, *The Mende of Sierra Leone* (1951), "The Role of the Secret Society in Cultural Specialization," in S. Ottenberg and P. Ottenberg (eds.), *Cultures and Societies of Africa* (1960); W. Lloyd Warner, *A Black Civilization: a Social Study of an Australian Tribe* (1937); R. H. Lowie, *Social Organization* (1948); W. H. R. Rivers, *The History of Melanesian Society* (1914); H. Schurtz, *Altersklassen und Männer-Bünde* (1902); C. G. Seligmann, *The Melanesians of British New Guinea* (1910); H. Webster, *Primitive Secret Societies* (1932); L. A. White, "The Pueblo of Santo Domingo, New Mexico," *Mem. Amer. Anthrop. Ass.*, vol. 43 (1935); J. L. Gibbs, "Poro Values and Courtroom Procedures in a Kpelle Chiefdom," *Southwestern Journal of Anthropology*, 18:341–50 (1962). (G. I. J.)

SECULAR GAMES were celebrated at Rome to mark the commencement of a new saeculum, or generation. It is important to note that there was a *saeculum civile*, the length of which was definitely fixed at 100 years, and a *saeculum naturale*, which, under Greek and Etruscan influence, came to be accepted by the quindecimviri as 110 years. According to tradition the secular games (*Ludi Saeculares*, originally *Terentini*) had their origin in

certain sacrificial rites of the gens Valeria, which were performed at the Terentum, a volcanic cleft in the Campus Martius. According to the Roman antiquarians themselves, they were derived from the Etruscans, who, at the end of a mean period of 100 years (as representing the longest human life in a generation), presented to the underworld deities an expiatory offering on behalf of the coming generation. The first definitely attested celebration of the games took place in 249 B.C., on which occasion a vow was made that they should be repeated every 100th year (their name being also changed to *Saeculares*), a regulation that seems to have been immediately disregarded, for they were next held in 146 (not 149, although the authorities are not unanimous); in 49 the civil wars prevented any celebration. They would probably have fallen entirely into oblivion had not Augustus revived them in 17 B.C. In explanation of the selection of this year, it is supposed that the quindecimviri invented celebrations for the years 456, 346, 236, 126, the saeculum being taken as lasting 110 years.

In later times various modes of reckoning were adopted. The dates were: A.D. 47 (under Claudius), celebrating the 800th year of the foundation of the city; 88 (under Domitian), an interval of only 105 instead of 110 years; 147 (under Antoninus Pius), the 900th year of the city; 204 (under Septimius Severus), exactly two saecula (220 years) after the Augustan celebration; 248 (under Philip the Arab), the 1,000th year of the city; 262 (under Gallienus), probably a special ceremony in time of calamity. In 304 (which should have been 314) Maximian intended to hold a celebration but does not appear to have done so. From this time nothing more is heard of the secular games until revived in 1300 as the papal jubilees instituted by Boniface VIII.

At the beginning of the harvest, heralds went around and summoned the people to the festival. The quindecimviri distributed to all free citizens on the Capitol and in the temple of Apollo on the Palatine various means of expiation—torches, sulfur and bitumen. There, and in the temple of Diana on the Aventine, wheat, barley and beans were distributed, to serve as an offering of first fruits.

The festival then began, at which offerings were made to various deities for three days and nights. On the first night the emperor sacrificed three rams to the Parcae at an underground altar on the banks of the Tiber, while the people lighted torches and sang a special hymn. On the same or following nights, a black boar and a black sow were sacrificed to Tellus, and dark victims to Dis (Pluto) and Proserpine. On the first day white bulls and a white cow were offered to Jupiter and Juno on the Capitol, after which scenic games were held in honour of Apollo and Diana.

On the second day noble matrons sang supplicatory hymns to Juno on the Capitol; on the third, white oxen were sacrificed to Apollo, and 27 boys and maidens sang the "secular hymn" in the temple dedicated to him by Augustus on the Palatine.

Originally the gods of the underworld had been the objects of the ceremony. The introduction of Apollo, Diana and Leto was attributable to Augustus, for whom they had become patron deities; hence their imposing position in the rites and in the *Carmen Saeculare* of Horace, written for the festival of 17 B.C. and still surviving.

BIBLIOGRAPHY.—The above particulars are from Zosimus (vol. ii, pp. 5–6, which contain the Sibylline oracle), who, with Censorinus (*De Die Natali*, 17), Valerius Maximus (vol. ii, p. 4) and Horace, is the chief ancient authority. The inscription commemorating the ludi of 17 B.C., discovered in 1890, relates in fragmentary form the details of the festival and preserves some of the expiatory prayers. It is printed in the *Ephemeris epigraphica*, vol. viii, with a full commentary by Mommsen. *See* also M. Warde-Fowler, *Roman Essays*, pp. 111–126 (1920); J. Gagé, *Recherches sur les jeux séculaires* (1934); H. Wagenvoort in *Studies in Roman Literature, Culture and Religion*, pp. 193–232 (1956); and the classical dictionaries and encyclopaedias. (T. V. B.)

SECULAR INSTITUTE, in the Roman Catholic Church, is a society whose members profess the evangelical counsels in the world in order to attain Christian perfection and the full exercise of the apostolate. Secular institutes were given papal approval in 1947 by the apostolic constitution of Pius XII, *Provida Mater Ecclesia*, further confirmation and explanation by his *motu proprio Primo Feliciter*, and juridical directives for establishment by the instruction *Cum Sanctissimus Dominus*. The approval of this new form of the Christian state of perfection was the result of a long historical development. The Gospel records the invitation to adopt permanently a more perfect way of life by the practice of poverty, chastity or celibacy, and obedience. The historical forms of responding to this vocation in societies ("institutes" in canon law) approved by the church and erected in the juridical "states of perfection" are: orders and congregations, in which the counsels are professed by public vows and life in community; societies of common life without public vows; and now secular institutes without public vows, required common life or distinctive garb. Members of the latter pursue personal Christian perfection and the apostolate while living "in and of the world." Subject to the Sacred Congregation of Religious, a secular institute may be clerical or lay or mixed, and may adopt a diocesan or interdiocesan form. Among the major secular institutes are Missionaries of the Kingship of Christ (priests, men, women); Opus Dei (priests, men, women); Caritas Christi (women); Society of the Heart of Jesus (priests); Our Lady of the Way (women). For a list of secular institutes of men, *see* ORDERS AND CONGREGATIONS, RELIGIOUS.

BIBLIOGRAPHY.—Jean Beyer, *Les Instituts séculiers* (1954); Salvador Canals Navarrete, *Secular Institutes and the State of Perfection* (1959); Joseph E. Haley (ed.), *Apostolic Sanctity in the World* (1957); *De Institutis saecularibus* (1951); J. M. Perrin, *Secular Institutes* (1961). (J. E. HA.)

SECUNDERABAD, a town and cantonment of Hyderabad district in Andhra Pradesh, India, now administratively forming part of Hyderabad city. The town was one of the largest cantonments in India during the British regime. Its area when handed over by the British on Dec. 1, 1945, was 3.6 sq.mi. but had more than doubled by the 1961 census, when its population was 187,471. There are some fine old buildings including St. John's church and St. Mary's cathedral (erected in 1847). St. Joseph's orphanage with a school, opened in 1856 and run by Loreto nuns until 1889, is managed by the nuns of St. Ann's convent. The town is still a stronghold of Christian missionary endeavour and education. The large and well-equipped King Edward VII Memorial hospital was renamed Mahatma Gandhi hospital. Founded in the name of Sikander Jah (Asaf Jah III) in 1806, Secunderabad is a junction on the Central railway and a centre of wholesale trade and commerce. *See* also HYDERABAD. (S. AH.)

SECURITIES REGULATION, a term commonly used to designate legislation designed to protect the investing public from misrepresentation in the sale of stocks, bonds, and other negotiable instruments.

Until the close of the 17th century, most industries and businesses were financed by the same persons who managed them. The Industrial Revolution, however, with its complex enterprises requiring more capital than could be contributed by few individuals, led to the formation of British and American companies for which necessary funds could be obtained only from the public. To protect investors against sharp practices used by some company promoters, legislation was enacted in both countries.

United States.—Although various statutory and constitutional provisions affecting sales of securities were enacted by the states beginning about 1850, Kansas, in 1911, adopted the first comprehensive "blue-sky law," so-called because it was directed at speculative schemes that "had no more basis than so many feet of blue sky." By 1917 the constitutionality of such regulation had been upheld by the U.S. Supreme Court, and by 1933 every state except Nevada had a blue-sky law.

State regulation is classifiable into three types, with most state systems constituting a combination of all three:

1. *Antifraud statutes.*—These statutes expand the usual definition of "fraud" in ordinary commercial transactions by declaring various unfair securities sales practices actionable by private suit, criminal penalty, or injunction. One important expansion is the inclusion of half-truths, promises without basis, and omissions or concealments of material facts within the definition of fraud in a securities transaction.

2. *Broker-dealer registration.*—Many states require registration and the filing of periodic reports by all brokers and dealers in securities, who must furnish in an application for registration evi-

dence of the financial responsibility and good repute of the applicant. Some states require that an independent audit of the registrant be made and reported annually. Most states outlaw "bucket shop" operations, whereby a broker anticipating a price decline would "bucket" a customer's purchase order instead of executing it and would thereby gamble on making a secret profit after purchasing at a lower price. The common requirement that specific confirmation slips be sent to all customers is aimed at "bucketing." Provisions in many states also require separate registration of each securities salesman.

3. *Securities registration.*—About 40 states have some requirement for registering each individual issuance of securities. Most states permit nonexempt securities—usually those of well-established companies traded on a national exchange or already registered under federal law—to be registered by mere notification to the state agency. Otherwise, securities must be the subject of a detailed registration statement disclosing the history and business of the issuer; names, addresses, and compensation of officers, directors, and large shareholders; the underwriting and promotional arrangements; and other information from which the security may be evaluated. Certified financial statements and the principal corporate documents governing legal organization and structure must be furnished with the registration statement. Many states also regulate the character and content of any prospectus or other advertising material used to solicit sales. Among the practices at which these provisions were aimed was the transfer of property, such as mining claims, patents, and other undeveloped assets, of speculative, unproved, or fictitious value, for securities having a face or market value equal to the inflated value of the assets. Such securities, sometimes called watered stock because not backed by solid asset value, would then be sold to the public without adequate disclosure of the background.

While many state statutes undertake no more than a full disclosure, on the theory that the government interest is limited to protecting investors against purchasing securities which they have no opportunity to evaluate, a few states, notably California, give the securities commissioner or other agencies power to prohibit the sale of securities deemed by the commissioner to be unfair or inequitable. These states usually issue rules indicating what promotional arrangements, in the form of compensation, share options or bonuses, will be considered so unduly favourable to the promoters and unfair to the public as to preclude sales within the state or require protective orders, such as holding promoters' shares in escrow until the company is successfully established.

Exemptions and exceptions in state statutes normally include small issues, variously defined as issues offered to less than 10 to 25 persons, or in the amount of $25,000, and for private offerings to institutional or other investors.

In 1956 a Uniform Securities Act containing all three types of state securities regulation was proposed by the National Conference of Commissioners on Uniform State Laws. It is a disclosure-type proposal and includes provisions whereby state regulations may be more closely coordinated with federal registration procedures. At least ten states adopted statutes modeled substantially after the Uniform Act in the first four years after its promulgation.

Federal securities regulation first came in the wake of the 1929–32 stock market decline. The Securities Act of 1933 required that all nonexempt securities be registered with the government and that the registration statement disclose all material facts about the issuer and the issuance. A prospectus, filed as part of the registration statement, must be furnished to all purchasers prior to sale or upon delivery of the securities. A proposed registration statement prepared by the issuer is reviewed by government lawyers and accountants, who frequently suggest revisions necessary to put it in form to be effective. No sale may be made until the statement is effective, and prospectuses may not be used except in the form approved by the government.

The principal exemptions from the registration process are for government securities, securities of charitable and similar nonprofit organizations, securities sold only intrastate, private offerings, and issuances of less than $300,000, for which a short-form notification procedure is provided. The small business exemption may, however, be withheld if there is inadequate disclosure of certain facts in the notification procedure.

Under the Securities Exchange Act of 1934, which created the Securities and Exchange Commission (SEC) to administer federal securities regulation, all securities exchanges must be registered with the SEC, and before shares are traded on the exchanges registration statements must be filed with the exchanges and the SEC. The act contains provisions governing a variety of practices, such as short-selling, floor trading, concerted buying and selling for the purpose of artificially manipulating prices, hypothecation of customer's accounts, and short-swing (within six months) "insider" trading in shares of a corporation by its officers, directors, and large shareholders. Brokers and dealers must register with the SEC, and supervision of their activities is carried out by the National Association of Securities Dealers, a private body which must be registered with the SEC and through which disciplinary action of its members is taken in accordance with SEC findings.

The 1934 act also gives the SEC broad rule-making powers, which it has exercised, to penalize fraud, as broadly defined in the statute and rules, in sales made by mail or in interstate commerce; to require that periodic reports be made to shareholders and filed with the SEC by corporations listed on exchanges; and to supervise the solicitation of proxies of securities listed on exchanges. Under its antifraud rules the SEC also undertakes to curb the activities of "boiler room operators," a term applied to organizations of securities salesmen using high-pressure sales tactics by telephone. The jurisdiction of the SEC over proxy solicitation, reporting, and "insider" trading as to listed companies was greatly expanded by the Securities Act Amendments of 1964 to include, after 1966, all corporations having assets of more than $1,000,000 and 500 or more shareholders.

Great Britain.—The regulation of securities in Great Britain began in the 19th century with enactment in 1844 of the British Companies Act requiring registration of company prospectuses. By mid-20th century it was largely governed by the Prevention of Frauds (Investment) Act, 1958, and the Companies Act, 1948. The Investment Act permits only members of a recognized stock exchange, dealers licensed by the board of trade, or exempted persons (such as banks, insurance companies, and issuing houses) to deal in securities. The activities of the former two are tightly regulated by the stock exchanges and the board of trade, respectively. The act also makes it a criminal offense to induce investment by the knowing or reckless use of misleading, false, or dishonest statements or the dishonest concealment of material facts. The act further prohibits the circulation of any written invitation to subscribe for, buy, or sell any securities, unless it is: (1) made through a permitted dealer; (2) approved by the board of trade; (3) made by a company only to its own shareholders, or (4) accompanied by a prospectus complying with the Companies Act.

The Companies Act requires every written invitation to the public to subscribe for or purchase shares made by a company, or by an issuing house which markets the issue, to be made in the form of a prospectus containing matters specified in the act. Copies of this prospectus must accompany all application forms sent to prospective investors. The prospectus must reveal full details of the promotion arrangements and the company's history and business, director, and promoters, capital structure, financial results for the past five years, certified by its auditors, and all important contracts entered into during the previous two years. Little, however, need be disclosed when the company makes a "rights issue" (in which existing shareholders are offered the opportunity to take up further shares), which is now the usual method used in England by an established company to raise additional equity capital.

An investor induced to subscribe for shares by misrepresentations in the prospectus may repudiate his subscription contract and claim the return of any money paid; or sue the directors for damages for deceit if he can prove fraud; or sue the promoters, directors, or the issuing house for compensation for any loss. Certain criminal penalties may also be imposed.

The prospectus must be filed with the registrar of companies, but this official does not check the accuracy of its contents nor does he comment on the merits of the company or of the investment. This absence of surveillance, however, has not generally resulted in abuses. The realities of corporate financing require, first, that the shares be traded on one of the stock exchanges, usually the London Stock Exchange, and, secondly, that they be offered to the public through an issuing house. Compliance with prospectus regulations of the stock exchanges, and the thorough investigation made by the issuing house of all aspects of the company and of the issue, together generally ensure that the public is invited to invest only in reasonably sound ventures. *See* also Stock Exchange.

Bibliography.—Laurence C. B. Gower, *Principles of Modern Company Law,* 2nd ed. (1957); Louis Loss, *Securities Regulation* (1951; supplement, 1955); Louis Loss and E. M. Cowett, *Blue Sky Law* (1958); Sir Francis Beaufort Palmer, *Company Law,* 20th ed. (1959). (G. T. Fr.)

SEDAINE, MICHEL JEAN (1719–1797), French dramatist who played an important part in the development of both light opera and drama, was born in Paris, June 2, 1719. Sedaine, a mason's labourer turned architect, was self-educated. He began by writing songs and poems, *Poésies fugitives* (1752), then turned to light opera with *Le Diable à quatre* (1756). He collaborated first with the musician François Philidor in *Blaise le savatier* (1759), etc., then in a more serious vein with Pierre Monsigny, gaining a signal success with *Rose et Colas* (1764). Sedaine's collaboration with André Grétry in *Aucassin et Nicolette* (1779; Paris, 1780), *Amphitryon* (1786; Paris, 1788) and *Guillaume Tell* (1791) brought light opera nearer still to opera and serious drama. Their best work, *Richard-Coeur-de-Lion* (1784) ensured Sedaine's election to the Académie Francaise (1786). Sedaine's *Le Philosophe sans le savoir* (1765) is important in dramatic history because, though entitled a "comedy," it is the best example of the *drame bourgeois,* which Diderot had expounded some years earlier. The play defends the businessman who is also a man of honour like any noble. *La Gageure imprévue,* a lively one-act play, followed in 1768, but censorship held up the performance of *Maillard,* a prose tragedy, and *Raymond V,* until 1788 and 1789. Sedaine died in Paris, May 17, 1797.

See L. Gunther, *L'Oeuvre dramatique de Sedaine* (1908); L. P. Arnoldson, *Sedaine et les musiciens de son temps* (1934). (D. Ks.)

SEDALIA, a city of west-central Missouri, U.S., about midway between Kansas City and Jefferson City; the seat of Pettis county. It was established in 1860 and was chartered in 1889. It became the western terminus of the Missouri Pacific railroad until after the American Civil War and was an important railhead for the Texas cattle drive of 1866. Later the Missouri-Kansas-Texas railroad built through Sedalia and large railroad shops became its most important industry. The city is a thriving agricultural shipping and distributing centre. Industries are diversified, including flour mills, ice and poultry-packing plants, and firms manufacturing glass blocks, trailers and prefabricated houses. Whiteman air force base is nearby. Missouri's state fair is held in Sedalia annually in August. During the Civil War the city was a Union post except for a one-day raid by Sterling Price. Generals Nathaniel Lyon and John C. Frémont outfitted forces there in 1861. For comparative population figures *see* table in Missouri: *Population.* (Jo. L. H.)

SEDAN, a town of northeastern France in the *département* of Ardennes, lies on the Meuse river 12 mi. (19 km.) by rail E.S.E. of Mézières. Pop. (1962) 20,261.

The old town is situated on the right bank of the river and in spite of many wars still has some old buildings, including a 15th-century castle locally claimed as the biggest in Europe. The Place Turenne contains a statue of the great soldier born in Sedan, the marshal de Turenne. South of it is the Place Goulden with its gateway separating the old town from the later additions, and the Avenue du Général Leclerc lined entirely with buildings erected since World War II. Other features are the town's oldest building (1 Rue du Ménil) which houses the museum, St. Charles' church (begun in 1593, severely damaged in 1940), the Dijonval or old royal serge factory and the botanical gardens. Across the river is the industrial suburb of Torcy.

Sedan has woolen, metallurgical, chemical and electrical industries; a factory making steel pipes employs about 1,700 workers. There are good road and rail connections and heavy freight is carried on the Meuse.

In the 14th century Sedan was a dependency of the abbey of Mouzon. United to the French crown by Charles V, it was ceded by Charles VI to Guillaume de Braquemont. Guillaume's son Louis sold it in 1424 to his brother-in-law, Évrard II de La Marck (d. 1440), who began the fortification of the town. Robert I, lord of Sedan (d. 1487), became chatelain of Bouillon in 1482. Robert IV (d. 1556), marshal of France, styled himself duc de Bouillon and was recognized as sovereign in Sedan. His son Henri Robert (d. 1574) was designated prince of Sedan. He publicly professed Calvinism and henceforth Sedan became a Protestant centre. Charlotte, the last La Marck sovereign of Sedan, married, in 1591, Henri de La Tour d'Auvergne, vicomte de Turenne (1555–1623). He and his successor, Frédéric Maurice (1605–52), conspired against the French government, to which Sedan was finally lost in 1642. (*See* also Bouillon.)

Battle of Sedan, 1870.—The battle of Sept. 1, 1870, in the Franco-German War, led to the overthrow of the French second empire. The French army under Marshal M. E. de MacMahon should have retreated toward Metz to relieve the other French army there under Marshal Achille Bazaine. On Aug. 31 MacMahon assembled his forces in and around Sedan to face the German 3rd and Meuse armies under the princes Frederick William of Prussia and Albert of Saxony, who approached from opposite sides to carry out an encircling movement, successfully cutting off the French lines of retreat.

In the course of the battle MacMahon was seriously wounded, whereupon confusion ensued as to the right of Gen. A. A. Ducrot or Gen. E. F. de Wimpffen to assume command. The German gunfire broke up the ground and inflicted enormous losses on the French cavalry under Gen. J. A. Margueritte who was killed in the fight. In the afternoon of Sept. 1 the Germans launched their main attack from east and west. The emperor Napoleon III realized that the position was then hopeless and surrendered together with nearly 82,000 officers and men.

Grave miscalculation of German strength led to the total collapse of French military power. At Sedan 17,000 French were killed or wounded. German losses were less than 9,000. Meanwhile, the other French army surrendered Metz and in Paris a provisional government of national defense was formed to face the imminent German siege. (*See* also Franco-German War.)

Battle of Sedan, 1940.—The battle of May 13, 1940, inaugurated the German invasion of France in World War II. Between the two world wars France had built the Maginot line along its eastern frontier. But the area around Sedan and the Belgian frontier was left poorly defended. In 1939 the 1st army group, under Gen. G. H. G. Billotte, was placed along the eastern line in accordance with the French Plan D, which assumed that the Germans would attack through Belgium and the northern section of the frontier. The French 2nd army under Gen. Charles Huntziger held the stretch from Longwy to Sedan.

The German high command, having learned of this plan, decided, however, to strike at three points along a 50-mi. line between Dinant and Sedan, the main blow falling on Sedan. Armoured divisions, under Gen. Paul von Kleist, reached the Belgian Ardennes through Luxembourg on May 10, pushing back the old-fashioned French cavalry. The German advance continued past Bouillon and by the evening of May 12, Gen. Heinz Guderian's 1st panzer division had reached the outskirts of Sedan.

On May 13, under cover of dive bomber attack, the Germans effected crossings north and south of Sedan. By evening the breakthrough was complete and Sedan was destroyed. Huntziger held out south of Sedan, but meanwhile Guderian had routed Gen. André Corap's forces to the west and the two other German thrusts at Houx and Monthermé farther north had succeeded. The rest of Plan D became unworkable. In June, the Franco-German armistice was signed. *See* also World War II: *The Western Front, First Phase.*

SEDBERGH, a small, ancient market town of the West Riding of Yorkshire, Eng., close to the Westmorland border, is 11 mi. (18 km.) E of Kendal in the basin formed by the junction of Dentdale and Garsdale with Lunedale. It lies under high fells within the Yorkshire Dales National Park and gives its Norse name to a rural district of three parishes and a deanery. Formerly an important stage on the roads through the Pennines, it is now popular for its majestic scenery. The local interests are agriculture, grazing, and cheese making. Sedbergh School, founded in 1525 by a native, Roger Lupton, canon of Windsor and provost of Eton, was disendowed under the Chantries Act, 1548, and refounded in 1552. Its constitution was revised in 1874 and the buildings greatly extended. (C. K. C. A.)

SEDDON, RICHARD JOHN (1845–1906), New Zealand statesman, Liberal prime minister from 1893 to 1906, was born at Eccleston, Lancashire, on June 22, 1845, the son of a schoolmaster. Brought up in the engineering trade, he emigrated to Australia in 1863, spent a short luckless period on the Bendigo goldfields, then joined the railway workshops at Williamstown near Melbourne. Early in 1866 he went to New Zealand, was moderately successful at goldmining on the west coast, and put the money he made into storekeeping. At the same time he became prominent and popular in local politics. Standing as a radical, he was elected to Parliament in 1879. He was a member continuously for west coast constituencies till his death. Verbose and tedious as a speaker, Seddon's energy and mastery of procedure made him notable, nevertheless, and when the Liberals, led by John Ballance, came into power in December 1890, he was given the important portfolio of mines and public works. When Ballance died (April 1893) Seddon was determined to have the succession, and got it; and forthwith began the masterful process that firmly established his power in the country. Powerful in body and voice, forceful, tireless, shrewd in his appreciation of men, a good administrator and a consummate politician, astute without being subtle, unable to tolerate equals around him, able not merely to jump on a bandwagon but to make the bandwagon his, with a rough geniality that went straight to the heart of the electorate, it was little wonder that long before his death he was known as "King Dick." Under his leadership the Liberal Party put through the great mass of social legislation that made New Zealand for a time an object of interest all over the world, though the measures were generally originated by others, such as William Pember Reeves (acts concerning industrial matters), John McKenzie (land), or William Hall-Jones (old-age pensions). In the imperial sphere, at home or abroad, Seddon was active and outspoken, and under him the Cook Islands were annexed to New Zealand (1901). After a last great electoral victory (December 1905) he died suddenly at sea on June 10, 1906, while returning to New Zealand from a visit to Sydney.

See R. M. Burdon, *King Dick* (1955). (J. C. Be.)

SEDERHOLM, JAKOB JOHANNES (1863–1934), Finnish geologist and an authority on Precambrian time, was born in Helsinki on July 20, 1863. He was educated at Helsinki, Stockholm, and Heidelberg. He was appointed geologist at the Geological Commission of Finland in 1888, was promoted to director in 1893 and retained this position until his retirement in 1933.

Sederholm made two major contributions: First, he vigorously promoted and directed the commission which produced numerous first-rate maps, reports, and scientific treatises. Second, he pioneered the investigation of the Precambrian rocks of Finland, studying especially the origin of granites and gneisses, and the stratigraphy, time classification and correlation of the rocks. He traveled in Europe, North America, and Asia and published many papers. He died at Helsinki on June 26, 1934. (Er. An.)

SEDGE, a grasslike plant of the family Cyperaceae (*q.v.*), especially any of the large genus *Carex*, which contains perennial species of temperate and subarctic zones, frequently growing in damp, boggy, marshy, or stream habitats.

SEDGWICK, ADAM (1785–1873), English geologist who introduced the terms Cambrian system and Devonian system (*qq.v.*), was born on March 22, 1785, at Dent vicarage, Yorkshire. He was educated at the grammar schools of Dent and Sedbergh, and at Trinity College, Cambridge, where in 1810 he was elected a fellow. In 1818 he was ordained and succeeded John Hailstone as professor of geology. He was elected president of the Geological Society in 1829–30, and in 1831 he commenced fieldwork in North Wales. This led to Sedgwick's application of the name Cambrian to the oldest group of fossiliferous strata. With Sir Roderick Impey Murchison he made a special study of the folded and faulted sandstones and shales of Devonshire known as the Culm Measures, showing that they belong to the age of the true Coal Measures (Carboniferous). Sedgwick and Murchison also introduced the geological term Devonian. These researches were published in the great memoir "On the Physical Structure of Devonshire," *Trans. Geol. Soc.* series 2, vol. 5 (1839). (*See* also ORDOVICIAN SYSTEM.) Among other works *A Synopsis of the Classification of the British Palaeozoic Rocks* (1855) may be mentioned. Sedgwick died at Cambridge on Jan. 27, 1873. The Sedgwick Museum at Cambridge (opened 1903) was built as a memorial to him.

See J. W. Clark and T. McKenny Hughes, *Life and Letters* (1890). (C. E. T.)

SEDGWICK, ADAM (1854–1913), English zoologist, best known for his researches on the wormlike *Peripatus*, was born at Norwich on Sept. 28, 1854. A grand-nephew of the geologist Adam Sedgwick (*q.v.*), he entered Trinity college, Cambridge, in 1874, there becoming attracted to zoology. He was one of the group of students inspired by Francis Maitland Balfour, leader of scientific research in comparative anatomy and embryology. In 1878 Sedgwick became Balfour's lecturer and in 1882, when Balfour was killed in the Alps, took charge of his young school of morphologists. Sedgwick's greatest researches were concerned with that remarkable animal *Peripatus* (*see* ONYCHOPHORA), an ancient link between the segmented worms and the insects. Later work confirmed Sedgwick's brilliant demonstration that the development of this animal holds the key to the relationships of several major divisions of the animal kingdom. From 1897 to 1909 Sedgwick devoted himself to his outstanding *Text-Book of Zoology*, and its publication marked the close of the great era of evolutionary morphology which began after 1859. In 1907 he succeeded Alfred Newton as professor of zoology at Cambridge, and two years later became professor of zoology in the new Imperial College of Science and Technology at South Kensington, London. He died in London on Feb. 27, 1913, after a long-standing pulmonary ailment. (C. F. A. P.)

SEDIMENTARY ROCKS. Although the term rock is commonly understood to be something durable and unyielding, and is so understood even in engineering circles, in geology it denotes only a body of solid materials forming a significant part of the crust of the earth. Included here are the unconsolidated clays and sands as well as the truly consolidated, or lithified, materials. Unconsolidated sediments, however, are not just soil. Soil is a somewhat ambiguous term and is usually applied to the very thin surficial materials, generally unconsolidated or only weakly indurated, as opposed to the bedrock upon which these materials rest and from which, in many cases, they were derived. Soil may be considered as one variety of sedimentary rock—and indeed has been called the mantle rock (saprolith or saprolite if it was produced by weathering *in situ* of the bedrock).

The sedimentary deposits form an important class of rocks. They are defined as those rocks formed by sedimentation—a process of deposition of mineral matter from (fluid) suspension or from solution at the relatively low temperatures and pressures which prevail at or near the surface of the earth. The systematic study of the properties and origins of these materials is the science of sedimentology.

This article deals with the history of sedimentology and its applications in exploration for gas and oil, the volume, origin and classification of the sedimentary rocks and the process of sedimentation. In addition to the cross references given under the various headings of this article, for explanation of the geological terms and concepts referred to *see* GEOLOGY; MINERALOGY. Modes and locations of occurrences and distinctive characteristics of various accumulations of sedimentary rocks are dealt with

in separate articles on geologic systems and epochs, as CAMBRIAN SYSTEM; PLEISTOCENE EPOCH; etc. *See* also *Geology* sections and sections dealing with physiographic regions of articles on countries and states, as CANADA; CONNECTICUT; etc.

SEDIMENTOLOGY

Petrology is concerned with the properties, classification and origin of rocks. Sedimentary petrology is that branch of petrology which deals with rocks of sedimentary origin. Some investigators have applied the term sedimentation to the body of knowledge, both observational and theoretical, concerned with the origin of sediments and sedimentary deposits (and their lithified equivalents). As the term, in the strict sense, only applies to a process, other workers have rejected it and used the term sedimentology for the science of sediments. As so used, it is more or less equivalent to the term sedimentary petrology but has tended to replace both sedimentation and sedimentary petrology.

History of Sedimentology.—Primitive man knew and made use of many sedimentary materials: flints for his spears, clay for his pottery. He soon learned where to look for these things and as his technology evolved he made many observations on and applied terms to these materials of sedimentary origin. Many of these terms derived from the common language remain and have been more precisely defined and incorporated into scientific nomenclature. Others have become obsolete and discarded. Such terms as cobble, sand, flint and limestone are among those retained; terms like hornstone or pudding stone are no longer used.

Geology had its real beginnings as a science in the early part of the 19th century. A significant event in its development was the publication in 1815 of the geological map of England by William Smith, a self-taught land surveyor and engineer. Smith's map embodied the first real attempt to portray the distribution and relative ages of the sedimentary strata of a region. Smith's contribution was in the establishment of the stratigraphic order and the discovery of the usefulness of fossils in stratigraphic correlation. The first scientific work on sedimentary rocks was essentially stratigraphic—namely the distribution of the sedimentary rocks in space and time. Field studies were directed, therefore, toward determining the order of superposition and the gross geometry of the sedimentary bodies—their thickness and lateral extent.

The science of sedimentology thus had its beginning in the science of stratigraphy to which it is closely related (*see* also STRATIFICATION). The central problem in stratigraphy is the establishment of the temporal order of deposition; the central problem in sedimentology is to determine the origin of the deposit. The early stratigraphers, however, made many observations on the character of the sedimentary strata with which they were concerned. The collective knowledge thus obtained became embodied in many scientific papers and reports. Such works as A. W. Grabau's *Principles of Stratigraphy*, published in 1913, contains an extended summary of such knowledge.

The justly famous "Challenger" expeditions of 1872–76 did much to put the study of the oceans on a scientific basis—did in fact establish the science of oceanography. The data collected by this expedition which appeared in 1880–95 in the monographic *"Challenger" Reports* by Sir John Murray and A. F. Renard shed much light on marine geology and especially on marine sedimentation. The impetus which this work gave to the study of modern marine sediments has continued. The influence of the *"Challenger" Reports* on speculations concerning the origin of the ancient sediments cannot be overestimated.

The science of sedimentary petrology or sedimentology can be considered to have been established as a separate geological discipline with the publications of two papers by Henry Clifton Sorby. Sorby's paper on the structure and origin of limestones which was delivered as a presidential address before the Geological Society of London in 1879 is a paper of such importance and merit that it can still be read with profit. A second paper "On the Structure and Origin of the Non-calcareous Stratified Rocks" which appeared in the *Proceedings* of the Geological Society of London the following year is a companion study of equal rank. Sorby can justly be called the father of sedimentary petrology. The last paper published by him, in 1908 at the age of 82, "On the Application of Quantitative Methods to the Study of the Structures and History of Rocks" foreshadowed work which was to come several decades later.

Sorby is best known for his early use, perhaps the first use, of the polarizing or "petrographic" microscope for the study of rocks and minerals in thin sections (*see* PETROLOGY: *Methods of Investigation*). This method of study, now routine for petrographers, was applied by Sorby to both minerals and rocks, especially the sedimentary rocks. But it remained for Lucien Cayeux (1864–1944), the French petrographer, to carry the microscopic study of the sedimentary rocks forward to maturity. Cayeux's monographs on the sedimentary rocks of France, especially those on the siliceous and the calcareous rocks, have never been excelled. Likewise his earlier studies of the ironstones and phosphates of France are without peers.

The study of sediments and sedimentary rocks entered a new era with the publication, in 1919, of "A Field and Laboratory Study of Cobble Abrasion," by C. K. Wentworth. This study, a master's thesis submitted at the University of Iowa, inaugurated an era of quantitative studies and laboratory experimentation. Except perhaps for the early work of Gabriel Auguste Daubrée (*q.v.*), the great French experimenter, sedimentary geology was largely an observational field science. Even the laboratory studies of Cayeux and others were primarily observational and nonexperimental. Wentworth's contribution was to formulate operational definitions for specific properties of clastic materials (*i.e.*, rocks made up of fragments of pre-existing rocks), such as roundness of cobbles, and devise methods whereby these properties could be measured. It became possible, then, to study the effects of abrasion in laboratory mills or "tumbling barrels" as well as to make a quantitative analysis of the action of streams and waves in the field. Wentworth's original paper was followed by similar quantitative field and laboratory studies by him and others. As a consequence of the collection of numerous measurements, the need for statistical summarization of these data became acute.

Although Wentworth himself attempted to utilize statistical methods, the first generally successful use of statistics in sedimentology seems to have been that of Parker D. Trask who in 1930 devised simple statistical parameters to characterize grain-size distributions in clastic sediments. The making of grain-size analyses of sediments—another quantitative procedure—had been done, however, as early as 1899 by J. A. Udden of Augustana college, Rock Island, Ill.

The study of sediments and sedimentary rocks was greatly accelerated by the establishment of a National Research Council Committee on Sedimentation in 1920. Under the chairmanship of W. H. Twenhofel of the University of Wisconsin, this committee prepared the well-known *Treatise on Sedimentation*, published in 1926, revised and republished in 1932.

Exploration for Oil and Gas.—Much of the interest in the sedimentary rocks and the accelerated development of sedimentology as a science has arisen because of the close relations between sedimentation and the formation of oil and gas. To promote the use of microfossils and heavy minerals in stratigraphy and the exploration for oil and gas (*see* also FORAMINIFERA: *Role in Petroleum Geology;* PETROLEUM), the Society of Economic Palaeontologists and Mineralogists was established in 1927. This society founded the *Journal of Sedimentary Petrology* in 1931, and this journal has been a popular outlet for researches on sediments and sedimentary petrology since that date.

Although oil and gas are caught in natural traps—mainly structural—it has been shown that their occurrence within a sedimentary basin can be better understood from studies of the distribution, both in space and time, of the various sedimentary lithologies (systems of rock) within the basin. Such studies are promoted by the preparation of lithofacies maps using data obtained both from outcrops and subsurface borings—large numbers of which have been made in exploration for gas and oil. The data so obtained, such as the number of sand beds, net sand thickness, and sand-shale ratio, can be plotted and contoured. Facies mapping of this type had to wait until adequate subsurface data from borings were avail-

able; as a result interest in the subject did not develop until after World War II.

The facies symposium, sponsored by the Geological Society of America, the proceedings of which were published as a *Memoir* of the society in 1949, did much to stimulate interest in the quantitative study of facies.

In 1938 Hans Cloos, of Bonn university, Ger., and his students began mapping of the primary sedimentary structures—cross-bedding, ripple marks, oriented fossils and the like—which record the direction of current flow during the deposition of the Devonian strata of the Rhine valley. Similar studies of cross-bedded sandstones in Arizona and in England were in progress at about the same time. Such paleocurrent studies combined with facies studies have proved not only important in the search for petroleum, but also are an aid to paleogeographic reconstruction—location of axis and trend of the ancient sedimentary basins, location of margins of basins and probable sources of the sedimentary fill, location and trend of the shore lines, etc.

The mapping of facies and directional properties of sediments has led to fusion of field and laboratory studies of sedimentary rocks and to a new phase in the development of sedimentology. This development has been furthered greatly by the study of the sedimentary framework—location and orientation of sand bodies—in modern sedimentary basins, most notably the delta of the Mississippi and adjacent gulf coastal areas.

International interest in sedimentology culminated in the calling of the first International Congress of Sedimentology in 1946 in Belgium. With the appearance of the *Treatise on Sedimentation*, various other reference and textbooks (see *Bibliography*), the establishment of a separate journal and a sponsoring society, and an international congress, sedimentology has come of age as a geological subscience.

The volume of research, represented by published papers, theses and courses taught in the universities, has grown very rapidly since the early 1930s.

THE SEDIMENTARY ROCKS

Because of the fossils entombed in them, the sedimentary rocks contain the whole of the life-record of the past. And by means of their fossil content they can be arranged in a chronological order. Moreover they also carry the record of past climate and past geography and, therefore, are of the greatest importance in unravelling the geologic and life history of our planet.

In addition, 85% to 90% of the mineral wealth of the world is derived from the sedimentary deposits or rocks. They are the reservoirs for petroleum, natural gas, and ground water. Indeed the oil and gas are as much a product of sedimentary processes as are the rocks in which they are found.

Coal is likewise a sedimentary rock and, together with petroleum and natural gas, supplies most of the energy for the industrial machine as well as raw materials for the chemical industries. Even the ores of many metals are of sedimentary origin. The principal ores of iron, aluminum, manganese and magnesium and some copper ores are of such origin. The alluvial deposits are, of course, sedimentary and many ores of tin, gold and platinum as well as gemstones occur as alluvial deposits. Most nonmetallic materials, such as building stone, the mineral fertilizers, the raw materials of the ceramic and cement industries, and rock salt, are of sedimentary origin. Clearly an understanding of these deposits, their discovery and their exploitation is basically a problem in sedimentary geology.

Volume.—The total volume of sediments, both ancient and modern, is very large. The thickness of sedimentary rocks varies greatly; locally accumulations in excess of 60,000 ft. are known. The average thickness on the continents, however, is much less and has been estimated to be about 7,000 ft. The thickness of the sediments over the floor of the ocean is highly uncertain and probably quite variable. The total volume of the sedimentary materials has not been measured and can only be estimated from geochemical calculations. If the sodium in the sea is derived from weathering of igneous rocks, at least 50,000,000 cu.mi. of average igneous rocks would be needed to furnish the quantity of sodium now contained in the oceans. But the volume of sediment released during this weathering, owing to porosity and to volume increases due to hydration and oxidation, would be somewhat greater. Inasmuch as not all the sodium released has been stored in the oceans, the total volume of igneous rock weathered (and hence the total volume of sediment produced) has been estimated to be in excess of 85,000,000 cu.mi. Some estimates, based on other considerations, suggest that the volume of sedimentary materials may be even greater. Even so, the sedimentary rocks form only a relatively small part of the outermost ten-mile shell of the earth. Of this shell, commonly called the "crust" of the earth, only 5% is of sedimentary origin.

Although there are many kinds of sedimentary rocks, some 99% can be classed as some type of sandstone, shale or limestone. Estimates based on measurement of exposed sedimentary sequences assign from 44% to 56% of the total to shale; from 14% to 37% to sandstone and the remaining 19% to 29% to limestone. The proportions of average shale, sandstone and limestone required to make the average igneous rock (from which they are presumed to have been derived by weathering) are computed to be 70%–82% shale, 12%–16% sandstone and 5%–14% limestone. The discrepancies between the observed and computed proportions are believed to be due to loss of the finest materials (which would form shale) to the deep oceanic basins.

Origin and Classifications.—Close examination of many sedimentary rocks shows that they consist of a detrital framework fraction that can only be debris derived from pre-existing rocks. This framework consists of close-packed, worn and rounded detrital elements: pebbles or cobbles of various rocks in the case of the gravels and conglomerates, and sand grains, most commonly single mineral grains of quartz or feldspar, in the case of the sands and sandstones (*see* CONGLOMERATE; SANDSTONE).

Some such deposits consist exclusively of one kind of detritus; others contain detrital material of diverse kinds. The lithified rocks, such as sandstone and conglomerate, contain a mineral cement which fills or partly fills the voids or interstices in the framework. This cementing mineral matter must be introduced therefore, after the deposition of the detrital elements and be the cause of the lithification of the rock as a whole. Commonly cementing materials are silica (quartz) and calcium carbonate (calcite). The total volume of cement may be as much as one-third or more of the whole rock.

The sedimentary materials described above are clearly the waste products of the disintegration of earlier rocks. They have been called the clastic (=broken) sedimentary rocks in recognition of the fragmental character of their detrital elements. The origin of these materials is not difficult to surmise. Remove the cement from a sandstone and one obtains a sand identical in all details with that found in streams or on beaches. The stream sands are clearly seen to form by erosion of the weathered residues of various older or pre-existing rocks. They are, however, washed residues markedly sorted and thus enriched in quartz and impoverished in the clay materials.

Other sedimentary rocks are not so formed. Some consist of a more or less cemented accumulation of shell or other skeletal debris and form a veritable fossil "hash." Such rocks must have originated from the accumulation of the skeletal hard parts of invertebrate organisms, either *in situ* or by current transport without introduction or addition of other mineral matter. Such accumulations can be found today in the coral reef areas and elsewhere. As these materials are mainly calcium carbonate, their accumulation leads to the formation of limestones and related rocks (*see* LIMESTONE; CHALK). The cliffs of the English channel are cut in thick and extensive deposits of chalk of Cretaceous age. The skeletal materials of which the chalk is composed are mainly of microscopic dimensions similar in character to those forming the calcareous muds found today in some places on the sea floor. Siliceous skeletal debris, though less common, may also form extensive and thick deposits (*see* DIATOMACEOUS EARTH).

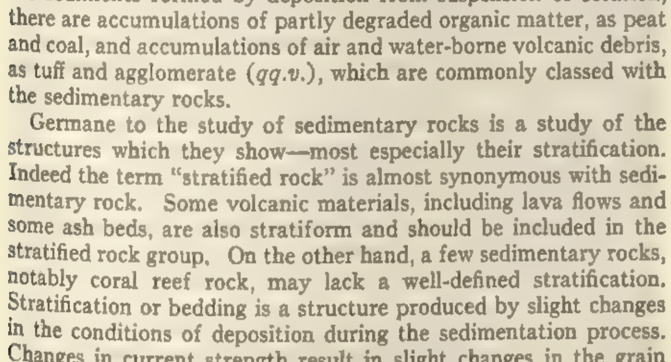
The shales appear to be the finest waste products of weathering and erosion. Under the microscope they are seen to consist of a mixture of silt particles and clay materials. The silt, like sand,

is the unaltered residue of weathering or abrasion; the clay is mainly the product of the decomposition of the feldspars of the parent rocks from which the detritus came. Not uncommonly to this silt-clay mixture is added skeletal and other biochemically precipitated carbonate and silica as well as organic materials. The hybrid sediment thus produced, upon compaction, becomes mudstone or shale (*q.v.*).

The sedimentary rocks are, therefore, accumulated sediments. Their constituents are of varied nature and their composition depends on the relative proportions of the materials of the diverse origins (*see* figure). Those consisting mainly of the products of abrasion ("rock flour") and the washed residues of weathering (sand, silt and clay) are the clastic sediments. Those consisting mainly of the biochemically or chemically precipitated materials (calcium carbonate and silica) are nonclastic sediments. Not uncommonly the sediment has a multiple origin and is in truth a hybrid deposit. Very commonly also the original composition is modified by reaction with materials in solution in the medium from which it was deposited or with those in solution in the ground waters with which it may later come in contact. The postdepositional alterations which result from such reactions with dissolved materials may lead to cementation of the sediment or drastic modification of its composition (as in the conversion of limestone to dolomite). The postdepositional changes are collectively considered diagenetic changes (*see* DIAGENESIS). Other products of diagenesis include concretions, stylolites, nodules and petrefactions.

A few rare sediments appear to be precipitated wholly from solutions. Thus formed are the extensive evaporitic sediments—the products of an evaporating brine (*see* GYPSUM; SALT: *Rock Salt*). Other sediments appear to be precipitated by reactions in more dilute solutions, as phosphorite (*q.v.*) and chert. In addition to the sediments formed by deposition from suspension or solution, there are accumulations of partly degraded organic matter, as peat and coal, and accumulations of air and water-borne volcanic debris, as tuff and agglomerate (*qq.v.*), which are commonly classed with the sedimentary rocks.

Germane to the study of sedimentary rocks is a study of the structures which they show—most especially their stratification. Indeed the term "stratified rock" is almost synonymous with sedimentary rock. Some volcanic materials, including lava flows and some ash beds, are also stratiform and should be included in the stratified rock group. On the other hand, a few sedimentary rocks, notably coral reef rock, may lack a well-defined stratification. Stratification or bedding is a structure produced by slight changes in the conditions of deposition during the sedimentation process. Changes in current strength result in slight changes in the grain sizes transported and deposited; changes in the chemical or physical environment may result in slight to marked changes in composition in the sediments forming by precipitation. Depending on the relation between the frequency of these changes and the rate of sedimentation, the stratification planes or surfaces of discontinuity may be close together or widely spaced. Sediments accordingly exhibit thin bedding or thick bedding. If the changes responsible for bedding planes are seasonally induced, the strata are then annual layers or varves (*see* VARVE ANALYSIS).



FROM F. J. PETTIJOHN, "SEDIMENTARY ROCKS"; REPRODUCED BY PERMISSION OF HARPER & BROS.

Though there are minor exceptions, the planes of stratification are initially horizontal, or substantially so. The sedimentary rocks in many places, notably in the mountain tracts, are no longer horizontal but are tilted at various angles, or even vertical or overturned (*see* FOLD). The recognition of these deformed stratified rocks as the ancient equivalents of sediments now being deposited in horizontal position, is one of the major achievements of geology for it leads to the conclusion that vast elevation and deformation of the earth's crust has taken place.

Stratification is not to be confused with rock cleavage which is a structure that commonly transects or cuts the true bedding and is a structure, therefore, that is postdepositional in origin. Another structure, also commonly confused with bedding is colour-banding—a phenomenon most characteristic of sandstones—which consists of rhythmically spaced precipitation surfaces of reddish iron oxide. These bands commonly, but not everywhere, transect the bedding.

SEDIMENTATION

Although the term sedimentation has been used as a synonym for sedimentary petrology and sedimentology, it is more properly a process. It is the process of deposition of a solid material from a fluid—usually air or water—from a state of suspension or solution. Broadly defined it would also include deposits from glacial ice and those materials collected by gravity alone—as in talus deposits, or accumulations of rock debris at the base of cliffs.

The physics of the most common sedimentation process—the settling of solid particles from fluid media—has long been known. The settling velocity equation formulated in 1851 by G. G. Stokes, is the classic starting point for any discussion of the sedimentation process. Stokes showed that the terminal settling velocity of spheres in a fluid was directly proportional to the difference in densities of fluid and solid, the radius of the spheres involved, and the force of gravity and inversely proportional to the viscosity of the fluid. Stokes' equation is valid, however, only for spheres of very small size (under 0.04 mm. in diameter) and hence various modifications of Stokes' law have been proposed for nonspherical particles and particles of larger size.

No settling velocity equation, no matter how valid, provides a sufficient explanation of even the basic physical properties of natural sediments. The grain size of the clastic elements and their sorting, the shape and roundness of these elements and their fabric and packing are the results of complex processes related not only to the density and viscosity of the fluid medium but also to the velocity of the forward motion of the depositing fluid and to the turbulence resulting from this motion and the roughness of the beds over which it moves, to various mechanical properties of the solid materials propelled, to the time or duration of the transport action and to other little-understood factors. Present knowledge, most of it empirically obtained, is inadequate to account for the known facts.

Chemical sedimentation is understood in terms of chemical principles and laws. Although the famous physical chemist, J. H. van't Hoff, applied the principles of phase equilibria to the problem of crystallizing brines and the origin of salt deposits as early as 1905, only a little effort was made to apply physical chemistry to the problems of chemical sedimentation. More recently, however, there has been investigation of the role of the redox (mutual reduction and oxidation) potential and pH (acidity/alkalinity) in the precipitation of many chemical sediments, renewed effort to apply known thermodynamic principles to the origin of anhydrite and gypsum deposits, the chemistry of dolomite formation, the problem of the ironstones and related sediments. New developments in isotope chemistry and nuclear physics are reflected in sedimentology in direct age determination of the potash-bearing minerals, notably glauconite, by the potassium-argon method, and in paleotemperatures or the determination of temperature of formation of the carbonate skeletal materials of marine invertebrates.

The geochemist also considers the sedimentation process in terms of the chemical end-results. To him sedimentation is like a gigantic chemical analysis in which the primary constituents of the silicate crust of the earth are separated from one another in a man-

ner similar to that achieved in the course of a quantitative analysis of rock material in the laboratory. The results of this chemical fractionation are not always perfect, for nature uses poor methods, but by and large the results are remarkably good. The greatest concentrations of many elements found in nature are those of the sedimentary rocks. No igneous or metamorphic process can match the sedimentation process in effective isolation and concentration of these and other elements. Geochemical differentiation or fractionation, which began in Precambrian times, perhaps 3,000,000,000 years ago, has resulted in an enormous accumulation of sodium in the sea, calcium and magnesium in limestone and dolomite, silicon in the bedded cherts and orthoquartzitic sandstones, carbon in the carbonates and carbonaceous deposits, sulfur in the bedded sulfate, iron in the ironstones, etc.

Sedimentation is most generally considered by the geologist in terms of the textures, structures, and fossil-content of the deposits laid down in different geographic and geomorphic environments. Great efforts have been made to differentiate between continental, littoral and marine deposits of the geologic record. The further classification of environments and the criteria for their recognition is still a subject of lively interest.

The analysis of the record of ancient basins of sedimentation has been promoted by the study of present-day basins. Oceanographic and limnologic expeditions have shed much light on the sedimentation in such basins as the Gulf of Mexico, the Black sea, the Baltic sea, various estuaries, lakes and fluvial basins.

See also references under "Sedimentary Rocks" in the Index.

BIBLIOGRAPHY.—A. W. Grabau, *Principles of Stratigraphy* (1913, 2nd ed., 1924); F. W. Clarke, "Data of Geochemistry," U.S. Geological Survey Bulletin 770 (1924); W. H. Twenhofel, *Treatise on Sedimentation,* rev. ed. (1932); F. H. Hatch, R. H. Rastall, and Maurice Black, *The Petrology of the Sedimentary Rocks,* 3rd ed. (1938); F. J. Pettijohn, *Sedimentary Rocks,* 2nd ed. (1957). (F. J. P.)

SEDIMENTATION RATE, BLOOD: *see* DIAGNOSIS.

SEDITION: *see* TREASON AND SEDITION.

SEDLEY, SIR CHARLES (1639–1701), English poet, dramatist, wit, and courtier, was baptized at St. Clement Danes Church, London, on March 5, 1639. He was the second son of a Kentish baronet, Sir John Sedley, and his wife Elizabeth, daughter of Sir Henry Savile, the great Elizabethan scholar. Admitted to Wadham College, Oxford, on March 22, 1656, he left without taking a degree. He inherited the baronetcy on the death of his elder brother in April 1656. After the Restoration he was a prominent member of the group of court wits called "the merry gang." Charles II delighted in his conversation and declared that "Nature had given him a patent to be Apollo's Viceroy." In June 1663 he took part in a wild frolic at the Cock Tavern in Bow Street, Covent Garden, for which he was fined 2,000 marks by the Court of the King's Bench. Dryden and Shadwell were among his friends and Dryden introduced him into his *Essay of Dramatic Poesy* under the name of Lisideius. His original comedy *The Mulberry Garden* was staged by the King's Company at Drury Lane Theatre in May 1668. In March 1677 his rhymed tragedy *Antony and Cleopatra* was produced at the Duke's Theatre and in May 1687 his second comedy *Bellamira or the Mistress* was acted with great success by the King's Company.

Sedley married Katherine Savage, daughter of Earl Rivers, in 1657. She became insane and was sent to a convent at Ghent probably about 1670. In 1672 he went through a form of marriage with Ann Ayscough, the daughter of a Yorkshire gentleman, and he seems to have lived with her for the rest of his life. Katherine Sedley, his daughter by Katherine Savage, though no beauty, was clever and witty. She became the mistress of James, duke of York (later James II), by whom she was created countess of Dorchester in 1686. Her father was an active supporter of William and Mary at the time of the 1688 revolution, when he is said to have remarked "Well I am even in point of civility with King James, for as he made my daughter a Countess, so I have helped to make his daughter a Queen." By Ann Ayscough he had a son who took the name of Charles Sedley and was knighted by William III. In the latter part of his life Sedley seems to have been transformed from a Restoration wit into an Augustan gentleman. He sat in all the parliaments of William III as member for New Romney and his speeches were thoughtful and sensible. He died at Hampstead in Aug. 1701.

Sedley's best play is undoubtedly *Bellamira,* a racy, amusing rehandling of the theme of the *Eunuchus* of Terence in terms of Restoration life. *The Grumbler,* his adaptation from the French *Le Grondeur* of D. A. de Brueys and J. Palaprat, is also a sparkling and lively performance. His literary reputation, however, rests on his lyrics and verse translations. His best lyrics such as the well-known "Phillis is my only Joy" have an exquisite grace and charm, while the poem beginning with the following stanza expresses deep feeling with a success only equalled among contemporary poets by his friend Rochester:

> Not *Celia* that I juster am
> Or better than the rest,
> For I would change each Hour like them
> Were not my Heart at rest.

His versions of the eighth ode of Book ii of Horace and the 4th *Georgic* of Virgil have been highly and deservedly praised.

A number of Sedley's best lyrics appeared in Hobart Kemp's *A Collection of Poems* (1672). The first collected edition of his works was published in 1702.

See V. de S. Pinto, *The Poetical and Dramatic Works of Sir Charles Sedley,* 2 vol. (1928) and *Sir Charles Sedley, a Study in the Life and Literature of the Restoration* (1927), where an account of early editions may also be found. (V. DE S. P.)

SEDUCTION, a term generally used in the special sense of wrongfully inducing a previously chaste woman to consent to sexual intercourse. A seduction action at common law stands in an anomalous position. The woman seduced has no right of action because of her consent. The action is instead given to her parent, the relationship to be protected, however, being that of master and servant rather than that of parent and child; a parent, as such, has no right of action, but must show his status as master or employer and that the seduction has deprived him of his child's services. The "loss of services" requirement is merely technical; any services actually rendered, however trifling, will suffice, and a minor daughter residing at home with her parent is without proof presumed to render such services. Furthermore, it is everywhere agreed that once a technical loss of services has been established, the parent may recover damages for medical and other expenses resulting from his daughter's pregnancy, for the loss of her society, and for his own outraged feelings. Punitive damages may likewise generally be recovered. Some states of the United States have abolished the "loss of services" requirement completely.

By the mid-20th century, the common-law seduction action in the United States had undergone significant statutory change. Thus statutes in a number of states authorize the seduced woman to sue in her own name, abolishing the common-law defense of consent. Such statutes, furthermore, have been interpreted by some courts as allowing adult married women to recover for their own seduction—a particularly unfortunate result, since the seduced plaintiff's husband likewise has a criminal conversation action against the seducer. Other courts deny relief to all women who have once been married, even though they are validly divorced when seduced. The theory of these cases is twofold: that the statutes are designed to protect virginity and not the Judeo-Christian concept of chastity, and that a previously married woman is simply too experienced to permit herself to be seduced in the statutorily required sense.

Criminal statutory-rape statutes in almost all jurisdictions have similarly been interpreted to give the seduced girl a cause of action in her own name, though only, of course, to those girls under the statutory age of consent, commonly fixed at 16 years. At the other extreme, several states, reacting to the blackmail possibilities of the seduction action and recognizing the questionable social policy involved in paying young women to abandon their virtue, have abolished the action altogether.

Seduction is not as a rule a criminal offense in England, but in the United States statutes in most states impose criminal liability upon the seducing male. The elements of the crime vary, consisting in one or more of the following: intercourse with a female

by trickery or under a promise of marriage; a requirement that the female be previously chaste; a requirement that the female be under a certain age, ranging from 18 to 25; and a requirement that the defendant be over a certain age. A seduction prosecution is barred in most states by marriage with the seducer and in others by an offer of marriage. Several states refuse to allow a conviction for seduction merely on the uncorroborated testimony of the seduced female. (D. W. Br.)

SEDULIUS SCOTTUS, poet and scholar, belonged to a group of Irish savants at Liège, where he can be traced from *c.* 848 to 860 (or 874?). His poems, mostly in classical Latin metres, include panegyrics on the emperor Lothair II, on Charles the Bald, and especially on his protector, Bishop Hartgar of Liège. A master of versification, Sedulius did not lack the human touch. His ingenious elegy on the death of Hartgar's ram blends mock-heroic elements with genuine affection and culminates in the bold comparison of the "martyred" ram with the Lamb of God. Some of his verse foreshadows the goliard songs of later times. More abstract poetry, alternating with prose, is found in his treatise *De rectoribus Christianis* ("On Christian Rulers"), one of the earliest among medieval "mirrors for princes." His commentary on the Pauline Epistles makes use of the commentary by Pelagius. As a scholar Sedulius cultivated grammatical, classical, and biblical studies. Connected with his circle are the Leiden and St. Gall copies of Priscian, with Old Irish glosses, and the miscellaneous Codex Bernensis 363, of the late 9th century, which ranks high among early manuscripts of Horace. Sedulius took considerable interest in the Greek text of the Bible. Four Greco-Latin biblical manuscripts are attributed to his circle; one of them was possibly written by his hand.

Bibliography.—His poems are edited by L. Traube, *Monumenta Germaniae historica. Poetae aevi Carolini,* vol. iii, pp. 151–240 (1896); other works in J. P. Migne, *Patrologia Latina,* vol. ciii (1851); *De rectoribus Christianis,* ed. by S. Hellmann, *Sedulius Scottus* (1906). *See* also J. F. Kenney, *Sources for the Early History of Ireland,* pp. 553–569, with bibliography (1929). (L. Br.)

SEDUM, a genus of the stonecrop family (Crassulaceae), mostly perennial herbs with succulent leaves. There are about 300 species, natives chiefly of the north temperate and frigid regions. The white, yellow, or red, rarely pink or blue, flowers are usually small and grouped in terminal clusters. They have a calyx of five sepals, as many petals, usually ten stamens and five distinct carpels, which have as many glands at their base and ripen into as many dry seedpods. More than 30 species are native to North America, widely distributed across the continent. In addition, some Old World species have become widely naturalized in the eastern states and Canada. Several species are found in Britain, including some with tuberous roots and large leaves like the orpine or live-forever (*S. telephium*), and others of smaller size, chiefly found on rocks, walls, and dry banks; *S. acre* is the wall pepper or golden moss, well known also in gardens, a variety of which, *aureum,* is in cultivation with golden-yellow tips to the leaves and shoots. Many species are cultivated for the beauty of their foliage or flowers, and many are remarkable for their vitality. They succeed on rockwork, old walls, or in garden borders; some, *e.g., S. lydium,* a native of Ásia Minor, are excellent for carpet bedding. *S. spectabile,* up to 2 ft. in height, with pink flowers in large heads, is fine for borders.

SEECKT, HANS VON (1866–1936), the officer who remodeled the German Army as the *Reichswehr* after World War I. He was born at Schleswig on April 22, 1866, and began his military career in 1885 as an ensign in an infantry regiment. After a period at the Military Academy he joined the general staff in 1899 and was sent to the Danzig Army Corps. By 1913 he was chief of staff to the Brandenburg Army Corps with the rank of lieutenant colonel. Throughout World War I Seeckt served on the staff. In February 1915 he was promoted colonel and became chief of staff to the 11th Army under Gen. August von Mackensen. In this capacity he played a leading part in the successes of the Central Powers against the Russians at Gorlice-Tarnów in May 1915 and in the subsequent invasions of Serbia (1915) and of Rumania (1916–17). After serving as chief of staff of the Turkish Army in the field from December 1917, he returned to Germany in November 1918 and took command of the *Grenzschutz Nord,* in East Prussia. His war service had not been spectacular, but he had won the reputation of being an exceptionally able staff officer who, unlike many of his contemporaries, always had an eye to the political aspect of a military problem.

It was after World War I that Seeckt rendered his most notable service. By the military clauses of the Treaty of Versailles, the German Army was limited to 100,000 men and, to prevent the creation of a reserve, service was to be for 25 years for officers and 12 for noncommissioned ranks. The Great General Staff was abolished, and the Germans were forbidden to have any aircraft, tanks, or heavy artillery. It fell to Seeckt to fashion to the best advantage the small, lightly equipped force that Germany was allowed to maintain as the *Reichswehr.* The extent of his success was seen in the speed with which this force was expanded after 1933 when Hitler came to power and the military clauses of the Versailles Treaty were disregarded. Seeckt created an army that was in fact the nucleus of a force of 35 divisions; and its leaders, trained to assume at short notice much higher ranks and responsibilities than those they actually held, were instructed by theoretical means to handle the modern equipment they were forbidden to possess.

In November 1919, Seeckt became chief of the *Truppenamt* or Troops Bureau, the organ whose relation to the new Ministry for the *Reichswehr* was equivalent to that of the former Great General Staff to the old War Ministry. On June 15, 1920, three months after the failure of the *Putsch* led by Wolfgang Kapp and Gen. Walther von Lüttwitz, he was appointed chief of the army command (*Heeresleitung*) in succession to Walther Reinhardt. The *Putsch* had produced disturbances, some of them Communist-inspired; and there had been clashes between the *Freikorps* (independent volunteer bands) and *Reichswehr* units. It was in these unpropitious circumstances that Seeckt began to remodel the *Reichswehr.* He had two principles in mind: to make it the kernel of a future national army, and to make a substantial part of it into the core of a mobile shock force. Instinctively distrusting mass armies, and convinced that in any future war trench warfare must be avoided at all costs, he probably regarded the second of his principles as the more important. In this remodeled army men such as Fedor von Bock, Gerd von Rundstedt, Albert Kesselring, and Heinz Guderian, who were to lead the German armies in their early successes in World War II, served as comparatively junior officers.

In the years immediately following his appointment Seeckt showed acumen in emerging on the right side in any political crisis or disorder, including Adolf Hitler's unsuccessful *Putsch* of November 1923. His policy was always to maintain the power and prestige of the *Reichswehr* by avoiding internal dissension and open conflict with Germany's former enemies. He supported a mild liaison with the Red Army, but discouraged internal left-wing activities, because they clashed too violently with the imperial tradition. In military matters his policy was to circumvent the more irritating restrictions imposed by the Treaty of Versailles, without arousing the Allies' suspicions. His measures included the setting-up of armament factories in the U.S.S.R., the training of German tank crews and air pilots in the U.S.S.R., and the training of "temporary volunteers" in the *Reichswehr.*

In 1926 Seeckt, who had many enemies, made two blunders: he issued an order regularizing dueling among officers; and he approved an application by Prince William of Prussia (son of the former crown prince) to take part in training with the 9th Infantry Regiment. As a result of these indiscretions, and the political pressure of his opponents, he was dismissed by President von Hindenburg on Oct. 8, 1926. Thereafter he spent much of his time in political activities, serving as a conservative member of the *Reichstag* in 1930–32. In 1934–35 he was sent on a mission to China to advise Chiang Kai-shek on the reorganization of his army. He died suddenly of a heart attack in Berlin on Dec. 27, 1936.

Seeckt's *Gedanken eines Soldaten* appeared in 1929 (Eng. trans. *Thoughts of a Soldier,* 1930). In 1933 he published *Deutschland zwischen West und Ost,* in which he declared himself against the notion of a Reich extending "from Lake Peipus to Calais," but

recommended the destruction of Poland and union of Germany with Austria. His memoirs, edited by Gen. F. von Rabenau, were published posthumously in two volumes, *Aus meinem Leben* (1938) and *Aus seinem Leben* (1940).

See E. von Schmidt-Pauli, *General von Seeckt* (1937); O. E. Schüddekopf, *Das Heer und die Republik* (1955). (C. N. B.)

SEED, a new generation in the life of a plant. Typically the ripe seed consists of a partially developed young plant (the embryo) provided with an abundant food supply (endosperm) and enclosed in a protective seed coat. In size seeds range from the dustlike seeds of begonia and orchid to the large nutlike seeds of the double coconut palm or Coco de Mer (*q.v.*), the fruit of which often reaches a foot in length and 40 lb. in weight. The mature seed usually is air dry and in a resting condition. In nature the seed is important as a means of carrying plant life over periods unfavourable for growth and as a means for distributing the plant in space and in time. In agriculture the seed is important in two ways: (1) the food stored in the seed for the development of the new plant can also serve as an important food source for man and animals (wheat, beans); (2) with most annual crop plants the seed is the start of the new crop each year. To a great extent the yield and quality of the crop depend on the quality of the seed planted. The quality of the seed depends on both its genetic background and its ability to germinate and to produce strong plants.

BIOLOGY

Development.—The seed develops from an ovule in the ovary of the flower (*q.v.*). The male germ nucleus from the tube of the germinated pollen grain unites with the egg cell within the ovule. This fertilized egg divides and develops into the embryo, which is the beginning of a new plant. The stage of development of the embryo at seed maturity varies from the well-formed plantlet of pea or corn to the slightly developed embryo of an orchid, a small spherical mass of cells in which the beginnings of embryo plant parts, cotyledons, hypocotyl and epicotyl (*see* below), are barely distinguishable. As the embryo develops, the other parts of the ovule also grow and develop until the seed is mature. The integuments of the ovule, which are a part of the mother plant, develop into one or more layers to form the seed coat, or testa. Details of the development of the seed from the ovule are found in articles on Gymnosperms and Angiosperms.

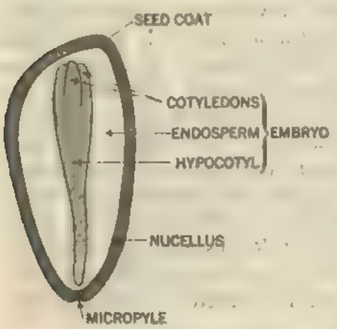


FIG. 1.—LONGITUDINAL SECTION OF PINE SEED

Structure.—The structure of seeds may be described first in general terms, although the details for different types of plants differ greatly. The embryo has one or more cotyledons (seed leaves) below which is the hypocotyl, with the growing point of the root at its tip, and above which is the epicotyl, with a growing point for new stem tissue. The embryo is surrounded by the endosperm with its food reserves or occasionally by the perisperm developed from the nucellus of the ovule. In the seeds of some plants, *e.g.*, the pea family, Leguminosae, the food reserves of the endosperm are absorbed to form enlarged cotyledons. The endosperm is enclosed by a seed covering (largely the seed coat). In some seeds thin layers remaining from the nucellus or the endosperm may form an important part of the seed covering. The details of seed structure differ in gymnosperms and in angiosperms. In gymnosperms the embryo with two to many cotyledons is buried in the endosperm. In pine seeds 6 to 15 needlelike whorled cotyledons, the number varying with the species, surround the small epicotyl. The cotyledons and epicotyl are attached at the top of the cylindrical hypocotyl. The endosperm is enclosed by the comparatively thick and tough seed coat.

In angiosperms the embryo may have one cotyledon (monocot) or two (dicot). Corn is fairly typical of many monocot seeds. The well-developed embryo is at one side of the seed. The hypocotyl and root tip are enclosed by a thin sheath (coleorhiza) and the epicotyl, with two or three young leaves formed, is enclosed in its sheath (coleoptile). The large shield-shaped scutellum almost encloses the axis, and one surface is in contact with the starchy endosperm. The scutellum is attached to the axis between the hypocotyl and the epicotyl, and is considered by many to be a special development of the single cotyledon. A thin seed coat and the adhering fruit coat surround the embryo and the endosperm.

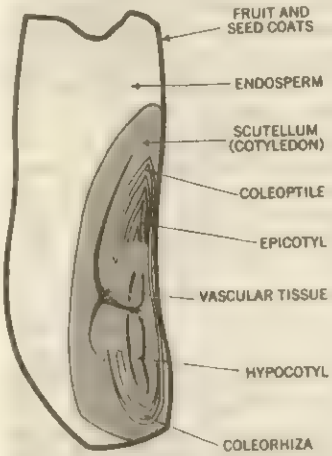


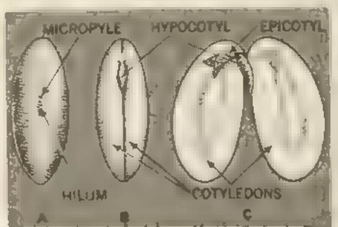
FIG. 2.—LONGITUDINAL SECTION OF GRAIN OF CORN, A MONOCOT SEED

The castor bean is representative of a dicot seed with much food stored in the endosperm. The embryo with two thin, broad cotyledons enclosing the small epicotyl and the short cylindrical hypocotyl are embedded in the endosperm. The whole is covered by the hard, brittle seed coat. In the bean, typical of many dicots, a well-developed endosperm of the immature seed is gradually absorbed by the enlarging cotyledons until at maturity only a thin membranous layer remains. At maturity the large food-filled cotyledons protect an axis consisting of a cylindrical hypocotyl and an epicotyl with two or three leaves already evident. This large embryo is enclosed by a comparatively thin covering of combined testa and endosperm.

The developing seed, attached to the placenta of the ovary by the funiculus, receives food through its vascular strand. At maturity the seed separates from the funiculus, leaving a scar called the hilum. The shape and position of the hilum often are a distinguishing character useful in the identification of seed kinds. The micropyle is a small porelike opening in the ovule, where the ovule coverings, or integuments, meet and the pollen tube usually enters. In some seeds the micropyle is visible when the seed is mature, but usually it is closed by cuticle or wax. In the bean it can be seen near the hilum. The raphe is seen as a slight ridge along one side of some seeds. The raphe is the result of the fusion of the funiculus to one side of the ovule (bean and pansy). The area at the end of the vascular tissue of the funiculus is known as the chalaza. Sometimes a rough outgrowth called a caruncle develops near the hilum (castor bean). The additional or partial covering outside the seed coat that develops on some seeds (water lily) after fertilization is called an aril.

The outer surfaces (epidermis) of some kinds of seeds may have characteristic modifications. Cotton fibres are formed as extensions of some of the epidermal cells of the seed coat. In contrast, a smooth cuticle forms on the epidermis of the lima bean seed and makes the surface hard and shiny.

The seeds of flax have thickenings of the epidermal cell walls that swell in water and become gelatinous. In turnip, cabbage and related seeds, a network of ridges, formed as a result of differences in heights of epidermal cells, gives the seeds a characteristic pitted appearance.



FROM EDMUND W. SINNOTT, "BOTANY: PRINCIPLES AND PROBLEMS," REPRODUCED BY PERMISSION OF MCGRAW-HILL BOOK COMPANY

FIG. 3.—BEAN, A TYPICAL DICOT SEED

(A) Face view of seed; (B) face view of the embryo after the seed coat has been removed; (C) the two cotyledons spread apart

The general form of different kinds of seeds is in part determined by the orientation of the ovule with respect to the placenta on which it is borne. Four types of orientation have been described. The orthotropous, or atropous, ovule is straight and erect, with the micropyle in line with the funiculus. The anatropous ovule is completely inverted, with the funiculus fused along one side so that the micropyle is close to the hilum. The campylotropous ovule is curved,

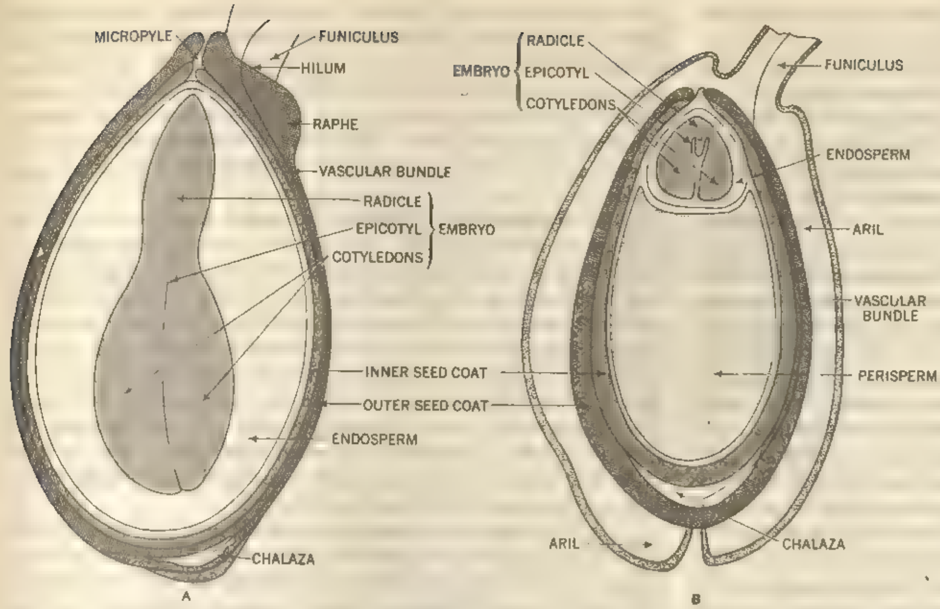


FIG. 4.—VERTICAL SECTIONS THROUGH TWO SEEDS: (A) PANSY; (B) WATER LILY

bringing the micropyle near the chalaza. The ovule is called hemitropous when it is half inverted but straight and has the hilum on the side of the ovule.

Composition.—The compositions of different kinds of seeds show great variation. The cotyledons of soybeans and peanut have a high oil content, but in bean the oil is replaced by starch. The seeds of all legumes have a high protein content. In cereals (corn, wheat, etc.) the embryos are high in oil and protein, but the endosperm, which occupies the greater portion of the seed, is high in starch. The oil of the castor bean seed is in the cells of the endosperm. A few seeds, *e.g.*, onion, contain sugar as a reserve food. The immature seeds of sweet corn and garden pea contain much sugar, but most of it changes to dextrin or starch by maturity. Layers of hemicellulose are deposited as thickenings of the cell walls of the endosperm of carrot, date and other seeds, and serve as food reserve.

Adaptations for Distribution.—The seed represents an inactive state in the life history of the plant and thus is itself an adaptation for distribution in space. It is relatively insensitive to changes of surroundings and lends itself admirably to transportation from place to place by wind, water, and animals. It was for this reason also that the seed made possible the development of agriculture by primitive man—the seed could be kept from one crop season to the next and could be carried from one region to another. As man selected plants for food, the seeds of many crop plants tended to lose their natural ability for survival in the wild and became largely dependent on man for preservation from season to season.

Special structures of seeds and fruits, adapting them for wide distribution, have developed during the evolution of plants. These adaptations vary greatly in nature, as do the means used for distribution.

Wind Distribution.—Wind is an important factor in seed distribution. The very small size of many seeds may be considered as an adaptation that allows them to be borne long distances in the air. Small grass seeds have been collected by airplanes at elevations from 200 to 3,000 ft.

The dandelion bears many one-seeded fruits to which are attached tufts of fine hairs that act as parachutes or sails. The seeds of the milkweed, when they burst out of their pods, have similar hairs that aid the wind in carrying them abroad. The cattail head has countless closely packed, tiny seedlike fruits, each with fine hairs attached; when the fully ripe head is disturbed, the expansion of the closely packed hairlike attachments pushes the seeds out into the wind to be carried away.

The much heavier seeds, or seed-bearing fruits, of pine, maple and ash trees have large, flat wings that permit the seeds to glide a moderate distance from the mother tree.

Among some plants the adaptation for distribution by wind is found in the form and structure of the entire plant rather than in the form of the seed or fruit. Several species of plants, belonging to different families but all known as tumbleweeds, are bushy and when mature nearly spherical. The main stem at the soil surface is weak and breaks readily. As the wind rolls these detached plants across fields and plains, the seeds gradually are loosened and fall to the soil. The Russian thistle and the tumbling mustard are common examples.

Many wild plants shed some of their seeds in winter after snow has fallen, and these may be blown considerable distances on the frozen surface of the snow.

Water Distribution.—Water is another natural agent for the dissemination of seeds and fruits. Aside from resistance to germination with the first absorption of water, the principal adaptation for water distribution is low specific gravity, permitting many seeds and fruits to float for long periods. One of the most noteworthy examples is the fruit of the coconut, which has a fibrous outer coat with many air spaces and a smooth outer surface. Shortly after a new island had been formed in the Pacific ocean by a submarine volcano in the Krakatoa group of islands, many germinating coconuts were found on the volcanic sand of the beach; some had sent down roots and become established plants. In addition to the coconuts, seedlings of several other tropical plants were found, all having developed from seeds washed up on this barren beach from distant islands. The seeds had floated long distances undamaged by sea water because of their protective coatings.

Animal Distribution.—Animals and birds are other agents for seed dispersal. After a walk in the fields in autumn, a person may find many seeds and burrs attached to his clothing. These burrs cling equally well to the fur of passing animals. Different species of plants have evolved various special forms for attachment, but all depend on projections with recurved hooks at the tip or with tiny reversed teeth. The devil's-pitchforks, the flattened fruit of a plant of the sunflower family, has two long prongs with many reverse pointing barbs. The sticktight is the pod of a plant of the pea family that has tiny spines with hooked tips; the pods are jointed and easily break apart into one-seeded segments. The cocklebur is a two-seeded burr with comparatively long hooked spines.

Fleshy fruits and berries may be considered as special adaptations of the plant that serve to aid in the distribution of seeds by animals and birds. The seeds of most fleshy fruits have tough coverings that resist digestion. The seeds remain in the excrement, ready to grow when they are deposited in some new location. The fruit of the mistletoe has a sticky covering, so that the seeds stick to the beaks of birds, to be scraped off later on the bark of a tree.

Distribution by Expulsion.—Some fruits have special adaptations that violently expel the seeds to distances of several feet from the mother plant. The pansy and the violet have three-parted pods that split into three boat-shaped segments. As the ripe pod dries, the sides of each segment close together and violently expel the seeds, one at a time. The two halves of the pod of the vetch and of other legumes split apart, and then each half twists and curls, thus violently throwing the seeds for some distance. The seeds of the touch-me-not or impatience are borne in a pod that develops strong differences of tension. When the fruit is ripe these tensions are so great that the least touch will cause

the pod to split apart; the segments curl violently and expel the seeds to a distance of several feet.

Adaptations for Preservation.—Seeds may be adapted not only for distribution in space but also for preservation of their viability over unfavourable conditions, thus providing distribution in time. When mature, the seeds of many plants, especially weeds and native species, are dormant and require special conditions before they will germinate. Some seeds, notably those of the pea family, have seed coats impervious to water (hard seeds). This impervious coat must be weakened by bacterial action in the soil, by mechanical scarification (scratching the seed coat) or by other treatment before the seeds can absorb water and germinate.

The seeds of many winter annual weeds (*e.g.*, chickweed, shepherd's purse and upland cress) that ripen in late spring or summer do not germinate while the weather is still warm, but they start to grow with cool fall temperatures. Other seeds germinate only after the moist seeds are exposed to light. As long as they remain buried in the soil the seeds are dormant, but when they are brought to the surface and receive light they germinate. It has been found that this response to light is controlled by a reversible photoreaction that also controls stem elongation, leaf expansion, flowering and other plant responses (*see* PHOTOPERIODISM). Germination of light-sensitive seeds is promoted by red light and prevented by radiation near the limit of visibility (far red). The seeds of many temperate zone plants have deep-seated and prolonged dormancy. The embryos must undergo chemical and growth changes (afterripening) before germination is possible. Most such seeds (many trees and shrubs) afterripen when held moist and cool for varying periods. In nature, afterripening occurs as the seeds lie in moist ground for one or more winters. It may be accomplished artificially by holding the seeds in moist sand or other media at temperatures a little above freezing for one to several months, the period depending on the species. (*See* PLANT PROPAGATION.)

Longevity.—The longevity of seeds of different kinds of plants varies greatly. Hard impervious coats protect the embryos so that they may persist for many years. Seeds of legumes taken from herbarium specimens known to be more than 100 years old have germinated. A very striking example of long life is shown by the seeds of the oriental lotus (*Nelumbo nucifera*); these were recovered from an ancient peat bog in Manchuria and germinated and produced flowering plants. Radiocarbon dating of seeds from

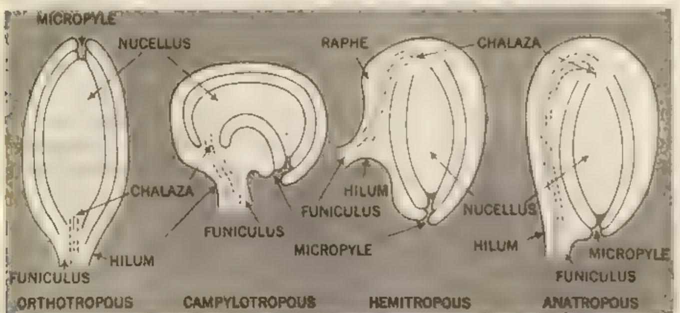


FIG. 5.—TYPES OF OVULES

the same source indicated an age of approximately 1,000 years. In contrast to this, the seeds of willow and poplar remain viable only a few weeks. The seeds of many plants, especially weeds, may remain viable for many years when buried in the soil. In an experiment in which seeds of 107 species were buried in the soil so as to permit recovery, those of 51 kinds grew after 20 years and those of 36 after 39 years. Seeds of both crop plants and weeds were included. The longevity of field and garden seeds depends on the moisture content of the seeds and on the temperature at which they are held. Low moisture and low temperature ensure long life. (E. H. T.)

AGRICULTURE

SEED TRADE

Factors affecting the production of satisfactory food crops include seed, soil type, fertilizers, cultural methods, weather, and control of insects and diseases. Unless seed true to name and of high quality is sown, the crop may be a commercial failure even though all other factors are at an optimum level. For greatest success not only should the seed be of high quality and true to name, but it should also have high vitality and be free of disease organisms and weed seeds. To ensure that the demand for high-quality seeds is met is the constant concern and duty of the seed trade.

Early Beginnings.—Man has grown and saved seed ever since he began cultivating plants. Seed growing as an industry, however, was hardly recognized as such until the early 18th century. The first commercial production of seed occurred in northern European countries, but this was followed within the same century by similar production in North America. The Shakers in New York state developed a business of growing and distributing to home gardeners over a dozen different kinds of vegetable and herb seeds. In the early 19th century some acreage of garden pea seed was grown in Ontario for the D. Landreth company, a U.S. firm. By 1847 this same company had 250 ac. of pea seed in Pennsylvania. Seed acreages increased steadily after that.

Prior to actual seed production in the United States, a number of seed dealers, of whom David Landreth was the first (in 1784), became established, principally in Philadelphia, but also in New York and Boston. These merchants imported seeds from European producers, and, until World War I, Europe continued to be the chief source of vegetable seed for most of the world. With the advent of that war, and again as a result of the dire need for seed in World War II, commercial seed production in the United States was greatly increased, as was seed production in Australia and New Zealand.

Climate and Seed Production.—During the 20th century, climate has become more of a determining factor in the location of areas where seed is grown than earlier in the history of the seed trade. This has been especially true within the United States, where there has been, and still is, a definite movement toward greater seed production in the western states. Low precipitation supplemented by adequate supplies of irrigation water has made it possible to produce vigorous crops free from disease. During the period of seed maturity and harvest, dry air and absence of rain facilitate the collection of high-quality seed. In Europe some northern seedsmen arrange for certain crops to be grown in countries to the south, where ideal conditions for seed production exist.

Various areas of the world have become associated with the production of certain seed crops. Among the producers of field seeds, Canada is known for the production of alfalfa, various clovers and grasses. Many of these same crops, except alfalfa, are grown in Denmark in sufficient quantity for export. France has long been a big producer of various field seeds; so also has Italy. All the field seeds are grown in large quantities in the United States, and, in spite of the movement in production to the western part of the country, over one third of the principal field seeds were produced in the north central states in the 1950s. Before World War II Czechoslovakia, Rumania and Poland supplied large quantities of these seeds, and Hungary was noted for production of clovers.

Such vegetables as cabbage and related crops, as well as spinach and beets, thrive best under conditions of high humidity in areas adjacent to the sea; hence they are widely cultivated for seed in the northern European countries such as the Netherlands, Denmark and England, as well as around the Puget sound area of the Pacific northwest in the United States. Japan also produces such types of seed extensively. In both Europe and Japan, curing of the seed often presents difficulties because of inclement weather. In such regions the moisture content of seed sometimes has to be reduced by artificial drying even after it has been cleaned.

In southern France and Italy, as well as in parts of adjacent countries, such vegetable seeds as lettuce, endive, onion and carrot, which require dry atmospheric conditions prior to and during harvest, are widely grown. Similar crops, and in addition peas, beans and sweet corn, are grown in the western United States except in the Pacific northwest. California, with a wide range of climatic conditions, grows a greater variety of such seeds than any other state, but they are also widely cultivated in Idaho and

the eastern sections of Oregon and Washington and to varying degrees in other western states. Some seed crops such as tomatoes, peppers, eggplants and watermelons, and to a lesser extent the other vine crops, are often grown for seed in the same regions where they are cultivated for processing and marketing. While tropical countries are not normally considered as a source of seed for temperate climates, there is some evidence that such seed can often be successfully grown if the crop is planted at the correct time at a high elevation where temperate climatic conditions exist.

Many of the flower seeds are grown in the Netherlands, France, England and the United States, particularly California, and to a lesser extent in the Pacific northwest and some eastern states.

Size of Industry.—The seed industry has considerable economic value aside from its importance as a source of seed for other varied businesses. Seedsmen, or at least seed dealers, are located in practically every country of the world. In those countries where seed is grown on a commercial scale, there are large capital investments in office buildings, warehouses, seed and many types of farm and specialized machinery necessary for the handling of seed crops, as well as trucks, other vehicles and the innumerable items associated with a highly diverse type of business. Figures on actual acreages and production are not available for the seed industry as a whole. Statistics collected by the agricultural marketing service of the U.S. department of agriculture give an idea, however, of the extent of production in that country. In general total seed acreage remains fairly constant from year to year, although acreage of individual kinds of crops may fluctuate greatly. Field seeds (alfalfa, grasses, clovers, etc.) excluding small grains were harvested from well over 3,000,000 ac. every year during the 1960s. At the same time 170,000 to over 200,000 ac. were devoted annually to the production of vegetable seeds. Of this latter acreage approximately 95% was occupied by the large-seeded vegetables—beans, peas and sweet corn—and 5% by more than 35 small-seeded vegetables. In contrast to the large acreages devoted to farm and vegetable seeds, there are about 6,000 to 7,000 ac. of flower seeds in the United States. Seed acreages in Europe, while differing as regards specific crops and varying from year to year like those in any seed producing area, are in general of the same over-all magnitude.

Types of Activity.—Seed production is a rather complex business enterprise in which the actual growing of seed is only one of many activities. Comparatively few seed companies are involved in seed-growing operations. Most of them as jobbers or retailers handle seed as a commodity rather than as a product that they produce. Many producers are also wholesalers and as such often operate large organizations among whose employees are plant breeders, production supervisors and fieldmen, warehouse men, seed analysts and salesmen. Often among warehouse employees are men especially trained to operate the numerous seed cleaning machines. Jobbers aid in the distribution of seed through their buying and selling operations. Retailers are most numerous in the trade. Most of them operate small stores, the business of which is not limited to seed. Some large producers retail their products by mail through seed catalogues. In such catalogues field seeds may be merely listed as to name and price, but the varieties of vegetables and flowers are often illustrated and described in detail. Still other producers market their seed by means of commission packets placed in grocery and other types of stores.

Growing the Seed.—In general two types of seed are produced and handled by all seed growers—stock seed and market seed. Stock, foundation or "mother" seed is seed grown and used by seedsmen to produce market seed. Market seed is the product that is sold. Stock seed must be of the highest quality, and in particular it must be absolutely pure as regards variety. It is the seedsman's most prized possession, for on it much of his reputation as a seedsman rests, and without good stock seed no producer can long stay in business and make a profit.

In growing the seed, and particularly stock seed, the best cultural practices must be employed and care must be taken to ensure adequate isolation from growing crops of a similar kind in order to prevent cross-pollination with related plants. Plants of the same species found growing along the margin of the field are cut down before they develop flowers and endanger the purity of the seed crop by cross-fertilization. Weeds particularly should be kept to a minimum in seed crops. They not only compete with the crop for moisture and nutrients, but their seeds become mixed with those of the crop and later can be separated only with difficulty. During the growing season, especially when the seed crop is approaching the flowering stage, all plants that show characteristics dissimilar to those of the particular variety under cultivation are carefully removed—a process known as roguing. With root and bulb crops that have to be dug and replanted after the root or the bulb has matured, off-type plants are usually rogued at that time. Severe roguing is more essential in a crop of stock seed than it is in market seed. The latter often needs no roguing if grown from strictly selected stock seed. Harvesting at the proper stage of maturity calls for great care and experience. If the seed is harvested too early, yields and quality may be reduced by immature seed; if harvested too late, seed may be lost by shattering or shedding. Methods of harvesting, curing and threshing vary according to the kind of seed produced, as well as the country or section in which it is grown. Machinery is commonly used wherever possible in harvesting and threshing operations. However, some of the flower seeds are grown on such a small scale that even if a machine could be used it would hardly be feasible. Where machines are used, they have to be thoroughly cleaned of seed between each crop. Although production and handling procedures are fundamentally alike for all seeds, those followed with field seeds do differ in some respects from those required for vegetable and flower seeds.

Field Seeds.—Trade in field seeds flows in general in two directions, particularly within the United States. This comes about largely because in certain regions the production of field seeds was a regular part of the cropping system prior to the period when specialty growers were becoming more common. The farmer producing clover or grass seed commonly sells to a local jobber or elevator man, who is either an agent of some wholesaler or is in touch with the general market. Such lots are shipped to a large dealer by whom the seed is cleaned, graded, bulked and bagged in accordance with quality and designated by the various grade names used by the dealer. Sales are then made to country merchants who, in turn, sell to the farmer. In some seed-producing sections, especially those far removed from terminal markets, large dealers have established plants in which field seeds are cleaned prior to shipment to a central warehouse.

Many field seeds are a world commodity. Price thus often depends on the available supplies in countries offering seed for sale. There is consequently a more or less constant movement back and forth, and a given lot of clover seed, for example, may be exported and later imported, depending on price fluctuation.

Production of Field Seeds.—Although field seeds are grown extensively by general farmers, there is a definite trend toward production by special growers. With the increasing development of new varieties and strains of alfalfa, grasses and clovers specially selected for definite regions, growers having the proper aptitude find it profitable to devote much of their farming operations to seed production. The steadily increased use of certified varieties of field seeds has done much to encourage growers to specialize in seed production. Within the United States, crop improvement associations operate in most of the states, and these organizations establish certification services in conjunction with public agencies such as experiment stations, state extension services or state departments of agriculture. Through a system of inspections and detailed records, officials are able to certify that crops grown by properly authorized farmers are true to name, free of varietal and mechanical mixtures, noxious weeds, etc. Similar certification practices have been established in Canada as well as in some European countries. An International Crop Improvement association has been in existence since 1919. Certification procedures for growing seed became necessary with the release of improved varieties from experiment stations and similar agencies. Certification tags on each bag of seed attest to the variety and quality of seed.

Verification of Origin.—Before the development of improved varieties, it was discovered through experimentation that commercial alfalfa and red clover seed grown in one country or in one section within a country were not always adapted to other areas having different soil and climatic conditions. In 1927 a seed verification service was inaugurated in the United States. This service verifies the origin of much of the domestic uncertified alfalfa, red clover and open-pollinated field corn and requires that the seed is so labeled. Imported alfalfa and red clover seed are artificially coloured to indicate country of origin. Verification of origin will continue to be a necessary service so long as the old common varieties are offered for sale.

Vegetable and Flower Seeds.—The production of vegetable and flower seeds tends to be more specialized than that of the field seeds, and control of the industry is maintained more within the trade itself. Practically all vegetable and flower seedsmen have their own stock seed of each variety that they handle. Such seed is usually grown by the seedsman on his own farm under his direct supervision. In contrast, most market seed, the volume of which normally exceeds by many times that of the stock seed, is grown under contract. Seedsmen make contracts with farmer-growers who grow the market seed from stock seed supplied by the seed company. Such a practice is common in all the large seed-producing areas of the world. Any unused stock seed, together with all the market seed produced, is the property of the seed company. The grower is paid on the basis of a certain price per pound of clean seed produced. Contract growing has come about partly because of the great diversity of vegetable and flower seed crops, and partly because many of these are grown for seed in areas other than those where they are commonly grown as a food or an ornamental. There are over 40 different kinds of vegetables and literally thousands of varieties. The number of kinds of flowers is even greater, and each of these in turn has many varieties. Much of the competition within the vegetable and flower seed industry is the development of new varieties and the maintenance of superior strains of standard ones. Certification of vegetable seeds is extremely limited and that of flower seeds practically unknown.

Many companies, almost by necessity, operate on an international scale. Northern European companies may produce their stock seed in their home country but, through branches or through contract arrangements with other companies or individuals, grow much of their market seed in a country more suited to such production. A number of U.S. firms have production stations in widely separated sections of the United States, and in addition sometimes make arrangements for market seed to be grown for them in Europe, Japan, Mexico or elsewhere. Likewise, European seedsmen have some of their supplies grown in North America.

Breeding and Selection.—With the tremendous increase in knowledge of plant genetics and modern techniques of plant breeding, the specialist in this field has become of increasing importance to progressive seed companies. Not only do such companies often wish to develop new varieties of their own, but they also need to increase new releases from experiment stations and other government agencies. Since such varieties have often been developed for resistance to such factors as diseases, insects, heat, cold and drought, they may also wish to transfer if possible such characteristics to already established varieties. Such programs require skilled plant breeders and often considerable investment on the part of the company. The increasing production of first-generation, or F_1, hybrids of corn (both field and sweet), onions, tomatoes, sorghums, petunias and other crops often necessitates elaborate breeding and production programs. *See* also PLANT BREEDING.

Processing Seeds.—Preparing seeds for market is a most important function of the wholesale dealer. All seeds as they come from the producer contain a certain proportion of dirt and weed seeds, all of which must be removed. Large dealers are equipped with cleaning mills of various kinds. Although screening and fanning mills are still the basic equipment required for seed processing, many ingenious and specialized machines for handling seeds have been developed. Equipment is available so that seeds can be separated on the basis of specific gravity, friction, adhesion, light and colour (by electric eye), as well as handled by the old standard procedures. Processing includes more than cleaning. With the increasing use of precision planting machines, seeds often need to be bulked on the basis of rather exact sizes, and this some seedsmen are prepared to do. In all plants of any size, seed is processed in line, which means that from beginning to end of the operation, the seed is mechanically transferred from one machine to the next without having to be set aside to await further handling. Such co-ordination of machinery to maintain an even flow requires skillful supervision by trained personnel. Throughout such operations a complete record of the various seeds being processed has to be maintained so that no mixtures occur. As the processing of each is completed, it is labeled as to its lot number, variety, strain, grower, etc., so that the seedsman can, if necessary at a later date, easily trace its history.

Legislation Affecting the Trade.—Many countries and most states within the United States have laws regulating the sale of seed, but these laws vary in detail. In general, such laws require the giving of certain information covering percentage of purity, of germination and of weed seeds present, place of origin and similar facts. In some states, the sale by trade of seed containing more than a specified number per unit of weight of certain noxious weed seeds, such as dodder, is prohibited. Owing to the fact that seed laws are not uniform in the various states, large dealers are compelled to attach various tags to shipments in accordance with the laws of the state to which the seed is consigned. Some weeds are considered noxious in one state or region but not necessarily so in another.

Laws in many countries regulate the importation of seeds and affect the movement of seed in international trade. Import licences are required in some countries. By such means, domestic producers are often protected. Domestic prices may be maintained at a higher level than otherwise and domestic production is encouraged. There are regulations in a number of countries to limit the sale of seed to those varieties that have been tested and approved by governmental agencies. A seedsman's problems are therefore by no means disposed of when he has grown and processed a crop of seed. He may still have the task of placing it in trade. (L. R. H.)

SEED TESTING

The scientific testing of seeds before planting, recognized as an essential aid to agriculture, is performed to determine the value of seeds for planting. Such testing reveals, for example, the proportion of seeds labeled white clover that are actually white clover seeds, the percentage of these that could produce plants, and the presence of seeds of noxious weeds.

History.—Seed testing developed first in Europe with the rapid expansion of agriculture in the second half of the 19th century. The practical application of botanical knowledge to determine the value of seeds for planting apparently developed more or less simultaneously in several countries, but leadership usually is credited to Frederick Nobbe, who established a laboratory for testing seeds in 1869 at Tharandt, Ger. Within a few years seed-testing laboratories were started in Denmark, the Netherlands and Switzerland. E. H. Jenkins, of Connecticut, while studying in Europe, was much impressed with the seed-testing work of Nobbe. Jenkins included in the 1876 report of the Connecticut Agricultural Experiment station a statement of the possible advantages of seed studies to American agriculture. Late in the 19th century botanists and agronomists at various state agricultural experiment stations tested seeds for farmers and determined the quality of those on the market. During this period, W. J. Beal of Michigan and C. E. Bessey of Nebraska were especially active in laying the foundations for seed testing in North America and interesting their students in seed studies. G. H. Hicks, a student of Beal, organized seed-testing work in the United States department of agriculture in 1896.

Governmental Controls.—The early work by several states and by the department of agriculture showed that much of the seed in commerce had low germination capacity, contained seeds of many weeds and often was adulterated with seeds of similar ap-

pearance but with low agricultural value. Much seed that had no other use than for the adulteration of other seed was imported. The generally low quality of seeds led, early in the 20th century, to the passage by several states of laws regulating the sale of seeds. These laws all were based on the premise that the farmer should have the details as to the quality of the seeds he bought. They required that all agricultural seeds be labeled as to percentage of pure seeds, inert matter, other crop seeds and weed seeds and as to the percentage of pure seeds capable of germination. Most state laws also designated the weeds especially troublesome in the respective states as "noxious weeds" and required a listing of the seeds of any such noxious weeds and a statement as to their rate of occurrence.

Each state controls the sale of seeds within its borders, just as it does the sale of other materials, but the federal government has power to control importations and interstate shipments. The first federal law concerning seeds, enacted in 1912 and administered by the United States department of agriculture, referred only to the purity and the weed-seed content of a few agricultural seeds. After several successive amendments, a new Federal Seed act, which prohibited the importation of all important agricultural and vegetable seeds unless they met certain minimum standards of purity and germination, was passed by congress in 1939. This act also required the truthful labeling as to purity and germination of agricultural and vegetable seeds entering into interstate commerce. The Federal Seed act is brought up to date by amendments from time to time. The foregoing is only a general outline of the practices of seed control and does not cover all features.

As might be expected, the seed-testing practices used in the United States followed in general those developed in continental Europe. Although the general methods of seed testing are similar, the details of the control of seed sales vary greatly in the different European countries. In Canada a system of seed grades has been developed to replace the labeling for purity, germination, and so forth. These grades represent an attempt to evaluate the different qualities of seed lots. Seed control in Canada is under the dominion government. The laboratory seed-testing procedures used in Canada, like those of the United States, are for the most part similar to those developed in Europe.

Testing Procedures.—The first step in seed testing is to be certain that the sample of seed accurately represents the lot of seed from which it is drawn. If the sample tested is not representative, the results of the test cannot accurately indicate the value of the entire lot of seed. Procedures for adequate sampling of seed lots have been worked out.

Purity.—The separation of four components of a seed lot is known as the "purity test." A representative portion of the seed sample of suitable weight is separated first into seeds and other material called "inert matter" (chaff, stems, fungus bodies, dirt, etc.). The seeds are separated into pure seeds (as labeled), seeds of other plants grown as crops, and seeds of weeds. Each portion is weighed and the percentage of each is calculated. Each kind of seed is identified and listed. Because even very small proportions of noxious weed seeds are important, a much larger sample than is used for the regular purity test is examined for the presence, the identification and the rate of occurrence of each kind of noxious weed seed found. Seeds of many kinds of weeds occur in the different crop seeds, and many of these are superficially similar to the crop seeds or to each other. The seeds of some related crop plants are very much alike and separable only by special study. Identifications are based mainly on the analyst's familiarity with the seeds but may be based on comparison with correctly identified samples or with illustrations. An analyst needs thorough knowledge of plant classification and must be familiar with the appearance of many kinds of seeds.

Seed as it comes from the harvester contains much foreign material and may carry excessive weed seeds. This seed must be processed to make it as clean as possible. Many seed companies have their own analysts who make many special tests at different stages of processing to evaluate the success of the cleaning operations. Also, successive germination tests may be made to learn the effectiveness of attempts to remove immature or injured seeds.

Viability.—The viability of seeds is determined directly by finding whether the seeds will germinate. This measurement of viability is not as simple as it may appear to be. Seeds are young plants whose growth processes have been suspended. In general these growth processes are resumed when the seeds are supplied with moisture and placed at a suitable temperature for germination. Nature, however, has developed various special provisions in certain kinds of seeds to prevent germination under conditions unfavourable for seedling development. Even in the absence of special requirements, seeds of different kinds have different moisture and temperature needs. Living seeds are subject to many hazards during harvesting, processing, storage and shipping. A seed does not lose life all at once, but certain parts may be injured or certain organs may die before others. Because the purpose of seed testing is determination of the value of seeds for planting, a seed cannot be counted as germinated unless it is able to develop into a plant capable of continued growth. A seed may show some signs of life and yet be worthless to the farmer. The evaluation of seedlings (determination of which seedlings can develop into useful plants) is an important part of the germination test and it requires much experience as well as knowledge about plant structure and normal plant development.

In order to handle many samples and to provide standard reproducible conditions, analysts carry out most viability tests under artificial but controlled conditions. They provide a suitable material to supply moisture (the substratum). This may be blotting paper, absorbent paper, moss, sand or suitable sterile soil. The substratum used varies with the kind of seed, the test conditions and the space available. Counted seeds are spaced on a suitable moist substratum and placed in a cabinet maintained at sufficiently high humidity to prevent drying of the substratum and at a temperature optimum for the particular type of seed. Many seeds, including clover and onion, must be kept cool to germinate; others, such as watermelon and cotton, must be kept warm. For good germination, seeds of bluegrass and many other plants must be kept cool at night and warm during the day. At suitable intervals, as the seeds germinate, the normal seedlings, which have all essential parts for a plant, are removed and the number recorded. Standard times for making counts of each kind of seed have been worked out.

Dormancy.—For varying periods after harvest, seeds of some species will not germinate under conditions usually favourable for germination. Such seeds are called "dormant" and must not be confused with dead seeds. The potential value of dormant seeds for planting can be determined by special procedures. For example, freshly harvested wheat will germinate if the temperature is lowered 5° to 10° C. below the usual temperature for germination, or if the seeds on the moist substratum are held at a few degrees above freezing for three to seven days. Freshly harvested bluegrass seeds and the seeds of some varieties of lettuce will germinate only after a suitable exposure to light.

Hard Seeds.—Seeds of some plants of the pea and morning-glory families have impermeable coats (the seeds do not absorb water when put on a moist substratum). These hard seeds are reported separately on the analysis label. Hard seeds usually are viable, and their viability can be checked by scratching the coat in some manner and then placing the seeds under suitable conditions for germination. The proportion of impermeable seeds of some crop plants increases with increased dryness of the surrounding air. The agricultural value of hard seeds varies greatly with the species or variety and the proposed use of the seeds.

Chemical and Other Tests.—Many attempts have been made to estimate the viability of seeds without actually germinating them. These methods depend on measuring by physical or chemical means some characteristic of a viable or of a dead seed. Because the development of a useful plant depends on many processes in many tissues of the seed, none of these quick methods has been entirely successful. The most nearly successful quick test, developed by Georg Lakon of Germany, depends on soaking the seeds in a weak solution of a salt of tetrazolium. Certain enzymes of the viable embryo reduce the colourless soluble salt to an insoluble red compound. The method has definite value for obtaining the approximate viability of seeds that require a long time to germinate,

but it does not promise to supplant the usual germination test.

Seed tests of other types are often required. With the development of many new varieties with special adaptations, it is increasingly important for a grower to know whether a seed lot is of the variety shown on the label. Usually the variety cannot be determined by examination of the seeds. Often it is necessary to plant the seeds and grow mature plants. Sometimes, however, the variety can be identified in the seedling stage, either directly from seedling appearance or from differential responses to special growing conditions such as temperature or length of day. For example, commercial seeds of annual rye grass having awns broken off cannot be separated by appearance from the awnless seeds of perennial rye grass, but they can be distinguished by a chemical property of the seedlings. The roots of young seedlings of annual rye grass growing on filter paper change the paper in contact with the roots so that it shows a blue fluorescence under an ultraviolet lamp in a dark room. Seedlings of perennial rye grass do not cause this change.

The region of origin of a seed lot can often be determined from the presence in the sample of weed seeds, incidental soil or rock particles characteristic of the region. It may be important to know the origin because seeds reproduced for many years in a given region may, by natural selection, have developed qualities that make them of little or of great value for crop production in some other region.

Some seed laboratories can determine whether seed-borne plant diseases occur in seed samples. The fungi causing the diseases are identified from spores washed from the seeds or developed on young seedlings. Sometimes distinctive symptoms develop on the seedlings.

The moisture content of seeds may be determined in connection with other seed tests. A sample for a moisture test must, of course, be submitted in a moisture-proof container. Interest in seed moisture arises from the knowledge that high moisture content is an important factor in the loss of seed viability in transit and in storage.

Statistics.—The special problems of statistics as applied to seed testing have received much attention. A knowledge of statistics and of normal distribution is important because the tests are made on a small sample of the seed lot. A second small sample will differ in composition from the first, and therefore it is important to know the amount of variation that normally may be expected. "Tolerances" have been applied in seed-law enforcement because of these expected variations of small samples.

Seed-Testing Associations and Rules.—Early in the development of seed testing it became obvious that many definitions and methods of procedure must be established and that these should be uniform in different laboratories. This need led to the organization of seed-testing associations that agreed upon rules and procedures. In 1896 a committee of the Association of American Agricultural Colleges and Experiment Stations described standard seed-testing apparatus and outlined methods of procedure for testing seeds. This information was published in 1897 as *Circular 34* of the United States department of agriculture. The formation of the Association of Official Seed Analysts of North America in 1908 assured more formal co-operation. This association, which consists of workers engaged in seed testing in the several states, the United States department of agriculture and the Canadian department of agriculture, continues to be increasingly active. The development of rules for the testing of seeds was one of the main objectives of this association. The first formal rules, adopted in 1917, have repeatedly been revised, brought up to date and published by the association. The United States department of agriculture also publishes rules and regulations as required for the enforcement of the Federal Seed act. These two sets of rules are in very close agreement.

Those doing seed testing for commercial seed companies have formed the Society of Commercial Seed Technologists. This group co-operates closely with the Association of Official Seed Analysts.

An International Congress of Seed Testing was held in Europe in 1905 and at several subsequent times. In 1924 the International Seed Testing association was organized; it continues to bring about uniformity of seed testing to facilitate international commerce in seeds. Procedures for testing seeds have been similar in the different countries, but differences in definitions and in interpretations have developed. As a result, reports from different countries on tests of the same lots of seed have not always been in agreement. The International Seed Testing association adopted international rules for seed testing at a congress in 1931, but it was not until 1953 that complete agreement was reached and truly uniform international rules were adopted.

Developments in Seed Research.—The improvement of seed testing to make it increasingly useful to agriculture has been based on seed research. Not only those engaged in testing seeds but also botanists interested in classification and plant physiologists interested in life processes have advanced knowledge about seeds.

Detailed studies of similar-appearing seed types have revealed characteristic differences of form, markings or structure by which the seeds can be distinguished. The results of such studies are the basis of illustrations, descriptions and keys used in the identification of seeds.

Other studies have shown that high seed moisture and high temperature during storage are causes of weakened seeds that eventually die. The weakened seeds produce seedlings of reduced vigour and also abnormal seedlings that cannot grow into useful plants. The difficulties of the interpretation of the results of germination tests are increased by seedlings of borderline value. It has been found that mechanical injuries to seeds during harvesting and processing often result in abnormalities and death.

The research of many workers has shown that the germination of different kinds of seeds is benefited by exposure to light during germination and also that the germination of a few kinds is prevented by exposure to light. It was found that the promotion of germination is caused by a comparatively narrow band of light in the red portion of the spectrum and that seeds promoted by this red light can be prevented from germinating by exposure to far-red light (near the limit of visibility). Whether a seed is promoted or inhibited by white light depends on the relative sensitivity of the particular kind to red or to far-red radiation.

Studies of the germination requirements of seeds of many species that are very difficult to germinate have demonstrated that many of these grow promptly after they are held on a moist substratum at a temperature just above freezing for from one to several months. Rose and apple seeds are typical examples.

Further research may reveal physical or biochemical methods of distinguishing similar-appearing kinds of seeds. The moisture and temperature conditions needed to keep seeds viable are known, but the changes within seeds that lead to weakness and death have had little study. The results of such studies would aid in developing better means of maintaining seeds at high vitality, thus reducing the amount of low-quality seed on the market. Very little is known about the changes within seeds that result in dormancy or about the internal processes that accompany the overcoming of dormancy. Further knowledge about such changes would make it possible to devise more dependable methods of determining seed viability. (E. H. T.)

See also references under "Seed" in the Index.

BIBLIOGRAPHY.—*General.*—A. S. Foster and E. M. Gifford, *Comparative Morphology of Vascular Plants* (1959); D. A. Johansen, *Plant Embryology* (1951); C. W. Wardlaw, *Embryogenesis in Plants* (1955); E. H. Toole *et al.*, "Physiology of Seed Germination," *Ann. Rev. Pl. Physiol.*, vol. 7 (1956); W. Crocker and L. V. Barton, *Physiology of Seeds* (1953); U.S. Department of Agriculture, *Seeds: The Yearbook of Agriculture 1961* (1961); J. P. Mahlstede and E. S. Haber, *Plant Propagation* (1957).

Seed Trade and Testing.—W. A. Wheeler and D. D. Hill, *Grassland Seeds* (1957); L. R. Hawthorn and L. H. Pollard, *Vegetable and Flower Seed Production* (1954); A. M. Lute *et al.* (History Committee) *History of the Association of Official Seed Analysts* (1941); Association of Official Seed Analysts, *Fifty Years of Seed Testing* (1908–1958) (1958); U.S. Department of Agriculture, *Manual for Testing Agricultural and Vegetable Seeds,* Agricultural Handbook, 30:1–440 (1952); Association of Official Seed Analysts, "Rules for Testing Seeds," *Proc. Ass. Off. Seed Analysts N.Am.*, 44:31–78 (1954); International Seed Testing Association, "International Rules for Testing Seeds," *Proc. Int. Seed Test. Ass.*, 21:1–80 (1956).

SEELEY, SIR JOHN ROBERT (1834–1895), English historian, apologist of the British Empire, and author of *Ecce*

Homo, a life of Christ that proved a landmark in liberal theology. Born in London on Sept. 10, 1834, he was educated at the City of London School, and (1852–57) at Christ's College, Cambridge, at which he became a fellow and classical lecturer. In 1863 he became professor of Latin at University College, London, and in 1869, as professor of modern history, returned to Cambridge, where he died on Jan. 13, 1895.

As a historian, Seeley learned accuracy in the Ranke school, but differed from Ranke in viewing history as the servant and teacher of the statesman, with a duty to take up an attitude to contemporary issues. His most solid historical work was probably *The Life and Times of Stein* (1878), but deservedly better known are *The Expansion of England* (1883) and *The Growth of British Policy* (2 vol. 1895; with memoir by G. W. Prothero). These won wide recognition for their treatment of the burning issue of colonial expansion: Seeley's view of the role of the colonial powers as trustees inspired succeeding generations.

In *Ecce Homo* (1865; anonymous but later acknowledged) Seeley presented Christ in human terms, ignoring, but not explicitly denying, the doctrine of his divinity; and in *Natural Religion* (1882), he set out the possibility of a Christianity without the supernatural, which he saw as an excrescence. Although he protested that these omissions did not imply disbelief, his works provoked controversy.

SEFERIS, GEORGE (Gr. Giorgos) (pen name of Georgios Seferiadis) (1900–), Greek poet, essayist and diplomat, who won the Nobel prize for literature in 1963, was born on March 1, 1900, in Smyrna (Izmir), Turk. He was educated in Athens, after his family went to live there in 1914. Later he studied law in Paris. In 1926 he joined the Greek diplomatic service, and served in London and Albania before World War II, during which he was with the free Greek government. After the war he held posts in Ankara and Beirut, and was Greek ambassador in London (1957–62).

Seferis was at once acclaimed as "the poet of the future" on the publication of *Strophe* ("Turning Point," 1931), his first collection of poems. It was followed by *Sterna* ("The Cistern," 1932), *Mythistorema* ("Mythical Story," 1935), *Himerologion katastromatos 1* ("Log Book I," 1940), *Tetradio gymnasmaton* ("Exercise Book," 1940), *Poïimata* ("Poems," 1940), *Himerologion katastromatos 2* (1944), the long poem *Kichle* ("Thrush," 1942), *Poïimata 1924–1946* (1950) and *Himerologion katastromatos 3* (1945). Selections of his poetry have been translated into English, French, German, Italian, Spanish and Swedish, the fullest English version being that by Rex Warner (*Poems*, 1960), which was awarded the William Foyle prize for poetry. Seferis also translated poems by T. S. Eliot and others into Greek, and wrote essays, of which the chief are *Dokimes* ("Essays," 1944) and *Erotokritos* (1946). He received the Kostis Palamas prize for poetry from the Academy of Athens in 1947. He is the most distinguished Greek poet of "the generation of the '30s," which introduced full-scale symbolism to modern Greek literature. His refined lyricism and the freshness of his diction brought a new breath of life to Greek poetry. His work is permeated by a deep feeling for the tragic predicament of the Greeks, as indeed of modern man in general. (Ce. A. T.)

SEGANTINI, GIOVANNI (1858–1899), Italian painter, known as a painter of the Alps and hard peasant life, was born at Arco in the Trentino on June 15, 1858. His mother died when he was four and his father went to Milan, leaving Giovanni in the care of a poor relative. At the age of seven the child ran away to the mountains and was employed by peasants as a herdsman. He spent long hours of solitude in drawing. His fame became known to the local authorities, who sent him to the school of art at Milan. In 1882 he settled in Brianza near Como and four years later sought the Swiss Alps, finally settling in the Engadine among the humble shepherds and farmers. There he gave himself up to the study of mountain life, and became in truth the painter of the Alps.

The "Ave Maria" took a gold medal at the Amsterdam exhibition (1883). The atmosphere of this picture is clear and crystalline, and his figures stand out in sharp relief. "The Drinking Trough" received a gold medal in Paris (1889) and "Ploughing in the Engadine" (Munich Pinakothek) gained a gold medal at Turin (1892). Besides those works in which he studied simple effects of light and Alpine scenery, such as "Midday on the Alps" and "Winter at Savognino," he also painted symbolical subjects: "The Punishment of Luxury" and the "Unnatural Mothers." He died on Sept. 28, 1899, at Maloja, Switz., where a museum contains many of his works, including the great unfinished triptych of "The Alpine World."

Bibliography.—L. Villari, *G. Segantini* (1901); F. Servaes, *G. Segantini* (1907); Gottardo Segantini, *G. Segantini* (1918), *Scritti e Lettere* (1910).

SEGESTA (Gr. Egesta), an ancient city of Sicily, about 2 mi. NW of modern Calatafimi. It was the chief city of the Elymi, a people for whom a Trojan origin was claimed by Thucydides; they are archaeologically indistinguishable in the Early Iron Age (first half of the 1st millennium B.C.) from their Sicanian neighbours (*see* Sicani). Segesta became in its material culture almost a Greek city, but it generally took the Phoenician side against its Greek neighbours; boundary disputes with Selinus (*q.v.*) were frequent from 580 B.C. onward. One of the ostensible objects of the Athenian expedition to Sicily in 415 was to aid Segesta against Selinus in a dispute about boundaries and rights of intermarriage. After its failure (413), the Segestaeans turned to Carthage; when in 409 Hannibal, son of Gisgo, invaded Sicily and sacked Selinus, Segesta became a Carthaginian ally. It was besieged by Dionysius in 397. In 307 Agathocles marched on the city, massacred many of its inhabitants, sold the rest into slavery, and changed its name to Dicaeopolis; but it soon recovered its old name and returned to the Carthaginians. Early in the First Punic War, however, the inhabitants massacred the Carthaginian garrison and allied themselves with Rome; they had to stand a severe siege from the Carthaginians (260 B.C.). Segesta was treated with favour by the Romans; it was a free city and the territory of Eryx may have been assigned to it.

The theatre, perhaps of the 3rd century B.C., and an unfinished Doric temple of the late 5th century B.C. are well preserved, and there are fragmentary remains of houses and the town walls.

See R. van Compernolle, "Ségeste et l'Hellénisme," *Phoibos*, vol. v, pp. 183–228 (1950–51); for the temple, *see* A. W. Lawrence, *Greek Architecture*, pp. 180 ff. (1957). (T. A.; T. J. Dn.; X.)

SEGHERS, CHARLES JEAN (1839–1886), Belgian Roman Catholic priest, missionary, twice bishop of Vancouver Island and archbishop of Oregon City, was born at Ghent on Dec. 26, 1839. Although his parents died while he was a young boy, Charles Seghers received a good education, and his preparation for the North American missions was made at the American Seminary of Louvain. He was ordained on May 31, 1863, and soon embarked for the diocese of Vancouver Island. His labours in this diocese were interrupted by his attendance at the first Vatican council as theologian to his bishop, Modeste Demers, whom he succeeded in March 1873. The new prelate continued his missionary activity in the northwest and Alaska until his appointment as coadjutor to the archbishop of Oregon City in 1878. His several visitations in his new archdiocese took him to Idaho, Montana, northern California and throughout Oregon. When the archbishop, Francis N. Blanchet, retired, he was succeeded by his coadjutor, who received the pallium at Portland, Ore., on Aug. 15, 1881. Archbishop Seghers' immediate task was the development of the Catholic school system. In 1883 he journeyed to Rome with other U.S. archbishops to prepare the schema for the proposed Baltimore council, which he attended in 1884. While he was in Rome, the see of Vancouver Island became vacant, whereupon Seghers offered to return to his former diocese since a successor was more readily available for Oregon than the more northern diocese. He repossessed the see of Vancouver Island on April 2, 1885. During a visitation to Alaska, he was murdered by a demented companion on Nov. 28, 1886.

See Maurice de Baets, *Apostle of Alaska: Life of C. J. Seghers*, Eng. trans. by Sister Mary Mildred (1943). (F. G. McM.)

SEGHERS (Segers), **HERCULES PIETERSZ** (1589/90–*c.* 1638), Dutch painter and etcher of stark, fantastic landscapes, was born in 1589 or 1590, probably in Haarlem, and became a

pupil of Gillis van Coninxloo in Amsterdam. In 1612 Seghers entered the painters' guild in Haarlem but was back in Amsterdam by 1614. In 1631 he resided in Utrecht, and in 1632 or 1633 he moved to The Hague. Early reports indicate that he was lonely and poor, even though several artists, including Rembrandt, gave evidence of their admiration for his work. Seghers' style contrasts strongly with the main aspects of the Dutch output of that period; most of his works would have been difficult to understand prior to their rediscovery in the 20th century.

The majority of Seghers' works represent forbidding mountain scenes with jagged cliffs, desolate valleys, broken tree trunks, and scant traces of human habitation, but he also painted and etched a smaller number of serene panoramas of his native countryside which deeply influenced Dutch painters of the "national" school. His etchings belong to the most original and impressive experiments in the history of print making. He used different-coloured inks and often printed on coloured or dyed paper and even on canvas; the diversity in the appearance of each individual print was increased by his adding accents by hand. The one large collection of original etchings, in Amsterdam, indicates Seghers' colouristic daring and technical skill. One plate was later owned by Rembrandt, who reworked Seghers' figures of Tobias and the Angel into those of a *Flight into Egypt*. Seghers' paintings are rare; few are documented, and a large number of forgeries exist. His *Mountain Landscape* in the Uffizi in Florence was formerly attributed to Rembrandt who seems to have altered parts of it.

Bibliography.—J. Springer, *Die Radierungen des Herkules Seghers* (*Graphische Gesellschaft, Veröffentlichungen*, no. xiii, xiv, xvi) (1910–12); L. C. Collins, *Hercules Seghers*, with many debatable attributions (1953); articles by J. G. van Gelder in *Oud Holland* (1950, 1953); Catalogue, *Hercules Seghers Exhibition*, Museum Boymans, Rotterdam (1954). (W. Sw.)

SEGOVIA, ANDRÉS (1893–), Spanish musician acclaimed as the foremost guitarist of his time, was born at Linares, near Jaén, Andalusia, on Feb. 17, 1893. He was self-taught as a musician, and first appeared as a guitarist at Granada in 1909 and later gave concerts at Barcelona and Madrid and in South America. In 1924, he won wide praise following a concert in Paris. Under the influence of Manuel de Falla and from the desire to establish the guitar as an instrument for the performance of cultivated music, he extended its repertory to include transcriptions of lute and harpsichord works by Couperin, Rameau, J. S. Bach, and others. He also inspired works for the guitar by several modern composers including Mario Castelnuovo-Tedesco, Joaquín Turina, and Albert Roussel. His repertory was extensively recorded.

See B. Gavoty, *Andrés Segovia* (1955).

SEGOVIA, a town and province in Old Castile, Spain. Pop. (1960 mun.) 33,360. The town lies north of the Cordillera Central (Cordillera Carpetovetónica), 3,297 ft. above sea level and 87 km. (54 mi.) N.W. of Madrid by road. Founded about 700 B.C., it was an important place in Roman times. It was taken at the beginning of the 8th century by the Moors, from whom Alfonso VI recaptured it in 1079. Thereafter Segovia enjoyed prosperity and a position of some importance in medieval Castile. During the reign of the emperor Charles V the revolt of the *comuneros* (common people) caused some destruction. An outbreak of plague at the end of the 16th century ushered in a long period of decadence, but the town's fortunes began to revive with the railway-building era in the 19th century.

There are two well differentiated areas: an upper town encircled by ancient walls situated on a narrow limestone ridge between the Eresma and its tributary the Clamores; and a lower part outside the walls. The centre of the old town is the Plaza Mayor (or del Generalísimo Franco) on one side of which stands the cathedral and from which streets lead into the outer suburbs. Crossing the Plaza del Azoguejo, the focal point of the town's communications, is the Roman aqueduct known as El Puente. It is one of the finest and best preserved examples of its kind, built probably in the reign of Trajan, from dark coloured Guaderrama granite without the use of mortar. Still in use, the aqueduct consists of a double tier of arches and carries water from the Río Frío for about half a mile across the valley outside the old town. At its highest point above the ground it stands 93½ ft. high.

J. ALLAN CASH—RAPHO GUILLUMETTE

ALCÁZAR, A CASTILIAN STRONGHOLD BUILT MAINLY IN THE 14TH AND 15TH CENTURIES AND EXTENSIVELY RESTORED AFTER 1862

The 16th-century Gothic cathedral was designed by Juan and Rodrigo Gil de Hontañón to replace the church destroyed during the revolt of the *comuneros*. Construction began in 1525 and was virtually finished by 1577. There are also several other fine ecclesiastical buildings including the Romanesque churches of San Esteban (12th–13th century) with a superb tower; San Martín (12th century); la Trinidad (11th–12th century); San Lorenzo (12th century); San Millán (12th century); Vera Cruz (13th century), the former church of the Knights Templar; and San Juan de los Caballeros (11th and 12th centuries), now the Museo Zuloaga of ceramics. Across the Eresma is the Gothic Hieronymite monastery of El Parral and also outside the walls is the Dominican convent of Santa Cruz (15th century) founded by Ferdinand and Isabella and now a welfare centre. In the Carmelite convent (17th century) is the tomb of St. John of the Cross.

The Alcázar, mentioned as early as the 12th century, was the fortified palace of the kings of Castile. Most of the original palace was destroyed by fire in 1862 but has since been extensively restored. There are many houses and palaces built between the 15th and the 17th centuries. Modern buildings include the palace of justice and the offices of the civil government. Segovia has an artillery academy in the former convent of San Francisco, a provincial museum and an archaeological museum.

In the late middle ages Segovia was the centre of a flourishing textile industry. As this declined agriculture took its place. The town now has factories for rubber, pottery, flour, biscuits, artificial fertilizers, cement, chemical products, etc.

Segovia is linked by road and rail with Madrid and also with the north of Spain.

Segovia Province (area 2,683 sq.mi.; pop. [1960] 195,602), formerly part of Old Castile, is bounded north and northeast by Burgos and Soria, southeast by Guadalajara and Madrid, southwest by Ávila and northwest by Valladolid. It is primarily an agricultural tableland, 2,500 ft. above sea level, producing wheat, rye, barley, hemp, flax and vegetables, and also rearing sheep, cattle, mules and pigs. The sierras in the province are quarried for granite, marble and limestone. The province is traversed by the Madrid-La Coruña railway which passes through its capital, the only city of importance.

Segovia was the scene of heavy fighting during the first months of the Spanish civil war. (M. G. Sa.)

SEGRÈ, EMILIO GINO (1905–), U.S. physicist, cowinner of the Nobel Prize in physics for 1959 with Owen Chamberlain for discovering the antiproton, was born in Tivoli, Italy on Feb. 1, 1905. He took his doctor's degree in physics at the University of Rome in 1928. He served on the staffs of the Universities of Rome and Palermo. After 1946 he was professor of physics at the University of California at Berkeley. After 1948 he

was also at the Lawrence Radiation Laboratory of the University of California. Segrè became a U.S. citizen in 1944.

An associate of Enrico Fermi when the epoch-making discovery of slow-neutron effects was made in Rome in 1934, he discovered (1937) the first artificial element, technetium, while at Palermo, using materials irradiated in one of the Berkeley cyclotrons. At Berkeley he made the first chemical separation of nuclear isomers, changed by chemical means the radioactive decay constant of beryllium and took part in the discovery of the element astatine and of plutonium-239, of which he determined the slow-neutron fissionability. He conducted extensive researches in high energy physics, including the discovery, with Chamberlain, Clyde E. Wiegand, and Thomas Ypsilantis, of the antiproton in 1955. In the 1960s he was contributor of several articles to *Encyclopædia Britannica*.

See also ANTIMATTER.

See T. N. Levitan, *Laureates: Jewish Winners of the Nobel Prize* (1960). (E. M. Mc.)

SEGREGATION, RACIAL, the practice of restricting people to certain limited areas of residence or to separate institutions (schools, churches) and facilities (parks, playgrounds, restaurants, restrooms) on the basis of race. In the United States such restrictions developed primarily as one aspect of relations between whites and Negroes, especially in the South, although other groups, such as Chinese and Japanese, have at times been affected. But segregation and the establishment of colour bars are not confined to the relations between white and coloured populations; coloured races—Asian Mongols, African Bantus, American Aztecs—have been great conquerors throughout the ages and have practised discrimination including segregation of subject races. The most extreme system of segregation, in which occupation as well as race and religion were equally important elements, prevailed in the caste system of Hindu India (*see* CASTE [INDIAN]). The history of racialist theory and thinking and studies of racial attitudes are dealt with in the article INTERRACIAL RELATIONS. (*See* also APARTHEID and related material under ANTI-SEMITISM; CLASS, SOCIAL; GHETTO.) For the U.S. Supreme Court decision in the case of *Brown* v. *Board of Education* (1954) declaring racial segregation in public schools to be unconstitutional, *see* EDUCATION, HISTORY OF. (*See* also *Education* sections of articles on individual states.)

Segregation appears always as a means of maintaining the economic advantages and the superior social status of the politically dominant racial group. This does not lead necessarily to spatial or physical segregation because the social segregation of the subordinate group may be maintained through a system of etiquette and symbols and exclusion from intimate social contacts that would reduce the social distance between the races. In an urban industrial society it becomes more difficult to maintain the physical and spatial segregation of races, and more dependence must be placed upon institutions to maintain it. Consequently, as the result of segregation, the races live in different social worlds and communication between them is restricted no matter how close they may be physically.

Racial segregation as an ecological process should be differentiated from racial segregation which is the result of a conscious or deliberate social policy. As an ecological process racial segregation is in a sense a natural process since it results from a relatively impersonal competition between races for space or land. An example of racial segregation as an ecological process is the concentration of Negroes in the areas of the U.S. South known as the Black Belt. These areas were congenial to development of the plantation system of cotton production (*see* SOUTH, THE). Ecological segregation may also be seen in cities where in the absence of legal enactments or social pressure various racial and cultural groups, because of their economic status, have created Little Sicilies, Jewish ghettos and black belts. It should be noted, however, that racial segregation as an ecological process is not a purely biotic or symbiotic phenomenon as among subhuman living organisms. Among human beings there is always social interaction involving the influence of customs and laws and human sentiments. During the process of ecological segregation human beings seek close association with those who have the same interests or similar traditions and values. Even within the same racial and cultural communities there is ecological segregation on the basis of age, sex, education, occupation and income. Hence, the social process can never be completely separated from the ecological process except for the purpose of analysis.

As far as the economic relations of races are determined solely by competition, a racial division of labour may emerge which results in the "segregation" of racial groups in different occupations. The racial division of labour is rarely due to an impersonal competitive process but more generally reflects the distribution of power in a community. Thus, the system of Negro slavery and the economic organization of life in the U.S. South as well as in the colonial areas in Africa represented a racial division of labour, but in all these cases the relation between that division and the distribution of power is clear.

When, however, members of different races have access to the same education the racial division of labour tends to break down. If competition on an individual as opposed to a racial basis is not permitted racial conflicts arise and the politically dominant race attempts to maintain its superior status by establishing a system of racial segregation.

The distribution of power in a biracial community is not only important in determining the economic relations of races but it is of equal importance in determining their social relations or their status in the social order. In fact, race relations can only be said to exist where, as sociologist Robert E. Park has stated, "racial differences enter into the consciousness of the individuals and groups so distinguished, and by so doing determine in each case the individual's conception of himself as well as his status in the community."

The phenomenon of racial segregation has appeared in all parts of the world where there are biracial communities except where racial amalgamation has occurred on a large scale, as in Hawaii and Brazil. In Brazil, where there has never been a system of racial segregation, the dominant economic and political position of the whites was never challenged by the rise as a group of persons of Negro derivation to economic and political power. In the transformation from a feudal to a middle-class society in the 19th century the mulatto was the most mobile element in the military, economic and political sectors of Brazilian society. (*See* BRAZIL: *The People*.) On the other hand, in the southern states of the United States the emancipation of Negroes brought them into economic competition with the white working class, and during Reconstruction following the American Civil War it was the political power of the Negroes that helped to consolidate the power of northern industrial capitalism. Since the proportion of Negroes in the population of the South has decreased and their concentration in the Black Belt has declined there has been some relaxation of segregation in public relations. But the white working class continued to refuse to permit Negro workers to compete, especially in white collar occupations where status is important, and whites of all classes continued to exclude Negroes from sharing in political power. The Civil Rights Act of 1964 established an Equal Employment Opportunity Commission, extended the Commission on Civil Rights and barred job discrimination in federally assisted programs. The act also contained provisions against discrimination and segregation in voting, education and the use of public facilities; and the Negro Voting Rights Act of 1965 was designed to assure Negroes the right to register and vote by, among other provisions, suspending literacy tests used to discriminate against them.

The problem of racial segregation is present in the multiracial communities of central Africa and east Africa but it has been most acute in South Africa. In South Africa the National party has attempted to maintain the dominance of the white minority through the policy of apartheid or racial segregation in the spatial, economic, political and social relations of the races. The white minority had already undertaken to maintain its economic position and social status by a colour bar in industry. The urbanization of the native or Bantu population and the dependence of industry upon native labour tend to nullify the spatial and economic segre-

gation. But the whites have been able through force and the threat of force to exclude the natives and other coloured peoples—Cape Coloureds (*q.v.*) and Asians—from political power, and to maintain a system of racial segregation in all social relations.

Whereas in the United States it appears that racial segregation in the public relations of the races tends gradually to disappear where Negroes offer no threat to the economic and political power of whites, in South Africa racial segregation appears as a desperate attempt on the part of the white minority to maintain its dominance despite the economic and political forces in the situation which are opposed to racial segregation.

See also NEGRO, AMERICAN and references under "Segregation, Racial" in the Index.

BIBLIOGRAPHY.—R. E. Park, "Human Ecology," *Human Communities*, the collected papers of Robert E. Park, ed. by E. C. Hughes *et al.*, vol. ii, pp. 145–158 (1952), "The Nature of Race Relations," *Race and Culture*, pp. 81–116 (1950); H. Blumer, "Social Science and the Desegregation Process," *The Annals of the American Academy of Political and Social Science*, 304:137–143 (March 1956), Arnold M. Rose (ed.), "The Negro Protest," section, *ibid.*, 357:1–126 (Jan. 1965); E. W. Burgess, "Racial Segregation in American Cities," *The Annals*, cxxxx, pp. 105–115 (Nov. 1928); R. C. Weaver, *The Negro Ghetto* (1948), *The Urban Complex* (1964); R. P. Warren, *Segregation* (1956); O. R. Williams, *Segregation and Common Sense* (1961); J. G. Cook, *The Segregationists* (1962); M. M. Grossack (ed.), *Mental Health and Segregation* (1963); N. Glazer and D. P. Moynihan, *Beyond the Melting Pot* (1963); M. M. Gordon, *Assimilation in American Life* (1964); P. I. Rose, *They and We* (1964); W. D. Avra, *Segregation, Integration, and the Constitution* (1964); H. H. Humphrey (ed.), *Integration vs. Segregation* (1964); K. B. Clark, *Dark Ghetto* (1965); I. A. Newby, *Jim Crow's Defense* (1965); O. Schroeder and D. T. Smith (eds.), *De facto Segregation and Civil Rights* (1965); Loren Miller, *The Petitioners: the Story of the Supreme Court of the United States and the Negro* (1966); A. I. Waskow, *From Race Riot to Sit-in: 1919 and the 1960s* (1966). *See* also Ellen Hellmann (ed.), *Handbook on Race Relations in South Africa*, ch. vii, xi (1949); M. Hailey, *An African Survey, Revised 1956* (1957); L. Kuper *et al.*, *Durban: a Study in Urban Ecology* (1958), *African Bourgeoisie* (1965); A. Blaxall, *Suspended Sentence* (1965). (E. F. F.)

SÉGUIER, PIERRE (1588–1672), chancellor of France in the critical period during which absolutism was consolidated. He was born in Paris on May 29, 1588, into a very prosperous family of lawyers. In 1618 he became one of the *maîtres des requêtes de l'hôtel*, the body from which the intendants (*q.v.*) were mostly recruited; and in 1621 he was appointed intendant, first for Auvergne, then with the army in Aunis and Saintonge. In 1624 he succeeded his uncle, Antoine Séguier, as one of the presidents in the Parlement (*q.v.*) of Paris.

The chancellor (*q.v.*) in France was an irremovable dignitary; but if a chancellor displeased the royal government, he could be relieved of his functions by a keeper of the seals. Séguier succeeded the marquis de Châteauneuf (Charles de l'Aubespine) as keeper of the seals in 1633 and became chancellor on the death of Etienne d'Aligre in December 1635.

Industrious subservience to the cardinal de Richelieu (*q.v.*) kept Séguier in function. In 1637 he directed the enquiry into the Affair of Val de Grâce, which compromised the queen of France, Anne (*q.v.*) of Austria. In 1639 he was, extraordinarily, sent to supervise the repression of the revolt of the *Va-nu-pieds* in Normandy. In 1642, at Lyons, he presided over the trial of Cinq-Mars (*q.v.*).

When Richelieu died (1642), Séguier succeeded him as official "protector" of the Académie Française, the sessions of which thereafter took place in his house. To the surprise of many, he was retained in function by Anne of Austria when she became regent for the child Louis XIV in 1643. Serving Cardinal Mazarin (*q.v.*) as he had served Richelieu, he was hated by the malcontents in the Parlement and by the populace. During the Fronde (*q.v.*), he was assailed by a mob in Paris in August 1648; Anne was obliged to transfer the seals back to Châteauneuf in 1650; and though Séguier recovered the seals after the momentary replacement of Châteauneuf by Mathieu Molé (*q.v.*) in April 1651, they were given to Molé again in September. In the last phase of the Fronde in Paris, Séguier was aligned with the rebel princes till August 1652. On Molé's death (1656) the seals were restored to Séguier.

Under the personal rule of Louis XIV, Séguier in December 1662 was put in charge of the trial of Nicolas Fouquet (*q.v.*), which he conducted brutally enough to secure a verdict against Fouquet. From 1665 he presided over the new Council of Justice for the reform of the legal system: though this was in fact dominated by J. B. Colbert, the detail of the Civil Ordinance or "Code Louis" (1667) and of the Criminal Ordinance (1670) owed much to Séguier.

Séguier died at Saint-Germain on Jan. 28, 1672. His voluminous papers are preserved partly in Paris, partly in Leningrad (numerous manuscripts were taken to Russia by the diplomat P. Dubrovski in 1800). The Leningrad papers were exploited by the Soviet historian B. F. Porshnev for his work on "Popular Risings in France before the Fronde" (1948), in which he sought to interpret 17th-century France solely in terms of "the class war." This gave rise to a resonant controversy between Porshnev and the French historian Roland Mousnier, whose monumental *Lettres et Mémoires adressés au chancelier Séguier, 1633–1649*, appeared in 1964.

SÉGUR, SOPHIE ROSTOPCHINE, COMTESSE DE (1799–1874), Russian-born French writer of famous books for children. Born at St. Petersburg on July 19, 1799, the daughter of Fedor Rostopchin (*q.v.*), she became a member of the French nobility through her marriage to Eugène, comte de Ségur, in 1819. In Paris she wrote several volumes for the *Bibliothèque rose*, a popular series of children's books started in the mid-19th century by the publisher Hachette. These were based on stories told to her children and grandchildren, with whom she was unable to play because she had become an invalid. The central character of many of her stories is "Sophie," a young girl whose trials and tribulations are related in a simple prose that is eminently suitable for reading aloud (*e.g.*, *Les Malheurs de Sophie*, 1859). Other favourite volumes, especially among French children, include *Nouveaux Contes de fées* (1857), *Mémoires d'un âne* (1860), and *Après la Pluie le beau temps* (1871). She died in Paris on Feb. 9, 1874.

BIBLIOGRAPHY.—J. Chenevière, *La Comtesse de Ségur, née Rostopchine* (1932); A. de Pitray, *Sophie Rostopchine, comtesse de Ségur, racontée par sa petite-fille* (1939); M. de Hédouville, *La Comtesse de Ségur et les siens* (1953).

SEGURA, a river of southeast Spain, rises in the Sierra de Segura and flows east for 212 mi. (341 km.) through the driest region of the Iberian Peninsula to enter the Mediterranean south of Alicante. Much water is drawn off from the Segura and its major tributary the Guadalentin (Sangonera) to irrigate the *huertas* (market gardens), especially the Vega de Murcia. Winter floods are notorious on both rivers but during the summer the stream beds are almost dry. Eight dams are being built to control the runoff, the largest being at Fuensanta. About 300,000 ac. (more than 121,000 ha.) are irrigated or under development, with further plans for irrigation on the Cartagena plains. (J. M. Ho.)

SEHESTED, HANNIBAL (1609–1666), Danish statesman who laid the foundations of a modern administrative system in Denmark, was born at Arensborg Castle on the island of Ösel (now Saaremaa). Educated at Sorø Academy he traveled abroad from 1629 to 1632, and on his return to Denmark was attached to the court of Christian IV. In 1635 he conducted the unsuccessful negotiations with the Swedish chancellor Axel Oxenstierna for an alliance with Sweden. Soon one of the favourites of Christian IV, in 1642 Sehested married the king's daughter Christiane. He became a member of the *rigsraad* in 1640, and in 1642 was appointed *statholder*, or viceroy, of Norway. There he displayed great energy in developing Norway's material resources and reorganizing its armaments and fiscal system, intending to make Norway more independent of Denmark.

During Christian IV's second war with Sweden (1643–45) Sehested invaded Sweden several times and defended Norway from attack. After the Peace of Brömsebro (1645) he continued to develop Norway's resources and during the years 1646–47, with Christian's support, he obtained for Norway partial control of its own finances. The revenues were used to cover the expenses of the Norwegian militia and to pay off the war debt. Sehested's

success and his personal accumulation of money and honours, however, aroused the distrust and envy of other members of the *rigsraad*. After the death of Christian (February 1648), Sehested was left without support when charged with embezzlement and peculation, and he was forced to surrender his private property in Norway to the crown. In 1651 he retired from the *rigsraad*, and although Frederick III granted him a pension, Sehested lived abroad from 1651 to 1658.

In the summer of 1657 at the time of the war with Sweden Frederick refused the offer of Sehested's services. During the crisis of the war in 1658 (*see* FREDERICK III, [king of Denmark and Norway]) Sehested was at the headquarters of Charles X of Sweden. In thus seeking the protection of Denmark's worst enemy, Sehested approached the verge of treason, but when he saw that Denmark was successfully resisting the Swedish attacks, he began to work in Sweden for Danish interests. In 1660 Frederick III allowed him to return to Copenhagen, and made him plenipotentiary to negotiate with the Swedes. The Treaty of Copenhagen (May 1660), which saved the honour of Denmark, was very largely Sehested's work. He was also prominent during the crisis in the autumn of 1660 which resulted in the conferring of absolute power on Frederick. The king now trusted him completely. Appointed lord high treasurer (*rigesskatmester*) and councilor of state, Sehested proceeded to lay the foundation of a modernized central administration, establishing boards in the different branches of the government (*see* DENMARK: *History*). He died at Paris on Sept. 23, 1666.

BIBLIOGRAPHY.—J. A. Fridericia, *Adelsvaeldens sidste Dage* (1894); C. O. Bøggild Andersen, *Statsomvaeltningen i 1660* (1936), *Hannibal Sehested* (1946). (As. F.)

SEHORE, a town and district in the Bhopal division of Madhya Pradesh, India. The town, headquarters of the district, lies near the confluence of the Siwan and Latia streams, 20 mi. W.S.W. of Bhopal on the Bhopal-Ujjain railway. Pop. (1961) 28,489. It became the headquarters of the British agent of the former princely state of Bhopal in 1818. It is an important trade centre, and an annual fair called Hardaul Lala mela is held in December.

SEHORE DISTRICT (area 3,600 sq.mi.; pop. [1961] 754,684) was formed by the merger of the former Bhopal state with Madhya Pradesh. It extends to the eastern margin of the Malwa (*q.v.*) plateau drained by the Betwa river. The soil is mainly black cotton soil; 44% of the total area is arable and 24% under forests. Wheat, jowar, cotton and linseed are the principal crops. There is a heavy electrical equipment plant at Bhopal (*q.v.*), the capital of Madhya Pradesh. Other industries include cotton textiles, sugar, paper and straw products. (S. M. A.)

SEICHE, a free oscillation of fluid in an enclosed or semi-enclosed basin. The word originated centuries ago (supposedly from Lat. *siccus*, "dry," hence exposed) to describe the occasional tidelike rise and fall of water at the narrow end of Lake of Geneva, Switzerland. Its rhythmic movement was noted first by N. F. de Duillier (1730); that it was common to other lakes and related to atmospheric conditions was observed by Jean Pierre Vaucher (1803). But to F. A. Forel (1869–95) belongs the credit for recognition of its essential character and origin. Forel's classic work incited continuing observational and hydrodynamic studies of lake and sea oscillations of which G. Chrystal's (1904–09) are most noteworthy. Refinements to Chrystal's theory were contributed by A. Defant (1929) and J. Proudman (1953).

Forel recognized that a pure lake seiche is in effect an interaction of two equal long waves, traveling in opposite directions. Their resultant is a standing wave that remains fixed in position while the water surface seesaws about nodal line or lines, at still water level. Between nodes, peaks and troughs are known as antinodes. If the distance between antinodes is a multiple or submultiple of the length of the lake (for uniform depth), the standing wave fits exactly into the length. It is then capable of resonating since repeated reflections contribute additional synchronous standing waves which accumulate amplitude. Only bottom friction and viscous turbulence prevent indefinite amplitude-growth. Such a resonant standing wave system constitutes a seiche. Equal opposing progressive waves, that are not commensurate with the lake length, produce interfering standing waves, on repeated reflection, tending to annul each other. Thus when the water in a basin is disturbed, incommensurable wave motions die out rapidly while commensurable ones develop as persistent seiches.

The fundamental seiche in a closed basin is uninodal. Forel discovered that binodal, trinodal and other multinodal seiches were often coexistent with the fundamental. These several modes of oscillation are higher harmonics of the fundamental seiche, the binodal seiche, for instance, being second harmonic. The periods of harmonics (time intervals between successive antinodal peaks) are usually integral submultiples ($\frac{1}{2}$, $\frac{1}{3}$, . . .) of the period of the fundamental; departure from this is a feature of irregular, variable-depth basins. Periods of seiches depend exclusively on basin-dimensions and mode of oscillation. As examples, the fundamental, second and third harmonic periods of Lake Constance (Switz.) are 55.8, 39.1 and 28.1 min.; of Loch Earn (Scot.) 14.5, 8.0 and 6.0 min. Lake Erie has a fundamental period of 13.1 hours; the Aral Sea 22.8 hours. In general, periods are proportional to basin length L, and inversely proportional to the square root of water depth d.

Unlike lakes, coastal bays and inlets require the node of any seiche to be at the mouth where the basin communicates with a larger body of water. The even harmonics of such basins are normally suppressed and periods thus tend to the harmonic order of 1, $\frac{1}{3}$, $\frac{1}{5}$, ...

An important property of pure seiches is that water-particle movements are all linear synchronous oscillations. At the nodes surge movements are horizontal, and greater in range than the antinodal vertical range by a factor of $L/\pi d$. At the antinodes, heaving movements are entirely vertical. Between these extremes, magnitudes and directions change gradually.

Seiches tend to develop in bodies of water during or after imposition of agitating forces. Forel, Chrystal and others identified forms of excitation as (a) release of pent-up water at a leeward shore through lapse of onshore winds, (b) heavy rain, snow or hail over a portion of a lake, (c) rapid change of air pressure through passage of a squall, (d) flood discharge from rivers at one end of a lake, (e) impacts of wind gusts on the water surface, (f) disturbances from earth tremors (*see* EARTHQUAKE). More gradual energizing sources are (g) variations in wind velocity and pressure and (h) passage of small barometric fluctuations, both synchronizing approximately with seiche periods.

Seiches in certain coastal ports have caused ships to break moorings and sustain damage under action of nodal surge currents. Such seiches are usually of two distinct kinds: those of fairly long period (15 to 60 min.) attributable to (h), and those of much shorter period (0.3 to 15 min.) attributable to ingress of long period waves originating from distant cyclonic storms through causation (g). "Surf beats" afford another mechanism for coastal seiches as result of greater water transport shoreward from groups of high waves than from groups of low waves.

Internal seiches at the stratified interface (thermocline) of lake waters were discovered by E. R. Watson and E. M. Wedderburn (1904–12). Amplitudes are larger than surface seiches and periods longer—Lake Baikal, for instance, has an internal seiche period of 38 days, amplitudes up to 75 metres.

See J. Proudman, *Dynamical Oceanography* (1953); Basil W. Wilson, "Origin and Effects of Long Period Waves in Ports," *Communication 1, Section II, XIXth International Navigation Congress, London, 1957* (1957). (B. W. W.)

SEIGNIORAGE is in general a term denoting something claimed or taken by virtue of sovereign prerogative. Specifically, it is the charge over and above the expenses of coinage, or brassage, deducted from the bullion brought to a mint to be coined. From early times coinage was the prerogative of kings, who prescribed the total charge and the part they were to receive as seigniorage. Seigniorage, as a royal revenue, was varied according to the sovereign's caprice, avarice, or necessity. The deduction was sometimes supplemented by replacing part of the bullion with base metal, resulting in debased coinage. As the seigniorage and coinage charges were collected by withholding part of the bullion brought for coinage, the currency value of

the coins received in exchange was often less than the bullion's market value. In consequence, merchants did not bring bullion to the mint and the supply of coins was inadequate. In England, therefore, all charges for coinage were abolished in 1666.

From early times coins of gold or silver had an intrinsic value close to their face value. In fact, as pieces of bullion of guaranteed weight and fineness, they served as a medium of international as well as internal exchange. Nowadays, bills of exchange and bank notes perform this function, and coins are merely token money, issued for domestic purposes only. Thus, coins need no longer possess a high intrinsic value, and low-standard silver or certain base metal alloys provide all the qualities required. A substantial margin usually exists between the cost of producing a coin and its statutory currency value; this margin or profit is known as seigniorage. (H. G. St.)

SEINE, the second longest river of France (482 mi.) and economically the most important, rises on the Plateau de Langres 18 mi. NW of Dijon, and reaches the English Channel in a wide estuary which passes between Le Havre and Honfleur. With its several large tributaries, it represents a relatively mature river system, with gentle gradients throughout. It has a drainage area of more than 30,000 sq.mi. and for most of its course flows transverse to the general geological structure of the Paris Basin. Since it rises on Jurassic (Oolitic) limestone (at about 1,500 ft.), the uppermost part of its course is not permanent, and in dry summers the valley may be dry as far down as Châtillon, where it leaves the limestone and crosses a narrow strip of Oxford clay. It then enters a narrow wooded valley through the escarpment of the Corallian limestone; the valley widens temporarily at the succeeding outcrop of Kimmeridge clay, where the Ource River enters from the east. At Bar-sur-Seine, the river cuts through the Portland limestone, beyond which it enters the broad lowlands of *Champagne humide* (mainly the outcrop of the Gault clay), before encountering the Chalk escarpment at Troyes. The course then follows a broad entrenched valley across the chalk area, *Champagne pouilleuse*, until it is joined by its right-bank tributary, the Aube River (*q.v.*). From there the Seine follows a southwesterly direction at the foot of the Tertiary escarpment to Montereau, where the Yonne River joins it, and then passes into the Tertiary limestone uplands, which rise in steep wooded hillsides 200–300 ft. above the river. From there to the sea, the Seine receives only minor left-bank tributaries, but larger rivers such as the Marne (at Paris) and the Oise on the right. Below Paris, the river begins its series of huge swinging meanders, and gradually reencounters the Chalk, which appears as cliffs wherever the valley sides are undercut by the river. The meanders add about 120 mi. to the direct distance between Paris and Le Havre. Below Rouen, the valley widens with extensive alluvial flats. In the estuary, which begins near Tancarville (the lowest bridging point), there are strong currents, and a tidal bore associated with exceptionally high tides.

The Seine carries the heaviest tonnage of shipping of all the French rivers; the most important section is Le Havre-Rouen-Paris. The lower-middle course of the river has a regular regime resulting from the large drainage area. Low water occurs in summer, while the maximum discharge is in winter and early spring, aided by snow melt and reduced evaporation losses. The concentration of several large rivers in the Paris district brings the danger of winter flooding. The approximate maximum discharge at Paris is about 100,000 cusecs. The head of navigation is at the Aube confluence, though from there upstream to Troyes a lateral canal provides a minimum depth of 5 ft. Between Paris and the Aube confluence, there are numerous cuts to straighten the river, and 39 locks maintain a minimum depth of 6.5 ft. At Paris, the minimum depth increases to 9.8 ft. Rouen can receive sea-going craft with 25 ft. draft at mean high-water springs. Below Tancarville, river craft tend to use the canal to Le Havre to avoid the strong currents of the estuary. The Seine is linked by canal with the Somme, Scheldt (Escaut), Meuse, Rhine, Saône, and Loire rivers.

Paris is the third largest French river port (after Rouen and Strasbourg). The river port begins at Nanterre, and wharves and quays extend upstream for more than 25 mi. Within Paris the river is divided by several islands, and the 27 bridges of the city are a serious obstacle to some ships at high water. The main incoming traffic consists of coal, iron and steel, building materials and agricultural produce. Rouen is the outport of Paris, 145 river mi. downstream; oil and coal are the principal imports, the former proceeding to the important Shell refineries there. Le Havre on the north shore of the Seine estuary is capable of handling large ocean-going liners, and possesses shipbuilding and engineering industries. (C. Em.)

THE SEINE ALONG QUAYS OF THE ÎLE DE LA CITÉ, PARIS. THE RIVER DIVIDES INTO TWO ARMS AROUND THE ISLAND

SEINE-ET-MARNE, a *département* of northern France formed in 1790 from parts of the Île-de-France and Champagne, and surrounded by the *départements* of Oise to the north, Aisne, Marne, and Aube to the east, Yonne and Loiret to the south, and to the west by Essonne, Val-de-Marne, Seine-Saint-Denis, and Val-d'Oise, four of seven new *départements* created in 1964 from the former Seine and Seine-et-Oise. Seine-et-Marne itself has not been altered in the regrouping of administration areas in the metropolitan region. Area 2,286 sq.mi. (5,917 sq.km.). Pop. (1962) 524,486. The *département* lies east of Paris, near the heart of the Paris Basin, and is traversed by the great valleys of the Seine and Marne rivers as they converge toward Paris. Between the two main valleys lies the Tertiary limestone platform of Brie, dissected by the Grand Morin, Petit Morin, Yerres (Yères), and other tributaries of the Seine and Marne. North of the Marne valley stretches the similar limestone platform of Valois, but south of the Seine are sandy tracts traversed by the Loing River. On its left bank as it approaches the Seine is the forest of Fontainebleau, occupying the park of an old royal palace, much frequented now by Parisians. On its right bank another infertile sandy area, the Gâtinais, extends toward the Yonne River. Elsewhere the *département* consists of rich farmlands and, in spite of its proximity to the capital, it remains primarily agricultural. The grainfields of the limon-covered limestone platforms have long supplied Paris. Wheat and sugar beet are the main crops, with oats, potatoes, and green fodders also important, and a heavy density of livestock is maintained on the large farms. Brie cheese is well known and is made especially at Meaux and Melun. The river valleys are lined with lush meadows, market gardens, and nurseries; and orchards surround the substantial villages at the foot of the steep valley sides. The valley towns are

market centres and river ports that have become the nuclei for residential development and a modest growth of industry. Chief among them are Montereau and Melun on the Seine, Coulommiers on the Grand Morin, and Meaux on the Marne, while in the east, toward Champagne, is Provins. The limestone has been extensively quarried and is much in evidence in local building. In the southwest the Fontainebleau sands have been worked for glass-making, and elsewhere gypsum has been quarried. Melun (*q.v.*) is the *préfecture* of the *département*, which is divided into three *arrondissements*, centred upon Melun, Provins, and Meaux. It forms the diocese of Meaux, and comes under the court of appeal as well as the *académie* of Paris. (AR. E. S.)

SEINE-MARITIME, a *département* of northern France, formed in 1790 from the eastern part of the ancient province of Normandy. Pop. (1962) 1,035,844. Area, 2,449 sq.mi. (6,342 sq.km.). The *département*, formerly called Seine Inférieure, includes the mouth of the Seine and its lower valley below Elbeuf, together with the districts along the English Channel coast to the east. It is bounded east by the *départements* of Somme and Oise, and south by Eure. The Seine, approaching its estuary, flows in sweeping meanders which cut across rolling chalk plateaus. The main river valley and the side valleys that wind back into the plateau are heavily industrialized, forming the Rouen textile region, but the chalk plateau to the north, the Pays de Caux, remains basically agricultural. The chalk is extensively covered with fertile *limon* and is largely arable, devoted especially to wheat and sugar beet. Potatoes are important and some flax is still grown. Behind Dieppe the chalk has been breached in an anticlinal structure to expose the underlying clays along the Béthune Valley. This diversified district, the Pays de Bray, has rich pastures that support dairy farms and is well known for its cheese. The farming country of eastern Normandy for the most part turns its back upon the sea, which is fronted by high chalk cliffs, but there are some fishing ports, notably Fécamp (*q.v.*), which is engaged in distant trawling as well as in herring fishing, and resorts such as Le Tréport and Yport, while Dieppe (*q.v.*) provides an important passenger link across the Channel with England (via Newhaven). Rouen (*q.v.*), situated at the lowest bridging point on the Seine, although it is 90 mi. (145 km.) upstream from the sea, is still an important port. Le Havre (*see* HAVRE, LE), created in 1517 as a planned town on the alluvium that flanks the north coast of the Baie de la Seine, has developed as the deepwater port of this ocean gateway of Paris. Besides its great commercial docks, it has an important shipbuilding industry and engineering and chemical works. The strip between Le Havre and Rouen has become the major area of oil refining in France, with large refineries at Gonfreville, Gravenchon, Port-Jérôme, and Le Petit-Couronne, which are connected with Paris by pipeline as well as by the river. The old established textile industry of the lower Seine, centred upon Rouen, has become especially concerned with cotton goods, its raw material imported through Le Havre, but Elbeuf (*q.v.*) is important for woolen manufacture, and the new man-made fibres are also represented.

Rouen, the historic capital of Normandy, is the *préfecture* of the *département*, which is divided into three *arrondissements* (Rouen, Le Havre, and Dieppe), and is the seat of its court of appeal. The *département* forms the diocese of the archbishop at Rouen, but for educational administration comes under the *académie* of Caen. The ports and towns suffered severe damage from bombardment during World War II, which took toll of Rouen's fine Gothic architecture as well as of modern housing and factory buildings. Apart from its magnificent cathedral, the law courts and the church of Saint-Ouen are especially noteworthy at Rouen, which also has many art treasures. Among several fine monastic buildings in the *département* are St. Martin-de-Boscherville, Valmont, Montivilliers, and Jumièges. The castle of Tancarville occupies a commanding site on a promontory overlooking the Seine Valley. Varengeville has the beautiful Renaissance manor house of Jean Ango (*see* MANOR HOUSE). (AR. E. S.)

SEINE-SAINT-DENIS, a *département* of France formed in 1964 of the northeastern part of the former *département* of Seine, together with Aulnay-sous-Bois and Le Raincy, which formed the northern part of the corridor of Seine-et-Oise that separated the Seine and Seine-et-Marne *départements*. Bounded east by the *département* of Seine-et-Marne, south by Val-de-Marne and Ville de Paris, west by Hauts-de-Seine, and north by Val-d'Oise, it comprises the northward and eastward extensions of Paris outside the line of the 1841–45 fortifications. Area, 91 sq.mi. (236 sq.km.). Pop. (1962 census of 1964 units) 1,083,724. The *département* is heavily urbanized, except in the extreme northeast beyond the airport of Le Bourget, where cornfields still lie close to Paris.

Beyond the northern gates of Paris, the flat plain of Saint-Denis was occupied during the 19th century by factories, housing, and a close mesh of lines of communication. The canals and railways that converge to enter Paris by the col between the hills of Montmartre and Chaumont are lined with establishments of heavy industry, especially chemical and engineering works, and gas and electric power stations; and a great variety of other factories are closely intermingled with workers' housing in the communes of Saint-Denis, Aubervilliers, and Pantin that lie immediately outside Paris. In this inner zone a large cemetery between Pantin and Aubervilliers provides the chief open space. Factories and various types of housing extend in more open association in La Courneuve, where the *préfecture* has been established, and Le Bourget, and market gardens have not altogether disappeared from these northern suburbs. In the west the *département* reaches the Seine at the tip of its meander, where it is joined by the Saint-Denis Canal, which bypasses the great loop of the river below Paris. The river and canal banks are crowded with port installations and factories, making Saint-Denis (*q.v.*) one of the major concentrations of industry in Greater Paris. Beyond Pantin, a compact belt of housing and industry extends eastward along the Canal de l'Ourcq and the historic road (N3) to Lorraine, but south of this the hilly country that occupies the southeast of the *département*, north of the Bois de Vincennes and the Marne valley, has been the scene of more open, predominantly residential development. Dormitory suburbs have rapidly expanded there since World War I.

The northern industrial suburbs of Paris in Seine-Saint-Denis suffered severe aerial bombardments by the Allies during the German occupation. (AR. E. S.)

SEIPEL, IGNAZ (1876–1932), Roman Catholic priest who was chancellor of Austria from 1922 to 1924 and from 1926 to 1929. He was born in Vienna on July 19, 1876. Ordained priest in 1899, he was appointed professor of moral theology at Salzburg University in 1909 and was called to Vienna University in 1917. In 1916 he had published *Nation und Staat*, in which he advocated cultural autonomy within the existing political structure for the various nationalities of Austria-Hungary and also gave expression to a moderate pacifism. Publication of these views led to his being chosen for a post in Heinrich Lammasch's government (Oct. 27, 1918). After the collapse of Austria-Hungary at the end of World War I, he prevented the Social Christian Party, to which he belonged, from splitting between monarchists and republicans. Elected to the Constituent Assembly (1919), he entered Karl Renner's coalition government. Party leader of the Social Christians from 1921, he ended their coalition with the Socialists and brought them into alliance with the Pan-Germans.

Seipel formed his first government on May 31, 1922. His personal diplomacy obtained for Austria a loan of $100,000,000 through the League of Nations; but the Geneva Protocols of Oct. 4, 1922, on which the loan was conditional, were bitterly criticized by the Socialists. Though the loan served to stabilize the Austrian currency and to balance the budget, the economy remained unsound, and Seipel's efforts to restore it met with objections from provincial governments and led to dissension within his own party. On June 1, 1924, he was severely wounded by a shot from a labourer. Seipel resigned office on Nov. 7, 1924.

Chancellor again from Oct. 20, 1926, Seipel hoped to confirm his mandate by holding an election in April 1927. The result went against him, but he reshaped his cabinet and remained in office. The Socialists' rioting in Vienna (July 15, 1927) started a period of strife in which Seipel tried to use the Fascist *Heimwehr* against the Socialists' *Schutzbund*. He resigned on April 3, 1929.

In *Der Kampf um die österreichische Verfassung* (1930) Seipel

argued for a reform of the constitution that would establish a corporate state with a federal president elected by national suffrage and endowed with greater powers. He himself sought to become president in 1930, but his party did not support him. He was foreign minister in Karl Vaugoin's cabinet of Sept. 10–Nov. 29, 1930. Seipel died at Pernitz on Aug. 2, 1932. (H. Pf.)

SEI SHŌNAGON (966/967–1013?), Japanese diarist and poet, a witty and learned member of the court, whose *Pillow-Book*, apart from its high value as the vehicle of one of the most original and scintillating Japanese prose styles, is the best modern source of information on Japanese court life in an unusually brilliant period. She was the daughter of the poet Kiyohara no Motosuke and in 991 entered the service of the empress Sadako. Her *Pillow-Book* begins in that year and continues until the year 1000. It consists in part of vividly recounted memoirs of Sei Shōnagon's impressions and experiences, in part of such categories or judgments as "Annoying Things" or "Things Which Distract in Moments of Boredom."

Sei Shōnagon was apparently not a beauty, but managed to hold her place at court by virtue of her ready wit and intelligence, qualities that won her numerous enemies as well, as is witnessed by a passage from the diary of her contemporary, Murasaki Shikibu (*q.v.*). Although capable of great tenderness, Sei Shōnagon was often merciless in the display of her wit, and she showed little sympathy for those unfortunates whose ignorance or poverty rendered them ridiculous in her eyes. Her ability to catch allusions or to compose in an instant a verse exactly suited to a particular occasion also led her to affect, if her detractors can be believed, an air of superiority that others found intolerable. Perhaps by way of poetic justice, legends state that Sei Shōnagon spent her old age in misery and loneliness.

Bibliography.—Arthur Waley, *The Pillow-Book of Sei Shōnagon* (1928); André Beaujard, *Les Notes de Chevet de Séi Shōnagon* (1934) and *Séi Shōnagon, son temps et son oeuvre* (1934). (Dd. K.)

SEISIN, a term from medieval English land law, of little modern significance, signifying possession of a freehold estate. Seisin was basic in the feudal theory of estates (*see* Real Property and Conveyancing, Laws of). The old form of conveyance, known as feoffment with livery of seisin, required a symbolic delivery of possession by handing the transferee a clod of earth, a twig or the like. *See* Feoffment. (A. Dm.)

SEISMOLOGY, the study of vibratory motions of the earth and other planetary bodies. A branch of geophysics (*q.v.*), seismology originated in the study of earthquake waves, but later was enlarged in scope to include vibratory motions due to other sources, including explosive charges. Records of the vibratory motions are made by seismographs; the principal types and their uses are described in a separate section, below. These records are carefully studied and compared to learn about the nature of earthquakes, and to obtain knowledge about some of the physical properties of the earth's interior. Both of these are possible in principle, since recordings at a seismic observatory represent the influence of the event at the earthquake focus, as well as the influence of the path through the interior of the earth over which seismic energy has been communicated to the seismograph. The properties of the seismograph are further impressed upon the recording, but this represents an influence which is presumed to be known.

Sources of Vibrations.—Sources of vibrations of the earth sufficient to cause measurable response upon sensitive seismographs include earthquakes, man-made explosions, volcanic eruptions, microseisms, the influence of wind upon surface structures, and, in urban areas, traffic. Microseisms are largely the vibratory motions due to storms at sea; the principal cause is differential loading of the ocean floor by large waves in the deep sea, and a secondary cause is the breaking of surf upon shore. For the purpose of interpreting earth structure, earthquakes and artificial explosions are most useful, since the focus of their occurrence can be located more or less accurately, and, as they are impulsive in character, a discrete time of occurrence can be ascribed to them. The remaining types of sources assume the character of a background "noise"; it may be possible to study seismic events of interest, such as earthquakes and explosions, if these events are recorded sufficiently strongly above the noise level.

Detectors.—Seismic recordings are made by seismographs (*see* below). Seismic instruments have varying characteristics that are specially selected in order to study features of seismic waves of interest. Pendulum seismographs are usually separated into short and long period instruments; the separation between these two classes of instruments is usually gauged by the relatively narrow band of periods in which microseismic energy is strongest; microseisms will usually mask earthquake recordings made in this band. Instruments having sensitivities at periods longer than 6 to 8 sec. are called long-period instruments; short period instruments have their greatest sensitivity at periods shorter than 6 sec. Geophones, or seismometers used in prospecting for oil and in some crustal studies, have their greatest sensitivities at periods of the order of .01 to .1 sec. in order to detect the motion of the earth due to the high frequencies obtained in the artificial explosions used in this type of work. Strain seismographs and gravimeters are seismic instruments with very long period responses. Part of the purpose of restricting the band of observable frequencies is to improve as much as possible the ratio of desirable signal energy due to earthquakes or explosions to noise energy due to microseisms, traffic, and other noise sources.

SEISMIC WAVES

Waves emitted from a seismic source fall into two main categories; each of these categories is further divisible into several types. One category is that of body waves, and these are of two main types, each of which travels with a different velocity. The body wave with the higher velocity is a compressional motion and is called a primary, or P, wave. The motion of the particles of the earth in a P wave is longitudinal or parallel to the direction of propagation. The secondary, or S, body wave has a shear motion, transverse to the direction of propagation. The velocity of S waves is always less than that of P waves at the same point. Thus an S event on a seismograph always follows the P event in time, and the farther the seismograph from the focus of the earthquake, the greater the time interval between the P and S events.

In general, the increase in velocity with depth in the earth causes the ray path of body waves to be curved (fig. 1). In addition to P and S waves that travel directly from the focal point of the earthquake along curved paths to the seismograph, other seismic events can be observed since, wherever discontinuities in physical properties exist in the earth's interior, reflection or refraction of body waves can take place, analogous to the reflection and refraction of light (*see* Refraction). Conversion from P to S motions and the reverse can take place upon reflection, as well as reflections without change of wave type. Thus a wave traveling downward from the earthquake focus as P will reflect off the core boundary as both S and P. These will appear on the seismogram as later events, labeled PcS and PcP, where the labels indicate the history of the event, with PcP occurring before PcS. From the multiplicity of events recorded on a seismogram for a single earthquake, considerable information can be obtained concerning the nature and subsequent history of the event.

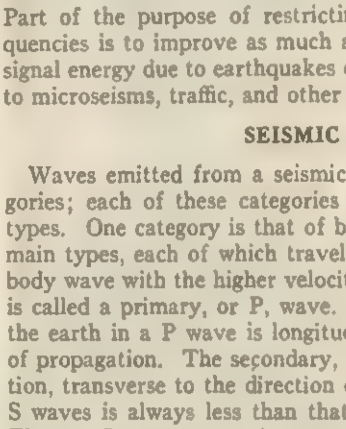


FIG. 1.—SOME OF THE PATHS FOLLOWED BY EARTHQUAKE WAVES. WITH A SCHEMATIC REPRESENTATION OF A SEISMOGRAM WITH BODY (P, S) AND SURFACE WAVES

The second main category of seismic waves is that of surface waves. In this case the seismic energy is propagated around the surface of the earth rather than through the body of the earth. Surface waves are of several types. Among the principal types are Love waves in which the particles vibrate parallel to the surface of the earth, and Rayleigh waves in which the particles move in vertical planes containing the great circle paths from epicentre to seismographs. (The existence of these waves had been predicted by A. E. Love and Lord Rayleigh, respectively.) Surface

wave energy generally diminishes with increasing depth, most of it being found within a depth of the order of the wavelength of the surface waves.

Surface waves are dispersed due to inhomogeneity in the earth's interior so that surface waves with different periods travel with different velocities. Rayleigh waves of 30-sec. period travel faster than do those of 20-sec. period, for example, and will appear earlier on the seismogram (*see* fig. 2).

On a seismogram showing both body and surface waves, the body waves show abrupt onsets. The surface waves show no sharp onsets, because the different frequency components of these waves travel with different velocities. The amount of the dispersion of surface waves is diagnostic as to the distance from the earthquake focus and as to the type of inhomogeneity in the earth.

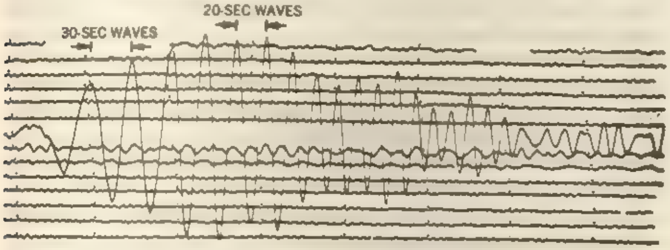


FIG. 2.—LONG-PERIOD SEISMOGRAM FOR THE MID-NORTH ATLANTIC OCEAN EARTHQUAKE OF OCT. 6, 1960, RECORDED AT CHUR, SWITZ., ILLUSTRATING DISPERSED RAYLEIGH WAVES

Abrupt onsets are characterized by large amounts of high-frequency constituents; hence, body waves are easily identified on short-period instruments. Surface waves can be more easily studied on long-period instruments because the dispersion is usually greatest for the crust and upper mantle in the long-period band.

See also EARTHQUAKE: *Earthquake Waves.*

SEISMOLOGICAL STUDIES

Seismologists study the recordings from earthquakes and explosions with several main purposes in mind. First, they wish to study the nature of the earthquake mechanism. From studies of seismograms, inference is made as to the characteristics of the source and the focal mechanism.

A second purpose is to use the seismic energy which has passed through the earth's interior as a means for studying the degree of heterogeneity of the earth's interior. Comparative studies of recordings from many stations to determine the changes in the seismic records from station to station, representing the influence of different parts of the earth on seismic waves, can lead to information about the structure of the earth. (*See* EARTH: *Structure of the Interior: Seismic Exploration;* GEOPHYSICAL PROSPECTING: *Seismic Methods.*)

Times of Arrival.—Universal time-distance relations have been constructed from compilations of the observations of many observatories upon the times of arrival for many earthquake waves. These relations indicate the time of travel of earthquake waves of any type from an earthquake focus to an observatory. From the time interval between the abrupt onsets of P and S events on a single seismogram, the distance to the impulse-source can be estimated by comparison with the universal time-distance relations. By comparison of the events recorded on a set of three-component seismographs at one observatory, the direction of the source can be located in principle. The times of arrival of later body wave events lend substantiation as to the location. The comparison of times of arrival at many observatories leads to a more accurate determination of the location and depth of the focus of the earthquake.

From the universal time-distance relations it is possible to compute the distribution of P and S wave velocities in the interior of the earth. The time-distance relations for the body events that arrive at times later than P and S give additional information about the depths and properties of inhomogeneities at great depth in the earth. The P and S wave velocities generally increase with depth to 2,900 km. (1,800 mi.) where a marked discontinuity in properties is found to exist. This discontinuity provides a horizon for reflections. No S waves have been found to penetrate the region below 2,900 km., called the core of the earth; since liquids cannot be sheared or twisted, the core is presumed to be liquid. Other discontinuities have been found at greater and at shallower depths. Gradual changes in physical properties with depth are found from the interpretation of times of arrival.

Mohorovičić Discontinuity.—The Mohorovičić discontinuity, moho, or "M," marking the lower boundary of the earth's crust, was discovered by Serbian geophysicist A. Mohorovičić. In his studies of an earthquake in the Balkans in 1909 he observed that travel-time curves determined from the recordings made by nearby seismograph stations showed an abrupt increase in seismic velocity occurred at a depth he estimated to be about 60 km. (37 mi.). From this he postulated the existence of a surface layer, now called the crust, covering the earth's mantle. Subsequent investigations have indicated that the depth of this discontinuity is 7 km. (4 mi.) on the average under the oceans, and about 40 km. (25 mi.) on the average under the continents, as indicated below.

Surface Waves.—Additional detail about the properties of the earth's interior is obtained from the analysis of the dispersion of surface waves. The times of arrival of different periods of surface waves can be interpreted with the aid of mathematical models to give estimates of the P and S wave velocity variations and the density variation with depth of the upper parts of the earth. Because of the many modes in which surface waves can propagate, the versatility of this method of interpretation of the structure is great. The differences in times of arrival for a given period of surface waves at neighbouring observatories can be used to study the structure of the upper parts of the earth between the observatories.

Focal Mechanism Studies.—In addition to studying the location of the earthquake focus, both geographically and in depth, and indicating the times of occurrence, studies of records taken at observatories remote from the focus can indicate the characteristics of the earthquake at the focus. In one example of this type of work, the direction of motion of the seismometer upon the onset of a P wave can be used to determine the orientation of the earthquake fault upon rupture. Assume that we take for our model of an earthquake the abrupt relative motion of two blocks of the earth abutting upon a plane which we call the fault plane of the earthquake (fig. 3). When the blocks abruptly slip, the material in both regions A and C is initially compressed, and in regions B and D is initially dilated or expanded. Thus at observatories located in region A and at points remote from the earthquake, the onset of the P event will appear to move in a direction away from the earthquake focus. Similar conclusions can be drawn from observatories in the other quadrants. If observatories all over the world are polled as to the direction of the first observed motions of P, some stations will report initial compressions and some initial dilatations. The stations reporting dilatations and compressions will distribute themselves into quadrants on a map; one of the lines of separation of the quadrant is then the projection of the fault plane onto the map. Unfortunately, the solution is not unique, since a hypothetical fault plane oriented at right angles to the first and undergoing initial displacements opposite in sense to the first will show the same distribution of compressions and dilatations. A resolution of this ambiguity comes from other evidence such as the distribution of aftershocks or geologic evidence. The distribution of first motions of S as recorded at many observatories can also be mapped, although some difficulty is connected with this procedure since the onset of S is often not clearly

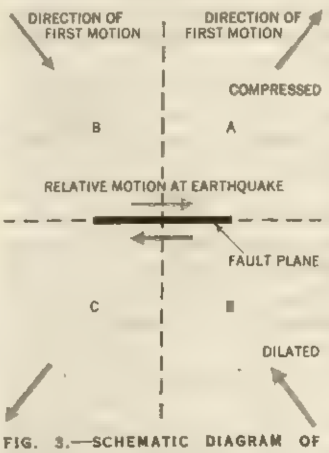


FIG. 3.—SCHEMATIC DIAGRAM OF THE RADIATION OF THE FIRST MOTION OF P WAVES FOR AN EARTHQUAKE FAULT

discernible in the presence of the oscillatory motion subsequent to the onset of P. By the above procedure, however, observations taken from remote observatories can be used to help determine the orientation of the fault plane of the earthquake.

In the case of an explosive seismic source, every observatory should move initially in a direction away from the focus. Thus all stations should report initial compressions; no quadrant separation between dilatations and compressions should be possible. This difference has been suggested as one diagnostic test for distinguishing between underground explosions and earthquakes.

The velocity of rupture of earthquake faults can be measured by studying the surface wave energy arriving at an observatory in various frequency bands. The predominance of energy at one frequency and absence at another can be interpreted as an interference phenomenon corresponding to a selection process associated with a moving source. The moving source, in this case, is the point of tear of the fault. In the case of the Chilean earthquake of May 21–22, 1960, the velocity of rupture was found to be 3½ km. (2 mi.) per second, a value roughly the same as the shear velocity for the crustal rocks in which the earthquake took place. This type of observation is possible only for the largest earthquakes.

Structure of the Earth.—Some of the physical properties of the earth can be determined by comparative studies of seismic recordings at several observatories; in some cases only one observatory is required if the assumption is made that the earthquake has certain ideal properties, such as that of being impulsive in time, and that its focus is located at a point in the interior. One example has already been given of the inversion of time-distance data to give the P and S wave velocity-depth relation for the earth.

Several other procedures can be listed; usually several methods are complementary since the results obtained by any one method are not absolutely precise.

The amplitudes of seismic waves at many stations may be compared. This is not always a reliable procedure, mainly because of varying instrumental properties from station to station, as well as a strong dependence of the amplitude of the response upon the type of soil upon which the instruments may be placed. Nevertheless, strong changes in amplitude in the range 11,700 to 15,900 km. (7,270 to 9,800 mi.) from earthquakes gave evidence for the existence of the outer core and later the inner core of the earth. Precise depth determinations of these discontinuities come from analysis of time-distance relations.

Another property that can be determined from the study of amplitudes is the ability of the earth to attenuate seismic waves. If the earth did not have some nonelastic mechanism whereby seismic waves were attenuated, the seismic energy released in earthquakes of the past would still be stored in the earth and the surface would vibrate continuously with very large amplitudes. The comparison of amplitudes on successive recordings at one station, as a surface wave from a large earthquake executes multiple circuits of the earth, leads to estimates of the attenuation factor. For Rayleigh waves the amplitudes are reduced by roughly 2% between two points a wavelength apart.

Other physical properties of interest have been discovered by seismic means, principally from time-distance relationships. Among these are: (1) a region of body wave velocities slightly lower than those above or below, centred at about 100–200 km. (62–124 mi.) below the surface; (2) a marked increase in the body wave velocity gradient at about 400 km. depth; (3) differences in the thicknesses of the crust of the earth between oceans and continents: under the deep ocean bottoms, the crust is 5 to 7 km. thick, while the continental crust is about 40 km. thick on an average; (4) a thickening of the crust under mountains. (*See* also EARTH; ISOSTASY.)

Free Oscillations.—The Chilean earthquake provided the first example of the recording of the free modes of oscillations of the earth, and these were observed by five groups of investigators using strain seismographs, gravity meters, and tilt pendulums. (For discussion and illustration of the recording made with the strain seismograph, *see* EARTHQUAKE: *Earthquake Waves.*) Before that time, observations of the free modes had not been made, principally because sensitive instrumentation had not yet been developed to measure the small displacements at the extremely long periods involved. The existence of the free modes had been predicted some time earlier but the exact values of the period were dependent upon the theoretical models used in the calculation.

The free modes of oscillation are a resonance phenomenon. A cast bell, when struck, rings in a fundamental and a large number of overtones. The relative amplitudes of the fundamental and overtones, the "overtone spectrum" of the bell, depend strongly upon where and how it is struck and where and how it is observed. The periods of the spectrum depend upon the shape and dimensions of the bell and the elastic structure of the bell. Since the earth is highly elastic (its attenuation factor is low as shown above), it too should ring in a fundamental and many overtones when struck, in this case by a large earthquake. From the models of the earth obtained by time-distance observations and from the dimensions of the earth, the longest period of the oscillation was predicted to be about an hour.

Harmonic analysis of the seismograms taken from the Chilean earthquake showed a fundamental of 54 minutes. At least 40 overtones were observed. The overtones could be divided into two groups corresponding to torsional and spheroidal oscillations. These two groups of oscillations correspond, in the torsional case, to the pure twisting of parts of the earth relative to one another without compression or dilatation, and in the spheroidal case, to the coupled compression torsion of parts of the earth relative to each other. In the torsional mode with longest period, one hemisphere is twisted clockwise while the other is twisted counterclockwise, and vice versa. In the lowest spheroidal mode the equator moves radially outward while the poles move radially inward, and vice versa; this is the "football" mode. The words hemisphere, equator, and poles as used here do not correspond to the geographic terms, but instead correspond to coordinates dependent upon the location of the earthquake.

The observed periods of the free modes from large earthquakes strengthen confidence in the use of models of the earth's interior derived from time-distance data. Comparative amplitude studies upon the spectra for the purpose of determining focal mechanism were not performed, because of the small number of observatories participating in the observations.

Planetary Seismology.—Many of the techniques and studies described here can be applied to studies of the planetary bodies in the solar system. The sources can be local quakes upon the body or artificial seismic sources such as explosives. In the case of the moon, measurements are simpler than on earth because of the absence of atmosphere and oceans; hence noise due to microseisms, wind, traffic, etc. is absent. (L. KN.)

SEISMOGRAPHS

Instruments for recording and measuring vibratory movements of the ground are known as seismographs. As indicated above, their use originally was restricted to movements produced by earthquake waves. Subsequently special types have been devised to record ground movements produced by artificial blasts for location or delineation of subsurface geological structures in prospecting operations for petroleum and other minerals. They have also been used for measuring the thickness of ice sheets covering polar regions (*see* INTERNATIONAL GEOPHYSICAL YEAR: *The Earth's Structure and Composition*).

Earthquake waves are propagated from the source in all directions throughout the earth and are affected or modified by the physical conditions occurring over their paths. Consequently study of their forms and the patterns of their occurrence over the earth as revealed by seismographs has been the principal source of knowledge concerning the nature of the earth's interior. The seismograph's contribution to geophysical and geological knowledge is thus comparable to the contributions of the telescope and microscope to astronomy and biology respectively.

(*See* also EARTHQUAKE; EARTHQUAKE. For their application in prospecting *see* GEOPHYSICAL PROSPECTING.)

General Types.—All seismographs fall into one or other of two general forms depending upon their initial responding element or mechanism which is either a pendulum or a device responsive to strain. The pendulum consists of a weight, called an inertia reactor, suspended or attached to the ground in such a way that it is free to oscillate. If displaced from its rest position it is acted upon by a force known as a restoring-force which tends to return it to its rest position and which is proportional to the displacement. The restoring-force may be provided by a spring or by gravity as in the gravity pendulum. The time for one complete oscillation of the pendulum is called the period of the pendulum. If the pendulum were well constructed and contained only a restoring-force, once set into oscillation it would continue to vibrate indefinitely and would thus be of little use in measuring ground movements.

Accordingly, an additional force is introduced into the system proportional negatively to the velocity of the pendulum and known as a damping force. The damping force is derived almost universally from reaction of electric currents generated in a moving conductor attached to the pendulum and immersed in the field of a magnet fastened to the supporting structure. The damping force is adjusted near or exactly to the critical value which allows the pendulum freely to just return to rest from a displaced position. When the ground vibrates, the inertia of the pendulum prevents it from moving identically with the ground and the difference in motion between the two is the quantity measured and recorded on the recorder, after suitable magnification and other modification. The written record is known as a seismogram.

In order fully to measure the ground movement, three perpendicularly oriented pendulums are required, one for the vertical or up and down component of motion and two for the horizontal components, usually oriented north-south and east-west. The simplest type of vertical component pendulum is represented by a weight suspended by a spring. With suitable transverse restraints, instruments of this type have found extensive use in prospecting seismology as well as in earthquake recording where such pendulums designed with periods of about 1 sec. are particularly well suited to the shorter period components of seismic waves ranging in period from about 0.2 sec. to 10 sec. Pendulums of this type are not constructed with periods much longer than 1 sec. because the extension of the spring by the weight of the inertia reactor is proportional to the square of the period and so becomes too large for practical use. In order to get around this difficulty Sir J. A. Ewing introduced (1880) a mechanical structure so arranged that a portion of the spring tension produces a negative restoring force; *i.e.*, a force which acts to drive the pendulum away from its rest position rather than toward it. Lucien LaCoste modified the original mechanical structure somewhat and introduced in addition a zero-length helical spring—one which when fully collapsed has a large built-in residual tension so that when extended just sufficiently to separate the coils, tension is exerted equal to that of an ordinary spring extended a distance equal to its own unextended length. The residual tension is due to the twist of the spring wire introduced during winding of the helix. The LaCoste pendulum is shown schematically in fig. 4A. With this form, periods up to 80 sec. have been attained with good stability.

The simplest horizontal component pendulum is perhaps the gravity pendulum such as used in pendulum clocks. Its use in seismographs has been limited, however, by a condition similar to that of the simple vertical component pendulum in that for periods much longer than 2 sec. the dimensions are impracticably large. A form in common use which gets around this difficulty is shown in fig. 4B. Here the inertia reactor consists of a weight attached to one end of a horizontal boom. The other end of the boom is pivoted against the vertical supporting column. The weight is supported by a wire fastened at the top of the column. If the weight is displaced horizontally a small distance, the boom rotates through a small horizontal angle and at the same time the geometry of the structure is such that the boom raises a slight amount. Gravity thus provides a restoring force which acts to return the boom to its rest position, and with careful design such pendulums can be made with periods extending to one minute or more. Another horizontal component form due to J. A. Anderson is known as the torsion pendulum (fig. 5). In this the inertia reactor is a small cylinder or vane supported eccentrically by a thin wire suspension. This type has found extensive use for periods from 0.1 sec. to 1 sec., although it can be constructed with longer periods.

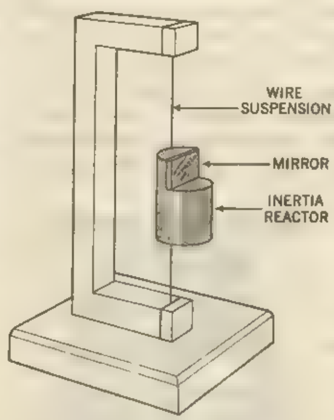


FIG. 5.—SCHEMATIC DIAGRAM OF TORSION PENDULUM

Recording Systems.—In the early seismographs movement of the pendulum was magnified and recorded on a strip of smoked paper using a system of levers ending in a pointed stylus. The useful magnification with these recorders was severely limited by the circumstance that any frictional or other force acting on the stylus tip is reflected back to the inertia reactor by a factor equal to the square of the magnification. To get around this difficulty John Milne and J. J. Shaw introduced a system using optical magnification with photographic recording. A small mirror was arranged to be rotated by a movement of the pendulum so that a beam of light reflected from the mirror produced a magnified replica of the motion of the pendulum on a strip of photographic paper wrapped around a slowly rotating drum. This system was brought to its ultimate efficiency in the torsion seismograph mentioned earlier. In this form the mirror is attached directly to the tiny inertia reactor, less than an eighth of an inch in diameter, so that a small linear movement of the ground produces a relatively large rotation of the inertia reactor and mirror. Magnifications of 3,000 times were attained with this instrument for short pendulum periods. For longer pendulum periods such direct coupled recording systems are severely limited as to useful magnification because of the horizontal pendulum's sensitivity to tilt of ground (varies as the square of the pendulum period) and because of creep and thermal response of the spring in vertical component pendulums.

Electromagnetic Seismographs.—To overcome this limitation B. B. Golitsyn developed an electrical system in which recording was accomplished photographically with a D'Arsonval galvanometer actuated by an electromagnetic transducer on the pendulum. The transducer takes the form of a coil of fine wire attached to the inertia reactor and arranged to move in the field of a strong permanent magnet fixed to the instrument. With movement of the pendulum the transducer thus becomes a small generator of electrical power which serves to actuate the recording galvanometer. Since the mirror system of the galvanometer was very small and of light weight in comparison with the pendulum weight of about 50 lb., the available magnification was increased to around 10,000 times. But of greater importance is the fact that the electromagnetic transducer output voltage depends upon the rate of movement of the coil relative to the field rather than the displacement. It thus produces negligible output for the slow movements of the pendulum caused by temperature

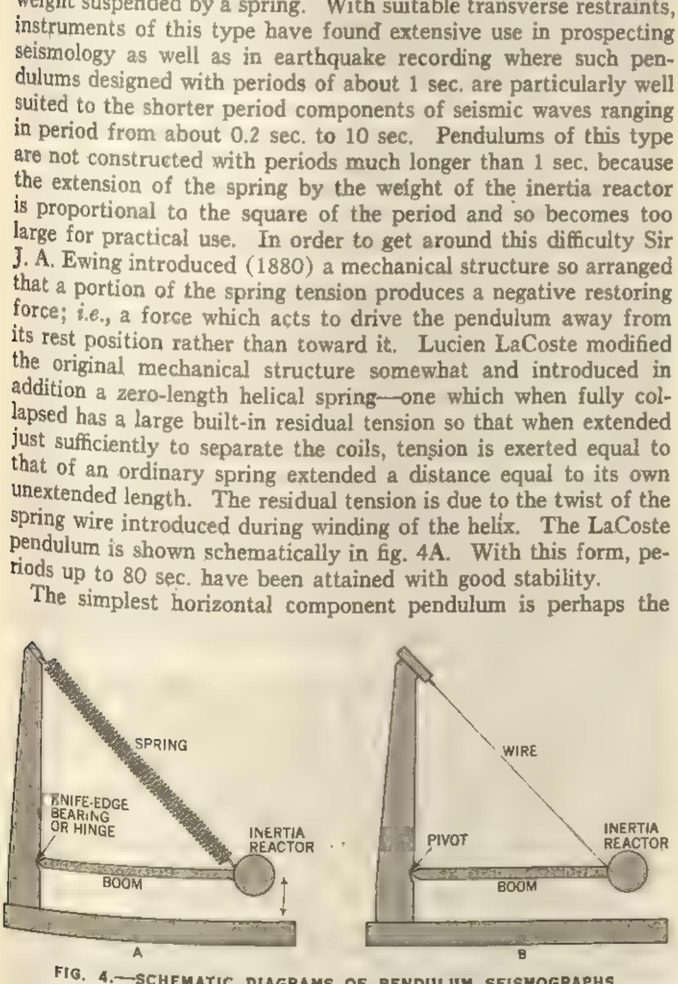


FIG. 4.—SCHEMATIC DIAGRAMS OF PENDULUM SEISMOGRAPHS
(A) LaCoste pendulum used to measure vertical component movement of the ground and (B) pendulum used to measure horizontal ground movement

changes or tilts of the ground and consequently higher magnifications for seismic movements can be used. Golitsyn adjusted his pendulum and galvanometer to the same period, about 12 to 15 sec. Later Frank Press and Maurice Ewing, and then Press, Ewing, and Francis Lehner made extensive modifications of the Golitsyn seismograph resulting in greatly improved performance, especially for very long period waves. They changed the vertical component pendulum to the LaCoste type and compensated it for barometric pressure variations which produce movements of the pendulum because of the accompanying changes in the buoyancy of the air on the inertia reactor.

In addition, they increased the pendulum period to about 30 sec. and the galvanometer period to 90 sec. Their instrument provides a maximum magnification of about 4,000 times in the period range from about 15 to 40 sec.

Another type of electromagnetic seismograph was developed by Hugo Benioff especially for recording the short period components of seismic waves in the range 0.2 to 5 sec. He used a cylinder of 200-lb. weight for the inertia reactor and suspended it directly by a helical spring (fig. 6), to give a free period of 1 sec. For such a large weight and short pendulum period it was impracticable at the time to use the Golitsyn type moving coil transducer because of the great size of the magnet required. He therefore used a variable reluctance magnetic transducer for providing the required electrical output. In this form magnetic flux from a permanent magnet flows across four air gaps formed between laminated iron alloy pole faces in a push-pull arrangement as shown in the drawing. Movement of the pendulum varies the length of one pair of gaps relative to the other with a resulting change in flux in the two circuits. These changes generate voltages in the two sets of coils wound around the laminated iron alloy armatures carrying the flux, and the induced currents operate two galvanometer recorders simultaneously with galvanometer periods of 0.2 sec. and 90 sec. respectively. The large weight of the inertia reactor and the high efficiency of the reluctance transducer allow maximum magnifications of about 1,000,000 times with the 0.2 sec. galvanometer for periods in the vicinity of 0.5 sec.

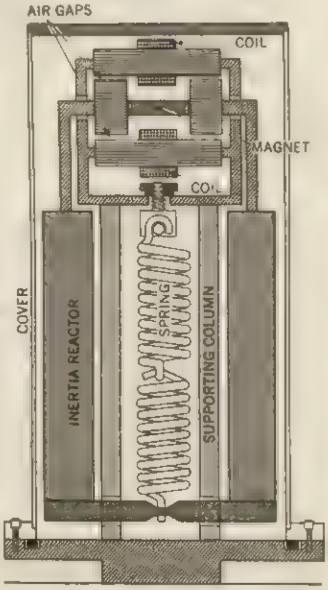


FIG. 6.—SCHEMATIC SECTION OF BENIOFF VERTICAL COMPONENT SEISMOGRAPH

Such high magnification can only be used in the few regions of the earth where the ground unrest is very small. For most regions the amplitudes of the ground unrest limit the useful magnification to 50,000 times or less. With the long period galvanometer the magnification is approximately 2,000 times at 1-sec. periods and gradually falls off to 30 times for periods of 90 sec. Because of a negative restoring-force introduced by the reluctance transducer, the spring and other constraints of this seismograph are much heavier than those of other types for the same pendulum periods. The instrument is thus extremely rugged relative to other seismographs.

Strain Seismograph.—The strain seismograph was developed by Benioff in 1935. Unlike pendulum seismographs which respond to vibration of the ground, the strain seismograph responds to linear strain (stretching or compression) of the ground. Strain is produced by seismic waves since in different parts of the wave the ground is displaced by different amounts or even in opposite directions. Thus during the passage of a seismic wave train past a given observing point, the ground is alternately compressed and stretched in conformity with the undulations of the wave. The strain produced by seismic waves is exceedingly small so that the measuring elements of the strain seismograph must be very sensitive. The principle of the strain seismograph is illustrated in fig. 7. The two piers, usually made of 12-in. steel pipe, are sunk into rock. A length standard in the form of a tube of fused quartz 60 to 100 ft. long is firmly attached to one pier and extends to within a short distance of the other pier. When the ground between the piers is stretched or compressed during passage of a seismic wave, the distance between the piers changes and this in turn alters the separation between the end of the standard tube and the adjacent pier. With a proper transducer and recorder this small change in separation can be magnified and recorded. Benioff uses two types of transducers. One is the variable reluctance electromagnetic type similar to the one described in the preceding paragraphs for his pendulum instrument. With this transducer, recording is accomplished with a galvanometer on photographic paper. The other transducer is a carrier-current resonant capacitance bridge type recording with an ink-writing galvanometer on a paper strip. Although the operation of this type of transducer is too highly technical to be described here, it is simple and stable in practice and records not only strains produced by seismic waves but also the strains produced in the solid earth by the tidal forces of the sun and moon. As to be expected, the response characteristics of the strain seismograph differ from those of pendulum instruments in a number of ways including different directional and period characteristics, and absence of response to tilt of the ground. (Hu. B.)

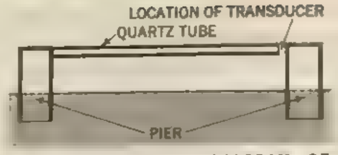


FIG. 7.—SCHEMATIC DIAGRAM OF STRAIN SEISMOGRAPH

BIBLIOGRAPHY.—Hugo Benioff, "Earthquakes and Associated Instruments," *Advances in Geophysics*, vol. ii (1955); K. E. Bullen, *Seismology* (1954), *Introduction to the Theory of Seismology*, 3rd ed. (1963); M. B. Dobrin, *Introduction to Geophysical Prospecting*, 2nd ed. (1960); F. W. Sohon, "Introduction to Theoretical Seismology," *Seismometry*, part 2 (1932); C. F. Richter, *Elementary Seismology* (1958); H. Odishaw (ed.), *Research in Geophysics*, vol. 2 (1964); J. H. Hodgson, *Earthquakes and Earth Structure* (1964).
(Hu. B.; L. Kn.)

SEISTAN (SISTAN), an extensive border region of eastern Iran and southwestern Afghanistan, centred on the Hamun-e Saberi about latitude 31° 30′ N and longitude 61° E. Its area is roughly 10,000 sq.mi. (28,000 sq.km.) of which about 40% is in Iran. The total population is approximately 300,000, of whom about half reside in Iran.

Physically, Seistan is a large depression at a general level of 1,500–1,700 ft. (450–520 m.) which receives the drainage of Harut Rud (Adraskan) and Farah Rud from the north, Khuspas Rud and Khash Rud from the east, Helmand (*q.v.*) River from the south, and the Rud-e Bandan and Shur Rud from the west. Of these only the Helmand has a permanent flow. These rivers fill a series of lagoons called *hamun* (*e.g.*, Hamun-e Saberi, Hamun-e Puzak) and in high flood inundate the extensive reed tracts (*naizar*) south of the lagoons, thus forming a great shallow lake which discharges its redundant waters through the Shelagh Channel into another southerly depression called Gaud-e-Zirreh. Three large deltas intruding into this lagoonal area form the main regions of settlement: Lash-Juwain on the Farah Rud, Chakhansur on the Khash Rud (both in Afghanistan), and Seistan proper on the Helmand River (see *History*, below).

Seistan has a true desert climate with a highly irregular rainfall averaging less than 4 in. (100 mm.), extreme heat in summer, and frost spells in winter. A special plague is the "wind of 120 days," which blows unceasingly from the north during summer, causing much erosion and the shifting of sand dunes.

The population is chiefly Tajik (*see* TADZHIK), though of mixed descent as many alien groups have more or less amalgamated with the original Iranian stock. There is a strong nomadic element of Balochi and Brahui, besides groups of Jat and Gujars, and Gawdars who breed strong, black, humped cattle.

Incessant internal strife and corrupt administration, together with its isolation, helped to keep Seistan's economic standard low. Irrigation was formerly practised by yearly rebuilding a dam (Band-e-Seistan) on the Helmand near where it enters Iranian

Seistan. The dam was annually swept away by flood. In 1952–53 two barrages, Kuhak and Zahak, were constructed to irrigate approximately 320,000 ac. (130,000 ha.) of land. The chief crops are wheat and barley with a little cotton and maize (corn).

In 1959 Iranian Seistan was united with Baluchistan (*q.v.*) to form an *ostan* (province). The only town of importance in Iranian Seistan is Zabol (pop. [1956] 12,221), the seat of the governor when Seistan and Qayen formerly together constituted a hereditary governorship. A motorable road (70 mi.) from Zabol links with the Meshed-Zahedan highway, and Zahedan with its railhead to Quetta is the international outlet. Chakhansur, 35 mi. (56 km.) E of Zabol, is the chief town of Afghan Seistan. (H. Bo.)

History.—Called Zranka in Darius I's inscriptions and Drangiana by Strabo, the country now known as Seistan is the reputed land of origin of the legendary Kayanian dynasty of Persia. After forming part of the Achaemenid Empire, Seistan was successively under Seleucid and Bactrian Greek control. Occupied by the Sakas, a Scythian tribe, *c.* 155 B.C., the country became known as Sakastan. The Sakas were later ousted by the Parthians, but their control over the country was often nominal.

In A.D. 226 the Sasanian Persians under Ardashir I conquered Seistan. The Sasanian monarchs did much to develop agriculture by constructing dams and irrigation canals. They introduced rule by satrap, often leaving control in the hands of members of the Kayanian family whose descendants still live in Seistan. In the reign of Khosrau I, the capital of Seistan was Zaranj, the site of which (now called Nad-i-'Ali) is 22 mi. E of the modern town of Zabol. The Zoroastrian religion flourished under the Sasanians.

In the mid-7th century A.D. the Muslim Arabs invaded Seistan and introduced Islam, but the fluctuating measure of control exercised by the caliphate probably enabled Zoroastrianism to survive longer there than in other parts of the country. In the latter part of the 9th century a Seistani named Yaqub ibn al-Layth al-Saffar ("the coppersmith") founded the Saffarid dynasty in Seistan. Yaqub and his brother and successor, Amr, extended their sway over the greater part of Persia. After the downfall of the Saffarids, Seistan became part of the empire of the Samanids. It was probably at this time that a new capital called Zahedan, 6 mi. SW of Zaranj, was founded. In 1003 Mahmud of Ghazni wrested Seistan from the Samanids, but under both the Samanids and the Ghaznavids the inhabitants retained some measure of governmental control through their own *maliks* (kings) of the Kayanian line.

Seistan twice suffered greatly at the hands of the Mongols in the 13th century. In 1363 Timur (Tamerlane) invaded Seistan. His rule was mild. After a revolt, however, he invaded Seistan again, in 1383, taking a terrible vengeance. He destroyed Zahedan and other cities and also the great dam over the Helmand River. In 1407–08, after a further revolt by the Seistanis, Shahrukh, Timur's son and successor, wrought further havoc to the dams and the canal system.

Seistan subsequently was independent under its own *maliks*. Although Shah Ismail I of Persia conquered Seistan, he and his successors left the *maliks* largely in control. In 1722 a *malik* Mahmud Seistani made himself sovereign over Seistan and much of Khurasan, but he was defeated and put to death by Nadr Quli Beg, the future Nadir Shah. In 1747 Seistan came under the control of the Afghan, Ahmad Shah Durrani, and after his death it was for long a bone of contention between Herat and Kandahar, but in 1865 Nasir ad-Din Shah of Persia took possession of it. Complications later ensued between Persia and Afghanistan which led to British arbitration and the delimitation of the Perso-Afghan frontier by the Seistan Commission under Sir Frederick J. Goldsmid in 1872. In accordance with the arbitration award, Persia handed over to Afghanistan the part of Seistan situated on the right bank of the Helmand River. The work of delimitation was finally brought to a close by the McMahon Mission in 1903–05. (L. Lo.)

BIBLIOGRAPHY.—C. E. Yate, *Khurasan and Sistan* (1900); E. Huntington, "The Depression of Seistan," *Bull. Amer. Geogr. Soc.* (1905); H. McMahon, "Recent Survey and Exploration in Seistan," *Geogr. J.*, vol. 28 (1906); G. P. Tate, *Seistan: A Memoir on the History, Topography, Ruins and People of the Country* (1910–12); S. A. Hedin, *Overland to India*, 2 vol. (1910); V. F. Büchner, "Sistan," in *Encyclopaedia of Islam*, vol. iv (1934); Sir Mark Aurel Stein, *Innermost Asia*, vol. ii (1928); J. Hackin, J. Carl, and J. Meunié, *Diverses Recherches archéologiques en Afghanistan* (1959); W. A. Fairservis, *Archeological Studies in the Seistan Basin of Southwestern Afghanistan and Eastern Iran* (1961).

SEJANUS, LUCIUS AELIUS (d. A.D. 31), close adviser of the Roman emperor Tiberius, was a son of Seius Strabo, a Roman knight from Etruria, and on his mother's side was related to the Cornelii Lentuli, a distinguished senatorial family. He was adopted by an Aelius, possibly Aelius Gallus who was prefect of Egypt. He accompanied Gaius Caesar, Augustus' grandson, to the East in A.D. 1; and in 14, directly after Tiberius' accession, was made his father's colleague as prefect of the Praetorian Guard, becoming sole prefect a year or two later. He concentrated his cohorts in a single camp, the Castra Praetoria, under the Viminal Hill just outside the walls of Rome. After the death in 23 of Drusus Caesar (*q.v.*), whom Sejanus was later accused of having murdered, he systematically attacked the party of Agrippina ("the Elder"; *q.v.*), whose sons were then Tiberius' probable heirs. In 25 Tiberius refused him permission to marry Livilla, Drusus' widow, but his power grew when next year he persuaded Tiberius to retire to Capreae (Capri), and in 29 he succeeded in having Agrippina and her son Nero exiled. In 31 Sejanus, hitherto not even a senator, became consul, and Tiberius was holding out hopes that he would be granted tribunician power. But suddenly the emperor, apparently warned by his sister-in-law Antonia, became suspicious of Sejanus' designs, and by a series of maneuvers had him arrested and executed on Oct. 18, afterward taking terrible vengeance on his party. The extent of Sejanus' plotting is a mystery, but it is not impossible that he planned to murder Tiberius and become emperor forthwith.

BIBLIOGRAPHY.—F. B. Marsh, *The Reign of Tiberius* (1931); M. P. Charlesworth in *Cambridge Ancient History*, vol. x, ch. 19 (1934); R. Syme, *Tacitus*, vol. 2, pp. 752 ff. (1958). (G. E. F. C.)

SEKONDI-TAKORADI, a seaport of Ghana, Africa, and a city from 1963, was formed of the old town and surf port of Sekondi and the modern deep-water port of Takoradi, opened in 1928. Pop. (1960) 123,313. Sekondi is a mixture of old and new buildings on a hilly site. Takoradi is planned as a modern town; in the harbour area is High Street with government and commercial offices and large stores, and there are estates for the workers shaded by nim trees, and a modern hospital. Two breakwaters enclose 220 ac. (90 ha.) of water area, with six quay berths, including one used only for manganese export, and moorings for several vessels. On the seaward side of the lee breakwater are berths for the loading of bauxite and the discharge of oil. The harbour is the terminus of the western branch of Ghana Railways, and motor roads link the city with the coast and the interior. The airport constructed during World War II is served by internal and external airlines. There are local light industries including timber and cocoa butter mills, railway repair shops, fishing-boat building, and cigarette making. The local inhabitants are peasant farmers or small fishermen, and market and street retail trade is conducted by women.

Nothing remains of Witsen Fort built at Takoradi by the Swedes in 1652–54. The Dutch and British both built forts at Sekondi in the 17th century which were destroyed in the 1690s by the Ahanta tribes, who inhabit the district. The Dutch rebuilt Fort Orange, which was bought by the British in 1872. It is the only surviving fort in the city and is used as a lighthouse. The rebuilt British fort passed to the French in 1779, was ceded to the Dutch in 1867, and bought back by the British in 1872, but has subsequently become a ruin. (E. A.-HA.)

SELANGOR, a state of the Malaysia federation and a constitutional sultanate, on the Strait of Malacca, bounded N by Perak, E by Pahang, and SE by Negri Sembilan. It is the second largest of the states of Malaya in population (1,012,929 in 1957) and the seventh in area (3,167 sq.mi. [8,203 sq.km.]). It consists of part of the western coastal plain, reaching inland to a mountainous limit along the central range which provides a natural boundary with Pahang. The Bernam River separates the state from Perak, and low hills form a southern divide from Negri

Sembilan. The history and economic development of the state have been closely bound up with the short Klang River, which has large mangrove-covered islands on its deltaic fringe. Along this river and the nearby Langat River the waterborne Malays who founded the state settled, as did the immigrants who pioneered the economy of the country. The foothill zone around the river near Kuala Lumpur (*q.v.*), the capital of the Malaysia federation and of Selangor, is Malaya's second largest tin-producing region. There, too, the tin is alluvial, the ore occurring as black, sandy grains deposited by streams debouching from steep courses in the hills which include Triassic limestones. The low-lying surface has been repeatedly turned over by gravel pump mines and dredges, drastically modifying the drainage pattern, diverting the Klang at many places, and hastening coastal sedimentation.

During the 18th century Selangor became an independent state ruled by Bugis from Makasar in the Celebes. In the late 19th century Chinese tin miners, mainly from Singapore, arrived in the state and civil war in their jungle camps led to British intervention, culminating in the establishment of protectorate status for Selangor in 1874. In 1895 Selangor joined the Federated Malay States. Occupied by the Japanese during 1942–45, after World War II it became part of the Federation of Malaya and subsequently of the Malaysia federation. (*See* MALAYSIA.)

At the 1957 census the Chinese formed 48% and the Indians 20% of the state's population, together outnumbering the Malays (29%), although Selangor has been the focus of Malay nationalist politics. The population of Kuala Lumpur was 316,230 (mun.). During 1947–50 there was a large-scale resettlement of urban and rural refugees, and 18 new villages were created (pop. [1957] about 43,000, mainly Chinese). According to the 1957 census about two-thirds of the state's population lived in about 40 towns.

The legislature consists of the sultan and the state legislative assembly. The ruler appoints an executive council, whose members must also be members of the legislative assembly. He acts on the council's advice except in certain matters. The legislature's responsibilities include land tenure and local government.

Selangor has a broad-based economy, its tin mining, which yields about 20% of the national output, being supplemented by about 46,000 ac. of paddy and about 410,000 ac. of rubber, a third of the last named in small holdings. Other agricultural crops include pineapples and palm products.

In 1890 a short railway was built between Kuala Lumpur and Klang, about 18 mi. away, as the river proved inadequate for the transport of goods. From this grew the rail and road links south to Singapore and north to Penang which handled the increasing output of rubber and tin while Klang declined as mining inland disturbed the river profile. Between World Wars I and II the Malayan railway embarked on developing Port Swettenham (now Malaya's second port), beside a broad channel on the delta below Klang, to handle bulky commodities. In the early 1960s the total net registered tonnage of ships entering Port Swettenham in cargo was about 6,100,000 or about 40% of the total for Malayan ports. (E. H. G. D.)

SELBORNE, ROUNDELL PALMER, 1ST EARL OF (1812–1895), English lawyer and statesman who was responsible for the passage of the Judicature Act of 1873, was born at Mixbury, Oxfordshire, on Nov. 27, 1812; his father was rector of the parish. Educated at Rugby and Winchester, and at Trinity College, Oxford, he was called to the bar on June 7, 1837, and soon had a good chancery practice. He wrote for *The Times* and the *British Critic*, and took an active interest in church affairs.

In 1847, and again in 1853, Palmer was returned as member of Parliament for Plymouth, as a Peelite, and in the House of Commons took an active and—in relation to the Crimean War, the Second Opium War with China, and religious questions—an independent part. This attitude was disapproved of by his constituency, and he gave up his seat at the election of 1857. In 1848 he married Lady Laura Waldegrave, and in 1849 became a queen's counsel. In July 1861 he accepted from Lord Palmerston the office of solicitor general, a knighthood, and a safe seat for the borough of Richmond in Yorkshire; in September 1863 he became attorney general and, as such, was adviser of the ministry, in the courts and in the House, on the questions which arose out of the American Civil War. In 1866 he advocated making household suffrage the basis of representation, an expression of opinion which probably influenced the Reform Bill of the following year—in the discussions on which Palmer took a prominent part, and especially in opposition to the so-called "fancy franchises" originally proposed by its authors. In April 1868 he refused to support Gladstone's measures for the disestablishment of the Irish Church, and after the election of that year he declined Gladstone's offer of the office of lord chancellor.

The Treaty of Washington cast a great duty upon Palmer. After the conclusion of the American Civil War very large claims were preferred against Great Britain for alleged breaches of her duty as a neutral power; and after long negotiations Great Britain and the United States agreed to arbitration.

In September 1872 Gladstone again offered him the great seal and he accepted it, with the title of Lord Selborne. In the following year Lord Selborne carried through parliament the Judicature Act, which produced a fundamental change in the judicature system. By the operation of the act one supreme court with several divisions was constituted; each division could administer the whole law; the conflict of divergent systems of law was largely overcome by declaring that, when they were at variance, the principles of equity should prevail over the doctrines of the common law. The details of this great change were embodied in a code of general rules prepared by a committee of judges, over which Lord Selborne presided week by week for two years. "If," he wrote in his memoirs, speaking of the act of 1873, "I leave any monument behind me which will bear the test of time, it may be this." This unification of the courts was more or less contemporary with the construction of a single building to house them; in 1882 Queen Victoria personally presided in the new law courts and handed them over formally to Lord Selborne. On this occasion he received an earldom. In 1885 he definitely broke with Gladstone over Home Rule and disestablishment. But though he never held office again, he continued to sit in the House of Lords both to hear appeals and for ordinary business.

In 1886 Lord Selborne published *A Defence of the Church of England* and in 1888 *Ancient Facts and Fallacies Concerning Churches and Tithes*. He died on May 4, 1895, at his seat in Hampshire after an attack of influenza.

See his *Memorials*, ed. by Lady Sophia Palmer, 4 vol. (1896–98)

SELBORNE, WILLIAM WALDEGRAVE PALMER, 2ND EARL OF (1859–1942), British statesman and high commissioner for South Africa (1905–10), was born in London on Oct. 17, 1859, the son of Roundell Palmer, who was created 1st earl of Selborne in 1882. Educated at Winchester and University College, Oxford, he took a first class honours degree in modern history. In 1883, when he was Viscount Wolmer as his father's heir he married Lady Beatrix Maud Cecil, eldest daughter of the 3rd marquess of Salisbury. He was M.P. for East Hampshire from 1885 to 1892 and for West Edinburgh from 1892 until 1895 when his father's death took him reluctantly to the Lords. Like his father he was a Liberal Unionist; and from 1895 to 1900 he was undersecretary to Joseph Chamberlain at the colonial office.

He entered his father-in-law's cabinet as first lord of the admiralty in 1900. He brought back Adm. Sir John (later Baron) Fisher to the board of admiralty in 1902, and with him revolutionized both the Navy's training system and its capital ships, turning it from an obsolescent force into the most powerful fleet in the world. The completion of this revolution, with the launching of the battleship "Dreadnought," came after Selborne had succeeded Lord Milner as high commissioner for South Africa and governor of the Transvaal and Orange River Colony in May 1905. When the Liberal government took office (December 1905), it decided to give both colonies self-government without delay. Selborne loyally accepted the new situation and in his memorandum of January 1907 proposed that the two colonies should unite with Natal and the Cape Colony to form the Union of South Africa; this policy was accepted. He returned to England in May 1910 just before the South Africa Act, establishing the union, came into force.

He then became one of the leading "die-hards" who opposed the Parliament Bill (*see* ENGLISH HISTORY: *The 20th Century: Struggle With the House of Lords*). During World War I he again cooperated with the Liberals, becoming president of the board of agriculture in the first war coalition of May 1915; but he resigned in June 1916 in protest at the compromise policy that H. H. Asquith pursued in Ireland after the Easter Rising in Dublin. Mistrusting Lloyd George, he refused to join his ministry in December 1916 and also rejected his subsequent offers of the viceroyalties of India or Ireland or of a marquessate. After 1918 he devoted himself partly to the affairs of the Church of England and partly to unsuccessful attempts to reform the House of Lords. Selborne died in London on Feb. 26, 1942. (M. R. D. F.)

SELBORNE, a village 5 mi. (8 km.) S of Alton, Hampshire, Eng., is celebrated for its associations with the naturalist Gilbert White (*q.v.*), who wrote his classic *Natural History and Antiquities of Selborne* (1789) while curate there. His house, "The Wakes," is open to the public and his grave is in the churchyard. The restored church contains interesting medieval work. At Priory Farm, about 1 mi. NE of the village, are the scanty remains of a medieval house of Augustinian canons. The National Trust owns Selborne Hill (241 ac. of common and freehold land) and the Long and Short Lythes (16½ ac., mostly hanging beech woods). Pop. (1961) 1,034. (A. G. G.)

SELDEN, JOHN (1584–1654), English jurist, legal antiquary, and Oriental scholar, was the acknowledged master of the Antiquarian Society, the centre of English historical research in the 17th century. He was born on Dec. 16, 1584, at Salvington, Sussex. His father, also John Selden, held a small farm. Selden was educated at Chichester grammar school and Hart Hall, Oxford. In 1603 he entered Clifford's Inn, London, and in 1604 migrated to the Inner Temple; in 1612 he was called to the bar. His practice was mostly conveyancing, and he rarely went into court.

Selden's early works were: *England's Epinomis* and *Jani Anglorum facies altera* (1610), which deal with the progress of English law down to Henry II; *Titles of Honour* (1614), which, in spite of some obvious defects and omissions, has remained to the present day the most comprehensive and trustworthy work of its kind; *Analecton Anglo-Britannicon* (1615); and *De diis Syriis* (1617), which immediately established his fame as an oriental scholar. For his *History of Tithes* (1618), Selden was summoned before the Privy Council and compelled to retract his opinions, or at any rate what were held to be his opinions. Moreover, his work was suppressed and he himself forbidden to reply to any of the controversialists who had come or might come forward to answer it.

He was first elected to Parliament in 1623. Even before his election he had shown his political sympathies by assisting in the preparation of the memorable protestation on the rights and privileges of the House of Commons affirmed by the House in 1621. For this he had been committed to the Tower of London. After his election he took a prominent part in the impeachment of the duke of Buckingham, was counsel for Sir Edmund Hampden in the "benevolence" case of 1627, and had a large share in drawing up and carrying the Petition of Right (*q.v.*).

In the session of 1629 Selden was one of the members mainly responsible for the tumultuous passage in the House of Commons of the resolution against the illegal levy of tonnage and poundage, and, along with Sir John Eliot, Denzil Holles, Benjamin Valentine, William Strode, and the rest, he was sent once more to the Tower. He was released by the intervention of Archbishop Laud. In 1628, at the suggestion of Sir Robert Cotton, he had compiled, with the assistance of Patrick Young and Richard James, a catalogue of the Arundel marbles. About this period he seems to have inclined toward the court rather than the popular party, and even to have secured the personal favour of the king. To him in 1635 Selden dedicated his *Mare clausum*, against the freedom of the sea, and under the royal patronage it was put forth as a kind of state paper. It had been written 16 or 17 years before but James I had prohibited its publication for political reasons; hence, it appeared a quarter of a century after Grotius' *Mare liberum*, to which it was intended to be a rejoinder. He was returned to the Long Parliament without opposition for the University of Oxford.

Selden joined in the protestation of the Commons for the maintenance of the Protestant religion according to the doctrines of the Church of England, the authority of the Crown and the liberty of the subject. In 1643 he participated in the discussions of the assembly of divines at Westminster and was appointed shortly afterward keeper of the rolls and records in the Tower. In 1646 he subscribed the Solemn League and Covenant and in 1647 was voted £5,000 by Parliament as compensation for his sufferings under the monarchy. He published in 1642 *Privileges of the Baronage of England when they sit in Parliament* and *Discourse concerning the Rights and Privileges of the Subject;* in 1644 *Dissertatio de anno civili et calendario reipublicae Judaicae;* in 1646 his treatise on marriage and divorce among the Jews entitled *Uxor Ebraica;* and in 1647 the earliest printed edition of the old English lawbook *Fleta.* In 1650 the first part of *De synedriis et prefecturis iuridicis veterum Ebraeorum* appeared, the second and third parts in 1653 and 1655, and in 1652 Selden wrote a preface and collated some of the manuscripts for Sir Roger Twysden's *Historiae anglicanae scriptores decem.* After the death of the earl of Kent in 1639 Selden lived with his widow. It is believed that they were married, although the marriage does not seem to have ever been publicly acknowledged. He died at Friary House in Whitefriars on Nov. 30, 1654, and was buried in the Temple Church, London. Of all the members of the Antiquarian Society, Selden was the acknowledged master.

BIBLIOGRAPHY.—A. Wood, *Athenae Oxonienses*, ed. by P. Bliss, 4 vol. (1817); J. Aikin, *The Lives of John Selden and Archbishop Usher* (1812); G. W. Johnson, *Memoirs of John Selden*, etc. (1835); S. H. Reynolds, *The Table Talk of John Selden* (1892); D. Wilkins, *Johannis Seldeni opera omnia*, etc. (1726); E. W. Brainerd, *Quaint Nuggets* (1900); O. H. G. Leigh, *English Belles-lettres from A.D. 901–1834* (1901). (X.; R. G. M. B.)

SELECTION, in biology, is that process whereby individuals with certain genetic endowments are chosen from a population of animals or plants apparently because they are more suited than others to a given environment. These "fitter" individuals are favoured by Nature and are selected from the widely variable pool of all genetic endowments produced. They survive and multiply more successfully than the others, and accordingly become more numerous with each succeeding generation, as long as the environment remains stable. This culling of fitter types, called natural selection, is believed by most biologists to have been the chief impelling and directing agent of organic evolution on earth.

When man began to choose one kind of animal or plant over another, either consciously (as in husbandry and horticulture) or unconsciously (as in extermination of unwanted forms like weeds and predators), he superimposed another selective influence on nature: this kind of interference is called artificial selection.

A third, strongly felt selective pressure is exerted through the attraction of one sex for the other. Simply stated, sexual selection means that among many species there is a constant preference for mates that are strongest, handsomest, most colourful, or attractive in any number of ways that are assumed to provide sexual stimulation, eventual mating, and ultimate continuation of those qualities in offspring.

NATURAL SELECTION

History.—Some historians profess to see the concept of natural selection adumbrated in the creation myths of Empedocles (5th century B.C.) and Lucretius (1st century B.C.), who believed that body parts appeared in the world before complete bodies, that these parts proceeded to combine at random, and that only viable combinations survived. P. L. M. de Maupertuis in 1745, J. C. Prichard in 1813, W. C. Wells in 1813, and especially Edward Blyth in 1835–1837, all expressed ideas implying processes that would be described at present as natural selection. But only Charles R. Darwin and Alfred Russel Wallace (*qq.v.*) in 1858, and especially Darwin in 1859, formulated a theory of evolution by natural selection. Both Darwin and Wallace credited their reading of Malthus' *Essay on the Principle of Population* (1798) as having suggested natural selection to them.

Malthus (*q.v.*) argued that human populations tend to grow in geometric progression, while their food supply increases only

arithmetically; and that more children are born than grow to maturity, because the rest are eliminated by hunger, disease, and war. Darwin and Wallace saw that the argument applies to all living species; they also perceived a similarity between the Malthusian elimination of the surplus population and artificial selection, whereby agriculturists and fanciers select for reproduction those individuals that possess the qualities considered useful or interesting by the breeders. In nature, the survivors in the "struggle for existence" are also likely to carry somewhat different genetic endowments (genotypes) from those who do not survive. Consequently, the genotypes of the succeeding generations will be different from those of the preceding ones, and the difference will usually be in the direction of a greater fitness of the descendants in the environments in which the species lives.

Meaning of Fitness.—Emotionally loaded metaphors, such as "struggle for existence," "survival of the fittest," "eat or be eaten," etc., were used by the early Darwinists, sometimes rather carelessly. This opened the way for misuse of the theory of natural selection as a pseudoscientific justification of war, exploitation, and cruelty. But who is the fittest? The strongest? The most aggressive? P. A. Kropotkin (1902) correctly pointed out that survival and biological success are promoted by cooperation at least as often as by combat.

The modern concept of natural selection overcomes these difficulties. The Darwinian fitness, adaptive value, or selective value (these terms are synonyms) of a genotype is measured merely in terms of its reproductive efficiency. Bodily vigour, strength, combativeness, or cooperativeness, etc., are not in themselves reliable measures of Darwinian fitness: they contribute to, or detract from, fitness only insofar as they influence the reproductive success of a genetic constitution. The adaptive value of a genotype is the mean contribution of its carriers, relative to those of other genotypes, to the gene pool of the succeeding generation.

Some examples will make the concept of Darwinian fitness clear. Mules are no less vigorous than their parents, horses and donkeys, under many conditions. Yet the selective value of a mule is zero, since mules are completely sterile and their genes are not perpetuated. By contrast, hereditary diseases that afflict their victims after the close of the reproductive period (diseases of old age) do not directly diminish Darwinian fitness, if children are born before the onset of the disease. In the house mouse, certain genes are favoured by natural selection because they are passed to more than 50% of the progeny of the heterozygous males, despite the fact that these genes are lethal in homozygous condition (*i.e.*, in a double dose).

How Selection Operates.—The evolution of man, like that of any other biological species, was governed in the past by natural selection acting upon natural variation (*see* VARIATION). Some writers have claimed that in modern mankind natural selection has been relaxed or even removed altogether. This claim is, at least in part, based on a misunderstanding. Natural selection is not necessarily selection taking place in nature untouched by human activities. No matter under what conditions our remote ancestors may have lived, the environments in which people live at present are largely those created by human culture. Even if these environments be regarded as unnatural, natural selection may operate to maintain or improve the adaptedness of our species for living in human societies with various forms and degrees of cultural development. Only if all existing human genotypes produced numbers of surviving children in exact proportions to their frequencies in the populations could selection be said to have been removed. This obviously is not the case. The inhabitants of different parts of the world have different reproductive rates, and everywhere some people die childless or have fewer children than do others.

In man as in all other organisms, the Darwinian fitness of different genotypes changes with time and varies in different places. Fitness is clearly a function of the environment. Certain bacteria are killed when placed on culture media containing the antibiotic streptomycin, except for carriers of a variant genotype, which are resistant to the drug. Moreover, some of the resistant variants die on media *without* streptomycin, because they "perversely" require this substance for growth. Obviously the fitness of the normal and the resistant genotypes in these bacteria depends upon the presence or absence of streptomycin in the environment.

Natural selection operates at present to perpetuate human genotypes that enable their carriers to live and reproduce in the environment created by man's culture and technology. Some of the genotypes that survive in these environments might have been incapacitating in our prehuman ancestors, and vice versa. This means that natural selection can enhance the adaptive value of the organism only in the environments that exist here and now; it has no prescience of the future. Natural selection is as incapable at present of maintaining the adaptedness of man to the conditions of the Stone Age as it is of preparing him for the conditions of the Atomic Age. In short, natural selection may be said to be thoroughly opportunistic.

The fact that natural selection continues to operate does not necessarily mean that all will be well with the human species in the future. A species may become too specialized to live in environments that may prove to be temporary, and may become extinct when the environments change (the rather sudden passing of the dinosaurs is a dramatic example). Furthermore, Darwinian fitness should not be equated with excellence in human estimation. It is not only possible but indeed probable that man will have to take the direction of his evolution in his own hands, and to replace natural by artificial selection.

Genetic Basis of Selection.—A gene may exist in more than one form, spoken of as alleles. Given allele *A* and allele *a* (in which *A* is dominant to *a*), a hybrid individual, or heterozygote (having received an *A* from one parent and an *a* from the other), will display only the character represented by *A*, that represented by *a* being masked and called recessive. There often exists a selection pressure favouring one allele over another. This pressure may be slight or intense. Selection will operate to preserve both alleles in the population, and tend to establish an equilibrium —even if one of the homozygous conditions (represented by *AA* and *aa*) bestows serious disadvantages upon the individual. Thus natural selection may operate to preserve in a population hereditary diseases or malformations, provided that the alleles that produce them in homozygotes confer hybrid vigour on the heterozygotes (*see* GENETICS: *Deleterious Genes in a Population*). The intensities of natural selection involved in the maintenance of some of these heterotic situations in nature are very large, and offer favourable opportunities for investigation of selection phenomena. The mathematical representation of natural selection is discussed in detail in GENETICS OF POPULATIONS; *see* also EVOLUTION, ORGANIC: *Statistical Transformation of Populations.*

ARTIFICIAL SELECTION

In contrast to natural selection, artificial selection is a purposeful process. The animal or plant breeder chooses some individuals to be the progenitors of the next generation, and limits or prevents the reproduction of others; in this way, breeders direct changes in the genetic composition of a population. The techniques of artificial selection were evolved by centuries of trial and error, and have been highly effective in developing domesticated animals and cultivated plants in accord with human needs and desires. With the progress of genetics the art became a scientific technique.

For any kind of selection to be effective, the individuals who leave relatively greater numbers of surviving progeny must be genetically different from those who do not. In other words, there must be some genetic variance in the traits selected. This was established in 1909 by W. Johannsen, who carried out experiments on selection for large and small seed size in beans. Selection was effective only when applied to a mixture of different genetic constitutions. It was without effect when applied to a pure line, *i.e.*, to progeny obtained by self-pollination of single individuals in a series of generations. Individuals that belong to a pure line are genetically identical, and any observable (phenotypic) differences between them are due to their having grown in varying environments.

One of the classical, and still widely used, methods of ensuring

a supply of genetic variance is hybridization (*see* HYBRIDISM). Many of the successful varieties and breeds of agricultural plants and animals have been obtained by selection within progenies of hybrids between two or more parental varieties. The parental varieties crossed often come from different countries and some of them may be initially quite remote from the ideal type that the breeder wishes to obtain. They may, nevertheless, supply the genetic building blocks from which selection may construct superior new genetic endowments. Such genetic building blocks can be obtained by using agents that cause mutations (*q.v.*), especially X-rays and other ionizing radiations.

See ANIMAL BREEDING; PLANT BREEDING. For a discussion of the artificial selection of human traits, *see* EUGENICS.

SEXUAL SELECTION

Females and males in many species of animals are more or less different in body size, coloration, or in the presence of various secondary sexual characters, such as beards in human males or enlarged breasts in human females. Darwin suggested that such differences evolved because the possessors of certain characteristics were generally more successful in mating or were usually more fertile than those with other characteristics. (*See* COLORATION, BIOLOGICAL: *Animal Colours: Function.*)

The range of applicability of the theory of sexual selection is rather uncertain. It is most plausible when applied to polygamous species, such as some monkeys or deer, in which the largest and strongest males may collect harems of numerous females, while weaker males are at least temporarily excluded from reproduction. The theory fares less well with monogamous species, in which fixed pairs are established for life or at least for a season. With males and females being about equally numerous, almost all individuals in such species find mates and consequently have a chance to reproduce. Secondary sexual traits may still be selected in monogamous species, if these traits enable their possessors to survive more often or to produce more offspring. (However, this is natural selection, rather than sexual selection as originally defined.)

A form of the theory of sexual selection may perhaps be used to explain the development of species signals and recognition marks. Several closely related species of animals often live in the same territory. And, hybridization of members of distinct species often results in production of poorly viable or sterile progeny. It is, therefore, advantageous for these species to develop characteristics that reduce the frequency of the "mistakes," and confine mating to individuals of the same species. Species differences in courtship rituals, and in the appearance, behaviour, and odours of the two sexes may well have arisen by way of sexual selection.

See COURTSHIP, ANIMAL; SEXUAL BEHAVIOUR: *Sexual Behaviour and Evolution; see* also references under "Selection" in the Index.

BIBLIOGRAPHY.—R. A. Fisher, *The Genetical Theory of Natural Selection* (1930); I. M. Lerner, *The Genetic Basis of Selection* (1958); J. Maynard Smith, "Sexual Selection," in S. A. Barnett, *A Century of Darwin* (1958); C. H. Waddington, *The Strategy of the Genes* (1957); C. Zirkle, "Natural Selection before the 'Origin of Species,'" *Proc. Amer. Philos. Soc.*, vol. 84 (1941); T. Dobzhansky and G. Allen, "Does Natural Selection Continue to Operate in Modern Mankind?" *Am. Anthrop.*, vol. 58, pp. 591 ff. (1956). (T. DY.)

SELENE, the Greek word for the moon, was personified in ancient Greece as the moon goddess. She had no important cults or temples but was worshiped to a minor extent at the new and full moon. In genealogy her parents were the Titans Hyperion and Theia, her brother was Helios (sometimes alternatively called her father), her sister Eos (Dawn), her husband Zeus, and her daughter Pandia. By the 5th century B.C. Selene was sometimes identified with Artemis. One of the few legends in which Selene appears tells how she was wooed by Pan, who enticed her with a white fleece. The best known legend concerns her love for Endymion (*q.v.*). In art she was usually represented as a woman with the moon (often in crescent form) on her head driving a two-horse chariot (as on the lost statue from the east pediment of the Parthenon in Athens). Selene was called Luna in Latin; she had temples in Rome on the Aventine and Palatine hills.

See Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft*, 2nd series, vol. 2, col. 1136–44 (1921). (H. W. PA.)

SELENGA-ORHON RIVER SYSTEM consists of the Selenga River (Mongolian Selenge Gol) formed by the junction of the Ideriyn (Iderüin) and Mörön rivers, which rise in the Mongolian People's Republic (Outer Mongolia) south of Lake Hövsgöl (Höbsogöl), and the Orhon River (Mongolian Orhon Gol), which rises farther south at about latitude 47° N on the slopes of the Hangai Range. Both rivers flow in a northeasterly direction to join at Sühbaatar (Sühebaatar); the river, from that point called the Selenga, continues north into the Russian Soviet Federated Socialist Republic, U.S.S.R., and bends eastward to Ulan-Ude, the capital of the Buryat Autonomous Soviet Socialist Republic; from there it continues north to Tataurovo, where it swings abruptly westward and flows on into Lake Baikal. From Sühbaatar to its delta the Selenga flows down a gently graded slope giving access from Outer Mongolia to the plains of Asiatic Russia. The Trans-Siberian Railway follows the valley upstream to Ulan-Ude; from there a branch continues to Sühbaatar, then follows the Orhon and the Haraa (Hara) Gol to Ulan Bator, capital of the Mongolian People's Republic, from where it extends southeast to Peking.

The Selenga is 620 mi. (1,000 km.) long and is navigable from its mouth to beyond Sühbaatar. Its basin covers an area of 172,600 sq.mi. (447,000 sq.km.), and its width varies from an average 20–50 yd. (18–45 m.) to 70–170 yd. (64–155 m.) at high water; it is fed mainly by the summer rains. High water is 6–12 ft. (2–4 m.) above normal level, and the absolute variation is 9 to 18 ft. (3–5 m.). The river freezes over from the first half of November until the end of April. The annual flow is estimated at 7 cu.mi. The main tributaries in the Mongolian People's Republic are, on the right, the Orhon and Hanuy (Hanuyn) Gol; on the left, the Egiyn (Egiin) Gol. The main ones in the U.S.S.R. are, on the right, the Chikoy, Khilok, and Uda; on the left, the Dzhida and Temnik. The Orhon is 698 mi. (1,124 km.) long, with a basin of 51,280 sq.mi. (132,800 sq.km.); its main tributaries are the Iro, Hara, and Tola Gol; it is navigable upstream to the Iro. Karakorum, the ancient capital of the Mongol Empire, was situated near the upper Orhon. (R. E. F. S.)

SELENIUM is a chemical element closely allied in physical and chemical properties with sulfur. It was identified in 1818 by J. J. Berzelius, in Sweden. It had been found a year earlier in a red deposit that had collected on the floor of a room in which sulfur, obtained as a by-product from a copper mine, had been used; but its similarity to tellurium had caused it to be mistaken for that element. Since the name tellurium had been coined from *tellus*, the Latin word for "earth," Berzelius named his discovery after the moon (in Greek, *selene*).

In 1873, Willoughby Smith discovered that crystalline selenium develops a tremendous increase in electrical conductivity when exposed to light, regaining its resistance immediately when the light is shut off. This property eventually was applied in motion-picture sound tracks, television, the telegraphic transmission of photographs and many other devices depending on light-sensitive materials. (*See* below, *Selenium Cells and Rectifiers; see* also SELENIUM CELL.)

The symbol of selenium is Se; atomic number 34; atomic weight 78.96. It is a complex element with six valence electrons ($4s^2$, $4p^4$) in the outer (N) main level; six stable isotopes with mass numbers (in order of decreasing abundance) 80, 78, 76, 82, 77 and 74; and eight radioactive isotopes, including Se^{75} (half life, 127 days) and Se^{83} (half life, 25 min.) have been prepared.

Occurrence.—Selenium is widely distributed throughout the world, although in small quantities (10%–7% in igneous rocks), being the 69th element in order of abundance, between silver and argon. It rarely occurs native; frequently it accompanies native sulfur, but is commonly found in combination with the heavy metals as selenides and accompanying the more abundant metal sulfides. The important minerals are clausthalite, PbSe; eucairite, $(Ag,Cu)_2Se$; crookesite, $(CuAgTl)_2Se$; naumannite, Ag_2Se and zorgite (a complex selenide containing 31% Se). When any of the selenium-containing sulfide minerals are subjected to chemi-

cal processing operations, selenium appears as a by-product.

Physical Properties.—Selenium exists in several allotropic forms including a glassy form, two monoclinic (metalloid) forms and hexagonal (metallic) selenium, all stable to some extent at room temperature. Glassy or amorphous selenium is obtained when any form is heated above the melting point and subjected to rapid cooling. No well-defined freezing point is observed, the mass gradually thickening until at 40°–50° C. a glassy mass develops. The cold glassy product is brittle, exhibits conchoidal fracture and is only slightly soluble in carbon disulfide. Rapid reduction of a cold aqueous solution containing a selenium compound yields red amorphous selenium which changes to the gray form upon heating the solution. Orange to red modifications of monoclinic selenium are obtained when the red amorphous powder is extracted with carbon disulfide and this solution is evaporated and cooled. These modifications may be changed to gray metallic selenium by heating, the orange at about 110° C. and the red at about 125° C. Gray metallic selenium may also be prepared by allowing molten selenium to cool slowly to room temperature or by annealing other modifications slightly below the melting point. This form is completely insoluble in carbon disulfide, but soluble at higher temperatures in naphthalene, aniline and ethyl benzoate. It is this form of selenium which possesses the remarkable property of having its conductivity increased by light. That freshly precipitated red selenium may be amorphous is a matter of doubt. The metallic gray selenium has a melting point of 217° C. and a specific gravity of 4.79 (25° C.). It crystallizes in rhombohedral crystals in the hexagonal system.

Production.—In the United States and Canada the chief source of selenium is the anode slime from copper refineries, lesser amounts from the roasting of metallic sulfides. In most cases the selenium is isolated as selenious acid and elemental selenium is obtained by reduction with sulfur dioxide.

The chief producing countries are the U.S. and Canada; other sources include Sweden, Japan, and Zambia.

U.S. production of selenium in the 1960s was about 1,000,000 lb. per year. Consumption of selenium has been affected by high prices as well as sharp inroads by competitive materials. Germanium and silicon were substituted for selenium in rectifiers; tellurium replaced some selenium in rubber and stainless steel; and mercury and cadmium are substituted for selenium in the glass and pigment industries.

APPLICATIONS

In the 1960s by far the largest amount of selenium was used in the manufacture of electronic equipment, particularly power rectifiers. The chemical industry was a large consumer, the most important applications being in pigments, pharmaceuticals and rubber. The glass and steel industries also used large amounts, and there was a wide range of minor applications.

Selenium Cells and Rectifiers.—In 1873 it was found that selenium changed its electrical conductance under the influence of light. Shortly thereafter, a simple electric circuit containing a selenium element was shown to generate current when the selenium was exposed to light. Not until 1930 was this idea developed to a stage where it became useful in the form of a layer-barrier cell. Such a light-sensitive cell consists of a conducting metal back of copper, aluminum or brass, etc., coated with a thin film of metallic selenium, this in turn coated with a thin translucent layer of gold. A transparent filter usually protects the gold surface. One lead of the circuit is connected to the metal back and the other to the gold surface, the two leads being attached to a sensitive galvanometer or other measuring device. When light shines on the gold surface of the cell, sufficient electromotive force is developed to operate a sensitive galvanometer without any external amplifying circuit. This cell was used for photographic exposure meters, photoelectric colorimeters and other instruments designed to measure light. Although these cells have been improved, they are in general the photovoltaic type.

The dry rectifier differs from the selenium cell in the substitution of a low melting alloy such as Wood's metal for the gold film. During application of the alloy coat on the selenium or in annealing of the selenium, a surface develops or is provided between the selenium and the alloy which resists backward flow of alternating current. By allowing current to flow in only one direction the rectifier converts alternating current to direct current.

Such rectifiers are used for charging batteries, electroplating and operating a variety of direct current devices. When the rectifiers are connected in multiple units in series or parallel, they provide a wide variety of sizes and capacities. The cells operate at high efficiency over a wide range of loads.

Glass.—A large consumer of selenium is the glass industry. Selenium has replaced manganese as a glass decolorizer. The objectionable green colour produced in glass by iron is neutralized by the addition of a small amount of selenium. The addition of a larger amount of selenium, either as such or as sodium selenite, imparts to glass a clear red colour useful for signal lamps.

Rubber.—Selenium is used with sulfur in the vulcanization of certain special rubber products. Crystallization of sulfur in rubber may be induced by selenium because it is isomorphous with monoclinic sulfur. In this manner, metastable solutions of sulfur are prevented and initial rapid bloom (haze surface) of uncured stocks is eliminated. Selenium exerts a distinct accelerating action upon rubber-sulfur mixtures and increased abrasion resistance of certain vulcanizates.

Selenium Oxychloride as a Solvent.—The specific solvent properties of this reagent were studied by V. Lenher. The reagent is a solvent for certain natural and synthetic resins, fish oils, etc.; it helps to distinguish between saturated and unsaturated hydrocarbons, attacking the latter but not the former. In a mixture of benzene and heptane, it dissolves the former, whereas the latter floats unchanged. Selenium oxychloride, $SeOCl_2$, is a selective solvent for Cb_2O_5 in the presence of Ta_2O_5 and MoO_3 in the presence of WO_3. A solution of molybdenum trioxide in selenium oxychloride exhibits a photochemical effect; in the dark, the solution maintains the yellow colour of selenium oxychloride; exposed to light it changes to blue. The reaction is reversible.

Miscellaneous Uses.—Cadmium sulfoselenide forms a red pigment and is produced in relatively large amounts. Colours ranging from orange to maroon are obtained, depending on the method of preparation. The addition of a small amount of selenium to stainless steel, copper and copper-rich alloys improves machinability and is a commercially established practice. A protective coat on magnesium alloys may be made with selenium. Chemical uses include that of selenium dioxide as an oxidizing agent in organic reactions, and the use of selenium and copper selenite as catalysts in oxidizing reactions, such as the Kjeldahl digestion. Ferrous selenide is used as a catalyst in cracking petroleum. The use of selenium compounds on battery plates prevents sulfation and therefore permits higher charging rates. The use of selenium in fungicides and insecticides is limited, because of the possibility of contaminating food products. With the common use of mercury in laboratories and plants, the detection of small concentrations of mercury has become important. A selenium-sulfur mixture blackens quickly in the presence of mercury vapour and is used to indicate exposure of workers to mercury fumes. Selenium has been used in flameproofing of textiles and certain cable coatings.

Some species of plants are not only tolerant of selenium but actually require selenium for growth and development. Since selenium occurs with uranium in certain types of deposits, the presence of plants which require selenium for growth may indicate the presence of uranium in a deposit. Such plants are sometimes called indicator plants. These plants may contain as much as 1.5% of selenium and have an offensive garliclike odour.

CHEMICAL PROPERTIES AND COMPOUNDS

Selenium reactions lie intermediate between sulfur and tellurium. At the boiling point selenium vapours consist of mixtures of Se_8 and Se_2. Heated in oxygen the element burns with a blue flame, selenium dioxide being formed. Direct combination of the element with hydrogen, oxygen, halogens and many metals is common. The halogen compounds are hydrolyzable but not so easily as comparable sulfur compounds.

Selenium is a weaker reducing agent than sulfur. Solution in concentrated sulfuric acid yields a green solution of $SeSO_3$. The acid-forming tendencies of the element are most pronounced in the higher valences. The valence types of selenium compounds are −2, 0, +2, +4, +6. Gray selenium has a melting point of 217° C.; heat of fusion 1.56 kg.cal. per gram; heat capacity 6.0 (25° C.) cal. per gram-atom; entropy at 25° C., eutectic units 10.0; and magnetic susceptibility -0.32×10^{-6}.

Inorganic Compounds.—*Hydrogen selenide,* H_2Se, is a colourless gas which has a disagreeable odour and is extremely toxic. It has a boiling point of −41.3° C. and a melting point of −64° C. It burns with a blue flame, giving either red selenium or white selenium dioxide, depending on the amount of oxygen present. The gas is prepared by dropping water on aluminum selenide, the latter being prepared by direct combination of selenium and powdered aluminum. Hydrogen selenide is less stable and is a stronger reducing agent than hydrogen sulfide. This behaviour may be associated with the fact that selenium has a larger atomic size than sulfur.

Metallic selenides are produced by direct combination of metals with selenium or by precipitation from solutions of metallic salts by hydrogen selenide. The alkali selenides are colourless but become red by addition of more selenium to form polyselenides such as Na_2Se_2.

Selenium dioxide, SeO_2, is easily prepared by burning selenium in an excess of oxygen. Impure selenium dioxide (called whispers) is obtained from dust chambers which catch the dust from roasting certain metallic sulfides. This material is contaminated with selenium and other impurities and may be purified by addition of nitric acid, removal of the excess nitric acid by fuming and finally obtaining the selenium dioxide by sublimation. The dioxide crystallizes in long colourless crystals, melts under pressure at 340° C. and dissolves in water to form selenious acid.

Selenium trioxide, SeO_3, has been prepared by the action of sulfur trioxide on potassium selenate and also by the vacuum sublimation of a 1:1 mixture of selenic acid and phosphorus pentoxide.

Selenium monoxide is a definite oxide but has been identified only in the gaseous state by spectroscopic means.

Selenic acid, H_2SeO_4, may be prepared by addition of bromine to silver selenite. This acid, the analogue of sulfuric acid, is usually seen as a sirupy liquid, but has been obtained solid (melting point 58° C.). The solid acid is prepared by desiccation of a strong selenic acid solution. It cannot be obtained pure by distillation since some of the selenic acid decomposes to form selenious acid. Selenic acid has the noteworthy property of dissolving gold in the presence of oxygen. Like telluric acid, it oxidizes hydrochloric acid to give free chlorine. The selenates are similar to the sulfates; the soluble selenates are stronger oxidizing agents.

Selenium tetrafluoride, SeF_4, was obtained in the pure state (1928 by E. B. R. Prideaux and C. B. Cox) by the interaction of selenium tetrachloride and silver fluoride. It is a colourless liquid boiling at 93° C. and melting at −13.2° C. Its specific gravity is 2.77. It attacks glass and is completely hydrolyzed by water to selenious and hydrofluoric acid.

Chlorides.—Selenium forms a number of chlorides. Diselenium dichloride, Se_2Cl_2, a brownish-yellow liquid with pungent odour, is decomposed by water into hydrochloric and selenious acids with elimination of red selenium. Selenium tetrachloride, $SeCl_4$, a colourless crystalline mass which vapourizes without melting, is produced by burning selenium in chlorine. Selenium oxychloride, $SeOCl_2$, a pale yellow liquid boiling at 177.2° C. and melting at 8.5° C., is prepared by direct chlorination of a mixture of selenium and selenium dioxide. The impure liquid is purified by distillation. Selenium substitutes for oxygen in a number of acids, forming such compounds as sodium selenocyanate.

Organic Compounds.—In many instances the equivalent of sulfur compounds are known. The following types of aliphatic compounds are known: types RSeH, R_2Se, RSeR′, and R_2Se_2, complex compounds of platinum and platinum halides with dialkyl selenides and types R_2SeX_2 and R_2SeX, R representing a variety of alkyl radicals and X various halogens. Compounds derived from aldehydes and ketones are similar to selenoacetaldehyde, CH_3CHSe, and diselenoacetone $(CH_3.CS_e.CH_3)_2$. Selenoacetic, CH_3COSeH, and ethyl selenoformic, C_2H_5COSeH, are examples of the relatively simple aliphatic acids, and 2:4 dinitrophenylselenoacetic acid is an example of the more complex. Nitric acid oxidizes dialkyl diselenides to alkyl seleninic acids.

A large number of aromatic derivatives of selenium have been prepared. Types in which R represents a variety of aryl groups are RSeH, R_2Se, R′SeR and RSe.SeR. The interaction of selenium tetrachloride and benzene in the presence of aluminum chloride yields diaryl selenides and diaryldiselenides. Diarylsulfones heated with precipitated selenium yield diaryl selenides, mixed selenides being formed if the radicals in the sulfones are different. Compounds of the types R_2SeX_2 and R_2SeO, R being an aryl group, are illustrated by diphenyl selenium dibromide and diphenyl selenoxide. Although bromine reacts directly with diaryl selenides in carbon disulphide to form bromides, the direct union of chlorine is less usual. Nitric acid reacts with diaryl selenides or their dibromides to form nitrates. The dibromides treated with moist silver oxide yield hydroxides. When sodium hydroxide is used, the diaryl selenoxides, RSeO, are obtained. Many mixed compounds are known. Examples are phenylmethylselenium bromoiodide and phenylmethylselenium dihydroxide. A comprehensive presentation of organometallic compounds was provided by Newton Friend and A. E. Goddard (1937).

Detection and Estimation of Selenium.—Selenium is precipitated from its inorganic compounds by hydrogen sulfide and is redissolved by ammonium sulfide. A characteristic test is the precipitation of red selenium from a hydrochloric acid solution by sulfur dioxide, hydrazine, hydroxylamine, etc. The gravimetric estimation is made by weighing this precipitate after it has been boiled to convert it to the gray modification. Selenium may be determined iodometrically, and selenious acid may be titrated with permanganate. Selenates may be estimated by distillation with hydrochloric acid. The liberated chlorine is collected in potassium iodide and the iodine thus liberated is titrated with thiosulfate. An excess of hydrazine can be added to selenious acid and the excess reagent titrated with potassium iodate to an ICI end point. Selenious acid may be reduced by hydriodic acid and the liberated iodine titrated with thiosulfate. Thiosulfate is used as a reducing agent for selenious acid and the excess reagent titrated with iodine. Tellurium must first be removed or accounted for. Instrumental methods include potentiometric titration, polarographic determination, coulometric and spectrophotometric measurements.

Medical Aspects.—*Selenium in Soils and Plants.*—Although the toxic nature of selenium was recognized as early as 1842, it was not until 1929 that a thorough study was initiated to determine the effects of the element in animal and human hygiene. It is recognized that plants grown on seleniferous soils may absorb selenium (*see* LOCOWEED). When used as food for animals or humans, the plants are toxic, causing either chronic or fatal poisoning. Affected areas have been surveyed and in some cases planting discontinued. Wheat samples grown in some areas contained 30 parts per million of selenium; only 10 to 20 parts per million are necessary to produce chronic poisoning in cattle. Although the hull, bran and patent flour become separated in milling operations, enough selenium can remain in the flour to be a hazard to humans, particularly if the wheat is grown on highly seleniferous soils. Selenium poisoning in horses is accompanied by loss of hair from tail and mane and by abnormal hoof growth; extreme cases produce blind staggers or even death. In humans selenium may concentrate in the lungs, liver, kidney or spleen. Occupational dermatitis is a mild form of poisoning. Few cases of death have been reported. Recent studies indicate arsenic as a possible inhibitor for selenium poisoning. It is known that soils which are leached by abundant ground water lose some selenium and that plants grown in areas of plentiful rainfall contain less selenium, even though the soil contains the element. The Colorado river, for example, carries away selenium from certain irrigated areas in the western United States. It is interesting to note that bot-

tom deposits and growths in the Gulf of California and the Gulf of Mexico contain notable amounts of selenium.

Protection Against Liver Necrosis.—Klaus Schwarz and C. M. Foltz and others have reported that liver necrosis in rats may result from low cystine and simultaneous deficiency in vitamin E and factor 3. As little as 13.33 micrograms of sodium selenite in 100 grams of diet afford complete protection against liver necrosis. The dose necessary for protection is less than 1% of the toxic dose.

See also references under "Selenium" in the Index.

BIBLIOGRAPHY.—B. S. Hopkins, *Chapters in the Chemistry of the Less Familiar Elements*, rev. ed. (1940); J. N. Friend, *Organometallic Compounds*, vol. ii, part 4 (1937); E. P. Painter, "Chemistry and Toxicity of Selenium," *Chem. Rev.*, 28:179 (1941); Fritz Ephraim, *Inorganic Chemistry* (1943); U.S. Bureau of Mines, *The Minerals Yearbook*, (1953), and *Information Circular 7715* (1955); G. R. Waitkins, A. E. Bearse and R. Shutt, *Industr. Engng. Chem. (Anal.)*, 24:899 (1942); D. M. Yost and H. Russell, *Systematic Inorganic Chemistry of the Fifth- and Sixth-group Non-metallic Elements* (1944); J. F. Gmelin, *Handbuch der Anorganischen Chemie*, 8th ed., System-Nummer 10, Teil A, Lfg. 1 (1942), Teil A, Lfg. 2, (1950); W. Conard (ed.), *Inorganic Syntheses*, vol. i, ii, iii, iv (1939, 1946, 1950, 1953); R. E. Kirk and D. F. Othmer (eds.), *Encyclopedia of Chemical Technology*, vol. 12, p. 1454 (1954). (V. W. M.)

SELENIUM CELL. The selenium cell is a photoelectric device used to generate or control an electric current through the influence of light falling upon the cell. Selenium photocells are commonly used in photographic exposure meters, burglar alarms, electronic door opening and counting devices, electronic control systems in factory assembly lines and industrial colour perceptors. In some ways (colour perception, sensitivity) the performance of selenium photocells is superior to that of vacuum-tube photocells. (For a discussion of the latter, *see* PHOTOELECTRICITY.) The fact that selenium cells do not respond instantaneously to changes in stimuli, however, makes them unsuitable for certain purposes, such as high fidelity sound reproduction on film.

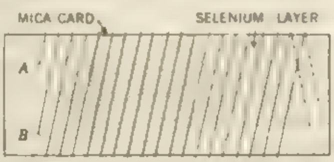


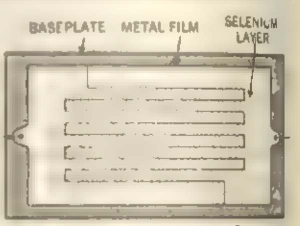
FIG. 1.—THE SELENIUM LAYER OF A SIEMENS PHOTOCONDUCTIVITY CELL CONDUCTS CURRENT BETWEEN THE TURNS OF WIRE (A AND B)

The photoelectric properties of selenium, a nonmetallic element, are exhibited in two ways: (1) When light falls on a specimen of crystalline selenium the electrical conductance of the selenium is increased. A device employing this effect is called a selenium photoconductive cell. If such a cell is connected in a circuit with a battery, a small current, called the dark current, will flow if the cell is not illuminated. If the cell is illuminated, the increased conductance of the selenium element permits a larger current to flow. The increase in current may then be used to operate a relay, produce an indication on a current meter or perform some other useful function in the circuit associated with the cell. (2) When selenium is placed in contact with an electrode made of a suitable metal, the contact has the ability to generate an electromotive force (E.M.F.) when exposed to light. The illuminated contact thus behaves like a battery, with the selenium becoming the positive terminal and the contacting metal becoming the negative terminal. A device employing this property of selenium is called a selenium photovoltaic cell. Such a cell can be used as a photoelectric battery in a practical electrical circuit.

The photoconductivity effect in selenium was first reported in Great Britain by Willoughby Smith in 1873. Selenium was the first solid material found to exhibit this effect. Following this discovery a number of photoconductivity cell designs were developed. The discovery of the photovoltaic effect at contacts between selenium and various metals was first reported by W. G. Adams and R. E. Day in England in 1877. Shortly thereafter, in 1883, Charles E. Fritts in the U.S. constructed a practical selenium photovoltaic cell which embodied most of the features found in cells of modern design.

Unfortunately, much of this early work either went unrecognized or was lost to sight; and the selenium photovoltaic cell had to be "rediscovered" later.

Selenium is one of the elementary chemical substances (*see* SELENIUM). It melts at 217.4° C. When cooled from the molten state it "solidifies" into a hard, brittle, noncrystalline material in which its physical condition is actually that of an extremely viscous liquid, like glass. In this condition it is called amorphous selenium. The electrical resistance of amorphous selenium is so high as to classify it as an insulator; and in the amorphous state selenium exhibits no usable photoelectric behaviour. If a specimen of amorphous selenium is heat treated at any temperature between about 120° C. and its melting point, however, conversion to the more useful gray crystalline form occurs. The electrical resistance of gray crystalline selenium between opposite faces of a 1-cm. cube is about 1,000 ohms, the exact value depending on impurity content and on the details of the annealing treatment by which the specimen was converted to the crystalline form. It is in this form that selenium is employed in the structures of photoconductive and photovoltaic cells.



ADAPTED FROM G. P. BARNARD "THE SELENIUM CELL", BY COURTESY OF CONSTABLE & CO. LTD., LONDON

FIG. 2.—PHOTOCONDUCTIVITY CELL IN WHICH THE SELENIUM LAYER BRIDGES THE GROOVE SEPARATING TWO SECTIONS OF A METAL FILM

Photoconductive Cells.—The basic parts of a selenium photoconductive cell consist of an active portion of gray crystalline selenium, so disposed that the light to be detected or measured can fall upon it, with metal contacts to the selenium permitting current to flow into, through and out of it. One of the simplest of these cells was devised by the German inventor and industrialist Werner Siemens in 1876. In the Siemens cell two fine copper wires are wound around a rectangular mica card as shown in the sketch in Fig. 1. The windings are separated so that the wires do not touch each other anywhere. This assembly is then placed on a hot plate and raised to a temperature a few degrees higher than the melting point of selenium. A piece of selenium placed on this assembly soon melts to a molasseslike liquid, in which condition it is spread in a thin film over the upper surface of the mica, completely embedding the copper wires. The assembly is then removed from the hot plate and allowed to cool, the selenium hardening into an amorphous layer. To convert the selenium to the gray crystalline form, the assembly is again heated for several hours at a temperature of about 180° C.

Electrical connections to this cell are made at the two contacts *A* and *B*; the other ends of these wires are left open circuited. Any current flowing between these terminals must flow through the selenium film between adjacent turns of the wires. Illumination of the selenium film then increases the total parallel conductance of the cell between these terminals. A Siemens-type cell with a selenium layer area of 1 sq.in. may exhibit a resistance of one or two megohms in complete darkness. When exposed to full sunlight the cell resistance drops to a few thousand ohms.

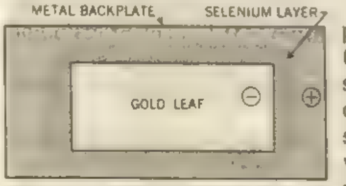


FIG. 3.—FRITTS'S PHOTOVOLTAIC CELL IN WHICH AN ELECTROMOTIVE FORCE IS GENERATED BETWEEN THE GOLD LEAF AND THE BACKPLATE UPON EXPOSURE TO LIGHT

A different form of selenium photoconductive cell, shown in the sketch in Fig. 2, employs a support plate made of slate, glass or other insulating material. One side of the support plate is provided with a film of deposited metal, which is then scribed through with a sharp tool, in a zigzag line, in such a way that the metal deposit is divided into two sections separated and insulated from each other by the zigzag line. A thin film of crystalline selenium is then laid down on the surface of the metal deposit, filling in the scribed line. Light falling anywhere on the selenium increases conduction between the two portions of the metal deposit which wires connect to the cell.

In their operation selenium photocells have some advantages over vacuum-tube photocells; they also display some disadvantages. A properly made selenium cell may be a hundred times more sensitive to a given amount of illumination than a vacuum-

tube photocell. Also, the selenium cell is responsive to light throughout the entire range of colours visible to the human eye, while many vacuum-tube photocells, particularly the more sensitive ones, are responsive only to light of the green, blue or purple regions of the colour spectrum. The response of a selenium photocell, on the other hand, is not directly proportional to the amount of light falling upon it. Instead, the change in conductance is proportional to the square root of the light flux on the cell. This relationship means that the cell is relatively more sensitive for the lower levels of light intensity.

Another disadvantage exhibited by the selenium photoconductivity cell is the time lag of its response. When a selenium cell is suddenly exposed to light, the change in conductance is not instantaneous. It may require a thousandth of a second or more for the cell conductance to rise to the new value appropriate to the illuminated condition. Similarly, if the light is suddenly cut off, a comparable length of time is required for the conductance to resume its dark value. This time lag in response means that the cell cannot follow the rapid fluctuations of light intensity which must be dealt with, for example, in the high fidelity reproduction of sound on film.

Selenium photoconductivity cells must find most of their uses, therefore, in applications where the cell must distinguish between light on and light off, or where the cell is exposed to light which is either steady or else varying at frequencies lower than those corresponding to speech or music.

The physics of the photoconductivity effect is easy to describe. The electrical conductance of any substance is proportional, among other things, to the number of electronic particles per unit volume which are free to move about within the substance and which, by their motion when voltage is applied, cause a current to flow. Selenium has a relatively small number of such free particles available in the dark. However, when light shines on a specimen of selenium and is absorbed by it, the energy of the radiation is expended in the freeing of additional numbers of conducting particles which increase the total number available and hence increase the electrical conductance of the illuminated specimen.

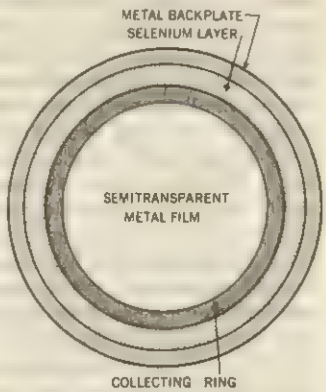


FIG. 4.—MODERN PHOTOVOLTAIC CELL SIMILAR TO FRITTS'S CELL (FIG. 3)

Photovoltaic Cells.—By far the most extensive modern use of selenium in photoelectric cells is in devices of the photovoltaic or self-generating variety. Fig. 3 shows a sketch of an early selenium photovoltaic cell constructed by Fritts. This cell made use of a thin layer of crystalline selenium spread out on the surface of a metal base plate. On the outer surface of the selenium layer was placed a translucent film of gold leaf. When the cell was illuminated from the gold leaf side, Fritts observed that an electromotive force (E.M.F.) appeared between the gold leaf and the base plate metal. If the gold leaf and the plate were connected through an external circuit, a current flowed as long as the illumination of the cell was continued.

Modern selenium photovoltaic cells are but little different from Fritts's early model. A thin layer of selenium is applied to one side of a supporting base plate of iron, nickel or aluminum. Instead of spreading the molten selenium onto the base plate at an elevated temperature, however, present-day manufacturers apply the selenium film by vapour deposition in vacuum. Following the customary heat treatment, an outer contact of gold is then applied to the surface of the selenium layer, again by vapour deposition in vacuum. This gold film must be thin enough to permit a substantial fraction of the light to penetrate to the gold-selenium contact, yet thick enough to permit good lateral conductance for the current which the cell is to deliver. Contact to the gold film is provided in modern design by a built-up collecting ring of painted-on graphite or more heavily deposited metal around the outer periphery of the gold film.

Fig. 4 shows a section sketch of a modern selenium photovoltaic cell showing these features. A cell of this construction will deliver an open-circuit E.M.F. of 0.3 or 0.4 volt in light of 100 foot-candles intensity. Under the same conditions of illumination, such a cell having an area of 1 sq.in. will deliver a current of 250 microamperes into a 1,000 ohm external resistance. Fig. 5 shows a family of characteristics for a modern cell having an effective area of about 2 sq.in. Each curve gives the output current as a function of light intensity for a particular value of external circuit resistance. Note that in no case is the current linearly proportional to circuit resistance, but is more nearly so for the lower values of resistance. By short-circuiting the cell, one could obtain a characteristic almost completely linear, but such a circuit arrangement would draw no useful power from the cell and would correspond, therefore, to a very inefficient use of the cell.

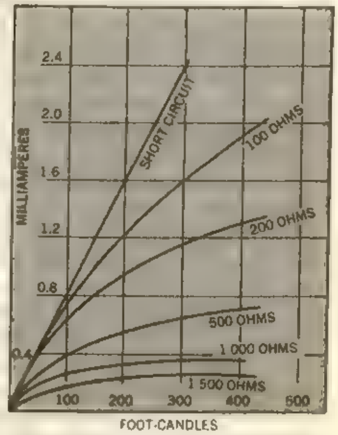


FIG. 5.—RELATIONSHIPS BETWEEN OUTPUT CURRENT AND LIGHT INTENSITY FOR A MODERN SELENIUM PHOTOVOLTAIC CELL WITH DIFFERENT EXTERNAL CIRCUIT RESISTANCES

The time lag exhibited by a selenium photovoltaic cell between sudden exposure to light and the development of the corresponding output E.M.F. is comparable with the time lags encountered in selenium photoconductivity cells. If a photovoltaic cell is exposed to interrupted light of constant intensity but variable frequency of interruption, the A.C. response of the cell drops to half its low-frequency value if the interruption frequency is somewhere in the range from 1 kc. to 10 kc.

The physics of the photovoltaic effect is somewhat more complicated than that underlying photoconductivity. The seat of the E.M.F. is the contact between the selenium and the overlying gold film electrode. The photoelectric property of this contact comes about because of the so-called contact potential difference between the selenium and the gold. This contact potential difference appears across a thin layer in the selenium immediately adjacent to the gold contact and within a distance of 10^{-5} cm. of the actual gold-selenium boundary. Inside this layer, called the barrier layer, the physical properties of the selenium are different from those in the rest of the selenium film. When the contact is illuminated, charged electronic particles are liberated by the absorption of light in the barrier layer. These particles, acting like ions in the electrolyte of an ordinary battery, are pulled by the electric field of the barrier contact potential difference in such a direction as to cause a negative charging up of the gold electrode. From this electrode, then, electrons can flow around the external circuit, back into the support plate metal, and thence into the other side of the selenium layer, as long as the illumination is continued.

The photoelectric properties of a selenium photovoltaic cell can be described in terms of the electrical equivalent circuit shown in Fig. 6. The gold-selenium contact barrier has three parallel properties: a barrier capacitance C; a barrier resistance R; and an effective barrier battery B having the property of generating not a voltage but rather a current proportional to the amount of light falling on the cell and numerically equal to the product of a constant of proportionality K by the light intensity

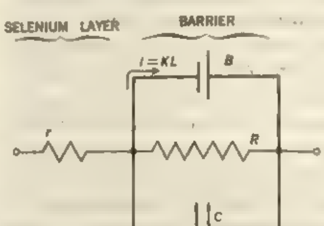


FIG. 6.—ELECTRICAL EQUIVALENT CIRCUIT EXHIBITING THE SAME TERMINAL BEHAVIOUR SHOWN BY AN ILLUMINATED PHOTOVOLTAIC CELL

L. For modern commercial selenium photovoltaic cells, *K* has values in the neighbourhood of half a milliampere per lumen. In series with the barrier is the resistance *r* of the selenium layer outside the barrier. When the cell is illuminated, the current generated by the equivalent barrier battery divides, part going around the external circuit and the rest leaking backward internally across the barrier resistance *R*. Unfortunately, this barrier resistance is not constant, but decreases for higher values of the E.M.F. appearing at the cell terminals. Thus, for a constant load resistance in the external circuit, if the illumination is made brighter *R* becomes smaller and a larger fraction of the equivalent battery current is lost by internal leakage across *R*. This variable loss explains the nonlinearity of the curves of Fig. 5.

The barrier capacitance *C* has virtually no effect on the performance of the cell for exposure to steady light; but for interrupted light or fluctuating light this capacitance has a bypassing effect which reduces the A.C. output of the photocell. The reduction is more severe the higher the frequency of interruption or fluctuation of the light. The magnitude of this capacitance is about one-half microfarad per square inch.

Incidental Properties.—Selenium photoconductive and photovoltaic cells have approximately the same colour responses. Selenium cells are responsive to all wave lengths in the visible spectrum and also to a short range of wave lengths in the near ultraviolet. The greatest sensitivity appears in the orange and orange-red spectral region.

The output of a selenium cell is dependent on temperature, the dependence being to a considerable extent determined by the nature of the associated circuit. In a selenium photoconductivity cell the magnitude of the dark conductance increases about 2% per degree rise in temperature, while the conductance change upon illumination decreases with increasing temperature by about 1% per degree. The chief temperature dependence of the selenium photovoltaic cell is exhibited by the barrier resistance *R*, which decreases with increasing temperature (6% or 7% per degree).

A selenium photovoltaic cell will deteriorate in sensitivity if it is exposed for long periods to a highly humid atmosphere. The modern practice of coating the cell surfaces with a protective lacquer film slows down this deterioration but does not prevent it altogether. Some manufacturers have experimented with complete hermetic sealing of the cell elements.

Uses of Selenium Cells.—Probably the most extensive use of selenium cells is in the construction of portable exposure meters used by photographers. In this application a photovoltaic cell is used in a self-contained package with a moving coil meter and a dial or table for converting the meter reading to the correct exposure. Selenium cells have a wide range of industrial applications for distinguishing between light on and light off. For example, a selenium cell may be placed near a light source so that the passage of an article along a conveyer belt interrupts the light beam and causes registry on a counter. A safety interlock for stamping and pressing machines can be arranged with a light source and photocell so placed that as long as a workman's hands are between the dies the light beam is broken and the circuit controlling the operation of the machine is inhibited.

Another on-off application of selenium cells is found in burglar alarm systems where the interruption of a beam of light across a doorway or window trips an alarm. In department stores and public buildings a closely allied application is coming into increasing use, in which a door is automatically opened when a light beam is broken by an approaching person. Similar systems are used for counting visitors at exhibitions or counting automobiles to obtain traffic data.

There are many industrial processes which can be monitored or controlled with the aid of selenium photocells in suitable circuits. Depth of colour is often an indication of the correct proportions of a mixture, of the thoroughness of mixing or of the stage of a chemical reaction. In these cases photocells have proved more vigilant and accurate as colour perceptors than the human eye, and their employment has resulted in better product control in the chemical, pigment and textile industries.

BIBLIOGRAPHY.—Willoughby Smith, *Journal of the Society of Telegraph Engineers*, 2:31–33 (1873); W. G. Adams and R. E. Day, *Transactions of the Royal Society of London*, 167:313–349 (1877); George P. Barnard, *The Selenium Cell* (1930); T. J. Fielding, *Photo-Electric and Selenium Cells* (1935); Trevor S. Moss, *Photoconductivity in the Elements* (1952). (J. N. SE.)

SELEUCIA (SELEUKEIA), the name of several Hellenistic cities named or renamed after the founder of the Seleucid dynasty, Seleucus I Nicator (ruled 312–280 B.C.).

SELEUCIA ON THE TIGRIS lay on the right bank of the Tigris River about 20 mi. SE of Baghdad, at the mouth of the "royal canal" connecting the Tigris and the Euphrates. The city was founded by Seleucus I as his eastern capital, and its building, mainly of materials brought from Babylon, marks the definite end of Babylon. Seleucia is closely associated with Hellenistic culture in Mesopotamia. It was peopled with Macedonians and Greeks, but its population also included Jews and Syrians; it was definitely a cosmopolitan city. Pliny gives the population as 600,000. During the Parthian domination the city continued to be the foremost city of the east in position and trade. Preserving its Greek sympathies, it was definitely opposed to and at times in open rebellion against the Parthian kings, who favoured the neighbouring Ctesiphon (*q.v.*) and founded Vologesias or Vologesocerta as a rival canal port. Seleucia eventually was burned by Avidius Cassius in A.D. 165, at which time it is said to have had at least 300,000 inhabitants. The destruction of the city marks the end of Hellenism in Mesopotamia. Septimus Severus, in his Parthian campaign of 198, found the site completely abandoned, and in 363 Julian's soldiers hunted wild game in marshes in its vicinity. Nothing remains above ground; excavation of the site (then called Tel Umar) during 1927–32 was sponsored by the University of Michigan and the Toledo and Cleveland museums of art, with interesting but unspectacular results.

SELEUCIA PIERIA, in Syria, a port and frontier fortress on the Cilician border (near modern Samandag, Turkey), lay 4 mi. N of the mouth of the Orontes River and was the port of Antioch (*q.v.*). With Antioch, Apamea, and Laodicea it formed the Syrian tetrapolis. The town occupies the rocky slopes of Musa Dagi, and the great Roman flood diversion, a deep cutting through the rock, still survives. The harbour is complete but high and dry. The city was of considerable military importance during the wars between the Ptolemies and the Seleucids and was recognized as an independent city later by the Romans.

SELEUCIA TRACHEOTIS or TRACHEA (modern Silifke, Icel il, Turk.), a city in Cilicia, lay on the Calycadnus River (modern Goksu Nehri), a few miles from the mouth, doubtless as a protection against attacks from the sea. There are ruins of a castle on the acropolis, and the city fortifications, a large rock-cut tank, and an extensive necropolis are all well preserved. The city was at one time a port serving Isauria and was entitled to strike its own coinage. It was built near an old site by Seleucus I. During the Third Crusade Frederick Barbarossa was drowned in crossing the river at this point (A.D. 1190). The city was captured by the Turks in the 13th century. It is now the centre of a fertile plain, which has been irrigated by modern works in the local river.

Several other towns also bore the name of Seleucia, among them Seleucia on the upper Euphrates, near Samosata; in the Persian province of Margiana, which was renamed by Antiochus I, having been previously called Alexandria after its original founder, Alexander the Great (*see* ALEXANDRIA: *Alexandria in Margiana*; and MERV); in Pisidia; in Pamphylia; Seleucia Sidera ("the iron") in Phrygia; on the Belus in Syria; and in Caria (Tralles; *q.v.*). (WM. C. B.; X.)

SELEUCID DYNASTY, a line of kings who reigned in nearer Asia from 312 to 64 B.C. as one of the successor dynasties of Alexander the Great.

The Seleucid Empire reached its greatest extent under the first of the kings, Seleucus (*q.v.*) I Nicator. At his death (280) it embraced the Iranian provinces as far to the east as the Hindu Kush Range and the valley of the Oxus River, Media, Babylonia, Mesopotamia, Cappadocia, the whole of Asia Minor (apart from Armenia, Pontus, Paphlagonia and Bithynia), Syria, and, in Europe, parts of Hellespontine Thrace. During the 3rd century B.C. (after 280) much ground was lost in Asia Minor either through

Greek cities on the Aegean coast throwing off their allegiance or through the emergence of new independent states (such as Pergamum and Cappadocia), or through the cession of territory to the Celtic invaders (*see* GALATIA) and the Bithynian and Pontic kings (*see* BITHYNIA; PONTUS). The hill country of the Taurus Range in southern Asia Minor was never properly subjugated, and the Seleucid possessions to the north and west of Cilicia were reduced, before the end of the 3rd century, to large parts of Caria, Lydia, Aeolis, and the Troad (all on the Aegean seaboard), connected by a narrow corridor running through southern Phrygia and Lycaonia to Cilicia and Syria.

In the northeastern area of the empire, Parthia, Bactria, and Sogdiana became independent after the middle of the 3rd century. The submission of these districts to Antiochus III later (209–206) was of only passing effect. Coele-Syria (Lebanon) and Palestine had been fought over by the Seleucids and the Ptolemaic kings of Egypt ever since the accession of Antiochus I (280); not until the reign of Antiochus III did they pass finally into Seleucid possession (*c.* 200). This acquisition and the short-lived subjugation of Armenia and the former eastern provinces of the empire counterbalanced the loss of all Asia Minor north of Mt. Taurus by the Treaty of Apamea (188), which resulted from the defeat of Antiochus III at the hands of Rome and Pergamum.

After the death of Antiochus IV (163) the small state of Commagene in northern Syria emerged as an independent kingdom, as did Judaea in Palestine. In the east all the provinces to the east of Media, Susiana, and Persis had been lost by the same date, either to the Parthians or to the kings of Bactria. In 145 Ptolemy VI of Egypt reoccupied Coele-Syria and Palestine. In 141 or 140 the Parthians annexed Babylonia and the southern part of Mesopotamia. By then everything to the east of the Euphrates, apart from a small foothold in northern Mesopotamia, had been lost; and the utter failure of the expeditions led by Demetrius II in 140–139 and Antiochus VII in 130–129 to recover the lost provinces meant that the kingdom was reduced to Syria and eastern Cilicia. Even this limited area was held only precariously in the final period of the dynasty (126–64 B.C.).

To the Romans the Seleucids were known as the kings of Syria, and many modern writers have followed Roman practice in so describing them. From the above account, however, it will be seen that this description is appropriate solely to the kings who ruled after the Parthian occupation of Babylon. At the same time it should be noted that Syria had generally been the royal residence ever since Seleucus I had founded Antioch on the Orontes as his capital. Seleucia on the Tigris, the residence of the governor of the eastern provinces (often the crown prince), held the position of a second capital. Significantly, the principal satrapy, or administrative division, of Syria was named Seleucis. Besides Antioch on the Orontes the Seleucids founded many other important new cities in Syria (*see* SYRIA) and made strenuous efforts to give it a large Greek population. Through the presence there of the court and by reason of its strategic position, Syria undoubtedly constituted the centre of the Seleucid Empire and surpassed in political importance all its other regions.

The Seleucids adopted all the outward form of Macedonian kingship (*see* HELLENISTIC AGE). Over their Asiatic subjects they claimed absolute power by the right of conquest. To the Greeks and Macedonians within their empire they were leaders rather than masters; but the traditional right of the Macedonian Army to confirm successions to the throne and declarations of war was ignored in practice, and the Greek cities, though many of them enjoyed the status of free allies of the king and all retained the forms of civic autonomy, were expected to carry out the king's wishes. The powerlessness of the later Seleucids led to the Syrian cities' asserting their independence in a manner that would have been unthinkable before the death of Antiochus IV (163). The absolutism practised by the kings is reflected in the establishment, not later than the reign of Antiochus II Theos (*c.* 261–*c.* 247), of a state cult of the living ruler.

The kings made themselves personally responsible for all matters of state. To assist them there was a small number of ministers in charge of particular departments of government. One of these, *ho epi ton pragmaton* ("the man in charge of things"), may have acted as chief minister under the king; possibly, however, the title merely denotes a regent acting in the king's absence. Indeed, virtually nothing is known of the central administration of the kingdom. In provincial government, wisely, no attempt was made at enforcing too strict a control over the numerous provinces, some of which lay at a great distance from the centre of the empire.

The large regional unit of the satrapy, inherited from the Persian Empire, was retained in the western half of the kingdom. Thus in Asia Minor the following Seleucid satrapies are known: Hellespontine Phrygia, Phrygia, Lydia, Caria, and Cilicia. East of the Euphrates the satrapies were divided into smaller areas—eparchies and hyparchies. There, the administration was in three stages: from the satrap's headquarters to the seat of the officer governing the hyparchy, the eparchy forming the link between satrapy and hyparchy. In some cases satrapies were governed by generals (strategi), in others the satrap himself exercised military power (though H. Bengtson argues that every satrapy contained both a satrap and a general). Matters relating to finance within the satrapy were assigned to a separate official.

Little is known of the forms of taxation in the Seleucid Empire. Perhaps from motives of expediency, the kings did not insist upon a uniform system. Though certainly less exacting than the Ptolemaic kings of Egypt in the demands they made on their subjects, the Seleucids must have drawn large revenues. The expenses involved in maintaining the court and army were heavy, but the surplus of revenue and the general prosperity of the kingdom were sufficient to make the kings and the leading men in the state very rich. All the categories within the empire—subject kings, dynasts, peoples, and cities—were liable to taxation in money or kind, but only in the case of the royal domains, where the land was cultivated by the royal serfs ("the king's people"), were taxes collected directly by royal officials; in all other cases collection was indirect and by stages. In accordance with their preference for a decentralized mode of government, and in order to encourage the growth of city life, the Seleucids commonly allotted large tracts of royal land to Greek cities and compelled feudal landowners and temple states to do the same. This process led to the gradual emancipation of the serfs on the land affected.

The Seleucid Army had as its nucleus the infantry phalanx, a traditional feature of all the Macedonian states. Its members were recruited from the military colonists who dwelt at various places throughout the empire, a hereditary class of Macedonian and Greek ancestry. To supplement the phalanx, the Seleucids relied upon mercenaries (Greeks, Galatians, and Jews, among others) rather than upon native levies, although contingents of the latter were also employed upon occasion. The army was large and frequently on active service; its upkeep was probably the principal charge upon the kingdom's revenue. By contrast, the fleet was of only slight importance; squadrons were maintained in the Mediterranean and the Persian Gulf.

While certain important regions on the confines of the empire broke away at various times and attained complete independence under their own rulers (*see* ATTALID DYNASTY; BACTRIA; CAPPADOCIA; COMMAGENE; PARTHIA), there were also many semiautonomous units in all parts of the realm, including those Greek cities that enjoyed full self-government. Temple states, all of very ancient origin, were particularly numerous in Asia Minor; examples occur too in Syria and elsewhere. The early Seleucids, as protectors of all varieties of religion, tolerated their existence while endeavouring to diminish the power of the priest-rulers by transferring temple land to neighbouring Greek cities or to new military colonies. The later Seleucids, however, were not strong enough to assert such control and the priests grew in authority once more, as they did in Egypt under the later Ptolemies.

Of all the Hellenistic dynasties the Seleucids were by far the most active in the foundation of new Greek settlements, whether military colonies or cities. In Asia Minor their foundations or refoundations included: Stratonicia, Thyateira, Blaundus, Seleucia (Tralles), Antioch (Alabanda) (all in Lydia and Caria); Apamea (Celaenae), Seleucia the Iron, Apollonia, Laodicea Combusta,

Seleucia Trachea, Antioch (Tarsus), Antioch (Adana), Seleucia (Mopsuestia) (all situated, in order from west to east, from Phrygia to Cilicia). (For foundations in other regions of the kingdom *see* PERSIAN HISTORY: *Hellenistic Period;* MESOPOTAMIA: *Seleucids, Parthians and Romans;* SYRIA: *History: Hellenistic and Roman Syria.*)

The purpose of the Seleucid kings in creating such an extensive network of Greek settlement was probably the limited aim of buttressing their authority over the masses of their native subjects by establishing among them centres of power upon which they could safely rely. They could not have expected that the centres of Hellenism so created would bring about the hellenization of the native peoples. In fact very few Asiatics became hellenized in any more than a superficial way; the cultural gap was too wide to allow easy assimilation to take place. The racial cleavage was recognized in the provision of separate courts of law for Greeks and Asiatics. Moreover, for all the unceasing efforts of the kings to maintain and increase Greek immigration into their kingdom, the flow of new settlers was never enough to reduce noticeably the great preponderance of the Asiatic peoples.

Of the other achievements of the Seleucids two may be noted. (1) their promotion of commerce; not only within their own kingdom but also internationally, by protecting and encouraging the caravan trade between central Asia and the Mediterranean and the trade by sea with India by way of the Persian Gulf; (2) their introduction of a calendar which supplied the universal need of a readily comprehensible system of dating. This calendar (the Seleucid era) made its starting point the reoccupation of Babylon by Seleucus I in 312 B.C. It was adopted by the Parthian kings after their conquest of the eastern Seleucid provinces and remained in use in many other areas long after the disappearance of the Seleucid kingdom.

The history of the later Seleucids is a record of rapid decline. The children of Demetrius II Nicator and Antiochus VII Sidetes formed rival branches of the royal house between which a more or less continuous struggle was waged for control of an increasingly smaller and weaker territory. The formal end of the dynasty came with the Roman annexation of Syria in 64 B.C., but before then the last real traces of Seleucid power had been destroyed by Tigranes, the king of Armenia, in his invasion and occupation of Syria (83–69).

See also references under "Seleucid Dynasty" in the Index; and ALEXANDER BALAS; ANTIOCHUS; DEMETRIUS; SELEUCUS.

BIBLIOGRAPHY.—E. R. Bevan, *The House of Seleucus,* 2 vol. (1902); A. Bouché-Leclercq, *Histoire des Séleucides,* 2 vol. (1913–14); C. B. Welles, *Royal Correspondence in the Hellenistic Period* (1934); E. Bikerman, *Institutions des Séleucides* (1938); M. Rostovtsev, *Social and Economic History of the Hellenistic World,* 3 vol. (1941); H. Bengtson, *Die Strategie in der hellenistischen Zeit,* vol. ii (1944); W. W. Tarn, *The Greeks in Bactria and India,* 2nd ed. (1951); W. W. Tarn and G. T. Griffith, *Hellenistic Civilisation,* 3rd ed. (1952). (R. H. Si.)

SELEUCUS, the name of six kings of the Seleucid dynasty (*q.v.*). The more important of these are listed below.

SELEUCUS I (surnamed for later generations NICATOR) (*c.* 358–280 B.C.), the founder of the dynasty, was the son of Antiochus, one of the generals of Philip II of Macedonia. He took part in the conquest by Alexander the Great of the Persian Empire, distinguishing himself in the Indian campaign. He did not, however, hold a leading position in the Macedonian army during Alexander's lifetime. In the arrangements made at Babylon upon Alexander's death (323) he was appointed senior lieutenant to the regent Perdiccas. During the attempted invasion of Egypt in 321, Seleucus was one of the officers who led the mutiny in which Perdiccas was murdered. In the settlement of Triparadisus in Syria (321), he was entrusted with the government of the Babylonian satrapy as a reward for his part in Perdiccas' overthrow, and in the subsequent war between Eumenes (*q.v.*) and Antigonus I (320–316) he supported the latter. Nevertheless, Antigonus regarded him with distrust, and after his final defeat of Eumenes (316), he ordered Seleucus to submit the accounts of his satrapy. Seleucus thereupon fled from Babylon and took refuge with Ptolemy in Egypt. In the war which followed between Antigonus and the other Macedonian chiefs, Seleucus played an active part as commander of naval squadrons placed under his charge by Ptolemy. The victory won by Ptolemy's army at Gaza (312) over Demetrius I (later Demetrius I Poliorcetes), the son of Antigonus, opened the way for Seleucus to return to Babylonia.

Though he brought with him only a small force of soldiers, Seleucus enjoyed the support of the Babylonian population whose favour he had won earlier in the time of his governorship. Accordingly, he successfully resisted intermittent attempts to crush him and gradually wrested the neighbouring provinces of Persis, Susiana, and Media from the nominees of Antigonus. Eventually he brought under his control the whole eastern part of Alexander's empire as far as the Jaxartes (modern Syr-Darya) and Indus rivers. But about 304 he felt obliged to make peace with the Indian king Chandragupta to whom he ceded the Indian provinces, receiving in exchange 500 war elephants. In 306, after the extinction of the old royal line of Macedonia, he had, like the other four principal Macedonian chiefs, assumed the style of king; the Seleucid era, however, which was officially regarded as the beginning of the Seleucid Empire, was dated from the time of his return to Babylon in 312. Invasions of Babylonia by Antigonus, though always ineffectual, had persisted until 302. In the campaign of Ipsus (302–301) Seleucus combined with Lysimachus to bring about the downfall of Antigonus (*see* IPSUS, BATTLE OF). A new partition of the Macedonian Empire was then made. Seleucus added to his kingdom Syria, where he founded his new capital, Antioch (*q.v.*), on the Orontes. For the eastern half of his realm Seleucia (*q.v.*) on the Tigris, another of his foundations, served as capital; there, about 293, he installed his son Antiochus as viceroy.

In the wars which followed the Battle of Ipsus, Seleucus sided at one time with Demetrius I (*q.v.*) Poliorcetes, whose daughter Stratonice he married (299 or 298); but later he joined with Lysimachus to defeat Demetrius, who became his captive in Syria (285). In the last war of his reign Seleucus defeated and killed Lysimachus at Corupedium in Lydia (281). Having occupied Asia Minor, he crossed into Europe with the intention of annexing Thrace and Macedonia. This object would certainly have been attained, for there was no one to oppose his advance, had he not been murdered by Ptolemy Ceraunus (who had earlier deserted Lysimachus in order to join him; *see* PTOLEMIES) near Lysimachia in the Thracian Chersonese (280).

In spite of his final conquests, as a result of which he came near to reuniting under himself the whole of Alexander's empire apart from Egypt, Seleucus must be reckoned, like Ptolemy I, to whom he owed a great personal obligation, as an upholder of the principle of separatism within the empire and a determined opponent of imperial unity. A fine general, he was above all a great organizer and administrator. He followed Alexander in promoting Greek and Macedonian colonization in Asia, and while he may have considered Alexander's policy of partnership between Macedonians and Persians impracticable or at least premature, he did not wholly reject it, for he was the only one of the Successors who did not repudiate after Alexander's death the Persian wife (Apama) he had taken in 324. His historical importance lies in his creation of the largest of the Hellenistic states.

SELEUCUS II CALLINICUS (*c.* 265–226 B.C.), son of Antiochus II Theos, reigned from 246. He fought against Ptolemy III of Egypt in the Third Syrian War (246–241), and in Asia Minor against his own younger brother Antiochus (*q.v.*) Hierax in a civil war which gravely weakened the Seleucid kingdom. Defeated by Hierax and the Galatians at Ancyra in Phrygia (*c.* 235), he was compelled to give up the country north of Mt. Taurus; but he struggled with some success to maintain his authority in the eastern provinces where he won a victory over the Parthians (*c.* 228).

SELEUCUS III SOTER (*c.* 245–223 B.C.), son of the above, took up the task of reconquering Asia Minor from Attalus (*q.v.*) I of Pergamum but fell by a conspiracy in his own camp.

SELEUCUS IV PHILOPATOR (*c.* 218–175 B.C.), second son of Antiochus III, succeeded his father in 187. His reign was uneventful. He was compelled by financial necessities, created in part by the heavy war indemnity exacted by Rome under the Peace of Apamea (188), to pursue an unambitious policy. Conflict with other powers was carefully avoided, everything being directed to im-

ternal consolidation and recovery. In 175 he was murdered by his minister Heliodorus, who usurped the throne. (R. H. St.)

SELF-DETERMINATION, a term widely used in contemporary international relations with a variety of meanings. Historically it meant the right of people in a state to choose their own government. In this sense it is opposed to an absolute monarchy based on the divine right of kings and is instead in accord with Rousseau's doctrine that sovereignty resides in the people and not in the ruler. The French Revolution gave currency to this idea. It was made explicit in the "Déclaration du droit des gens" that Abbé Grégoire submitted to the Convention on April 23, 1795. It declared that peoples are independent and sovereign and that every people has the right to organize and change its own government. The American Revolution provided another contribution to the doctrine of self-determination. It became the signal for similar anticolonial and independence movements in Latin America, though without the connotation of self-government. The congress of Vienna (1815) attempted to restore the old European system based on a balance of power and the principle of legitimacy. The Holy alliance of Sept. 26, 1815, was particularly concerned with the latter and did not shrink from intervention on its behalf. The Monroe Doctrine of Dec. 2, 1823, was in part at least designed to forestall any attempt at restoration of dominion over the former European possessions in the western hemisphere.

National self-determination as a political principle has evolved in the last 150 years as a by-product of the doctrine of nationalism in combination with other philosophical and political currents. Immanuel Kant's concept of the autonomy of the individual and of freedom as a condition of autonomy furnished an essential element. Johann Fichte's emphasis on the state as a condition of man's freedom and Johann Herder's idea that mankind was divided by God into different nationalities, each of which had a particular mission to fulfill, furnished the other elements. Schleiermacher, like Herder, stressed the distinctiveness of each nation based on language, character, history and culture. Nations should constitute their own sovereign state to preserve their distinctiveness and to make their preordained contribution to mankind. Mazzini argued that multinational states are artificial whereas national states are products of nature.

Nature, however, did not of and by itself produce national states. The persistent efforts of politicians acting from other motivations but drawing upon the ideological bases provided by philosophers, and the general drift of European politics resulted in a drastic redrawing of the map of Europe produced by the congress of Vienna. The Ottoman empire, a multinational and multireligious state, suffered first the impact of doctrine and politics: Greece gained independence in 1829; through the century other parts of the empire split off: Rumania, Bulgaria, Serbia. Italy was unified under the banner of nationalism and so was Germany; in the former case Napoleon III and in the latter the Prussian Chancellor Bismarck were active forces inspired by less than idealistic motives. The democratic-liberal elements looked with favour upon the spread of nationalism but, as later events showed, the new states formed on a national basis were not necessarily committed to liberal or democratic ideals or to the pursuit of peace.

World War I and the League.—In World War I the Allies, largely under the influence of Woodrow Wilson, accepted self-determination as a peace aim; the Bolshevik revolution in 1917 removed a multinational state, Russia, from their ranks and made it easier to espouse the claim of oppressed nationalities. The Bolsheviks themselves through Trotski's note to the Allied embassies in Petrograd, Nov. 21, 1917, advocated a democratic peace on the basis of "the self-determination of nations." Pope Benedict XV in his message to the heads of the belligerent governments of Aug. 1, 1917, favoured a territorial settlement taking account of the desires of the peoples. President Wilson in his address to the U.S. Senate, Jan. 22, 1917, proposed among the essential terms of peace that all nations accept the Monroe Doctrine and agree "that no nation should extend its polity over any other nation or people, but that every people should be left free to determine its own polity." In his Fourteen Points, Jan. 8, 1918, Wilson listed among the peace objectives the restoration of independence to Belgium, Rumania, Serbia and Montenegro, and of Alsace-Lorraine to France, "the freest opportunity of autonomous development" for the peoples of Austria-Hungary, and a similar opportunity for the nationalities of the Ottoman empire, and the resurrected Poland should "include the territories inhabited by indisputably Polish populations." By the time of the peace conference events had outdistanced these goals and independent states emerged in the territory of the Austro-Hungarian empire and the Ottoman empire. The principle of self-determination had to be adjusted to political exigencies and promises of territorial compensation made during the war to Italy. The principle was further compromised by strategic and economic considerations and the virtual impossibility of drawing boundaries along linguistic or ethnic lines. Some of the Ottoman provinces whose existence as independent nations was provisionally recognized in Art. 22 (4) of the covenant of the League of Nations but which were not yet ready to stand alone were placed under the mandates system (*see* MANDATE). Contrary to the assumptions of the liberal democrats, the new order based on the principle of national self-determination proved no basis for a stable peace. The principle itself became in the hands of Adolf Hitler the chief instrument for the destruction of peace.

World War II and the UN.—In World War II liberation of the territories enslaved and occupied by Germany, Italy and Japan became one of the chief peace goals of the United Nations. Self-government was envisaged for all, including those which came within the sphere of influence of the U.S.S.R., particularly Poland. Thus the Atlantic Charter of Aug. 14, 1941, declared the desire of the United Kingdom and the United States to "respect the right of all peoples to choose the form of government under which they will live." This was reaffirmed in the declaration by the United Nations of Jan. 1, 1942, and in numerous other official statements. Be it noted that self-determination here is used in the sense of self-government but because of the participation of the Soviet Union it could not necessarily mean democratic self-government nor could it mean national self-determination for all. The Baltic republics incorporated in the U.S.S.R. in 1940 were not restored to independence, though the U.S. did not recognize their annexation. In spite of the 1945 Yalta agreement (*see* WORLD WAR II CONFERENCES, ALLIED) Poland was not granted a genuine right of self-determination. Many Germans from Czechoslovakia and from territories under Soviet or Polish administration east of the Oder-Neisse line were moved to Germany.

Self-determination is affirmed in the charter of the United Nations organization. One of its major purposes is "to develop friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples" as one of the measures to strengthen universal peace (Art. 1, par. 2). Another is to promote co-operation in the economic, social, cultural and humanitarian fields "with a view to the creation of conditions of stability and well-being which are necessary for peaceful and friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples" (Art. 55, par. 1). In pledging themselves "to take joint and separate action in co-operation with the Organization for the achievement" (Art. 56) of those purposes the members appear to assume an obligation and thereby to elevate the principle of self-determination of peoples—neither of which is defined—from the plane of politics to the level of law. More specific obligations to develop self-government and free political institutions in nonself-governing territories and to promote self-government or independence in trust territories are assumed by the members that have responsibility for their administration (Art. 73, b, and 76, b). Moreover, the general assembly of the United Nations urged the members to speed up the process and specifically recommended that the wishes of the peoples be "ascertained through plebiscites or other recognized means" (Resolution of Dec. 16, 1952). It also recommended that "the right of peoples and nations to self-determination" be included in the projected covenant on human rights (Resolution of Feb. 5, 1952). The traditional political concept of self-determination was also declared by the assembly to apply to the economic sphere.

Responding partly to the continuous pressure of the United Nations and partly to other considerations of a political and economic

nature, western governments after the end of World War II embarked on a vast process of decolonization by divesting themselves of sovereignty over their dependencies in Asia and Africa and transferring it to the newly established states. By contrast the Soviet Union, clearly subordinating self-determination to the objective of promoting communism or socialism, in practice denied to the peoples under its sway either the right freely to elect their government or to form a state of their choice. The suppression of the Hungarian revolution in 1956 by armed intervention led to a condemnation by the United Nations (Resolution of Nov. 9, 1956) but not to a liberation of the Hungarian people. On the other hand, the Soviet Union adopted the policy of promoting self-determination and "national liberation" movements in other states provided they are deemed to offer sooner or later an opportunity for penetration by or are themselves bearers of the Communist ideology.

See A. Cobban, *National Self-Determination* (1945); S. Shaheen, *The Communist Theory of National Self-Determination* (1956).
(L. Gs.)

SELFRIDGE, HARRY GORDON (1857?–1947), founder of a great London department store, was born at Ripon, Wis. The son of a small storekeeper, he became an errand boy at 12. At 17 he joined Field, Leiter and Co., later Marshall Field & Co., Chicago, where he worked with great success for 25 years, becoming a junior partner. After using his fortune to buy and resell profitably another store, Selfridge went to England in 1906. He began building a large store in Oxford Street, London, but his partner withdrew from the venture. He then obtained support from a wealthy tea broker and in 1908 Selfridge & Company, Ltd., was registered (with £900,000 capital) to complete the project. Covering 42,000 sq.ft. (later twice as much), the store was opened in 1909. Imaginative advertising, ingenious publicity, and novel interior arrangements made Selfridge's a household word at once. In 1937 he became a British citizen. By 1939 he had lost grip on his great enterprise and because of his personal extravagance the banks caused him to be deposed. He died at Putney, London, on May 8, 1947.

See A. H. Williams, *No Name on the Door* (1956); R. Pound, *Selfridge* (1960).

SELIGMAN, CHARLES GABRIEL (1873–1940), a pioneer in British anthropology, was born in London on Dec. 24, 1873. Educated at St. Paul's School, London, he trained in medicine at St. Thomas's Hospital and specialized for a time in pathological research. Attracted, however, by anthropology, he joined a field expedition to the Torres Strait in 1898 and another to southern New Guinea in 1904. With his wife, Brenda Z. Seligman, he studied the Veddas of Ceylon in 1907–08 and then made three ethnographic survey tours in the Sudan. This work resulted in *The Melanesians of British New Guinea* (1910), *The Races of Africa* (1930), and (jointly with his wife) *The Veddas* (1911) and *Pagan Tribes of the Nilotic Sudan* (1932). In 1910 Seligman was appointed university lecturer and in 1913 part-time professor in ethnology at the London School of Economics and Political Science. Retiring in 1934, he was given the title of emeritus professor. He became visiting professor at Yale University in 1938, and died at Oxford on Sept. 19, 1940. Seligman's anthropological interests and scholarship were very wide, in physical anthropology, archaeology, cultural anthropology, and Oriental art. He was one of the first anthropologists to study the relationship between psychoanalysis and anthropology. He received many distinctions, including fellowship of the Royal Society in 1919.

See G. H. Brown, *Lives of the Fellows of the Royal College of Physicians of London, 1826–1925* (1955). (R. F.)

SELIM I (*c.* 1470–1520), sultan of the Ottoman Empire 1512–20, known as Yavuz ("the Grim"), was a son of Sultan Bayazid II. In 1511, in order to strengthen his position in the struggle between himself and his brothers for succession to the throne, he sailed from his province of Trebizond (Trabzon) to Kaffa (Feodosiya) in the Crimea, won the aid of the khan of the Krim Tatars, and then crossed the Danube into the Balkans. Bayazid, worried over a great Shi'i revolt in Asia Minor, assigned to Selim a province in the Balkans, but Selim, fearing that the partisans of his brother Ahmed might raise that prince to the throne if their efforts to crush the Shi'i rebellion were successful, turned against his father, was defeated near Corlu in 1511, and fled for refuge to the Crimea. His brother Ahmed, finding, however, that the janizaries (*q.v.*) would not accept him as future sultan, seized much of Asia Minor. The danger that he might reach an entente with the shah of Persia led to the recall of Selim from the Crimea and to the abdication of Bayazid II in favour of Selim (April 1512). Selim overcame and killed Ahmed in 1513. He then moved against the Safavid regime in Persia, which was the source of a Shi'i religious propaganda so successful as to threaten to undermine Ottoman control over much of Asia Minor. Selim defeated Shah Ismail at Caldiran in 1514 and then, in 1515–16, subjugated the border lands between the Ottoman Empire and the Mameluke sultanate of Syria and Egypt, along the Taurus Mountains and in Kurdistan. The Mamelukes, thereupon, marched northward in 1516, only to be routed by the Ottomans at Marj Dabiq in Syria. A further campaign led to the Ottoman conquest of Egypt in 1516–17. Selim had thus, in five years of relentless warfare, solved the dangerous problems which confronted the Ottomans in the East at the close of the reign of Bayazid II. Selim I died near Corlu in September 1520, while apparently about to undertake a campaign against the Knights of St. John at Rhodes. *See* also TURKEY: *History*.

See *The New Cambridge Modern History*, vol. i, ch. xiv (1957).
(V. J. P.)

SELIM II (1524–1574), sultan of the Ottoman Empire 1566–74, was a son of Sultan Suleiman the Magnificent. During 1559–61 he was involved in a bitter conflict with his brother Bayazid over the succession to the throne. The conflict ended with the execution of Bayazid. The major event of Selim II's reign was the conquest of Cyprus, then under Venetian rule. Nicosia fell to the Ottomans in 1570 and Famagusta in 1571. Venice, meanwhile, formed an alliance with the pope and with Philip II of Spain. A joint Christian armada destroyed the Ottoman fleet at Lepanto in 1571, but the divided aims of the Christian allies prevented the exploitation of this famous victory. Venice made peace in 1573, surrendering Cyprus to the Ottomans. Don John of Austria seized Tunis in this same year, but the Ottomans, having built a new fleet since Lepanto, captured the town in 1574. Selim II died at Istanbul in December 1574. *See* also TURKEY: *History*. (V. J. P.)

SELIM III (1761–1808), sultan of the Ottoman Empire 1789–1807, was born in Istanbul on Dec. 24, 1761, the son of Sultan Mustafa III who was succeeded by his brother Abdul-Hamid I. After his father's death Selim lived in the cage-prison (*kafe*). He was well educated and, free from traditional prejudices, he maintained a secret friendship with the French ambassador and sent a representative to the French court to investigate French institutions. Thus he became familiar with Western civilization and ways of thought. When, on April 7, 1789, Selim succeeded his uncle as sultan, Turkey was at war with Russia and Austria, but he was determined to modernize his empire. After the peace of Jassy (1792) he outlined a project of reforms, the *nizam-i-jedid* ("new organization"). This was based on a scheme drawn up by his ministers, at his order, particularly on a long report on Austrian institutions prepared in Vienna by his friend Ratip Efendi. To put the plan into operation Selim created a special treasury and secured the services of European officers and experts. With their help he formed a new army modeled on Western lines; he reorganized the school of navigation; opened a new engineering school, with French a compulsory language; and introduced European military textbooks. Selim also established permanent embassies in European capitals and instructed embassy staffs to learn the language of the country to which they were accredited and send regular reports on its political, social, and cultural life. These reforms, coupled with unsuccessful wars against France and Russia, stirred up the janizaries (*q.v.*), the ulamas and the provincial governors against Selim. At one time a janizary cried out that he would rather be a Moskof (a Russian) than a *nizam* (a new-type soldier). In 1807, with the army on the Danube facing the Russians, the moment seemed favourable for

an uprising. Encouraged by some reactionary ministers, the auxiliary troops (Yamaks) on the Bosporus, who were charged with the defense of the Black Sea coast, rose in revolt. To save his throne and his life Selim abolished the *nizam-i-jedid*. Nevertheless, he was deposed on May 29, 1807, and on July 28, 1808, he was murdered in Istanbul. *See* also TURKEY: *History*.

(E. Z. K.)

SELINUS (Greek SELINOUS, modern Italian SELINUNTE), an ancient city on the south coast of Sicily, just west of modern Marinella and 8 mi. (13 km.) SE of Castelvetrano, famous for its ruined Doric temples. It was traditionally founded in 651 or in 628 B.C. (pottery finds may support the earlier date) by colonists from Megara Hyblaea and from Megara in Greece (*see* SICILY: *History*). The name, which belonged both to the city and to the river west of it (modern Modione), was derived from the wild celery (Greek *selinon*) which grows there abundantly and appears on some of its coins (*see* NUMISMATICS: *Greek Coins*). Boundary disputes with Segesta (*q.v.*) are recorded as early as 580 B.C. Selinus soon grew in importance, and its territory extended from the Mazarus (Mazaro) to the Halycus (Platani) river. About 510 Peithagoras, and after him the Spartan refugee Euryleon, set themselves up as tyrants. Selinus was on the Carthaginian side when the Carthaginians were defeated at Himera (*q.v.*) in 480. Thucydides speaks of the treasures in Selinus' temples, and the city had a treasury of its own at Olympia. A dispute between Selinus and Segesta was one of the causes of the Athenian expedition to Sicily of 415 B.C. After its failure (413) the Segestaeans appealed to Carthage, and an overwhelming force under Hannibal, son of Gisgo, took and destroyed Selinus in 409; the walls were razed and only 2,600 inhabitants escaped to Acragas (Agrigento). In 408 Hermocrates (*q.v.*) occupied Selinus and rebuilt the walls, but he fell in an attempt to enter Syracuse, and as a result of the treaty of 405 Selinus became subject to Carthage and remained so with brief interruptions until its inhabitants were transferred to Lilybaeum (Marsala; 250 B.C.). It was never afterward rebuilt and the temples were all toppled, probably by an earthquake; several fallen columns have been reerected in modern times.

The ancient city occupied a sand hill running north and south; the southern part (overlooking the sea), which was the acropolis, is surrounded by fine masonry walls of rectangular stone blocks, which show traces of the reconstruction of 408 B.C. Two main streets run north and south and east and west, with others diverging at right angles.

The southeast corner contains several temples. The deities to whom they were dedicated are not in every case known, and they are conventionally known by letters. In all the large temples the cella is divided into two parts, the smaller and inner of which (the *adytum*) was intended for the cult image. There are traces of early temples on the acropolis, but the earliest and largest extant is C (mid-6th century; 209 × 78 ft.). It had 6 columns (the usual number) at each end (a double row in the front) and 17 on each side. From it came three archaic metopes in the National Archaeological Museum, Palermo. Some of its coloured terra-cotta decorations have also been discovered, including the fragments of an enormous Gorgon mask, over 8 ft. high, from the centre of one of its pediments. In front of it stood a large altar more than 60 ft. long. Next to it on the north lies temple D (late 6th century; 183 × 77 ft.), both having been included in one precinct with other buildings of less importance; immediately to the east of D is a large altar. B is a small late temple; A and O (early 5th century; both about 132 × 53 ft.) lie on the south side of the main east-west street in another precinct.

At the north end of the acropolis are extensive remains of the fortifications of Hermocrates, built across the narrow neck connecting the acropolis with the rest of the hill. In front of the wall lies a deep trench, and outside this a projecting semicircular bastion commands the entrance on the east, a winding trench approached by a pair of double gateways. The builders reused fragments from earlier buildings, mostly private houses, but including a small temple; fragments of four metopes from this temple (early 6th century) are at Palermo.

Outside the ancient city, on the west of the Selinus River at Gaggera, lie the ruins of a temple of Demeter Malophoros, with a propylon (monumental entrance) leading to the sacred enclosure. The temple itself (the third on the site) has a cella with a narrow door and no columns; outside and in front of it was a large altar 52 ft. long. A large number of offerings were excavated, including strange double-headed steles dedicated to Zeus Meilichios. The earliest temple must have been built soon after the foundation of the city, the latest shortly after 600 B.C. The propylon may date from after 409 B.C.

On the hill east of Selinus, separated from it by the flat Gorgo di Cottone, lie three huge temples. No other remains have been found around them; it is likely that they were outside the town but stood in a sacred enclosure. A peculiarity of the construction of the oldest temple, F (or S; late 6th century; 203 × 80 ft.), is that the outside columns were linked together by stone screens over 10 ft. high, possibly to keep the ceremonies secret. The huge temple G (or T; begun late 6th century but worked on well into the 5th; 361 × 164 ft.) was dedicated to Apollo; though it was never completed (many of the columns still remain unfluted), it was in use. Its plan is curious; it has 8 columns at each end and 17 along each side, and two lines of 10 small columns, probably in more than one story, divide the cella lengthwise. The blocks of stone were quarried from the Cave di Cusa, 8 mi. to the northwest, where similar blocks intended for it may still be seen. The third temple, E (or R; early 5th century; 222 × 83 ft.), was apparently dedicated to Hera. Its fine metopes are now at Palermo. The cemeteries outside the city on all three sides provided valuable material for dating.

BIBLIOGRAPHY.—G. Fougères and J. Hulot, *Sélinonte* (1910); T. J. Dunbabin, *The Western Greeks*, esp. ch. 9, 10 (1948); A. G. Woodhead, *The Greeks in the West*, esp. pp. 48 ff., 70–71 (1962).

SELJUKS, the name of the ruling family of the Oghuz (Ghuzz) Turkmen tribes who invaded western Asia in the 11th century and founded a powerful empire, later to be divided into the principalities of Iran, Syria, and Asia Minor. The history of these invaders forms the first part of the history of the Turks in the Near and Middle East.

During the 10th century, in the course of the migrations and struggles of the Turkish peoples of central Asia and southeast Russia, a group of nomadic Turks under a chief named Seljuk settled in the lower reaches of the Syr-Darya River, and later embraced Islam. They were incorporated in the frontier defense forces of the Samanids (*q.v.*) and moved to the vicinity of Bukhara *c.* A.D. 985. After the fall of the Samanids, they were established by Mahmud (*q.v.*) of Ghazni in the border regions of Khurasan, but it was difficult to induce warlike nomads to live peacefully within a settled empire. While Mahmud's son Mas'ud was busy waging war in northern India the two grandsons of Seljuk, Chagri (Chaghri) Beg and Togrul (Tughril) Beg, saw the chance to win a realm of their own with the support of Persian nobles and religious leaders who were anxious to meet the threatening progress of heretic Shi'ite sects reinforced by the Buyid princes of western Iran and Iraq and the Fatimid Isma'ilian caliphate of Egypt. When Mas'ud saw the danger, it was too late: he was utterly defeated at Dendenkan (Dandanqan) near Merv in 1040. Thereafter Chagri Beg remained in control of the greater part of Khurasan while Togrul, with a mixed army of Turkmen and regular Mameluke (slave) troops, set out to conquer new territories in the west.

The Great Seljuks.—For the details of the reigns of the first three Seljuk sovereigns, known as the Great Seljuks, reference should be made to the articles TOGRUL BEG; ALP ARSLAN, MOHAMMED IBN DA'UD; and MALIK SHAH. Suffice it here to say that Togrul Beg, a strong leader of men and a clear-headed statesman, achieved the creation of an empire which at his death (1063) included central and western Iran and Mesopotamia with Baghdad itself. There the caliph al-Ka'im welcomed him in 1055 as his deliverer from the rule of the Buyids and bestowed on him the title of sultan with the task of restoring the unity of the Muslim world under the religious sway of the Abbasid caliphate. Alp Arslan (1063–1072 or 1073), who succeeded both Chagri and Togrul, and Malik Shah completed his conquering work. When

Malik Shah died in 1092, the Seljuk Empire included the whole of Iran with the exception of the Ghaznavid realm of the Indian borderland, the whole of Mesopotamia, and Syria, including Palestine, to the frontier of Egypt. Furthermore, other Turkmen invaders closely akin to the Seljuks had occupied the Asiatic provinces of the Byzantine Empire up to the Aegean Sea. The organization of the empire was the task of the famous vizier Nizam al-Mulk (*q.v.*), who remained in power during the reigns of both Alp Arslan and Malik Shah.

The main problem facing the Seljuks was to persuade the Turkmen invaders, to whom they owed their victory, to live within the framework of the bureaucratic Iranian-Iraqi state. This problem was never completely solved. The Turkmen tribesmen had no interest in victory over heretics if their lands were unsuitable for stockbreeding or if pillage was not allowed; but on the northwestern frontier of Iran lay the Byzantine Empire, including half of Armenia, which was lacking religio-national homogeneity and the kind of military force able to fight against nomadic invaders. Love of plunder and zeal for a holy war drew the Turkmens into Byzantine territory, and, to prevent them from becoming quite estranged from him, the sultan had sometimes to lead them himself. Thus Alp Arslan won immortal fame in the Muslim world when he defeated a huge Byzantine army at Manzikert in 1071, capturing the Byzantine emperor Romanus IV Diogenes. Thenceforth the Turkmens were able to settle in Asia Minor undisturbed.

The main feature of the Seljuk Empire was its religio-political character. The regime had to fight not only external wars but also wars against heretics within its frontiers, and to the Seljuks Islam owes much of its later achievements. Before the Seljuk period education in the Muslim world had not been organized by the state. Now madrasahs (colleges) were founded, with the object of training administrators and the upper classes in Islamic science. The head of orthodox Islam was the Abbasid caliph, but, though he was permitted a degree of veneration and independence in Baghdad, he had no real political power in the empire. The sultans as well as the caliph considered it their duty to promote religious works; for this reason as well as to show their power they built many mosques, including the glorious mosque of Isfahan.

Strange as it may seem, the rule of the Turkish Seljuks fostered rather than hampered the progress of Persian cultural autonomy. Before the Seljuk era a new form of Persian literature had developed in the Samanid territories of northeastern Iran and Transoxiana. This was no longer written in the pure old language or in the alphabet of pre-Islamic Iran, but in the Arabic script, with a number of borrowed Arabic words. At the same time the Buyids had favoured a form of Persian patriotism which expressed itself in Arabic. Seljuk rule, on the other hand, led to the spread of literary Persian to the whole of Iran. The Turks had no cultural Islamic tradition and almost no written literature. As the Iranians had been their instructors in Islam, they knew no cultural language other than Persian and showed no interest in the achievements of scholars or poets writing in Arabic. From the time of the Seljuks onward, the Arabic language disappeared from the Iranian lands.

It is wrong to believe, as many have believed, that the pursuance of an Islamic policy and of conquest in Asia Minor led the Seljuks to persecute the Christians. The plundering of the Turkmen armies in the Byzantine Empire no doubt caused much suffering and loss to Eastern Christianity, but inside the Seljuk Empire, as soon as order was restored, the lot of Christians was much the same as it had been before: the crusaders, who thought it must be otherwise, were judging conditions in Jerusalem by those prevailing in Asia Minor.

The Later Seljuks.—Despite its successes there were weaknesses in the Seljuk Empire. It could not prevent the growth of the new terrorist sect, the Assassins (*q.v.*), which was to inspire such dread in the medieval East and which began its activities with the murder of Nizam al-Mulk (1092). Furthermore the Seljuks, like the Buyids before them, could not rid themselves of the idea that sovereignty belonged to a family rather than to a single man; thus even the most powerful sultans thought it necessary to give provinces as appanages to their relatives. In the early years of the Seljuk conquests a practically independent principality had been founded in Kerman by a son of Chagri Beg, Qavurt, who extended his conquests to Oman on the Persian Gulf. Another independent but short-lived principality was created in Syria, partly for military reasons, by Malik Shah in favour of his brother Tutush.

When Malik Shah died, leaving several sons by different wives, all of them very young, the quarrels between them and their supporters resulted in a partition of the Seljuk heritage among Barkiyarok, the eldest son; his brother Mohammed, who received northwestern Iran; and a third brother, Sanjar (*q.v.*), who was given charge of Khurasan. When Barkiyarok died in 1104 Mohammed was able to establish only a measure of control in central and western Iran, and this broke down at his death (1118). Sanjar's realm in Khurasan, however, attained a new power and glory, which were magnified in the eyes of posterity by the length of his reign and the disasters which befell him in his last years. It is necessary here only to recall his victories over the rival Turkish kingdom of the Qarakhanids in Transoxiana and over the Ghaznavids of the Indian borderland. This was followed by his defeat in 1138 by new invaders from central Asia, the Khara-Khitai, and by the revolt of his vassal in Khwarizm (Khorezmia). Finally came the general uprising of the Ghuzz (1153), which resulted in Sanjar's captivity, from which he escaped only to die in 1157. Years of disorder and misery ensued for Khurasan, until its final incorporation in a new Khorezmian empire in the reign of Takash (1172–99).

Thereafter no territory was left in the former Seljuk Empire where a sultan could enjoy real power. When pretenders quarreled they sought the help of high military officers, to whom they were obliged to abandon the government of whole provinces. This was especially so in the case of the atabegs, officers who were entrusted with a kind of tutorship over minor princes. Tribal chiefs also profited by the disorders to make themselves independent, and the caliph himself thought the time ripe, if not to recover the ancient power of the Abbasids over the whole of the Muslim world, at least to secure a principality of his own in Iraq.

The first of the Seljuk lines to disappear was that in Syria. There Tutush's sons, Ridwan of Aleppo and Duqaq of Damascus, had quarreled and were weakened by the attacks of the crusaders. Toghtekin, son of Buri, the atabeg of Duqaq, founded a dynasty of his own (the Burids) when his master died in 1104. In Aleppo the general feeling was against "Easterners," and the dynasty scarcely survived Ridwan's death (1113). Atabegs were also governing upper Mesopotamia around Mosul, and one of them, Zangi, united this province and Aleppo under his sway in 1128, thus creating a new dynasty there. Another atabeg dynasty, which was to last until the beginning of the 13th century, came to power somewhat later in Azerbaijan. These atabegs "protected" the last Seljuks of the Iranian line, who resided in Hamadan but had no powers except those granted them by their all-powerful tutors.

In the mid-12th century, the caliph again became independent, and when the last of the Iranian Seljuks, Togrul, tried to recover some of his lost influence, the caliph appealed to the shah of Khorezmia (the Khwarizm-shah) for help. As a result Togrul died on the battlefield in 1194. Meanwhile other parts of the Seljuk heritage had met their fate at the hands of tribal chiefs. Even Arab tribes, such as the Mazyadits on the Iraqi borders, had taken advantage of the situation to win at least temporary autonomy. The principality of Kerman did not long survive the death (1170) of its sultan Togrul. His sons quarreled and called in foreign aid, with the result that the country was utterly devastated and fell an easy prey to a band of Ghuzz from Khurasan under their chief, Malik Dinar (1185). Ten years later it was incorporated in the Khorezmian empire. Only in Asia Minor did Seljuk power survive, and there it did not reach its peak until the 13th century.

The Sultanate of Rum.—The invasion of Asia Minor began as an uncoordinated movement of Turkmen tribesmen from Azerbaijan, and the expeditions of Togrul Beg and Alp Arslan were aimed as much at controlling the tribesmen as at conquest. Alp Arslan's victory at Manzikert (1071) destroyed the Byzantine

frontier organization and enabled the Ghuzz to establish themselves in Byzantine territory, where they engaged as mercenaries in the struggles between the local commanders. Among a host of tribal chiefs the leading place was taken about 1075 by the sons of Togrul Beg's cousin Kutalmish (Kutlumush), who were violently hostile to Malik Shah. During the struggles of rival generals for the throne of Constantinople (1078–81), one after another of the contestants called for aid from the sons of Kutalmish and opened to them the gates of their cities, including Nicaea and Nicomedia (Izmit). By 1080 Suleiman ibn Kutalmish held the greater part of Asia Minor as the ally of the emperor and rival sultan to Malik Shah. In 1084 Antioch (Antakya) was surrendered to him, but in 1086 he was killed near Aleppo in battle with Tutush and a coalition of Syrian princes. Malik Shah attempted to take advantage of Suleiman's death, both by proposals for an alliance with the Byzantine emperor Alexius I Comnenus and by military expeditions into western Anatolia, but with little success; and on his death (1092) Suleiman's son Kilij Arslan (who had been captured by Malik Shah at Antioch), escaping from his captivity, reconstituted the sultanate. In the interval, however, northeastern Anatolia had been occupied by a rival Turkmen chief, Danishmend of Sivas, whose successors engaged in a prolonged conflict with the Seljuk sultans.

The new sultan was immediately confronted with the First Crusade. The first crusader bands, under Walter the Penniless, were defeated at Nicaea (1096), but the Turks were severely defeated before Nicaea and at Dorylaeum (Eskisehir) in 1097 and driven into the interior, while the emperor Alexius reoccupied western Anatolia. After his victory over the next wave of crusaders (1101), Kilij Arslan, tempted by the disorders which had weakened the Seljuks in Iraq, made a bid for the greater sultanate. He succeeded in capturing Mosul but was defeated by the forces of his kinsman of Aleppo and drowned in the Khabur (Habur) River (1107).

This event proved decisive in determining the future development of the sultanate of Rum (the name used by the Turks for the Byzantine empire; *see* RUM). The Seljuks, hemmed in between the Greeks and the crusading states in Syria and increasingly isolated from the east, gradually established an organized and settled Anatolian kingdom. Kilij Arslan was succeeded by his son Malik Shah (1107–16), and he by his brother Mas'ud I (1116–55), who established his capital at Konya (Iconium). Mas'ud's long reign was occupied by resistance to the encroachments of the Greeks and by engagements and negotiations with the Danishmends, the crusaders, and his Muslim neighbours. During the reign of his son and successor, Kilij Arslan II (1155–92), the Armenians established a principality in Cilicia, but the Danishmends were finally subdued and their territories annexed to the sultanate.

Kilij Arslan's numerous sons, each of whom held the command of a city of the empire, embittered his old age by their mutual rivalry, and the eldest, Qutb ud-din, tyrannized over him in his own capital, exactly at the time that Frederick I Barbarossa entered the sultan's dominions on his way to the Holy Sepulchre (1190). Konya itself was taken and the sultan was forced to provide guides and provisions for the crusaders. Kilij Arslan lived two years longer, finally under the protection of his youngest son, Kaikhosrau, who held the capital after him till 1196 when his elder brother, Rukn al-din Suleiman, having vanquished his other brothers, ascended the throne and obliged Kaikhosrau to seek refuge at the Greek emperor's court. Rukn al-din, who died in 1204, saved the sultanate from destruction and conquered Erzurum, which had been ruled for a considerable time by a separate dynasty; but Rukn al-din's son, Kilij Arslan III, was soon deposed, with Greek assistance, by Kaikhosrau.

After the establishment of the Latin empire of Constantinople (1204) the Turks were the natural allies of the Greeks and the enemies of the crusaders and their allies, the Armenians. Kaikhosrau, therefore, in 1207 took the important harbour of Attalia (Antalya) from the Italian Aldobrandini; but in 1211 he perished in battle with Theodore I Lascaris, emperor of Nicaea. His son and successor, Kaikaus, made peace with Lascaris and extended his frontiers to the Black Sea by the conquest of Sinope (Sinop; 1214). On this occasion he took prisoner the Comnenian prince Alexius, who ruled the independent empire of Trebizond (Trabzon), and compelled him to acknowledge the supremacy of the Seljuks, to pay tribute, and to serve in the armies of the sultan. Elated by this great success and by his victories over the Armenians in Cilicia, Kaikaus attempted the capture of Aleppo, at this time governed by the descendants of Saladin, but the project was defeated by Ayyubid resistance.

Kaikaus' brother, Ala al-din Kaikobad I (1219–37), was the most powerful and illustrious prince of this branch of the Seljuks. He extended his rule as far as Seleucia and desisted from further conquest only on condition that the Armenian princes would enter into the same kind of relation to the Seljuks as had been imposed on the Comnenians of Trebizond. But his greatest military fame was won by a war which, however glorious, was to prove fatal to the Seljuk Empire in the future; in conjunction with his ally, the Ayyubid prince Ashraf, he defeated the Khorezmshah, Jalal al-din, near Erzincan (1230). During this war Kaikobad put an end to a collateral dynasty of the Seljuks of Erzurum and annexed its possessions. He also gained the city of Khilat with dependencies which had recently been taken from the Ayyubids by Jalal al-din. This acquisition, however, led to a new war, as Kaikobad's ally, the Ayyubid prince, contested it. Sixteen Muslim princes, mostly Ayyubids, of Syria and Mesopotamia, under the leadership of Malik al-Kamil, prince of Egypt, marched with considerable forces into Asia Minor against him. Happily for Kaikobad, the other princes mistrusted the power of the Egyptian, and it proved a difficult task to penetrate through the mountainous, well-fortified accesses to the interior of Asia Minor. The advantage thus rested with Kaikobad, who extended his power in upper Mesopotamia. This expansion was pursued by his son Kaikhosrau II.

The Seljuk sultanate of Asia Minor can now be seen to have been one of the most important Muslim states of its age. Its population was a mixed one, including Christians, Armenians, Greeks, Syrians, and Iranian Muslims; but, when compared with the other Seljuk realms, of which the Turks had been only a small, if leading element, it was really "Turkey" and was so called by its contemporaries. The Seljuks of this line succeeded in establishing an administration largely based on the institutions of their Iranian-Seljuk neighbours, but it was modified by the Byzantine heritage and by its own evolution. With order and tolerance of all races and religions established, agriculture and mining activity revived, so that to foreigners Turkey seemed one of the richest of countries; meanwhile commerce was developed with the assistance of Italian merchants. In the cities beautiful buildings, several of which still survive, show the purity of an art related to, though not exactly identical with, the Iranian-Seljuk art of the same period. Literature, mostly in Persian, flourished, and, with it, mystical movements, such as that inspired by Jalal-ud-din Rumi (*q.v.*).

Yet this political organization and civilization were not really soundly based. The Seljuks failed to tame the Turkmens, particularly in the frontier districts, who, though partially settled, were impatient of centralized administration; they remained fond of petty wars against their Christian neighbours and still adhered to their traditional social and religious beliefs and practices. Newcomers from the countries invaded by the Khorezmians and Mongols added to their feeling of unrest. This helps to explain why, when the Mongol flood reached the frontiers of Turkey, which were no longer protected by the Khorezmian state, the Seljuk realm was no more able to resist than had been the Muslim principalities of Iran. At the Battle of Kosehdagh (1243), between Erzincan and Sivas, the independence of the Seljuks was lost forever.

The Mongol Protectorate.—The Mongols did not destroy the Seljuk state, but, in accordance with their plan, reduced it to vassalage. After the death of Kaikhosrau II (1245) the quarrels between his sons led to a division of Asia Minor, Izz al-din Kaikaus II taking the part west of the Halys and Rukn al-din Qilij Arslan IV the part to the east. When the former intrigued with the Mamelukes of Egypt and the Byzantine emperor, however, he was driven out and fled to Constantinople. The unity of the realm was thus restored, but the financial requirements of the

Mongols and the intrigues of the great chiefs with them against one another quickly brought about a breakdown of the Seljuk administrative system. In the central provinces and cities, the all-powerful minister, the *perwane* Mu'in al-din Suleiman, who in 1267 had Rukn al-din executed, succeeded in maintaining the Iranian-Islamic civilization of the recent golden age; to this end he co-operated with the Muslim viziers of the Mongol Il-khan of Iran, the suzerain of Asia Minor. In the distant and mountainous districts, however, the Turkmen amirs, free from any form of government, established small principalities of their own: in the Taurus; on the Byzantine frontier; on the south Aegean coast; later even near the Straits and the Black Sea. With their help and that of Baybars, the powerful Mameluke sultan of Egypt, the most dreaded enemy of the Mongols, a group of Muslim nobles in Asia Minor revolted during 1276–77 against the Mongol protectorate. The *perwane* himself had negotiated with them and although he finally refused to join them, thus assuring the victory of the Mongols, he was executed. For a time the Seljuk sultanate was then no more than a Mongol province. Seljuk sultans, Kaikhosrau III (d. 1283), Mas'ud, and Faramarz, still reigned nominally, but in the first years of the following century the dynasty ended in obscurity.

Thereafter the power of the Turkish amirates increased and the control of the Mongol Il-khans of Iran became less effective until it also finally disappeared, in the 1330s. For later Turkish history *see* TURKEY.

See also references under "Seljuks" in the Index.

BIBLIOGRAPHY.—No adequate general studies of the Seljuks exist, and the only important Oriental works to have been translated into a Western language are that of Nizām al-Mulk, Eng. trans. by H. Darke, *The Book of Government* (1960), and the chronicle of Ibn al-Bībī, German trans. by H. W. Duda, *Die Seltschukengeschichte des Ibn Bibi* (1959). Original works in Western languages include: three chapters by Claude Cahen in the *History of the Crusades*, ed. by K. M. Setton, vol. i (1955) and vol. ii (1962), where references will be found to other studies; P. Wittek, "Deux Chapitres de l'histoire des Turcs de Roum," in *Byzantion*, vol. xi (1936); W. Barthold, *Turkestan Down to the Mongol Invasion*, Eng. trans. by H. A. R. Gibb (1928, reprinted 1958); and M. F. Sanaullah, *The Decline of the Seljuk Empire* (1938). (H. A. R. G.; CL. CA.)

SELKIRK (SELCRAIG), **ALEXANDER** (1676–1721), Scottish sailor who was the prototype of Robinson Crusoe, was born at Largo, Fife, the son of a shoemaker. In 1695 he ran away to sea to join the buccaneers and by 1703, under the surname of Selkirk, he was sailing master in the galley "Cinque Ports," the consort of the "St. George," captained by William Dampier, on a privateering expedition in the Pacific. In September 1704, after a quarrel with his captain, Thomas Stradling, Selkirk was put ashore at his own request on the uninhabited island of Juan Fernández, west of Valparaíso, Chile. He remained there alone until Feb. 1, 1709, when he was discovered by Capt. Woodes Rogers in the "Duke," whose pilot was Dampier. Selkirk finally sailed from the island on Feb. 13. He was appointed mate to Rogers and in March was given command of a captured prize ship, the "Increase." He returned to England in October 1711 and met the essayist, Richard Steele, who wrote up his story in *The Englishman* (No. 26; Dec. 3, 1713). In 1719, a year after the publication of the second edition of Rogers' narrative of his voyage, Daniel Defoe made use of his story in *Robinson Crusoe*. Selkirk returned in 1712 to Largo, where he lived like a recluse, and later went to sea again. He died in 1721 when he was master's mate of HMS "Weymouth." A pamphlet, *Providence Displayed*, purporting to be by Selkirk, is a forgery.

BIBLIOGRAPHY.—W. Rogers, *Cruising Voyage Round the World* (1712); J. Howell, *Life and Adventures of Alexander Selkirk* (1829); P. K. Kemp and C. C. Lloyd, *The Brethren of the Coast* (1960).
(C. C. L.)

SELKIRK, THOMAS DOUGLAS, 5TH EARL OF (1771–1820), Scottish philanthropist who founded the Red River Settlement in Canada, was born at St. Mary's Isle, Kirkcudbrightshire, on June 20, 1771. He succeeded his father, Dunbar Douglas, in 1799. A Whig in politics, he believed that the hardships of the peasantry in the Scottish Highlands could only be alleviated by emigration. He went to Canada in 1803 and established a large settlement on Prince Edward Island. In 1810 he acquired control of the Hudson's Bay Company, which granted him (May 1811) a vast tract of land in the Red River Valley, near present-day Winnipeg, where he founded a settlement. However, the hostility of the North-West Fur Company seriously impeded the colony (*see* RED RIVER SETTLEMENT). Selkirk returned in January 1818 to Upper Canada where, after lengthy legal actions, his rivals won heavy damages against him. Broken in health and having lost a fortune, he returned to England in November 1818. He died at Pau, in France, on April 8, 1820. Selkirk laid the foundations of Winnipeg and Manitoba, where the town of Selkirk commemorates his name.

SELKIRK, a southern county of Scotland, bounded north and west by the county of Peebles, north by Midlothian, east and southeast by Roxburgh, and south and southwest by Dumfries. Area 268 sq.mi. (693 sq.km.).

Physical Geography.—Forming a rectangle 25 mi. by about 12 mi. (40 km. by about 19 km.) it lies southwest–northeast, and this trend echoes that of the three main rivers, Yarrow, Ettrick, and Ale, the first two joining above Selkirk and flowing to the Tweed which runs for some 10 mi. (16 km.) across the north of the county. Underlain throughout by folded Silurian grits and shales, the physical divisions are largely a function of differential erosion. The west and southwest comprise the highest land, much of it a rolling erosion surface at 1,800–2,000 ft. (550–600 m.), above which rise Broad Law (2,755 ft. [840 m.]) and Ettrick Pen (2,270 ft. [692 m.]) on the Peebles and Dumfries boundaries respectively. The deep trenches cut by the Megget Water and by the Yarrow and Moffat Waters open routes and provide farmland in this area of nardus grass sheep country, where precipitation is over 60 in. (1,524 mm.) and snow may lie for 60 days in the year. As it was once a local centre for ice dispersal there is much bare rock and thin podsolized soils, and glacial drift occurs only in the valleys. The beauty of the Vale of Yarrow inspired Sir Walter Scott and William Wordsworth. Eastward the rocks are less resistant and the landscape consequently softens, smaller remnants of the higher plateau surfaces remain, valley slopes are gentler, and there is more improved land, although the flat floors at about 600 ft. (180 m.) are liable to flooding. St. Mary's Loch (3 mi. [4.8 km.] long), separated by a narrow isthmus from the Loch of the Lowes (1 mi. long), lies in this belt. The eastern third of the county is the lowest. The Ale Water is entrenched in a surface at about 1,200 ft. (370 m.) along the Roxburgh border, with hills seldom above 1,600 ft. Ice action has molded the landscape into "swarms" of drumlins (*see* GLACIER) 100–300 yd. (90–275 m.) long and 30–50 yd. wide, lapped by peat bogs and lochans formed on thick deposits of glacial drift. The heart of the county lies in the north around the town of Selkirk, where the valleys widen and coalesce with Tweedside itself; precipitation is about 35 in. (889 mm.) and there is much glacial drift, so a farming landscape predominates. (A. M. LE.)

Bird and animal life include pheasant, grouse, partridge and duck, roe deer, and foxes. The rivers are noted for their trout and salmon.

History.—The site of a Roman camp has been confirmed near Oakwood, in the Ettrick Valley; but the natives were probably held in check from the station at Newstead near the Eildons. Of the few so-called British camps situated in the lower valleys the most important is the large work on Rink Hill in the parish of Galashiels. At Torwoodlee, 2 mi. (3.2 km.) NW of Galashiels, are the ruins of the only example of a broch (round tower) in the Border counties; some Roman remains were found there. An ancient earthwork of uncertain origin runs 4¼ mi. from Linglie Hill near Philiphaugh, northwest of Selkirk burgh, to Mossilee near Galashiels. It consists of a ditch, about 30 ft. wide and 3 ft. deep, and a bank made by the excavated earth, thrown out usually on its lower side. The history of the shire for six centuries following the retreat of the Romans is that of the whole of southeastern Scotland. The county formed part, first, of the British kingdom of Strathclyde, then of the Saxon kingdom of Northumbria, and finally, about 1020, was annexed to Scotland.

To the north of Hangingshaw, in the country between the Yarrow and Tweed, Sir William Wallace constructed an earthwork in 1297, still called Wallace's Trench, 1,000 ft. (300 m.) long and

terminating on the top of a hill in a large square enclosure. There he lay till his plans were completed, and at last departed, his forces including a body of Selkirk archers, for a raid into the north of England. During the prolonged strife that followed the death of Robert I, the Bruce (1329), the foresters were constantly fighting, and the county suffered more heavily at Flodden (1513) than any other district. The lawlessness of the Borderers was at length put down by James V with a strong hand. He parceled out the forest in districts and to each appointed a keeper to enforce order and protect property. The county was known at that time as Ettrick Forest and was a hunting ground for the Scottish kings. In 1529 the ringleaders of the Borderers, including William Cockburn of Henderland, Adam Scott of Tushielaw, and the notorious Johnnie Armstrong, were arrested and promptly executed. This severity had the desired effect, though in 1603 the freebooters and moss-troopers again threatened to be troublesome, until James VI's lieutenants ruthlessly stamped out disaffection. The Covenanters held many conventicles in the uplands, and their general, David Leslie (later Lord Newark), routed the marquess of Montrose at Philiphaugh in 1645.

Sir Walter Scott (*q.v.*) was sheriff of the county from 1799 until his death, and from 1804 to 1812 he lived in the mansion of Ashiestiel, on the south bank of the Tweed.

Population and Administration.—The population of the county in 1961 was 21,052, of whom only 46 could speak both Gaelic and English. Selkirk (*q.v.*) is the county town and only royal burgh (pop. 1961, 5,634), with a charter dating from 1535, but Galashiels (*q.v.*), the only other small burgh, is the biggest town (pop. 1961, 12,373). There are two county districts. The rural population is on the decline.

The county, along with the counties of Roxburgh and Peebles, sends one member to Parliament. It forms part of the sheriffdom of Roxburgh, Berwick, and Selkirk with a court in Selkirk.

Agriculture and Industry.—The hill land is divided into large farms of 1,000–3,000 ac. (400–1,200 ha.) carrying sheep and hill cattle, the sheep being mainly Black-face Highland and Cheviot, and the cattle Galloway or other hardy breeds. Arable land is largely used to produce fodder for stock, principally oats and turnips, with some barley. The numbers of hill cattle have markedly increased since 1946. The predominant industry in the towns is the woolen industry (tweeds, hosiery, and yarn), and there is a large tannery in Galashiels. (W. T. D.)

See D. L. Linton and C. P. Snodgrass, ***Peebles and Selkirk***, Land of Britain Series, no. 24, 25 (1946).

SELKIRK, a royal and small burgh and the county town of Selkirk, Scot. Pop. (1961) 5,634. It lies on a hillside at the eastern edge of Ettrick Forest and on the right bank of Ettrick Water, 39 mi. SSE of Edinburgh by road. There are statues of Sir Walter Scott in his sheriff's robes (he was sheriff 1799–1832) and of Mungo Park, the African explorer, who was educated at the grammar school; and there is a memorial of Flodden Field, where 80 townsmen fought and fell.

As its early name (*Scheleschyrche*) indicates, Selkirk originally consisted of a number of *shiels* (huts) in the forest beside which a church had been built by the Culdees of Old Melrose. David I, while prince of Cumbria, founded in the early 12th century a Benedictine abbey, which was removed in 1147–52 to Kelso. He also built a castle which was captured by Edward I, by whom it was enlarged and strengthened. It was retaken by William Wallace in 1297, from which time it remained in the hands of the Scots till the Battle of Halidon Hill (1333), when it was delivered to the English. It was probably destroyed in 1417 when Sir Robert Umfraville, governor of Berwick, set fire to the town. Nothing remains of the castle save some green mounds at Haining. The burgh charter granted by David I was renewed by James V in 1535. From an early period shoemaking has been the staple industry and the inhabitants have been known as souters (shoemakers). Though this craft has given way to tweed manufacture, the tradition of the shoemakers is still upheld when someone is made a "Souter" on receiving the freedom of the burgh, for he then has to "lick the birse." This entails drawing a small shoemaker's brush of mild boar's bristles, steeped in wine, between his lips. In memory of Flodden, the Common Riding takes place each June, when the bounds are ridden by a cavalcade of horsemen; at the end of the ride the standard-bearers cast their colours to the strains of Selkirk's lament, "The Flowers of the Forest."

SELKIRK MOUNTAINS, in southeastern British Columbia, Can., extend for about 200 mi. (322 km.) N from the United States border and, with the Purcell Mountains to the east, lie between it and the northern bend of the Columbia River. Selkirk, Purcell, Monashee, and Cariboo mountains form major subdivisions of the Columbia Mountains, a great wedge of rugged terrain lying west of the Rocky Mountain Trench and south and east of the Fraser River.

Lowest in the south where their summits average about 7,500 ft. (2,286 m.), the Selkirk Mountains increase in elevation northward, several peaks attaining heights of over 10,000 ft. toward their northern extremity. The highest peak, Mt. Sir Sandford, stands at 11,590 ft. In many places the mountains rise abruptly to over 8,000 ft. above adjacent valley floors, affording wild and magnificent scenery. Extensive glaciers and snowfields are found in Glacier National Park, in the north; perhaps the most celebrated is the Illecillewaet, near Glacier, on the Canadian Pacific Railway. The Selkirks are crossed by railway and highway at Rogers Pass, discovered in 1883. The engineering difficulties to be overcome included the building of five-mile-long Connaught Tunnel, east of Glacier. (A. L. Fy.)

SELLA, QUINTINO (1827–1884), Italian statesman who not only put the government's finances on a sound footing but also played a decisive role in the annexation of Rome. He was born at Mosso, near Biella, in Piedmont, on July 7, 1827. After studying engineering at Turin and in Paris, he obtained an academic post at Turin, and traveled widely in the course of his research. In 1860 he was elected to the Sardinian–Piedmontese Chamber of Deputies.

In March 1862 Sella was appointed minister of finance in Urbano Rattazzi's government for the new Italy; but this government fell in December. When he became minister of finance again in A. F. La Marmora's cabinet in September 1864, he dealt energetically with the deficit of 200,000,000 lire then existing, but was again forced to resign in January 1866. The grist (or corn milling) tax, which he had proposed in 1865, was adopted by L. F. Menabrea's government in 1868; and in December 1869 Sella became minister of finance for the third time, in a cabinet formed by himself but with Giovanni Lanza as prime minister. The grist tax and Sella's other fiscal expedients over the next three years restored the national finances. In 1870, during the Franco-German War, it was Sella who prevailed on Victor Emmanuel II to take advantage of France's predicament in order to occupy Rome (*see* ITALY: *History*).

The Lanza-Sella government fell in June 1873. Having retired from politics in 1881, Sella died at Biella on March 14, 1884. His *Discorsi parlamentari* fill five volumes (1887–90).

SELMA, a city of Alabama, U.S., the seat of Dallas County, is located on the north bank of the Alabama River, 50 mi. W of Montgomery. It is in an area of rich soil suitable for cotton, livestock, pecans, and diversified farm products, all of which are produced in abundance. The city is a lumber and manufacturing centre; products include cotton bags and yarn, clothing, food products, cigars, toys, and fertilizers. Craig Air Force Base is located nearby. The city is also noted as a medical centre. In 1965 it was the centre of a Negro voter registration drive led by Martin Luther King (*q.v.*). Terrorism and violence, culminating in the murder of James J. Reeb, a Boston minister, led to a massive, nonviolent protest march on the state capital (*see* MONTGOMERY).

The place was known as Moore's Bluff or Moore's Landing in 1815; it was incorporated in 1820 by a company headed by William Rufus De Vane King (vice-president of the United States in 1853), who named it Selma from a poem by Ossian.

It replaced Cahaba (capital of Alabama from 1820 to 1826) as county seat in 1865. Its chief pre-Civil War showplace is Sturdivant Hall, formerly the Gilliam home. For comparative population figures *see* table in ALABAMA: *Population.* (W. T. Jo.)

SELOUS, FREDERICK COURTENEY (1851–1917), English hunter and explorer in south-central Africa whose travels added to the knowledge of Rhodesia, was born in London on Dec. 31, 1851, and was educated at Rugby and in Germany. In 1871 he went to South Africa and in 1872 traveled to Matabeleland where he hunted big game. He then wrote *A Hunter's Wanderings in Africa* (1881).

For 18 years Selous explored and hunted over the country between the Limpopo and the Congo basin, collecting specimens of many kinds. In 1890 he entered the service of the British South Africa company and was active in the successful pioneer expedition to Mashonaland and in the arrangements whereby the district of Manica, near the Portuguese frontier, was brought under British control.

Selous summarized his travels in "Twenty Years in Zambesia," *Geogr. J.*, i (1893). His account of the Matabele War, in which he was wounded in 1893, was entitled *Sunshine and Storm in Rhodesia* (1896). Selous then settled in England but made hunting expeditions to many parts of the world. He was killed in action near Kisaki in Tanganyika on Jan. 4, 1917, during World War I.

See J. G. Millais, *Life of Frederick Courteney Selous* (1918). (R. L. Hl.)

SELWYN, GEORGE AUGUSTUS (1809–1878), Anglican bishop of New Zealand, in whose memory Selwyn College, Cambridge, was built by public subscription, was born in Hampstead on April 5, 1809, and educated at Eton and St. John's College, Cambridge. In 1841 he was appointed first bishop of New Zealand, then just declared to be under the British crown. A man of great heart and physical strength, he acquired the Maori language on the voyage from England and soon won the confidence and affection of the Maori people. The first Maori War made him unpopular with both parties, since he became suspect to the Maoris as an Englishman and was hated by many settlers as an upholder of native rights. But both sides came in the end to respect his courage and integrity and he had much influence in creating a New Zealand with room for both races. In 1867 he accepted the see of Lichfield, where he died on April 11, 1878.

See Lives by H. W. Tucker, 2 vol. (1879), G. H. Curteis (1889), and J. H. Evans (1964). (W. O. C.)

SEMANG, a nomadic Negrito hill people of the Malay Peninsula (Thailand and Malaysia); trustworthy population data were unavailable in the 1960s (best estimates put their total number at less than 3,000, with only about 100 in Thailand). Wandering in small bands, and living on game and forest products, they use the blowgun (*q.v.*) with poison darts. Men wear a waistcloth and women a wraparound (skirt) of bark beaten with a mallet. Scarification is practised by wounding the skin with the edge of a sugarcane leaf and rubbing in charcoal powder. They play bamboo instruments, the Jew's harp, and the nose flute; the dead are buried with food and drink in simple graves. Leaf-covered dwellings have a framework of saplings with no side walls; sleeping mats are laid on a floor of small crossed logs. The Semang (called Ngo in Thailand) have no tribal organization. The basic group, which includes an elder male with his immediate family (and perhaps married sons and daughters), breaks up when whim or necessity dictate. They speak one of the Austroasiatic languages (*q.v.*), which (along with their surprisingly well-developed religion) suggests outside influences. *See* also NEGRITO; MALAYSIA.

BIBLIOGRAPHY.—P. Schebesta, *Among the Forest Dwarfs of the Malay Peninsula* (1927), *Die Negrito Asiens*, 2 vol. (1952–57); I. H. N. Evans, *The Negritos of Malaya* (1937); F.-C. Cole, *Peoples of Malaysia* (1945); G. H. Lisitzky, *Four Ways of Being Human: an Introduction to Anthropology* (1956); F. M. Lebar, *Ethnic Groups of Mainland Southeast Asia* (1964). (J. H. H.; F. C. Ce.)

SEMANTICS, GENERAL, a doctrine and educational discipline developed by Alfred Korzybski (1879–1950), a Polish scholar and engineer who lived in the U.S. from 1917. The term "semantics" was used to indicate concern with the many ways in which meanings of words and other symbols influence the responses of human beings to their environment and to each other (called "evaluation" by Korzybski and his followers). The adjective "general" was added to avoid confusion with other logical and linguistic studies called semantics (*see* SEMANTICS IN LINGUISTICS and SEMANTICS IN LOGIC).

In his chief book, *Science and Sanity* (1933; 3rd ed., 1948), Korzybski argued that prevailing habits of thought have lagged far behind the linguistic and logical assumptions of modern science. He drew upon relativity theory, quantum mechanics, colloidal chemistry, neurology and mathematical logic as evidence of a revolution against so-called "Aristotelian evaluative habits." Current thought and practice in nonscientific situations were held by him to be seriously hindered by defective habits traced back to Aristotle's logic and its influence on western culture. The alleged Aristotelian assumptions attacked include: (1) the claim that subject and predicate are identical (the use of "the 'is' of identity"); (2) the exclusion of any but the two truth-values of truth and falsity (as shown by the law of excluded middle); and (3) the law of contradiction (nothing is both *A* and not-*A*).

Modern science is held to have shown that these assumptions represent at best oversimplifications of the true structure of reality. The resulting "evaluative habits," resulting from an "Aristotelian orientation," are marked by dogmatism, rigidity and lack of emotional balance. These defects show themselves in confusion between symbols and the realities for which they ought to stand, inattention to the limitations of abstraction, willingness to make sharp "either-or" distinctions and uncontrolled, "trigger" responses to stimuli. The condition of disorientation thus produced deserves to be called "un-sanity," and calls urgently for therapy.

On the theoretical level, Korzybski elaborated a complex doctrine that he called "non-Aristotelian" (abbreviated to the symbol "Ā" by general semanticists). Much emphasis was laid, however, upon practical training in better evaluative habits. As simple but useful aids to the learning of such habits, he advised the use of "indexing" (so that one speaks of "man_1," "man_2," etc., to distinguish different senses), "dating" (*e.g.*, "Roosevelt_{1930}," "Roosevelt_{1940}") and accompanying all statements by an implicit "et cetera" (shown by the symbol ".,") to indicate awareness of factors omitted in any process of abstraction. To facilitate "awareness of abstracting," he also invented an ingenious model known as the "Structural Differential."

An Institute of General Semantics (Lakeville, Conn.) was established in 1938 to give training in Korzybski's methods. The International Society for General Semantics established a quarterly journal *ETC*. Numerous attempts have been made to apply general semantics to the teaching of English, speech correction, psychotherapy and many other fields. The enthusiasm of many devoted adherents of general semantics is not shared by most academic scholars in related disciplines.

See I. J. Lee, "General Semantics 1952," *ETC* (winter, 1952), which contains useful references, and M. Black, *Language and Philosophy*, ch. x (1949). (M. Bk.)

SEMANTICS IN LINGUISTICS. The word "semantics" as a noun is a modern word, based on the Greek adjective *semantikos* ("significant") which was used by Aristotle in a linguistic sense in *On Interpretation*. The first known use of the word as a noun appeared in a review of Arsène Darmesteter's book *La Vie des mots étudiée dans leurs significations* (Paris, 1887) by Gaston Paris in the *Journal des savants*, p. 65 (Feb. 1887), but the expression used by Paris ("this part of linguistics, so subtle and yet so new, that is called semantics") shows clearly that the word had been used before. From French the word entered quickly into other languages.

The word "semasiology," however, had already been proposed by the German Latinist Karl Reisig in his *Vorlesungen über lateinische Sprachwissenschaft*, published in Leipzig in 1839 (reprinted in Berlin, 1881–90). A kind of study of lexicological semantics was written by the Neoplatonist Proclus in the 5th century in his commentary on Plato's *Cratylus*. The philosophical dispute between the exponents of realism and idealism (*qq.v.*) has gone on through the ages. In modern times Lady Welby (*Significs*), George Boole, C. S. Peirce and their successors (Bertrand Russell, Rudolf Carnap, and others) have pursued the problem of meaning with great acumen and originality. More recent is the view that there is a close connection between probability and

meaning, intimately related to variation in language, which is a constant process, and to selection, which repairs the ravages of variation. Selective variation in language tends to maintain an equilibrium in meaning as in all other linguistic features. A word used with great frequency is low in semantic value, important as it may be functionally or grammatically; content of meaning is enhanced by relative rareness, and affectivity is higher in words upon the use of which restrictions are placed by context or even by extralinguistic factors such as taboo. Meaning withal "means" something, namely intelligibility or understanding on the part both of speaker (or writer) and hearer (or reader) and this is perhaps the simplest way of describing it, as "linguistic goal-directed activity." The analogy of modern communication theory and even of servomechanisms and computing machines has pointed to and emphasized this view of what "meaning" is, and of how it is conveyed symbolically.

In the 19th century neogrammarians (*q.v.*) saw language in two aspects, a "psychological" and a "phonetic." The phonetic was considered as purely physiological, and phonetic changes, then labeled phonetic laws, were seen to be highly regular. Therefore the word "semantic" was at first opposed to the word "phonetic" as the signifying part of language, in opposition to the nonsignifying part—in the individual speech-sound. But the work of the Prague school of linguistics, since about 1920, has shown that each sound in itself—or more exactly each class of sound, *t, d, p, b,* etc.—is not indeed significant but distinctive, *i.e.,* a pair *b : p* distinguishes two meanings. If "bit" and "pit" are two different words in English, the difference in meaning hinges obviously on the difference between *b* and *p;* this difference, called opposition, is therefore distinctive. There are languages, such as the Swiss-German dialects, in which *p* and *b* are mere nondistinctive variations of the same phoneme. Some differences (technically, oppositions) that are such in English are not so in other languages. English distinguishes a velar ŋ (as in "sing," "wing") and a dental *n* (as in "sin," "win"); Italian or Spanish, although having the sounds ŋ and *n,* for example, in Italian *unto* and *ungo,* do not attach any really linguistic significance to this difference, which is mechanically determined by the adjoining sounds *t* and *g.* A distinction between semantic and phonetic cannot then be sharply maintained. The whole of language is a unity, and a nonsignificant part is not to be separated from the remainder.

The definition of semantics is then the study of the connection between a linguistic feature and the corresponding mental process or symbolism in the act of speaking. In English the form "I love" represents action in the present, "I loved" in the past, but the meaning of "I went" with respect to "I go" is identical to that of "I loved" with respect to "I love." The same holds true for "boy," "boys"; "mouse," "mice"; "ox," "oxen"; and for "clear," "clearer"; "good," "better." English expresses the comparative or the plural or the past in different ways, for grammatical, not semantic reasons. There is no necessary invariant correspondence between form and meaning in language, which is a systematic symbolism, the meaning of which depends almost entirely upon a conventional tie between a symbol and that which it symbolizes.

Thus language tends to eliminate homonyms when homonymy becomes destructive to mutual understanding. In those regions of France where Latin *molere* "to grind," and *mulgere* "to milk," came to coincide, *mulgere* was replaced by *trahere* (*traire*), "to draw"; and where the names of the "cat" and the "dog" or the "cat" and the "rooster" came to have the same form, one of the two (or sometimes both) was replaced by another word (*e.g.,* "rooster" by "pheasant").

The connection between language and human mental processes has been conceived in several ways. One school of thought simply ignores all mental processes as unnecessary and confusing. According to this view, language must be studied by itself, without any reference to meaning. Apart from the obvious fact that it cannot be explained how, without some kind of mental process, anything about language can be asserted or understood, or that discussion of abstract terms usually produces significant responses only in other linguistic terms, it is clear that the theory simply "solves" the problem by suppressing the fact that the solution is itself, like language as a whole, a response to linguistic (and mental) stimuli.

A quite different opinion has been defended by the Norwegian scholar A. Sommerfelt, who follows the French sociological school, in particular that of Émile Durkheim. "Language," says Sommerfelt, "is in principle independent of the individuals as such. It is a system of actions which the individual must learn from other older individuals, who are members of the same society as he is." However, he also admits:

> Words have become something more [than actions]: [they become] the means by which society not only acts on the surrounding world, but also conceives this world. . . . Malinowski is therefore quite right when he criticizes the usual definition of language as a means of expressing thoughts. . . . The categories of thinking are based upon language. Logic and grammar are not at all identical; the confusion of the two has damaged, it is well known, the development of language study. However, we think by means of language. In logic and in language, such as we know them in the European civilisation, we find the relationship: subject-predicate. When Aristotle, in his *Logic,* operated with the categories of substance, number, or relation, he could do so because these categories were represented in the Greek language. Our [current popular] logic is based on the categories of the Indo-European languages.

This theory cannot be accepted as it stands. Animals certainly think, in their own way and in their own world, although they do not "speak." It is moreover the daily experience of everyone to think before he speaks, or without speaking at all, and without always putting his thoughts into words; and it is the experience of every speaker that he sometimes has difficulty in putting into words just what he wishes to say. There must be, however, a connection between language and thinking. It is also certain that language to some extent influences a person's way of thinking and his whole conception of the universe, and that language helps tremendously to evolve, develop, and mature his thinking, especially during childhood. It is obvious that a child learns what a dog is, or what a tree or a bird is, by hearing them called "dog," "tree" or "bird." If he did not (and in some cultures he does not), he would not connect a Pekingese with a Dalmatian, or a cherry tree with an oak, or a sparrow with an eagle; or at least he would arrive at such an abstraction much later and with much greater difficulty. It is alleged that modern Americans have a somewhat different concept of the wolf from the Italians. To the Italian the wolf denotes a person who eats too much, to the American a man who pursues women; the two metaphors are far apart, and no Italian who has not lived in the United States would be likely to understand any of the many jokes about the American wolves, or vice versa. And there is hardly any doubt that without language Plato's or Giambattista Vico's or Immanuel Kant's philosophy could never have been elaborated, and the whole intellectual edifice of our modern culture could not have been built up.

Thus the sequence is from percept to concept to verbalization. There is a constant interaction between these three elements and necessarily language partakes of this mutual interaction and reflects it historically, not least in meaning. The fundamental mistake of ancient, medieval and Renaissance thinkers was not to consider that language is, among other things, the expression of human thinking (which of course was right) but to apply grammatical categories as if they were the only true logical categories, complete in themselves, and universal. This can be seen quite clearly even today in our grammatical terminology, which has been preserved from antiquity; *e.g.,* when the word "substantive" obviously derives from "substance." The philosophers who created the term thereby attributed not only to contemporary illiterates—which is already to go much too far—but even to Stone Age men of 5,000 years ago the idea of substance. Nor does this category of substance apply well to modern English or to Old English or to any other language living or dead, nor can it possibly apply, for language is just not built that way. "No language has been formed by an assembly of logicians," as Voltaire once wrote. It may be admitted that iron is a substance, but not freedom or light or beauty. While substances are certainly indicated by means of substantives, substantives are also

used to indicate many ideas that by no stretch of the imagination could be called substances. Likewise, while gender in French or Italian or German obviously indicates sex, it is also applied to the moon, a star, a tree, a stone, a revolution, a country, a bridge, a nose, a mountain and thousands of objects or concepts obviously devoid of natural sex. It was partly this difficulty that prompted Leonard Bloomfield and his followers to deny all connection between language and mental categories and to study language by itself, without any reference to the workings of the mind.

Language has a creative or poetic quality in the formation of philosophical concepts and of scientific constructs. Myth also has played, and still (under the name ideology) plays a part in the development of human thinking and of language. Abstract nouns—so difficult for a "primitive" mind to conceive—are originally often the names of gods or heroes; that is, of concrete, personal creatures, having human shape—legs, arms, head, teeth. They were only much greater, much more powerful than normal men and women, and they possessed to a maximum extent some particular quality that is found among men and women in greater or lesser degree. Aphrodite would be a most beautiful woman, Ares an extremely strong man, Athena immensely wise and so on. It was in part through myth that the human mind bridged the enormous gap from words such as "dog," "wolf," "man," "stone," to words such as "beauty," "generosity," "freedom," "width," "greatness," "virtue," "love," etc. And this helps to explain why such words, contrary to what should logically be expected, are endowed with gender—that is, with sex—in all languages that have kept such a distinction. It is the modern remainder, and reminder, of their original divine nature.

Names of actions—birth, murder, conspiracy, destruction, deliverance—also usually have gender in many languages and have a similar origin as abstracts; the suffixes are frequently the same: *cf.* Eng. "birth," "mirth," "strength." Objects or phenomena which are sexless—stone, water, thunder, star, moon—were also endowed with life for primitive man, as can easily be shown not only by the adoration of these objects or phenomena in ancient religions (Persian, Roman, etc.) but also by the fact that they are endowed with sex, and precisely in significant couples: sun, moon; earth, sky; fire, water. It is interesting to observe that in those languages where the sun and fire are masculine, the moon and water are feminine; but when the moon and water are masculine, then the sun and fire are feminine (*e.g.*, Lat. *sol* m., *luna* f.; but Ger. *die Sonne* f., *der Mond* m.).

The category of the noun, which is unintelligible as long as substances and other logical or philosophical concepts are used, becomes quite clear in the magical realm of poetry, where stones dance and trees sing; nouns indicate persons or objects or phenomena or qualities considered as persons (or deities, which is linguistically the same). But it is not only primitive man who is a poet: in the act of speaking, all are poets. The scholar, coming back from his library, will personify his automobile, just as any primitive man might. In modern English this process is repeated every day. The suffix *-er* was originally used for agents, for persons who are engaged in a certain activity: "love," "lover"; "kill," "killer"; "drive," "driver"; or for means of action, and such names are continually being formed with this suffix: "starter," "steamer," "liner," "sweater," "buzzer," "fighter," "lighter," "diner," "sleeper," "propeller," "silencer," "nutcracker," "breaker," "trailer"; and the same suffix is used more and more for acts or actions: "thriller," "chiller," "shocker," "puzzler," "opener," "starter," "howler," "screamer" to denote an action or event that thrills, chills, shocks, puzzles, that opens, that starts, that simulates howls or screams. English, so to speak, treats actions and events as living things, especially in vivid, colloquial speech, and in slang. The like is true for other suffixes ("rockette," "majorette," "usherette," "kitchenette," "dinette," "stopette").

That adjectives are closely connected with substantives the ancient grammarians knew quite well, since the two words are merely two adjectives qualifying a noun: *nomen substantiuum, nomen adiectiuum*. The *nomen adiectiuum* is the noun that is added (*adiectum*) to the *nomen substantiuum* as an indication of a special quality which the noun possesses in the particular circumstance to which the speaker refers: the white cow as opposed to the black cow or the yellow cow. While the noun, when it indicates a quality, indicates the totality of it, the adjective indicates the quality in a specific and concrete situation and with reference to another noun. Or again it may itself be used as a substantive: "the good," "the true," "the beautiful."

The concept of quality requires a much greater degree of abstraction than that of object or person. It is much easier to conceive of a cow or a stone than to conceive ideas such as big or white or good. The history of philosophy, which is after all the history of human thinking, starts with the philosophies of substance (*i.e.*, objects—water, fire, air, earth) and progresses only with George Berkeley and Kant to the ideas of quality (*i.e.*, philosophically, of sensations: an object has no inner reality, it is only a bundle of sensations; it can be defined only as being white, sweet, tender, etc.). The study of Indo-European grammar confirms perfectly this evolution. Adjectives in Indo-European languages that designate qualities are used freely as nouns. The endings of the adjectives are the same as those of the nouns, and even change of gender, which is considered as peculiar to adjectives, occurs frequently also with nouns: Lat. *lupus, lupa; equus, equa*, like *bonus, bona*. Latin preserves at least two examples in which the development is quite obvious. Latin *uber* means "udder," "breast," and it is certainly an old Indo-European word (*cf.* Eng. "udder," Gr. *outhar*). But in Latin it is also normally used as an adjective (with one ending) with the meaning "productive," "fertile," "abundant"; *uber ager* is a "fertile field"—originally an "udder field"; that is, a field comparable to an udder in fertility. Likewise *uetus* meant originally "year" (Gr. *wetus*, Sanskrit *vatsás* "year"), and was used at first of wine, then plants or animals or humans; *uetus uinum* was a "one-year-old wine," whence later the adjectival meaning "old." In other words, adjectives were once nothing but nouns, and they testify to the metaphorical, *i.e.*, poetic nature of language. The process is reproduced in our modern languages especially by names of colours: a "violet sky" is not a sky made of violets, but a sky which has its colour in common with violets; likewise a "rose opal" is an opal which has a rosy hue. In French and in Italian such adjectives are not declined for gender—in other words, they have still the form of nouns: Fr. *une robe rose, un ciel violet;* It. *un vestito rosa, un nastro viola.*

The article, both definite and indefinite, is a very late acquisition of the Indo-European languages (and, for that matter, of many other languages). Indo-European languages had no article of any kind, and the conservative Baltic and Slavonic languages still have no article (with the isolated exception of Bulgarian). Homeric Greek had no real article; the definite article was therefore a development of classical Greek (the indefinite article arrived even later). Classical Latin had no article either, but spoken Latin certainly had it in the imperial period, and from there it spread to all the Romance languages, and the modern Germanic languages have it (though Gothic still lacked it). This grammatical device, though not indispensable, contributed to the growth both of abstraction ("the good") and of identification ("the stone" as distinguished from "stone" in general).

There was a similar evolution in the verb system. While the verbal system of Indo-European languages was based on aspect (indicating the manner of the action—perfective, imperfective, iterative, unique, inceptive, etc.), in modern languages, especially the western ones, the verb is based on the notion of time (Eng "tense"). It is true that the English or French verb has frequently some aspectual characteristics ("I was going"), but the fact remains that no finite verbal form ("I write," "I wrote," "I shall write") in English, French, Italian or German can be conceived outside time; it must definitely be a present, a past or a future, whereas the aspectual character, rarely predominant, may even be entirely absent. The same trend from aspect to time may be observed in Semitic languages; it is impossible to reconstruct any real tenses for proto-Semitic any more than for Indo-European. This is easy to understand, for the concept of time is weak among primitives, as it is among children: "The category of aspect is more concrete than the category of time, and

in the course of the history of the Indo-European languages, we see aspect losing importance, time gaining it" (Antoine Meillet).

However, this "linguistic" time is, like every linguistic category, an imaginative, subjective quantity. Just as gender is not sex, so linguistic time is neither physical nor mathematical time. Language freely uses the future for the present ("This will be John"), the past for the present ("If I had two dollars . . ."), the present for the past (*praesens historicum*), the present for the future (*praesens pro futuro*), etc.; and these two last uses are so frequent that the Latin grammarians speak of a present present (*praesens praesens*).

Moreover, from a mathematical point of view, past and future are two infinite, equivalent quantities; linguistically they are quite different. In fact, the past is known; the future is unknown. The consequence is that many languages have a past, or even two or three or more, but no future (*e.g.*, Gothic); the past appears long before the future, and it is much more stable and more developed. The future not only appears at a very advanced stage of civilization, but remains shaky and evolves with great difficulty into a real prospective, objective future, equivalent to the past; it remains for a very long time either a potential or a volitive or a debilitive future, for man conceives with difficulty of future events as really happening in the same way that past events have happened; he thinks of them as events he wants to happen ("I will go"), or that must happen ("I shall go") or that may happen ("Tomorrow I may go to New York"). Even in English the future is not on the same level with the past. "I wrote" is a simple, factual form; "I shall write" or "I will write" are compound forms with verbs implying ideas different from those of simply expressing the future time.

Lexicological problems are not essentially different from morphological or syntactical problems. The history of words expressing the ideas of being and becoming furnishes a good example. It is well known that the human mind conceives much earlier and more easily essence rather than change, objects than processes. It has taken thousands of years of philosophizing to prove that the world is in perpetual flux. Ancient, medieval and Renaissance philosophers, with the exception of Heraclitus, investigated what the world is, of what substance or substances it is formed. Modern philosophy is a new philosophy of becoming. Language reflects this philosophical evolution. While the idea of being had been reached early in Indo-European (roots **es-* and **bhu-*, Eng. "I am," "I be"), the idea of becoming was late and therefore expressed in a different way in every single Indo-European group of languages: Lat. *fio*, Gr. *gignomai*, Ger. *ich werde* (akin to Lat. *uertere* "to turn," *cf.* Eng. "He turned red"). Modern Spanish and modern Russian do not have one verb to express such an idea, but use different words, proving that the concept is neither clear nor unified (Spanish *ponerse, hacerse, llegar a ser*, etc.). Even English, besides the usual "to become," has "to grow," "to turn," "to get," "to wax." Verbs meaning "to become" seem to be somewhat unstable: Old English *weorþan* disappeared, and so did Lat. *fio*, being replaced by Fr. *devenir*, It. *divenire*, etc. It is very difficult to express in Spanish the Hegelian concept of *das Werden*, as opposed to *das Sein*.

The study of lexicology shows clearly that language is not entirely the product of economic, technical, social or political and other environmental conditions. It is also the product of man's imagination, like poetry, with which it is closely allied, for both spring from the same source. Almost any word can give a shining example of this obvious truth. The American calls his automobile a car, using a word that the Celts used 3,000 years ago for a heavy, two-wheeled wagon drawn by animals; no amount of changes or technical improvements have been able to change the word. On the other hand, words for child, boy, girl, bad, small, etc., have changed several times in the last centuries, and are still changing ("kid," "bambino," "lad," "fellow," "lass," "gal," "doll," "baby," "stinker," "teeny"), although no new type of girl or boy or child has been imported from abroad or transformed through technical or genetic improvements. But the need for poetic expressiveness prompts the speaker to replace such words constantly, because the old ones soon appear prosaic and colourless, through overuse. There is an inverse ratio between frequency of usage and affectivity.

Language being a poetic creation, it is at first normally concrete, for poetry is by its nature concrete, describing things that can be seen, smelled, and touched. The process of linguistic change, although infinitely varied, is almost inevitably one that goes from the concrete to the abstract, again and again and again. It was noted above that aspect, which is concrete, was replaced progressively by tense, which is more abstract; language always uses concrete tools to express abstract thoughts. In the English "I have seen" the "have" is a purely grammatical tool, without any concrete value of its own; but its ancient concrete, plastic use can still be seen in phrases such as "I have two dollars" or even better "I have the key in my hand" (now being replaced by the addition of another more concrete verb, "I've got the key," "I've got two dollars"). Likewise, abstract words of philosophy or religion may easily be traced to concrete, sensuous expressions: "absolute" is that which is "untied" from something (like a dog); "abstract" is that which is "drawn out of somewhere" (like a loaf from the oven); "concept" is that which is "taken together" (an image reproduced in "comprehend" and now in "to grasp"); idea comes from Greek *widea*, that is, from the root **wid-* "to see" (Eng. "vision," "video," "providence," etc.). To take the verb "to become," discussed above, the English word is a compound of the preverb *by, be-* and the verb "to come"; the German *werden* is a verb once meaning "to turn," preserved in Russian *vertét'* and Lat. *uertere* "to turn," whence Eng. "convert," "pervert," "extrovert" (*cf.* for the meaning Eng. "He turned red"); Spanish *ponerse* is "to put oneself"; Gr. *gignomai* contains the root *gen-*, "to generate," "to engender," denoting a biological phenomenon.

Since language is a poetic creation, it follows that every semantic change may be classified under the categories called rhetorical figures, or figures of speech, under which all the metaphors of the poets were once catalogued. To say that a man is no good, meaning that he is a rascal, is to use a litotes; to say irons for fetters, or glasses for spectacles, is to use a metonymy; to say that a woman dancer is terrific or devastating, or that a man is a lady-killer, is to use a hyperbole—since ladies generally will survive such amorous experiences. All of language is a living metaphor, which is not noticed, precisely because metaphor is so amalgamated into it that it appears natural and inevitable.

See also LANGUAGE; LINGUISTICS.

BIBLIOGRAPHY.—G. B. Vico, *The New Science*, Eng. trans. from the 3rd ed. (1744) by T. G. Bergin and M. H. Fisch (1948); B. Croce, *Aesthetic, as Science of Expression and General Linguistic*, Eng. trans. by D. Ainslie, 3rd ed. (1953); M. J. A. Bréal, *Semantics, Studies in the Science of Meaning*, Eng. trans. by Mrs. H. Cust (1900); J. B. Greenough and G. L. Kittredge, *Words and Their Ways in English Speech* (1920); A. Darmesteter, *La Vie des mots*, 15th ed. (1925); A. Sommerfelt, *La Langue et la société, caractères sociaux d'une langue de type archaïque* (1938); S. Ullmann, *The Principles of Semantics* (1951); H. Kronasser, *Handbuch der Semasiologie, kurze Einführung* (1952); F. Dornseiff, *Der deutsche Wortschatz nach Sachgruppen*, 3rd ed. (1943); J. Whatmough, *Language: A Modern Synthesis* (1956).

(G. U. B.; J. WH.)

SEMANTICS IN LOGIC aims at the building of an abstract theory of the relation between signs and what they mean. *Semiotic*, the study of signs and languages, is divided into three parts. *Pragmatics* studies the way languages are used. In *semantics* we abstract from usage, and are interested only in the relations between signs of a language and their meaning. In *syntax* we further abstract, and consider signs aside from what they mean. Semantics can be further divided: We either study existing languages or aim at an abstract theory. The former study is a branch of empirical science (*see* SEMANTICS IN LINGUISTICS), while the latter (*pure semantics*) belongs to logic.

The semantic structure of an ordinary (or natural) language is hopelessly complex because of the fact that its syntactic structure is not precisely determined and that it employs many ambiguous and vague expressions. Hence pure semantics deals only with fully formalized languages. Such languages aim at a reconstruction of ordinary discourse, trying to free it from the pitfalls of natural languages. Pure semantics in turn aims at a rigorous

theory of the way in which signs in a formalized language have meanings, and at the solution of the many problems raised by the meaning-relations. It is hoped that this theory can also be formalized, and hence pure semantics aims at the establishment of (one or more) formalized semantical systems which are adequate for a comprehensive theory of the meaning-relations and of related concepts. The most important of these semantic concepts are discussed below.

For an explanation of the logical terminology and symbols used in this article *see* Logic.

PARADOXES

The next sentence is true. The previous sentence is false.

If you consider either one of the foregoing sentences, you note that it is a perfectly simple and clear sentence of the English language. And yet the two sentences together constitute a paradox; they are a version of the ancient paradox of the liar. A (declarative) sentence must be either true or false, and cannot—of course—be both. Let us suppose that the first sentence is false. That means that the second sentence is not true. But to say that "The previous sentence is false" is not true is equivalent to the assertion that the previous sentence is not false. This in itself is not a paradox; from the assumption that the first sentence is false we have derived a contradiction, which shows only that our assumption was wrong. Hence *the first sentence is not false.* That implies that it is true, which means that the second sentence is true. But if "The previous sentence is false" is true, then *the first sentence is false.* We have now derived two contradictory results using only the rules of logic.

Another type of semantic paradox is due to Kurt Grelling. Let us classify the adjectives of the English language as to whether they are self-applicable or non-self-applicable. An adjective is self-applicable if it has the property it expresses; *e.g.*, the adjective "short" is self-applicable since it is a short word, but "long" is non-self-applicable since it is not a long word. Every adjective is either self-applicable or non-self-applicable, and cannot—of course—be both. Which is the case for the adjective "non-self-applicable"? Suppose that it is self-applicable. Then it has the property which it expresses, *i.e.*, it is non-self-applicable, contrary to our supposition. Hence *it is non-self-applicable.* This means that it does not have the property it expresses, the property of non-self-applicability. But this is just another way of saying that *it is not non-self-applicable.* We have again arrived at two contradictory results.

Paradoxes are often treated as if they were harmless amusements. But this is far from a correct evaluation of them. The above paradoxes show that English allows the derivation of contradictory conclusions by means of the rules of logic; *i.e.*, that English is inconsistent. Since it can be shown that in an inconsistent system anything at all, true or false, can be proved (this is a consequence of the law of denial of the antecedent; *see* Logic), we have to conclude that ordinary English is a language not suitable for logical arguments. And it is not the case that some simple trick will remove this inconsistency.

The root of the difficulty lies in the fact that English, in common with other natural languages, allows us to formulate its own semantic theory. We can discuss such concepts as being a true sentence (of English) and prove the basic truths about these concepts. While this appears at first to be a highly desirable feature of natural languages, Alfred Tarski showed that any language having this feature is necessarily inconsistent. He has shown for a wide variety of languages that if they enable us to discuss semantic problems about themselves and to prove at least the most basic expected results, then a paradox can be formulated in the language. Indeed, if the reader will re-examine the way we arrived at the two sets of contradictions above, he will find that we have appealed only to the most elementary semantic truths.

We are forced to the conclusion that we must never attempt to formulate the semantic system of a formalized language within the same language; we must employ a second language which is used to talk about the first one. This second language must enable us to express and prove anything that can be expressed or proved in the first language, and in addition it must be able to do more, namely, allow us to formulate a semantic theory of the first language.

The lesson we learn from the semantic paradoxes is that in order to formulate a comprehensive semantic theory for a given language we must employ a second, stronger (or richer) language.

META-LANGUAGE

The second language introduced to talk about a given language is known as the *meta-language,* while the first language is called the *object language.* It is important to note that these are relational terms: It is nonsense to ask whether a given language is a meta-language; we can only say that it is being used as a meta-language of another language at the moment. The meta-language can in turn be studied, in which case it is used as an object language, and the previous object language can be the meta-language of a third language.

When a language is used to formulate a semantic theory for a given object language, we call it a *semantic meta-language.* Thus the lesson of the semantic paradoxes is that a semantic meta-language must always be stronger than its object language. This makes the task of formalized semantics very complex: If we want to formulate a semantic theory of a given language, we employ a second, stronger language. If we want to study the semantic structure of this second language as well, Tarski's result shows that we must use a third language even stronger than the second one. This new language can, in turn, only be studied in a fourth language, etc. We are led to an infinite hierarchy of stronger and stronger languages in which each (after the first) serves as a meta-language of the previous one, and in turn serves as an object language. We cannot escape the conclusion that if the aim of a comprehensive formalized theory of the semantic structure of languages is at all attainable, it can only be achieved by means of infinitely many formalized languages.

The Model Language L.—In the following discussion we will always suppose that we are discussing some given formalized language. Since the discussion will take place in English, English is our meta-language. We have already noted that this is an unsuitable language from the logical point of view, but it is used for the sake of ease of presentation. To make some of the points more concrete, we will employ a simple language **L** as an example of an object language. **L** is a simple applied singulary functional calculus of the first order (*see* Logic), which is designed for tasks like the taking of a census. It has predicates $A, B, C, \ldots$ which express properties of human beings, and individual names $a, b, c, \ldots$ which are names of human beings; *e.g.*, B, M, and R express the properties of being a bachelor, married, and red-haired, respectively, while b, r, and s are names of John Brown, Robin Robinson, and Jim Smith. The simplest sentences of **L** are exemplified by $B(r)$, expressing that Robin Robinson is a bachelor, and $R(s)$, expressing that Jim Smith has red hair. (We will assume that the latter is true and the former is false.) More complex sentences are formed by means of sentence connectives and universal quantification; *e.g.*, $R(b) \cdot \sim B(b)$ expresses that John Brown has red hair and is not a bachelor. (We will assume that as a matter of fact he is the only red-haired non-bachelor.)

While it will be helpful to illustrate various points throughout this article in terms of **L**, we must at all times keep such examples distinct from the general discussion which applies to any formalized language.

AXIOMATIC TREATMENT

Denotation, satisfaction, determination, and truth are some of the fundamental semantic concepts. We will use these to illustrate the axiomatic treatment of semantics.

Denotation is the relation between a constant of the formalized language and the object to which it refers; *e.g.*, an individual constant denotes the individual of which it is a name. In **L**, s denotes Jim Smith, while B denotes those human beings collectively who happen to be bachelors—the class of bachelors. Classes are taken in extension; *i.e.*, we do not distinguish between two classes having the same membership. Were it the case that all bachelors

hate women and all woman-haters are bachelors, then we could also say that B denotes the class of woman-haters.

Satisfaction and Determination are relations between a propositional form and an object. A given object satisfies a propositional form if putting a name of the object in place of the variable of the form turns the form into a true sentence. The form determines the object if the object is the only thing that satisfies the form. Thus, in **L**, Jim Smith satisfies $R(x)$, Robin Robinson satisfies $\sim B(y)$, and $R(x) \cdot \sim B(x)$ determines John Brown.

Truth is a property of sentences. This use of the word "true" agrees with common-sense usage, and must be distinguished from certain other recent uses of the word. We say, *e.g.*, that a certain theory is true when it is only highly probable. In semantics we want to admit the possibility that a highly probable theory may be false and an improbable one may be true. We must also distinguish truth from opinion, no matter what the grounds of belief may be; and we must not identify the concept with that of usefulness. We are using "true" in the time-honoured sense in which Aristotle uses it in the quotation: "To say of what is that it is not, or of what is not that it is, is false; while to say of what is that it is, or of what is not that it is not, is true."

The four concepts are interrelated. For example, we can define the other three in terms of satisfaction: (1) A form *determines* an object if the object satisfies the form, and if it is the only object satisfying the form. (2) Truth is a property of sentences rather than of forms, but it is easy to find a propositional form equivalent to a given sentence. We may, *e.g.*, replace the sentence S by the form $S \cdot [x=x]$. A sentence S is *true* if any object satisfies $S \cdot [x=x]$. (3) For a definition of denotation we must distinguish between various types of constants; *e.g.*, R *denotes* the class of all objects satisfying the form Rx, while r denotes the one object (person) satisfying $r=x$. (These definitions presuppose that the object language has an equality sign.)

Axioms for Truth.—First we must have axioms specifying the conditions under which individual sentences are true. We are indebted to Tarski for giving the first precise formulation. This is perhaps the most frequently misinterpreted result of formalized semantics. On the one hand the conditions have such an appearance of obviousness that we may think them trivial; but it is often the fate of great discoveries that after they have been made we forget that it took thousands of years to find them. On the other hand many philosophers have tried to assign too great a role to these conditions; they have been taken as a definition of truth, which they are not.

What are these conditions? For any given sentence it is easy to give the condition for its being true: $R(s)$ is true if and only if Jim Smith has red hair, $B(r)$ is true if Robin Robinson is a bachelor, and "all swans are white" is true if all swans are white. These conditions are of the form

(I) X is true if and only if Y,

where in place of X we have the name of a sentence, and in place of Y we have a statement (in the meta-language, of course) of what the sentence asserts. Thus, roughly speaking, a sentence is true if and only if what it asserts *is* the case. We have a complete set of conditions if in (I) we have an axiom for each sentence of the object language. Thus (I) is an axiom schema, which is normally infinite. In the exceptional cases where there are only a finite number of sentences in the object language, (I) may be replaced by a conjunction of its instances, which will serve as a definition of truth. But in the usual case, including all non-trivial languages, the list of axioms is infinite and no such conjunction can be formed. Only this case will be discussed from here on.

While (I) determines the truth or falsity of every sentence, even the simplest generalizations do not follow from it. We need

(II) For all sentences **S**, **S** is true if
and only if $\sim$**S** is not true,

as an additional axiom, for example. While for any one sentence we can prove that it is true if and only if its negation is not true, we are unable to derive the generalized result (II) from (I). We know from the schema (I) in the meta-language of **L**; *e.g.*, that $B(r)$ is true if and only if Robin Robinson is a bachelor, and that $\sim B(r)$ is true if Robin Robinson is not a bachelor, hence $B(r)$ is true if and only if $\sim B(r)$ is not true. But we can *prove* that (II) itself does not follow from (I): Let us suppose that it does follow; *i.e.*, there is a proof of (II), given (I). Since a proof contains only a finite number of steps, only a finite number of instances of (I) can be used in the proof. But then (II) must follow from a finite number of instances of (I). Such a finite list says nothing at all about most of the sentences, hence a result about all sentences cannot be a consequence. This proves that we were mistaken in supposing that (II) follows from (I).

Similarly we will have to add

(III) For all pairs of sentences **S** and **S′**,
Sv**S′** is true if and only if **S** is true
or **S′** is true.

Since all the remaining sentence connectives can be defined in terms of $\sim$ and v, we can derive corresponding results for them from (II) and (III). It is also worth noting that some famous "laws of thought" are consequences of these axioms; *e.g.*, if we define a sentence to be *false* if its negation is true—as is customary—then it is an immediate consequence of (II) that every sentence is either true or false. *See* THOUGHT, LAWS OF.

An additional axiom which we may want to introduce is: $(\mathbf{x})\mathbf{Fx}$ is true if and only if every sentence of the form **Fa** is true, where **Fx** is a propositional form, and **Fa** results from **Fx** by replacing the variable **x** of the form by the constant **a**. While this seems to have the same intuitive appeal as (II) and (III), it can only be introduced in exceptional cases. If in **L** every human being has a name, then the axiom is correct. But let us suppose that the individual names of **L** were taken from telephone books, and that P expresses the property of being listed in at least one phone book; then every sentence of the form P**a** is true, but $(x)Px$ is false! In this case, which represents the usual situation, the proposed axiom must be rejected.

These axioms are only the bare beginning of an axiomatic treatment of truth; *e.g.*, in general it will not follow that all theorems of the object language are true; we will need one or more additional axioms to assure this. These axioms can be further strengthened by combining them with axioms of *satisfaction*. From such a set of axioms it is possible to derive the equivalences that make up Tarski's definition (*see* the next section). But a complete treatment is not to be expected, since it can be shown that no matter what list of axioms is laid down, there will be true statements about truth which do not follow from the axioms.

DEFINITION OF SEMANTIC CONCEPTS

Tarski showed that the above semantic concepts can also be introduced by definition in the meta-language. It is convenient to start with a very broad notion of satisfaction. We consider not only forms with one free variable, but we allow any number of free variables. "x is white" is a propositional form with one free variable, "x is between y and z" has three free variables, while a sentence like "all swans are white" may be considered an extreme case of a form with no free variables. Since we allow several free variables, the relation holds not between a single object and a form, but between a list of objects and the form. Since forms are built up by step-by-step procedures, our definition also proceeds in this way.

Let us give the definition for **L**. Given a form **F**, we must define what we mean by having its various parts satisfied by a given list of people. We associate one member of the list with each variable (free or bound) of **F**. The simplest parts of any form are of the type **P**(**a**) or **P**(**x**). **P**(**a**) is satisfied by any list if the person whose name **a** is has the property expressed by **P**, while **P**(**x**) is satisfied by the list if the person associated with the variable **x** has this property. We may require that all more complex forms be formed by means of negation, disjunction, and universal quantification. The negation of a form is satisfied by the list if and only if the form itself is not satisfied. The disjunction of two forms is satisfied if and only if the list satisfies at least one of the

original forms. $(\mathbf{x})\mathbf{G}\mathbf{x}$ is satisfied by the list if and only if every list differing from the given one only in the person associated with $\mathbf{x}$ satisfies $\mathbf{G}\mathbf{x}$. (In this case it makes no difference which person was associated with $\mathbf{x}$ in the given list!) Thus we define step-by-step the satisfaction relation for all parts of **F**, and hence for **F** itself. We can do the same for every propositional form in **L**.

Given the definition of satisfaction, we can define the other concepts in terms of it, as indicated above; *e.g.*, we note that it makes no difference which person was assigned to a bound variable of **F**. Hence if the form happens to be a sentence—which has no free variables—then it is either satisfied by all lists or by none. In the former case it is true, in the latter false. From this definition we can derive all instances of (I), and also such generalizations as (II) and (III). We also see from this that our definition agrees with the intuitive meaning of "true," since only in this case can it yield all the instances of (I).

In most cases these definitions are the best basis for a theory of the concepts discussed so far.

ANALYTIC TRUTH

There is a second group of semantic concepts, related to each other, which are more difficult to define than the previous ones. No entirely satisfactory treatment of these is available, and there is still considerable controversy concerning them. Rudolf Carnap gave us the first systematic approach to semantics as a whole; *i.e.*, to both concepts. Many of his ideas are used in this article.

The central concept in this group is that of analyticity. A sentence is analytically true (false) if it is true (false) purely on the basis of the meaning of its words. Opposed to these are synthetic sentences, whose truth (falsity) depends on information about the physical world. "All white swans are white" and "all bachelors are unmarried" are examples of analytic truths, while "the earth is a planet of the sun" and "Jim Smith has red hair" are synthetic truths. The negations of these sentences are examples of analytically and synthetically false sentences, respectively.

In order to define these concepts we must consider not only the intended interpretation of the symbols, but all logically possible interpretations; *e.g.*, we intend to interpret R as denoting the class of red-haired people and r to denote Robin Robinson; but it is logically possible to let R denote the class of bachelors and r to denote John Brown, as long as no axiom of **L** is violated. It will turn out that some of the constants in the object language have only one possible interpretation (these are *logical constants*), but that the constants used to express facts (*extra-logical constants*) by their very nature have many interpretations. While we have a right to decide that R should express the property of red-hairedness, it takes factual information to determine which class it denotes. As far as logic is concerned, there is no reason why the people who happen to have red hair should be the redheads of the world, it could just as easily have happened that the people who happen to be bachelors fulfill this role. This possibility (and hence the corresponding interpretation) can be eliminated only by an axiom about the membership of these classes; *e.g.*, that John Brown belongs to the class of redheads, but not to the class of bachelors. But this is a factual assertion, and is therefore not suitable as a logical axiom.

In **L** an interpretation tells us (1) which class of human beings each predicate denotes; (2) which human being is denoted by each name; (3) which operation on a propositional form is denoted by negation; (4) by universal quantification; and (5) which operation on a pair of forms is denoted by disjunction. Given this information, we can define satisfaction—as we did above—relative to these meanings of the constants. A form satisfied by all lists is said to be *valid* under this interpretation. If all axioms are valid under the interpretation, if the rules allow us to infer only valid conclusions from valid premisses, and unless the interpretation is trivial in the sense that it makes all sentences valid, we then call the interpretation a *sound* interpretation, or a *model* of **L**. It is possible to show that negation, disjunction, and universal quantification have the same denotation in all models, while the predicates and individual names all have a wide range of possible interpretations.

A sentence that is valid in all models is *analytically true*; a sentence valid in no model is *analytically false* (or *self-contradictory*). An *analytic* sentence is one that is analytically true or analytically false. By the Law of the Excluded Middle, a sentence is either analytically true or not analytically true, but the latter by no means assures us that it is analytically false. In this "middle" lies the vast group of *synthetic* (or *factual*) sentences, which are valid in some but not all models. $R(r) \vee \sim R(r)$ is analytically true; no matter what class R denotes and no matter who r is, he either does or does not belong to this class. But $R(r)$ is factual; it is valid if R denotes the class of redheads and r denotes Jim Smith, and it is not valid if R denotes the class of bachelors and r denotes Robin Robinson.

It is a consequence of the above definitions that all theorems of **L** are analytically true, and that all analytically true sentences are theorems. But the latter does not hold for all formalized languages, due to the impossibility of giving a complete formalization of advanced branches of logic. (*See* LOGIC.) This is the source of many difficulties in the general definition of analyticity. Because of the incompleteness of our axioms certain logical constants admit more than one interpretation, and the definition of analyticity cannot be as simple as for **L**. There is every reason to hope, however, that semantics will overcome these difficulties.

Many concepts can be defined in terms of those just discussed. A few of these are: **S** and **S′** are *logically equivalent* if $\mathbf{S} \equiv \mathbf{S}'$ is analytically true. **S′** is a *logical consequence* of **S** if $\mathbf{S} \supset \mathbf{S}'$ is analytically true. **S** and **S′** are *consistent* if $\mathbf{S} \cdot \mathbf{S}'$ is not analytically false. Two constants are *synonymous* if they are given the same interpretation by every model. A formalized language is *complete* if every analytically true sentence is a theorem. (This last definition must be modified somewhat in languages having restrictions on quantification.)

Criticism of These Concepts.—The above concepts have been severely criticized from various points of view. W. V. Quine has even presented a strong argument to show that no sharp distinction can be made between analytic and synthetic sentences. The arguments of Quine, and of a number of other philosophers, are based on examples from natural languages in which we are unable to say whether a given sentence is analytic or synthetic. It is admitted by these critics that a certain class of sentences, the so-called *logical truths*, are clearly analytically true. Thus there is no objection to sentences like $B(r) \supset B(r)$. But what of a sentence like $B(r) \supset \sim M(r)$? How do we know that this is true purely on the basis of the meaning of the constants? We could interpret B to be the class of redheads, give M its usual meaning, and let r denote John Brown; then the sentence is false.

"All swans are white" is a good test case for these objections. The word "swan" was for a long time applied only to certain heavy-bodied, long-necked, aquatic birds with white plumage. Then a new species was discovered in Australia that was just like these, only the birds were black. The English language could have developed along either of two lines: We could use "swan" in a broader sense from then on, and speak of black swans (which is what has happened), or we could give a new name to the Australian species, say "gwan," and describe gwans as just like swans except in their colour. It certainly takes factual information to know which line was followed, but this type of factual information is no more relevant to the question of analyticity than information about how the axioms were chosen. What is relevant is that if the former line is followed, then "all swans are white" is factual and false; while under the latter procedure whiteness is part of the meaning of "swan," hence the sentence is analytically true. This sentence is analytic if and only if it is true, hence we cannot accept its classification as true or false without accepting its classification as analytic or synthetic.

The lesson to be learned from this example is that we must have complete information about the meaning of all the words in the object language, hence we must have axioms not only for the logical constants, but also for extra-logical constants; *e.g.*, we will have $B(x) \supset_x \sim M(x)$ as an axiom of **L**. Then $B(r) \supset \sim M(r)$ is analytically true, because any interpretation under which it is not valid fails to make one of the axioms valid, and hence

is not a model. Axioms which restrict the usage of extra-logical constants are called *meaning postulates*. It can be shown that the examples constructed to demonstrate that the concept of analyticity is unsound show rather that there is some question about the meaning postulates; *i.e.*, of the way certain words are used. This is not at all surprising since the examples are always chosen from a natural language. These difficulties are resolved by the fact that a language is not considered fully formalized unless all the axioms, including the meaning postulates, are explicitly stated.

TWO KINDS OF MEANING

We are indebted to J. S. Mill and even more to Gottlob Frege for showing clearly that names have two kinds of meaning. On the one hand we could take the meaning of R to be the object to which it refers, the class of red-haired people; on the other hand we could take as the meaning the idea which it calls to our minds, namely the concept of red-hairedness. The former is the *denotation* of R, the latter is its *sense* (or *connotation*). (*See* MEANING; CONNOTATION; DENOTATION.) Let R' denote the class of all those red-haired people who have no more than 1,000,000,000 hairs. Since it so happens that no one has more than 1,000,000,000 hairs, R and R' have the same denotation. But since their definitions are not logically equivalent, they have different senses. The denotation of R is a class of human beings; and its sense is a concept of this class. The sense of a term determines its denotation, but the same object may have many different concepts.

This distinction is closely related to the distinction between analytic and synthetic truth. If in an analytically true sentence we replace a term by another having the same sense, the result is also analytically true. If we substitute a term with the same denotation but a different sense, the result must again be true, but it need not be analytically true. This is due to the fact that while the sense uniquely determines the denotation, it takes factual knowledge to find out what this denotation is. Hence in order to establish the truth of the resulting sentence we may need factual information concerning the identity of the two denotations. "Woman" and "adult female human being" have the same sense, while "woman under ten feet tall" has the same denotation but a different sense. "All women are women" is analytically true, and so is "all women are adult female human beings"; but while "all women are women under ten feet tall" is true, we must have factual information to establish its truth.

Individual names also have two kinds of meaning, which is more easily seen for descriptions than for proper names. "The 32nd president of the U.S.A." and "the first person to be elected president of the U.S.A. four times" both denote Franklin Delano Roosevelt, but they have different senses as can be seen from the fact that we must know some American history to know that they have the same denotation. The distinction can even be extended to sentences: A sentence is said to denote its truth-value (*i.e.*, truth or falsity), and to have as its sense the proposition which it expresses. (In this usage propositions are both abstract and objective.)

Any theory of meaning must overcome certain difficulties raised by descriptions. "The so-and-so" has a clear-cut denotation only if there is one and only one so-and-so. What is the denotation, *e.g.*, of "the king of England in 1953"? And how are sentences containing such descriptions to be classified as true or false? A second type of difficulty is raised by sentences like "I know that Franklin Delano Roosevelt was the 32nd president of the U.S.A." This sentence differs from "I know that Franklin Delano Roosevelt was Franklin Delano Roosevelt" only in the interchanging of two terms having the same denotation, hence the former must be true, since the latter is. But if I am sufficiently ignorant the former sentence is false.

Bertrand Russell suggested a way of overcoming these difficulties by a new analysis of descriptions. He asserted that a sentence containing a description is a condensation of a more complex sentence; *e.g.*, "Franklin Delano Roosevelt was the 32nd president of the U.S.A." is an abbreviation of "there was one and only one person who was the 32nd president of the U.S.A., and Franklin Delano Roosevelt was this person." Under this analysis all sentences of the form "the so-and-so is . . ." are false if there is no so-and-so or if there is more than one. Thus the sentences "the king of England in 1953 was tall" and "the king of England in 1953 was not tall" are not contradictories, but are both false. According to Russell, descriptions do not function as names, and there is no need to assign a denotation to them. And if we rewrite the first "I know . . ." sentence according to Russell's analysis, we see that it has an entirely different structure from the second one.

Though this approach overcomes the difficulties without using the Frege distinction, it raises certain problems of its own which have induced other investigators to return to Frege's approach. The difficulty of the denotation of "the king of England in 1953" is overcome by Frege by assigning denotations to such descriptions by some arbitrary convention. The two major systems that employ this procedure differ as to the way in which they treat the second difficulty. Carnap maintains a distinction between the *extension* and *intension* of terms which is very similar to the distinction between denotation and sense. But Carnap's rule of substitution (of a term for another having the same extension) is restricted so that it cannot be applied in contexts like "I know. . . ." Alonzo Church worked out a suggestion of Frege's according to which names in contexts like "I know . . ." are not used in the customary manner, but *obliquely*. This means that they are used not as names of their denotation, as usual, but as names of their sense. Thus the sentence above is not about the person F. D. Roosevelt, but about two different concepts of this person.

While this analysis seems like the most natural one, it obligates us to build a formalized language in which a name is not used with different denotations at different times. This means that we need a name for the sense of each name. And then, of course, we need a name for the sense of the new name, etc. This leads to a very complex hierarchy of names and concepts. Such an *intensional* system was formulated by Church, and to anyone who accepts the Frege analysis this system represents as much progress over the usual (extensional) formalized languages, as those represent over unformalized languages. But, because of the very great complexity of the Church system, it remains to be seen whether it can stand the test of detailed scrutiny.

Abstract Entities.—The problem of treating abstract entities, like the sense of a name, is the most controversial topic in semantics. Many logicians have strong philosophical objections to the treatment of concepts as in any sense on a par with physical objects. Quine in particular has devoted considerable effort to a reconstruction of logic along nominalist lines (*see* NOMINALISM), trying to rid the foundations of what are usually called Platonic doctrines. So far the nominalist systems have only been able to cope with a small part of logic. On the other hand, the type of Platonist system advocated by Church or Carnap runs the serious risk of being inconsistent. It remains to be seen whether nominalist or Platonist systems prove more successful in the formalization of semantics.

BIBLIOGRAPHY.—Gottlob Frege, "Über Sinn und Bedeutung," *Zeitschrift für Philosophie und Philosophische Kritik*, vol. c (1892; a translation into English may be found in Feigl and Sellars collection mentioned below); Bertrand Russell, "On Denoting," *Mind*, vol. xiv (1905; reprinted in Feigl and Sellars collection mentioned below); Alfred Tarski, "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, vol. i (1936); Rudolf Carnap, *Introduction to Semantics* (1942), *Meaning and Necessity* (1947); C. I. Lewis, *An Analysis of Knowledge and Valuation* (1946); John G. Kemeny, "Models of Logical Systems," *The Journal of Symbolic Logic*, vol. xiii (1948); J. C. C. McKinsey, "A New Definition of Truth," *Synthèse*, vol. vii (1948–49); Alonzo Church, "A Formulation of the Logic of Sense and Denotation" in Paul Henle *et al.* (eds.), *Structure, Method and Meaning; Essays in Honor of Henry M. Sheffer* (1951); W. V. Quine, Max Black and Alonzo Church, part 3 of "Contributions to the Analysis and Synthesis of Knowledge," *Proc. Amer. Acad. Arts Sci.*, vol. lxxx (1951); W. V. Quine, *From a Logical Point of View* (1953); H. Feigl and W. Sellars (eds.), *Readings in Philosophical Analysis*, part 1 (1949); L. Linsky (ed.), *Semantics and the Philosophy of Language* (1952). (J. G. Ky.)

SEMAPHORE, a method of signaling by means of flags, lights or arms. Before the invention of the telegraph, semaphore signaling from high towers was used to transmit messages between

distant points. Movable arms, or rows of lights simulating arms, displayed from towers and used to signal railroad trains are modern semaphores. Semaphore signaling between ships is accomplished by men who hold a small flag in each hand and, with arms extended, move them to different angles to indicate letters of the alphabet. While such signaling has been largely superseded by radio it will probably continue in use at times when radio silence is necessary.

See SIGNAL COMMUNICATION, MILITARY. (M. R. D.)

SEMARANG, capital of Central Java province and Semarang residency and one of the chief towns and ports of Java, Indon., is a centre of commerce situated almost centrally on the north coast, 250 mi. (402 km.) E of Jakarta. Pop. (1961) 503,153.

The town, a short distance inland from the port, lies on the banks of the Kali Semarang, above where the Kali Baru branches from it. Like Jakarta, it is divided into two parts: the old native town near the coast, thickly housed, with narrow streets but with some good shops, hotels, business houses, and churches (in former times it was surrounded by a moat, with forts); and a newer town, farther inland, where there are various government and public buildings, including a military hospital, and also hotels, restaurants, and in the town square a mosque; the square also has recreation grounds. The two districts are connected by the fine Bojong road. South of this road is a picturesque residential hill suburb (Tjandi), 500 ft. (150 m.) above the sea. The town is linked by rail with Surakarta, Jogjakarta, Madiun, and Surabaja, and by steam streetcar with Tjirebon, where connection is made with the railway to Bandung and Jakarta. Of interest in the vicinity of Semarang are the towns of Demak and Salatiga.

Although Semarang is the third port in Java, its harbour, at the mouth of the Kali Baru, is unprotected against the northwest monsoon, which can cause suspension of port operations. Vessels anchor about 3 mi. (5 km.) out, and for the protection of the lighters which take their cargo, there is a harbour, the western pierhead of which extends 5,249 ft. (1,600 m.) into the sea, to prevent the mouth from silting up. This harbour has two customhouse basins, and there is a fishing vessel harbour. The port is connected by rail and road with the hinterland, and by cable with Jakarta, Surabaja, and Balikpapan in Kalimantan (Indonesian Borneo). The river between the town and the sea is canalized for traffic. Industries include the manufacture of machinery and textiles. Wooden lighters, motor boats, and tugs are built at a yard southwest of the port. Fishing is also an important industry. Rubber, coffee, sugar, and other agricultural products are exported. During World War II Semarang was occupied by the Japanese from 1942 to 1945.

SEMELE, in Greek mythology, daughter of Cadmus and Harmonia, and mother of Dionysus by Zeus. *See* DIONYSUS.

SEMENOV, NIKOLAI NIKOLAEVICH (1896–), Russian physical chemist, who was awarded the 1956 Nobel Prize in Chemistry, jointly with Sir Cyril Hinshelwood, for work on the kinetics of chemical reactions, was born at Saratov on April 15, 1896. He graduated from Petrograd University in 1917 and worked in the Leningrad Physical Technical Institute (1920–31) where he became professor in 1928. In the following year he became corresponding member, and in 1931 full member, of the Academy of Sciences of the U.S.S.R. Semenov later became director of the Institute of Chemical Physics in Moscow.

Semenov had long been known to chemists for his outstanding work on the mechanism of chemical transformations. In his book *Chemical Kinetics and Chain Reactions* (published in England in 1935) he gives an exhaustive analysis of the applications of the chain-theory to varied reactions, and especially to those involved in combination processes. He also put forward the idea of degenerate branching, in terms of which some of the mysterious phenomena associated with the induction periods of oxidation processes can be understood. Less well known is his work on the propagation of explosive waves, which appeared only in Russian. Semenov was the first Soviet citizen to win a Nobel Prize.

SEMINOLE, an American Indian tribe of Muskogean linguistic stock (*see* MUSKOGEAN INDIANS) formed in the 18th century by settlers from the Creek (*q.v.*) confederacy, primarily speakers of the Creek and Hitchiti languages. Many runaway Negro slaves joined them. In Creek the name means "wild," "nondomesticated." Early in the 1800s the Seminoles occupied former Apalachee and Timucua territory in north Florida. The Seminole wars (1817–18; 1835–42; 1855–58), caused by Seminole resistance to giving up their lands and Negro associates, resulted in the removal of most of the tribe to Indian Territory (later Oklahoma) where they became one of the Five Civilized Tribes (*q.v.*). The Second Seminole War was one of the most costly Indian wars ever fought by the U.S. Army. In the 1960s there were about 3,000 Seminoles in Oklahoma and 1,000 in southern Florida. *See* also FLORIDA: *History;* OKLAHOMA: *History;* OSCEOLA.

See J. R. Giddings, *The Exiles of Florida* (1964). (W. C. ST.)

SEMIPALATINSK, an *oblast* of the Kazakh Soviet Socialist Republic, U.S.S.R., was formed in 1939. Area 67,343 sq.mi (179,599 sq.km.). It lies just west of the Altai Mountain system and is drained by the Irtysh River in the north and by the Ayaguz in the south. These two basins are separated by the Tarbagatay and Chingiz-Tau ranges which cross the *oblast* from east to west. The climate is continental with a January average of −16° C (3.2° F) in the north and −12° (10.6°) in the south. The July average varies from 20° (68°) to 23° (74°). Parts of the large lakes of Sasykkol', Alakol', and Zaysan (fresh water) are within the boundaries of the *oblast*. Part of Zaysan Lake forms the Ust'-Kamenogorsk water reservoir.

The population, 520,299 in 1959, consists mainly of Kazakhs and Russians, but there are also Ukrainians, Tatars, and others. The *oblast* is divided into 13 *rayons*, 2 towns, and 9 settlements of town type. The two towns are Semipalatinsk (*q.v.*), the *oblast* centre, and Ayaguz, 165 mi. (266 km.) S of Semipalatinsk.

The economy is mainly agricultural, with a large number of collective and state farms. Of the total sown area in the early 1960s more than three-quarters was under grain and nearly one-fifth under fodder with small areas under technical crops (mainly sunflower) and vegetables. Livestock consists mainly of horned cattle and sheep with a smaller number of pigs, goats, and horses. There are valuable deposits of lime, graphite, gold, silver, copper, and manganese. The *oblast* is traversed by 350 mi. (563 km.) of the Turksib (Turkistan-Siberian) railway. A main road from Ayaguz runs southeast to the Chinese border. The Irtysh River is navigable. There are a number of special training and higher educational establishments in the *oblast*. *See* also KAZAKH SOVIET SOCIALIST REPUBLIC. (G. E. WR.)

SEMIPALATINSK, a town and centre of Semipalatinskaya (Semipalatinsk) Oblast' in the Kazakh Soviet Socialist Republic, U.S.S.R., lies on the Irtysh River, 560 mi. (900 km.) NNE of Alma-Ata and 445 mi. (716 km.) SE of Omsk. Pop. (1959) 156,110. The town was founded in 1718 on a site 11 mi. (18 km.) from the present one, to which it was transferred in 1778. The industry is mainly concerned with food processing and production and with building materials. The town has one of the largest meat-packing plants in the U.S.S.R., and there are leather, textile, lumber, and wool-processing works. There are a Russian-Kazakh theatre and two museums. Semipalatinsk is a junction on the Turksib railway to Arys', and a branch line links the town with the South Siberian system at Kulunda and with the Trans-Siberian at Tatarsk. (G. E. WR.)

SEMI-PELAGIANISM is a term introduced into theological discussions of the 17th century, designating the doctrine of an anti-Augustinian movement that flourished from about 429 to about 529 in southern France. As the name suggests, the spokesmen of the movement did not subscribe to pure Pelagianism but defended a position congenial to Pelagian principles (*see* PELAGIUS). Unlike Pelagians, the Semi-Pelagians believed in the universality of original sin as a corruptive force in man. They also believed that without God's grace this corruptive force could not be overcome, hence They admitted the necessity of grace for Christian life and action. But also insisted on the necessity of baptism even for infants. But against Augustine they taught that the innate corruption of man was not so great that the initiative toward Christian commitment was beyond the powers of man's native will.

This commitment was called by John Cassian *initium fidei* and

by Faustus of Riez *credulitatis affectus*. In this view man by his unaided will could desire to accept the gospel of salvation, but he could not be actually converted without help from on high. At least in later Semi-Pelagianism, divine help was conceived not as an internal empowering graciously infused by God into man but as purely external preaching or the biblical communication of the gospel, of the divine promises, and of the divine threats. The strong point for all Semi-Pelagians was the justice of God: God would not be just if man were not natively empowered to make at least the first step toward salvation; if salvation depended initially and unilaterally on God's free election of the saved, those not chosen could rightly complain that they were doomed by the mere fact of being born. The fathers of Semi-Pelagianism were monks who stressed the need of ascetical action. For them, the doctrine of Augustine made for quietism and passivity; it fostered sloth.

The Semi-Pelagian Writers.—The surviving evidences of the original movement are scanty. Yet it is clear enough that Semi-Pelagian doctrine was widespread in southern Gaul in the 5th century. The champions of the system were holy men and highly respected leaders in the church. No more than five or six can be named, and of these only three were positive influences in the history of this thought. The first was John Cassian (*see* CASSIANUS, JOHANNES), a monk formed in the East, who founded two monasteries in Marseilles. It was in his *Collations of the Fathers* (especially *Collation* XIII) that he formally proposed the earlier Semi-Pelagian theory. Because of Cassian the term *initium fidei* (initiative toward Christian commitment) became a Semi-Pelagian label. (The term was used in much the same sense before in Augustinian circles in Africa and is probably taken from the Vulgate Version of Ecclus. 25: 16 [11], or possibly from S. of Sol. 4:8 in some non-Vulgate translation of the Septuagint.)

The second Semi-Pelagian writer was Vincent, a monk of the celebrated Abbey of Lérins (*see* VINCENT, SAINT, of Lérins), who wrote, in opposition to the grace doctrine of Augustine, a work now lost, though its substantial content can be seen in the answer of Prosper of Aquitaine, *ad Obiectiones Vincentianas*. Vincent's important work, *Commonitorium*, probably was an indirect attack on Augustine, making him out to be an innovator. Vincent does not seem to have influenced Semi-Pelagianism beyond writing in its favour.

The third Semi-Pelagian theologian of importance was Faustus, bishop of Riez (*see* FAUSTUS, SAINT, of Riez), a former monk and abbot at Lérins. At the request of the Provence bishops he wrote *De gratia*, in which Semi-Pelagianism was given its final form, more naturalistic than in Cassian. Notwithstanding Faustus' insistence that he is not a Pelagian, and the fact that he uses orthodox language, his doctrine is close to Pelagianism. For Faustus, God cannot interfere with man's freedom before or after Christian conversion, and freedom itself is an inclination to virtue and faith. In this sense all faith is rooted in grace because human freedom is a grace.

The result of Semi-Pelagianism was the denial of the necessity of God's unmerited, supernatural, gracious empowering of man's will for saving action. It contradicted St. Paul and St. Augustine, the latter by papal declaration the approved Catholic doctor in the question of grace and so beyond attack.

Opposition and Condemnation.—In its early stages Semi-Pelagianism was opposed in Gaul by two polemicists, Prosper (*q.v.*) of Aquitaine and an otherwise unknown Hilary. Augustine wrote two treatises for his defenders in Gaul: *On the Predestination of the Saints* and *On the Gift of Perseverance*. But opposition did not in any way convert the Semi-Pelagians. The champions of Augustine, as Goth declared, were the victims of attack and persecution. What happened to Hilary is not known, but Prosper soon left Gaul to work at the Papal Curia; the result was that no one of stature led the Augustinian forces, which were numerically weak and of no prestige. At the time of Faustus, the symbol for Augustinianism was a certain priest, Lucidus, known only through the writings of Faustus; the latter wrote his book on grace against this obscure man, who could be safely attacked while Augustine could not. After Faustus' death, Semi-Pelagianism was still in high respect as may be seen in the biographical sketches of Gennadius, in which all Semi-Pelagians are praised and Augustine along with Prosper is disparaged.

Nevertheless the 6th century saw the decline of Semi-Pelagianism, mainly through the action of Caesarius of Arles (*see* CAESARIUS, SAINT, of Arles). Though Caesarius had been a monk at Lérins, his theology was not at all in line with the Lérins tradition, and at the instigation of Pope Felix IV (III) (526–530) he condemned Semi-Pelagianism at the so-called second Council of Orange (529). Caesarius presented the papal propositions with some of his own to a few bishops, priests, and important laymen gathered at the consecration of a church in Orange. They signed the document and it was sent to Rome for approval. In the Orange decrees the very words of Cassian and Faustus are used in the condemnation of their doctrines, but their names are not given. When the missive arrived at Rome, Pope Felix's successor, Boniface II, a friend of Caesarius, solemnly approved the conciliar decrees.

From that point on, Semi-Pelagianism was recognized as a heresy in the church. In Gaul, however, it was not entirely dead. Either before or shortly after the second Council of Orange, there was held at Valence another council which, according to the biographer of Caesarius, included bishops unfriendly to Caesarius and to his theology. The acts of this synod have not survived and it is not possible to say how far it favoured Semi-Pelagianism. In any case, the position of Caesarius with its papal confirmation was impregnable, and by the middle of the 6th century Semi-Pelagianism was no more. Future resurrections of it were renewals of Semi-Pelagian tendencies rather than the reproduction of Semi-Pelagian doctrine.

BIBLIOGRAPHY.—F. Wörter, *Zur Dogmengeschichte des Semipelagianismus* (1899); C. F. Arnold, *Caesarius von Arelate* (1894); G. Weigel, *Faustus of Riez* (1938); G. J. Beck, *The Pastoral Care of Souls in South East France During the 6th Century* (1950).

(G. WL.)

SEMIRAMIS, the Greek name of an Assyrian queen who was not only a great figure in her own age but later became the heroine of legendary exploits and left reputed monuments of herself in distant lands. The original Semiramis was undoubtedly Sammu-ramat, mother of the Assyrian king Adad-nirari III (reigned 810–782 B.C.). Her memorial stele has been found at Ashur, while an inscription at Kalakh (Nimrud) shows her to have been dominant there after the death of her husband, Shamshi-Adad V (824–810 B.C.). She seems to have been in fact the actual ruler of Assyria during the early years of her son's reign. It is also clear that she was devoted to the Babylonian god Nabu (Nebo), whose worship she substituted at Kalakh for that of its former patron-god. No more than this is known concerning the authentic Sammu-ramat.

The name and reputed acts of Semiramis are first found in Herodotus (i, 184), who ascribed to her certain earthworks, but no more, although he said that a gate of Babylon was named after her. Much more important in Herodotus are the great structures of flood protection and defense at Babylon which he ascribes to another queen, "Nitocris," who lived "five generations" after Semiramis. This name is quite unknown to Babylonian history.

On the other hand, the historian Diodorus Siculus relates the alleged exploits of a supposed Assyrian queen Semiramis, while making no mention of Nitocris. According to this story Semiramis was born of a goddess, and after being married to an Assyrian officer she captivated the king Ninus by her beauty and valour and became his wife. Soon afterward Ninus (himself a mythical king) died, leaving a son, Ninyas, by Semiramis, who herself assumed the power and reigned for many years. In that time she built Babylon and turned to the conquest of distant lands. At length, finding that Ninyas was plotting against her, she abdicated and disappeared.

The figure of Semiramis comprises, therefore, a kernel of attested fact with a later accretion of fabulous history. Some of its features indicate an Ishtar-divinity, with her mythical birth, her warlike and amorous qualities, but it has been observed that Semiramis, Ninus, and Nitocris have absorbed into themselves some of the history and the qualities of certain real characters in late Assyrian history, namely the king Sennacherib (*q.v.*), his wife Naqi'a, mother of his successor, Esarhaddon (*q.v.*), and an-

other remarkable woman of that age, Adda-guppi', the mother of "Labynetus" (Nabonidus; *q.v.*).

BIBLIOGRAPHY.—D. D. Luckenbill, ed. *Ancient Records of Assyria and Babylonia*, vol. i, par. 731, 745 (1927); *Cambridge Ancient History*, vol. iii, 27, 183 ff. (1925); Diodorus Siculus, ii, 1–20, Eng. trans. by C. H. Oldfather, Loeb Classical Library, (1933); H. Lewy, "Nitokris-Naqi'a," in *Journal of Near Eastern Studies*, vol. xi, 264 ff. (1952). (C. J. G.)

SEMITIC LANGUAGES. The "Semitic" languages, so named in 1781 by the German historian A. L. Schlözer because most of the people who spoke them were descended from Shem or Sem (Gen. x–xi), were spoken in Arabia, Mesopotamia, Syria and Palestine, from which they spread, beginning with the 1st millennium B.C., into Ethiopia and later into Egypt and northern Africa. The Semitic languages go back to a "proto-Semitic" language, the general structure of which can be derived from the historically attested features of the various Semitic languages. In all probability, proto-Semitic was at no time a unified language, but had dialectal variants. No single Semitic language can be said to be the representative of the proto-Semitic type. In phonology one language may come the closest to the proto-Semitic type (as is probably the case of epigraphic South Arabic) whereas for certain morphological features other languages may be considered representatives of it.

Since the Semitic languages go back to a common origin the question of the location of the speakers of this proto-Semitic language is of importance. Various regions have been so considered: Kurdistan, Mesopotamia, northern Syria (the country of ancient Amurru), Arabia and Africa. No definite answer, however, can be given to this question. The likeliest regions are those of Arabia and Mesopotamia.

Hamito-Semitic.—Semitic is a part of a larger language group, namely Hamito-Semitic. The language families that belong to Hamito-Semitic are: Semitic, Ancient Egyptian, Berber (spoken in North Africa) and Cushitic (that is, the non-Semitic languages spoken in Ethiopia). The genetic relationship among the various members of Hamito-Semitic is evident in the phonology, morphology and vocabulary. Whereas the relation between the various Semitic languages can be compared with that of, say, the various Germanic or Romance or Slavic languages, Hamito-Semitic would have more or less the role of Indo-European. (*See* AFRICAN LANGUAGES.)

Classification.—The Semitic languages are classified as North Semitic and South Semitic. North Semitic, in turn, is divided into Northeast Semitic with Akkadian as its only representative, and Northwest Semitic which includes Canaanite (Hebrew, Moabite, Phoenician), Ugaritic, Amorite and Aramaic. South Semitic is divided into Southeast Semitic, including South Arabic and Ethiopic, and into Southwest Semitic with Arabic as its representative. Of these languages, Arabic, Modern South Arabic, Hebrew, Ethiopic and Aramaic, to a limited extent, are still spoken. It is interesting to note that nearly all the Semitic languages continued to be employed as literary languages long after they had ceased to be spoken. The general features of Semitic will be discussed after the survey of the various languages.

Akkadian or Assyro-Babylonian.—This is the oldest attested Semitic language. The name Akkadian comes from Akkad, the ancient capital of Mesopotamia. The documents in the language range from 2500 B.C. to the beginning of the Christian era. There are various periods in its development. Old Akkadian is the language of the documents from *c.* 2500 to 1950 B.C. During that period the Akkadians lived side by side with the Sumerians. Toward the end of the Old Akkadian period the language divided into Babylonian and Assyrian dialects. Within the Babylonian dialect one can distinguish the following periods: Old Babylonian (*c.* 1950–1530), Middle Babylonian (*c.* 1530–1000), Neo-Babylonian (*c.* 1000–625) and Late Babylonian (after 625). In the last period Aramaic was the spoken language and Late Babylonian was the literary language. The various linguistic stages of Assyrian are: Old Assyrian (*c.* 1950–1750); Middle Assyrian (*c.* 1500–1000), strongly influenced by Babylonian; Neo-Assyrian (*c.* 1000–600). Inscriptions represent an important documentation of the literature, history, religion, magic, law, science and commerce of Mesopotamia. Because of the cultural prestige of Babylonia, Akkadian was also used in the neighbouring countries, such as Cappadocia, Elam, Canaan (Tell el Amarna letters), the Hittite empire (Bogazkoy) and Mitanni (Nuzi). Akkadian was also for a time the international language of the near east. (*See* AKKADIAN LANGUAGE.)

Canaanite.—This name came from the word Canaan, the ancient name for Palestine, Phoenicia and part of Syria. The languages that are considered Canaanite and known through direct sources are: Hebrew, Moabite and Phoenician. The older stage of Canaanite is known indirectly through the Tell el Amarna letters. (For a description of Hebrew, *see* HEBREW LANGUAGE.)

Moabite.—There is a single inscription in Moabite dating from the 8th century B.C. It is a report of Mesha, king of Moab (to the southeast of the Dead sea), on his relations with the king of Israel.

Phoenician.—This language was current in ancient times in the Phoenician cities of Tyre, Sidon, Byblos and neighbouring towns. Comparatively few inscriptions have been found in Phoenicia itself. The earliest known decipherable inscriptions are those of the kings of Byblos from the 10th century B.C. The bulk of material from Phoenicia proper consists largely of royal stelae dating from the 5th to the 2nd century B.C. Many inscriptions have been found at points along the Mediterranean shores—at the sites of the ancient Phoenician colonies, on Malta, in Carthage, and in other cities of the North African coast, as well as in Karetepe in Anatolia. The linguistic stage of Phoenician of the North African coast is called Punic and dates from the 5th to the 2nd century B.C.

Another important source of information are the transcriptions of Phoenician words into the scripts of other languages such as Egyptian, Akkadian and Hebrew. The most fruitful external source is to be found in classical literature. Phoenician words appear mostly in Greek literature; Punic in both Greek and Latin. Among the earliest are the names of the letters of the alphabet, which the Greeks took over together with the "Phoenician" alphabetic signs, perhaps in the 9th century B.C. A connected Phoenician discourse in transcription is found in *Poenulus* of Plautus (end of 2nd century B.C.). Since the "Phoenician" alphabet does not include vowels, these sources serve especially to give a picture of the vocalization of the language.

Ugaritic.—In Ras Shamra, on the coast of Syria, about seven or eight miles to the north of Latakia, another Semitic language was discovered in 1929. As a result of epigraphical and archaeological work, it soon became clear that the ancient name of the city of Ras Shamra was Ugarit, hence the name Ugaritic for that language. The history of Ugarit ended *c.* 1200 B.C. with the invasion of the Philistines. Ugaritic was written in an alphabetic cuneiform using about 30 simple signs which, on the whole, present single sounds rather than syllables or ideograms. A few tablets in Ugaritic script were also found in Palestine. The bulk of the texts is epical in content, and the language as well as the content throw considerable light on the Hebrew language and literature of the biblical period. The exact position of Ugaritic is not certain, some opinions being that it is "Early Hebrew," "Early Phoenician," "North Canaanite," an early dialect of Canaanite, an independent Northwest dialect or an Amorite dialect.

Amorite.—There is a term "MAR.TU" appearing in Sumerian texts of the 3rd millennium B.C. In the Akkadian texts of the 2nd millennium B.C., it appears as "Amurru." It is not yet definitely established whether at that time the term was to be limited to a "western" region or people or to a specific population element. Outside of Babylonia proper the term also appears in the texts of Mari (modern Tell el Hariri) on the Euphrates; these texts date from the 19th and 18th centuries B.C. In the Tell el Amarna letters of the second half of the 2nd millennium B.C., "Amurru" designates a district of Syria embracing the area of the Lebanon and Anti-Lebanon. The linguistic documentation of Amorite consists of proper nouns (with specific grammatical forms) and also of some words dispersed throughout the various Old Babylonian documents. These proper nouns and their morphological forms are interpreted as belonging to an "Amorite" dialect or, according

to some scholars, an "East Canaanite" dialect.

Aramaic.—The Aramaeans appear in the cuneiform texts of the 14th century as *aḫlamê*, later *aḫlamī armaya*. They extended their conquest from Mesopotamia to Syria-Palestine and northern Arabia. Aramaic superseded the various languages of the conquered countries and beginning with the 8th century B.C. it became the international language of the near east, as well as the official language of the Persian empire. The period of its greatest extension was from *c.* 4th century B.C. to the 7th century A.D. at which time it was supplanted by Arabic. Until the beginning of the Christian era there were no outstanding dialectal variations in the language called Common Aramaic. The language was then divided into West Aramaic and East Aramaic. The documents of Common Aramaic are various in kind. The inscriptions of principalities of Syria, such as the kings of Hamah between Damascus and Aleppo and of the kings of Samal found in Zinjirli to the north of Aleppo, date from the 9th to the 8th century B.C. The inscriptions of the Jewish colony of Elephantine (Aswan) in Egypt date from the 7th to the 4th century B.C. The Aramaic sections of the Bible (Ezra iv, 8–vi, 18; vii, 12–26; Daniel ii, 4–vii, 28, and isolated sentences) date from the 4th to the 2nd century B.C.

The dialects of West Aramaic are: Judeo-Aramaic, Samaritan, Palestinian-Christian, Nabataean, Palmyrene and Western Neo-Aramaic. Judeo-Aramaic is the dialect of certain Aramaic works found among the Dead Sea scrolls; of the Palestinian Targums, *i.e.*, the translations of the Bible into Aramaic; of the Palestinian Talmud compiled in the 5th century A.D.; and of some inscriptions. Samaritan is represented by the translation of the Pentateuch in the 4th century A.D., and by some other prayers and religious works. Palestinian-Christian is the dialect of sections of a translation of the Old and New Testaments, and of some religious texts translated by the Christian Melchites of Palestine in the 8th and 9th centuries A.D. Nabataean is the Aramaic dialect used by Arabs of Arabia Petraea and of Hauran to the east of Palestine, in the inscriptions found on the caravan roads through Sinai, northern Arabia, Transjordania dating from the 1st century B.C. to the 4th century A.D. Palmyrene is the dialect of the inscriptions of Palmyra, to the northeast of Damascus, dating mainly from the first three centuries A.D. Western Aramaic is still spoken in the mountainous regions of the Lebanon and Anti-Lebanon in the villages of Maʿlūla (Christian), and Baḫʿa and Ğubaʿdīn (Muslim).

East Aramaic includes Syriac, the Aramaic of the Babylonian Talmud, Mandaic, and Eastern Neo-Aramaic. Syriac was the language of Edessa (modern Urfa), the centre of Christianity at the end of the 2nd century. Since the 5th century A.D. owing to theological differences, Syriac-speaking Christians have been divided into Nestorians or East Syrians under the Persian sphere of influence, and Jacobites or West Syrians under the Byzantine sphere. These two groups became linguistically distinguished by certain differences in pronunciation. The greatest period for Syriac literature was between the 3rd and 7th centuries. (*See* SYRIAC LANGUAGE.) The Aramaic of the Babylonian Talmud, compiled in the 6th century, is another important dialect of the East Aramaic group. Mandaic is the dialect of a gnostic sect (also called Sabaeans) of lower Mesopotamia. An East Aramaic dialect is still spoken in the regions between Lake Urmia and Lake Van (by Nestorians, sometimes called Assyrians), in the district of Tur Abdin (Jacobites), and in the region north of Mosul. Catholic and Protestant missions have tried to develop the Mosul dialect as a literary language.

South Arabic.—*Epigraphic South Arabic.*—This language is represented by about 5,000 stone inscriptions found in the region of modern Yemen, Hadhramaut, and the Aden Protectorate. Since the tribe of Himyar of southern Arabia gained some importance before the Islamic period, the name "Himyarite" is also used for Epigraphic South Arabic. There are two main dialects in South Arabic: Sabaean and Minaean. The dialects of Qatabanite-Awsanite and of Hadhramaut are related to Minaean. The main difference between Minaean and Sabaean consists in the usage of the morphemes of the causative and of the independent and suffixed pronouns of the third person. These elements are *h* in the Sabaean group, and *s* in the Minaean group. Minaean inscriptions were also found in the northern Hejaz, in the neighbourhood of El ʿÖla and Tebuk. The Sabaean inscriptions, also discovered in the region of Aksum and of Yeha in Ethiopia, help reconstruct the political, cultural and religious history of the ancient kingdoms of "Arabia Felix," the most important being Maʿin, Saba, Qataban and Hadhramaut. The most ancient Minaean inscriptions are probably from the 8th century B.C., the Sabaean later.

Modern South Arabic.—There is a non-Arabic dialect cluster spoken in certain regions of southern Arabia between Hadhramaut and Oman. Conventionally called "Modern South Arabic," these dialects are: Mahri (spoken in Mahra); Botahari and Harsusi (close to Mahri); Shahari (also called Qarawi, Ehkili), east of Mahri; Kuria Muria (close to Shahari), a group of five islands off the coast of Arabia; and Sokotri (spoken on the island of Sokotra in the Gulf of Aden). Their relationship to Epigraphic South Arabic remains to be established. Since the Modern South Arabic dialects are surrounded by Arabic dialects, they are considerably influenced by them, especially in vocabulary.

For a description of Arabic, *see* ARABIC LANGUAGE.

Ethiopic.—The indigenous language of Ethiopia was not Semitic. Cushitic was the language group of the geographical area. It was some time in the 1st millennium B.C. that Semites (called Ḥabašat) from south Arabia entered Ethiopia. The region of their origin is not known nor is it known whether a single Semitic language was imported into Africa from south Arabia or several related tongues. The South Arabic speakers imported a Semitic language and a Semitic script, and it is this language that developed into Semitic-Ethiopic. Since the indigenous language group of Ethiopia was Cushitic, there was a very strong influence of Cushitic on the phonology, morphology, syntax and vocabulary of Semitic-Ethiopic. This influence is stronger in the south than in the north. The Semitic-Ethiopic languages are: Geez, Tigre, Tigrinya, Amharic, Argobba, Harari, Gurage and Gafat. From the descriptive point of view a division into North Ethiopic (including Geez, Tigre and Tigrinya) and South Ethiopic (including Amharic, Argobba, Harari, Gurage and Gafat) is appropriate.

Geez or Ancient Ethiopic is no longer spoken; it is the language of the liturgy. The oldest Ethiopic inscription, in unvocalized script, is that of Matara from the 3rd or 4th century A.D. The following inscriptions, in vocalized script, cover the period from the 4th to the 9th century. Between the 5th and 7th centuries the Bible was translated. No literary documents of the period from the 9th to the 13th century have come to light. Even though Geez ceased to be spoken sometime between the 10th and 12th centuries, it continued to be the literary language of Ethiopia. The classical period of Geez literature was between the 13th and 17th centuries. It is important to note that there is a traditional pronunciation of Geez used by the priests. This pronunciation, however, seems to be a reflection of the speech habits of the various vernaculars rather than any concrete evidence of the ancient pronunciation of Geez.

The languages closely related to Geez are Tigre and Tigrinya. Tigre is mainly spoken in the eastern and western lowlands of Eritrea, including the Massawa region and the Dahlak Islands in the east and up to the Kassala province and in the border regions of the Sudan in the west. The only literary documents are some religious texts printed by Swedish and Catholic missions. Tigrinya is spoken in the province of Tigre in northern Ethiopia. Tigrinya literature is still in its beginnings. As in the case of Tigre, various mission societies have printed biblical and religious texts in Tigrinya. In addition textbooks and other literary specimens are currently printed in the language.

Amharic (*q.v.*) is the national language of Ethiopia. Argobba is closely related to Amharic. It is spoken in the region of Ankober, to the north of Addis Ababa. It is decreasing in favour of Amharic. Argobba was also spoken to the south of Harar, but disappeared completely in favour of Galla, a Cushitic language. There is no literature in the language. Harari is spoken in the city of Harar in eastern Ethiopia. Because of its geographical position the Harari vocabulary has many Amharic and Galla loanwords. There are religious texts in Harari written in Arabic characters. Gurage is a dialect cluster spoken in the region of

Gurage, to the southwest of Addis Ababa. There are three main groups: (1) Eastern Gurage made up of Selti, Wolane, Ulbarag, Inneqor and Zway; (2) Western Gurage including Chaha, Eža, Ennemor, Endegeñ, Muher, Masqan and Gogot; (3) Northern Gurage with Aymellel as its only representative. Since Gurage is a Sidamo enclave, the vocabulary has a considerable number of Sidamo loanwords. There is no literature in Gurage other than a catechism in Chaha written in Ethiopic characters. Gafat was spoken in the region of the Blue Nile in the province of Godjam. The language disappeared completely in favour of Amharic and very few speakers remain.

GENERAL FEATURES

Phonology.—The consonant phonemes of proto-Semitic are: labials *b, p, m;* dentals *d, t, ṭ,* lateral (?) *ḍ;* interdentals *ḏ, ṯ, ṯ̣;* sibilants, *z, s, š, ṣ,* lateral (?) *ś;* velars *g, k, q,* spirants *ġ, ḫ;* liquids *l, n, r;* laryngeals and pharyngeals *ʼ, ʻ, h, ḥ;* semivowels *w, y.*

As can be seen, labials are not abundant. They are: *b, m, p.* In Hebrew-Aramaic *f* is a phonetic variant of *p;* indeed, *p* after a vowel becomes a spirant *f*. In South Semitic the original *p* became *f* in all positions. Ethiopic developed secondarily a *p* (beside *f*) and a glottalized *p̣*. There was no *v* in proto-Semitic.

The interdental series *ḏ, ṯ,* and *ṯ̣* remained only in Arabic and South Arabic. In the other languages the interdentals became either dentals or sibilants; this also occurred in some Arabic dialects. The correspondences for the ancient interdentals are as follows:

Proto-Semitic	Akkadian	Hebrew	Aramaic	Ethiopic
ḏ	*z*	*z*	*d*	*z*
ṯ	*š*	*š*	*t*	*s*
**ṯ̣*	*ṣ*	*ṣ*	*ṭ*	*ṣ*

The original laryngeal series (*ʼ, ʻ, h, ḥ*) is preserved only in some Aramaic dialects, ancient Hebrew, Arabic, South Arabic, and North Ethiopic.

Phonemes of uncertain pronunciation are *ḍ* and *ś;* they were perhaps laterals. They are preserved in Hebrew (*ś*), Geez (*ḍ*), South Arabic (*ḍ, s₂*), and in Arabic (*ḍ*).

The correspondences are as follows:

Proto-Semitic	Akkadian	Hebrew	Aramaic	Ethiopic	South Arabic	Arabic
ḍ	*ṣ*	*ṣ*	*ʻ(q)*	*ḍ*	*ḍ*	*ḍ*
ś	*š*	*ś*	*s*	*ś*	*s₂*	*š*

A characteristic feature of Semitic is the triadic system: voiced-voiceless-emphatic (mostly voiceless). Thus *d-t-ṭ; g-k-q; z-s-ṣ.* Ethiopic *b-p-p̣* is a secondary development.

Concerning the emphatic pronunciation there are two types: there is the emphatic-velarized type of Arabic (*ṭ, q, ṣ*), and the glottalized type of Ethiopic (*tʼ, kʼ, sʼ*). It is difficult to determine which was the original pronunciation of proto-Semitic.

In the vocalic system, proto-Semitic had the phonemic pattern of the vowels, *a, i, u* and *ā, ī, ū*. The various languages developed vocalic variants.

Morphology and Syntax.—*Roots.*—The meaning of a root lies in the consonants, the vowels serving to express shades of the basic meaning. This is unlike English, for instance, where *love, live* and *leave* have different meanings. In Semitic from the root *ktb* "write" are obtained *kataba* "he wrote," *kātib* "the writing one," *kutiba* "it was written," *kitāb* "book," the vowels expressing only shades of meaning without changing the basic meaning of the root. The Semitic root consists mainly of three radicals. There are, however, signs that there were more biradicals in ancient Semitic than at present. In the course of linguistic development many biradicals became secondarily triradicals. Ancient biradicals still preserved in the language are essential elements of the vocabulary, such as *ʼab* "father," *dam* "blood," *yad* "hand," and many more. Likewise many biradical verbs of the type *qām* "stand," and *mad(d)* "stretch" became perhaps only secondarily triradicals through the addition of a semivowel *w, y* (*qawama*) or the repetition of the last radical. A typical example of an ancient biradical "squeezed" into the triradical system is the Arabic diminutive *yudayy* "small hand" (of dialectal Arabic) from the biradical *yad* "hand" or the Ethiopic denominative *lab(b)awa* "understand" for *ləbb* "heart." In addition there are many roots whose basic meaning lies in the two radicals, as in *prq* "break," and *prṣ* "break," *prs* "separate," *pšr* "dissolve."

Nouns and Articles.—Proto-Semitic seems not to have had a definite article. It was formed secondarily in the various languages. The article is *al-* in Arabic, *ha-* (perhaps **hal-*) in Hebrew, a suffixed *-ā* in Aramaic, and so on. Akkadian has no definite article. While Geez has no definite article, the various modern Ethiopic languages employ various elements for the expression of the article.

Cases.—Proto-Semitic had three cases: a subject (nominative) case *-u,* an adnominal or possessive (genitive) *-i,* and a relational case mostly for the expression of the direct complement (accusative) *-a.* These case endings are also traceable in the plural. The case endings are completely preserved only in certain stages of Akkadian and in classical Arabic; the other languages preserve only traces of them.

Gender.—Semitic has a masculine and a feminine gender. The feminine is characterized by the suffixed morpheme *-(a)t* becoming *-ā* in the absolute state of Hebrew and *-a* in spoken Arabic. The feminine morpheme is most consistently used with the adjective. Special classes of nouns (such as parts of the body, geographical terms and others) are treated as feminine without having the feminine ending *-t.* Also words denoting female beings very often have no feminine ending (**ʼumm* "mother").

Number.—Proto-Semitic had three numbers: singular, dual and plural. The dual was originally used for parts of the body going in pairs. It is preserved in Akkadian, Hebrew, South Arabic and Arabic (where it extended to all nouns). Traces of the dual are to be found in the other languages. There are two kinds of plurals: an external plural and an internal plural. The external plural consists of the addition of a suffixed morpheme to the singular base. Its basic form for the masculine was probably *-ū* (or *-ī, -ā*) with the consonant addition *-n* or *-m,* the feminine plural morpheme is *-āt:* Arabic *malik* "king," pl. *malik-ūna;* Hebrew *meleḵ* "king," pl. *məlāḵîm;* fem. Arabic *malik-āt,* Hebrew *məlāk-ōṯ.* The internal plural consists of the vocalic change of the singular base; thus, for instance, *kitāb* "book," but *kutub* "books." The internal plural is preserved only in Arabic, South Arabic and in North Ethiopic. Traces of it are preserved in the other languages.

Numerals.—A puzzling feature in the usage of the numerals is the fact that the feminine form of the numeral is used with masculine nouns, while the masculine form of the numeral is used with the feminine nouns; thus, *ʼarbāʻā ʼaḥ-īm* "four brothers" (*ʼarbāʻ-ā* is the feminine form), *ʼarbaʻ ʼăḥāyōṯ* "four sisters."

Verbs.—The majority of the verbs are triradical (*ktb*) although there are some biradicals (*qām*) and quadriradicals (Omani *ḥarwas* "stamp"). The modern languages make extensive use of the quadriradicals, forming them either by total or partial repetition (1.2.1.2: Hebrew *gilgel* "roll"; or 1.2.3.3: Tigre *tbll* "envelop") or by inserting a liquid *n* or *l, r* into the triradical (Arabic *zanʻaq* "shout," root *zʻq;* Egyptian *ta-ʻarqal* "be bent," root *ʻql; ḥalbaṭ* "hit," root *ḥbṭ*).

The basic stem of a triradical verb in the perfect has the structure CaCaC- (C = consonant), as in Arabic *qatala* "he killed," and also CaCiC- and CaCuC-, as in Arabic *fariḥa* "he was joyful," *ḥasuna* "he was (is) beautiful." (In Hebrew the original *-i-* became *-e-* and *-u-* became *-o-,* thus *kāḇeḏ* "he was heavy," *qāṭon* "he was small".) Basically it would seem that the type CaCaC- expresses an action, be it transitive or intransitive (Hebrew *kāṯaḇ* "he wrote," Arabic *kaḏaba* "he lied"), whereas the types CaCiC- and CaCuC- express a state or condition that is completed (Hebrew *kāḇeḏ* "he was heavy," *qāṭon* "he was small") or uncompleted (*ḥāpeṣ* "he found delightful, he was delighted with a thing, he wished"). However, there are many exceptions to this principle.

Aspects or Tenses.—Proto-Semitic had aspects, not tenses. The aspects were: completed (perfect) and uncompleted (imperfect) action. Only at a later stage did the perfect come to express the past, while the imperfect expressed the present or future. Persons, numbers and genders were expressed in the perfect by

suffixes; in the imperfect by prefixes and suffixes. In Akkadian alone the completed and uncompleted actions are both expressed by forms with prefixes (*ikšud* "he has conquered," *ikaš(š)ad* "he conquers, he will conquer"), whereas duration ("the permansive") is expressed by a suffixed form (*kašd-ū* "they are, were, will be conquering"). Note that Ethiopic has the same syllabic structure of the prefixed forms but with different functions: *yĕnag(g)ĕr* "he speaks, he will speak," *yĕngĕr* "may he speak."

Stems.—An important feature of the Semitic verb is the formation of derived stems to express various modifications of the basic meaning. These stems are formed either by vocalic change (*qātala* against *qatala*), by gemination of the second radical (*qattala*) or by prefixed morphemes. The varieties in the formation of the derived stems in the various languages are considerable and no basic pattern can be given for proto-Semitic. An illustration will be taken from Arabic, but it should not be considered as representative of the general pattern of Semitic. Basic stem: *qaṭa'a* "cut." The intensive is formed by the gemination of the second radical: *qaṭṭa'a* "cut to pieces"; however, this form also expresses occasionally a causative: *ḥassana* "embellish," from *ḥasuna* "be pretty." The conative meaning is formed by changing the vowel *a* of the first radical into *ā*: *kātaba* "correspond" against *kataba* "write." The causative is formed by a prefixed *'a-*: *'a-ǧlasa* "cause to sit down, seat," from *ǧalasa* "sit down"; the other languages express the causative either by *'*, *h* or *s*, *š*. The reflexive-passive of the basic stem is formed either by prefixed *n* or by infixed *t*, from the intensive and conative it is formed by a prefixed *t-*, from the causative it is formed by a prefixed *st-*: *in-qaṭa'a* "be cut" (from *qaṭa'a*), *iǧ-ta-ma'a* "come together, be united" (from *ǧama'a* "gather"), *ta-qaṭṭa'a* "be cut to pieces" (from *qaṭṭa'a* "cut to pieces"), *ista-ḫbara* "inquire, take information" (from *'aḫbara* "inform"). The passive meaning is also formed in Arabic and in Hebrew (but not in the other languages) by an internal vocalic change: *kutiba* "it was written" against *kataba* "he wrote." The reciprocal is formed by *ta-* added to the basis of the conative: *ta-qātala* "fight with one another."

Vocabulary.—A short list of common words will illustrate the closeness of relationship.

TABLE.—*Common Words*

English	Akkadian	Hebrew	Aramaic	Ugaritic	Ethiopic	Arabic
brother	*aḫu*	*'aḥ*	*'aḥā*	*aḫ*	*'ĕḫᵘ*	*'aḫ*
master	*bēlu*	*ba'al*	*ba'lā*	*b'l*	*bā'ĕl*	*ba'l*
dog	*kalbu*	*keleḇ*	*kalbā*	*klb*	*kalb*	*kalb*
fly	*zumbu*	*zəḇūḇ*	*dabbāḇā*		*zĕmb (Amh.)*	*ḏubāb*
seed	*zēru*	*zera'*	*zar'ā*	*dr'*	*zarĕ'*	*zar'*
head	*rēšu*	*rō(')š*	*rēšā*	*riš*	*rĕ'ĕs*	*ra's*
eye	*ēnu*	*'ayin*	*'aynā*	*'n*	*'ayn*	*'ayn*
tongue	*lišānu*	*lāšōn*	*liššānā*	*lšn*	*lĕsān*	*lisān*
tooth	*šinnu*	*šen*	*šənānā*		*sĕnn*	*sinn*
sky	*šamū*	*šāma-yim*	*šəmayyā*	*šm(y)m*	*samāy*	*samā'*
night	*līlātu*	*laylā*	*lelyā*	*ll*	*lelīt*	*layla*
water	*mū*	*ma-yim*	*mayyā*	*my*	*māy*	*mā'*
house	*bītu*	*bayiṯ*	*baytā*	*bt*	*bet*	*bayt*
peace	*šalāmu*	*šālōm*	*šəlāmā*	*šlm*	*salām*	*salām*
name	*šumu*	*šem*	*šəmā*	*šm*	*sĕm*	*ism*

Scripts.—There are three scripts used in the Semitic languages: cuneiform writing, North Semitic writing and South Semitic writing. Cuneiform writing was used in Akkadian. It seems that the invention was that of the Sumerians of Mesopotamia and it was from them that the Assyrians and Babylonians took over the script. In the beginning the symbols were pictorial. At a later stage the symbols were used as word-signs (or ideograms) representing also abstract ideas; so, for instance, the solar disc came to indicate the idea of "day" and "time." As a further development the word-signs represented the phonetic value of words without any regard to their meaning as pictures.

Thus a syllabary was produced without achieving, however, an alphabetic system.

Cuneiform signs were also used in Ugaritic for which a particular cuneiform alphabet was developed, not dependent on the Sumerian-Assyrian-Babylonian. The symbols represented only consonants with the exception of three *alephs* with vowels (*'a*, *'i-'e*, *'u-'o*).

The North Semitic script has two main branches: the Canaanite and the Aramaic. The offshoots of the Canaanite branch are "Early Hebrew" and "Phoenician." (For the "Early Hebrew" script, *see* HEBREW LANGUAGE.) The script of the Moabite stone is closely related to Early Hebrew writing, the Samaritan alphabet is a descendant of Early Hebrew script.

The earliest written document in Aramaic is from the 9th century B.C. In the second half of the 1st millennium B.C. it became the most widespread script of the entire near east. The Semitic offshoots of the Aramaic alphabet are: Square Hebrew (which developed into the modern Hebrew script), Nabataean-Sinaitic-Arabic (with its two main branches of Naskhi and Kufic), Palmyrene, Syriac-Nestorian and Mandaean.

All these alphabets express only consonants. The vocalic symbols of Syriac, Arabic and Hebrew were probably introduced in the 8th century A.D.

South Semitic Alphabet.—The South Semitic alphabet was used for Epigraphic South Arabic and for the North Arabic inscriptions of Thamudenic, Lihyanite and Safaitic. The South Arabic alphabet passed into Ethiopia in the 1st millennium B.C. But whereas the South Arabic alphabet is consonantal, the Ethiopic alphabet is syllabic, that is, there is a single symbol for the consonant with vowel. The problems of the origin of the South Semitic alphabet and the invention of the Ethiopic vowels are still unsolved.

See also references under "Semitic Languages" in the Index.

BIBLIOGRAPHY.—C. Brockelmann, *Grundriss der vergleichenden Grammatik der semitischen Sprachen*, 2 vol. (1908–1912); W. Marçais and M. Cohen, *Précis de linguistique sémitique* (1910), translation of C. Brockelmann, *Semitische Sprachwissenschaft;* G. Bergsträsser, *Einführung in die semitischen Sprachen* (1928); Luis G. Gray, *Introduction to Semitic Comparative Linguistics* (1934); H. Fleisch, *Introduction à l'étude des langues sémitiques: Elements de bibliographie* (1947); S. Moscati (ed.), *An Introduction to the Comparative Grammar of the Semitic Languages* (1964); Th. Nöldeke, *Die semitischen Sprachen* (1899); W. Wright, *Lectures on the Comparative Grammar of the Semitic Languages* (1890); H. Zimmern, *Vergleichende Grammatik der semitischen Sprachen* (1898). (W. LU.)

SEMLER, JOHANN SALOMO (1725–1791)

SEMLER, JOHANN SALOMO (1725–1791), German Lutheran theologian who developed the principles of textual criticism of the Bible and the idea of a history of the Bible canon, was born at Saalfeld, Thuringia, on Dec. 18, 1725. The son of a pastor, he grew up in pietistic surroundings which left him with a lifelong aversion to Pietism. At Halle University he became the disciple, later the assistant, and finally the literary executor of the orthodox rationalist S. J. Baumgarten. After holding positions at Coburg and the University of Altdorf, Semler became professor of theology at Halle in 1753, succeeding Baumgarten in 1757 as head of the theological faculty. He died on March 14, 1791.

Semler's lectures and copious writings on biblical criticism provoked fierce opposition. He broke away from orthodox theology in evolving an undogmatic, strictly historical interpretation of Scripture, challenging the idea of verbal inspiration. His distinction between religion and theology was based on Luther's distinction between the Word of God and Scripture; revelation he understood purely as the message of truth related to man as a natural being. Despite contemporary accusations, Semler was no rationalist, in that faith remained essential for him to an understanding of religion; he proved his position by his reply (1779) to the deistic *Wolfenbüttel Fragments* of H. S. Reimarus (*q.v.*; *see* also LESSING, GOTTHOLD EPHRAIM).

On the practice of religion, however, Semler's ideas were by no means clear. He attempted to draw a distinction between private and public religion, the former to be a matter of private conscience, the latter a question of public order. This led to his unfortunate defense of the Prussian government's reactionary *Religionsedikt* (1788), an attack on enlightened tendencies in the church.

BIBLIOGRAPHY.—J. S. Semler, *Lebensbeschreibung von ihm selbst verfasst*, 2 vol. (1781–82); F. C. Baur, *Epochen der Kirchlichen Geschichtsschreibung* (1852); H. Hoffmann, *Die Theologie Semlers* (1905); P. Gastrow, *J. S. Semler in seiner Bedeutung für die Theologie* (1905); A. Schweitzer, *The Quest of the Historical Jesus*, 3rd ed. (1954); G. Hornig, *Die Anfänge der historisch-kritischen Theologie* (1961); W. Schmittner, *Kritik und Apologetik in der Theologie J. S. Semlers* (1963). (E. BI.)

SEMMELWEIS, IGNAZ PHILIPP (1818–1865), Hungarian physician, whose influence on the development of knowledge and control of infection was cited by Joseph Lord Lister, the father of modern surgery: "Without Semmelweis my achievements would be nothing. To this great son of Hungary Surgery owes most." He was born at Buda on July 1, 1818, and educated at the universities of Pest and Vienna, where he attracted the attention of Joseph Skoda and Karl von Rokitansky. He graduated M.D. at Vienna in 1844 and was appointed assistant in the first obstetric clinic under Johann Klein.

BY COURTESY OF THE LIBRARY OF MEDICAL HISTORY, BUDAPEST

IGNAZ PHILIPP SEMMELWEIS, PAINTING BY A. CANZI, 1857

Puerperal infection was the scourge of maternity hospitals throughout Europe. Although most women delivered at home, a few women who had to seek hospitalization because of poverty, illegitimacy or obstetrical complications faced a mortality rate of 3% to 25% because of childbed fever, or puerperal infection. Although its cause was unknown, it was thought that it was induced by overcrowding, poor ventilation, the onset of lactation or a "miasma."

Semmelweis, during his student days, became interested in this great killer of young mothers and as an assistant he proceeded to investigate its cause despite the serious objections of Klein, the chief of the maternity. Two or three times as many women who delivered in the first division died of childbed fever as those who delivered in the second division, although the two maternities were identical with the exception that students were taught in the first and midwives in the second. Semmelweis reasoned that perhaps the students carried something to the patients they examined during labour. The death of his friend Jakob Kolletschka, a pathologist, from a wound infection incurred during the examination of a woman who died of puerperal infection, and the similarity of the findings in his friend and in the women who died of puerperal infection, provided Semmelweis' first breakthrough. He concluded that students, who came directly from the dissecting room to the maternity, carried infection from mothers who had died of this dread complication to healthy mothers they examined during labour.

Semmelweis promptly instituted thorough washing of the hands with soap and water prior to examination of a woman in labour and later introduced ablutions with chlorinated lime. The mortality of the first division promptly dropped below that of the second. The outstanding success of this simple measure failed to impress Klein, who, blinded by ignorance, vanity, or jealousy, restricted Semmelweis' activities, underhandedly prevented his promotion to assistant professor, and drove him from Vienna.

In 1850 Semmelweis was appointed obstetric physician in the maternity department at Pest, where his ideas promptly reduced the high maternal mortality to 0.85%. He developed a private practice among influential people, married, and seemed reasonably content. However, he continued to be plagued by the failure of the profession to accept his doctrine and thereby save the lives of countless young women.

He died on Aug. 17, 1865, from a wound of the right hand, a victim of the very disease for which he had sacrificed his health, fortune, and peace of mind.

Semmelweis' chief publication was *Die Ätiologie, der Begriff und die Prophylaxis des Kindbettfiebers* (1861), Eng. trans. by F. P. Murphy.

BIBLIOGRAPHY.—Biographies by Hegar (1882), Bruck (1887), Duka (1882), Grosse (1898), and Schürer von Waldheim (1905). *See* also the interesting medical novel, *The Cry and the Covenant* (1954), by Morton Thompson, which depicts the life of Semmelweis.

(M. E. Ds.)

SEMMERING, Austria, the most easterly and lowest (3,215 ft. [980 m.]) of the great Alpine passes. In the watershed dividing the Mur and Leitha drainage basins, it facilitates Vienna's communications with Yugoslavia via Styria (Steiermark), with Italy via Carinthia (Kärnten), and via the upper Enns Valley with Innsbruck. Since the ancient routes skirted the Alps in the east, the Semmering remained unimportant until the foundation of the hospice (Spital am Semmering) in 1160 and the union of Austria and Styria. The first road was built in 1728 and rebuilt in 1839–42. The Semmering railway (1848–54) was the first mountain railway in the world. It passes about 282 ft. (86 m.) below the summit of the pass through a tunnel nearly one mile long. A second parallel tunnel was inaugurated in 1952. (K. A. S.)

SEMMES, RAPHAEL (1809–1877), U.S. naval officer who commanded the Confederate ship "Alabama" during the American Civil War, was born in Charles County, Md., on Sept. 27, 1809. He was appointed midshipman in the Navy in 1826, and while waiting for orders studied law and in 1834 was admitted to the bar. In the war with Mexico, Semmes was active in superintending the landing of General Scott's troops at Veracruz in March 1847, and later, as volunteer aid to General Worth, took an active part in the battles of the valley of Mexico. These experiences were interestingly described in his books *Service Afloat and Ashore During the Mexican War* (1851) and *The Campaign of General Scott in the Valley of Mexico* (1852). In 1855 he was promoted to commander and afterward was made naval secretary of the Lighthouse Board at Washington, in which service he was when the Civil War broke out. When Alabama, his adopted state, seceded, he resigned his commission and received an appointment of the same rank in the Confederate Navy. He fitted out the packet "Sumter," in which he captured as many as 17 Northern merchant vessels, chiefly along the South American coast. Later he commanded the "Alabama," a 1,016-ton ship built in England for the Confederacy, with which he made a series of daring and successful cruises lasting two years. Finally, he met the Northern ship "Kearsarge" in the English Channel, and after a 70-minute battle was forced to surrender. Twenty minutes later the "Alabama" sank and Semmes was rescued by an English yacht.

When Semmes returned home he was commissioned rear admiral and assigned to the Confederate fleet in the James River. When Richmond was captured he blew up his ships and, with his men, joined the army of General Johnston. When Johnston surrendered, Semmes returned to his home in Mobile and opened a law office, to which practice he devoted the greater part of the rest of his life. He died on Aug. 30, 1877.

In addition to his books on the Mexican War, he wrote *The Cruise of the Alabama and the Sumter* (1864) and *Memoirs of Service Afloat During the War Between the States* (1869).

SEMNAN (SAMNAN), the chief town of a *shahrestan* (district) and of the general governorship of Semnan-Damghan in northern Iran, lies at an elevation of 3,740 ft. (1,140 m.) on a large, mostly desert plain at the southern foot of the Elburz Mountains, and 145 mi. E of Teheran on the main road to Meshed. Pop. (1956) 29,036. In the town are a minaret (12th century) with rich ornamentation of raised brick and remains of a rebuilt mosque of the same period. Several large *Hosseiniyehs* (*Husainiyehs*, places of worship of Husain; *see* HASAN AND HUSAIN) are conspicuous. Semnan is a thriving market centre for the agricultural products of the district, chiefly grains, cotton, and tobacco. Besides a lively bazaar, there is some textile industry and carpet making. There are sulfur deposits 20 mi. (32 km.) SE of the town, and in 1959 oil was discovered at Khurian (formerly a Soviet-held concession), 25 mi. (40 km.) S. The town is also linked by rail with Teheran and Meshed.

Semnan is mentioned by Ptolemy as Samina, and cited by the Oriental geographers for the rivulets running through its streets, for the making of soft cotton stuffs, and for its sweet paste made from almonds and figs. The town was the administrative headquarters of the former province known as Semnan wa Damghan, which became part of the 2nd *ostan* (province) of Mazanderan in 1938. In 1961 Semnan-Damghan was formed into a separate general-governorship. *See* MAZANDERAN. (H. Bo.)

SEMO SANCUS DIUS FIDIUS, an obscure Roman deity; a god Fisios or Fisovios Sancios was also worshiped in Umbria, and appears to be the same. He had a temple at Rome on the Quirinal, in which was an ancient statue of a woman, said to be Gaia Caecilia, or Tanaquil, wife of Tarquinius Priscus. His functions are very obscure. The four parts of his name seem to be connected respectively with seed (*semen*), purity or holiness (*sancus* and *sanctus* are from the same root), Jupiter (or simply brightness, or celestial nature; Dius, root *Div*), and faith (*fides*). Hence perhaps "the spirit of sowing (or seed), pure, Jovian (or bright, or celestial), faithful." Of his ritual we know only that oaths by him were taken in the open air, the formula being *medius fidius, cf. mehercule*, "by Hercules"; that there was an opening in the roof of his temple; and that certain disks of metal were kept in the temple. The first two facts perhaps suggest a sky god, but we know that certain deities undoubtedly not celestial (for example, Terminus, the spirit of boundaries) had similar ritual. The disks might possibly be solar symbols. On the whole, the balance of the very scanty evidence is rather in favour of supposing him to be a god of the sky, perhaps connected, as Wissowa supposes, with Jupiter, who had power to influence seed corn (by sending rain in season?) and, being able to see what went on in the world, was a natural witness to solemn oaths. The ancients, wrongly supposing the last two members of his name to signify "son of Zeus," identified him with Hercules, an error which some moderns also have fallen into.

SEMPER, GOTTFRIED (1803–1879), German architect and writer on art, who initiated the neo-Renaissance style, was born in Hamburg, on Nov. 29, 1803. He studied in Munich and Paris and practised architecture in Dresden from 1834 until 1848 when, because of revolutionary activities, he was forced into exile to Paris and London. He headed the architecture department of the Zürich Polytechnikum (1855–71) and between 1871 and 1876 participated in the rebuilding of Vienna. His work marks the transition away from the classicism of his friend Karl Schinkel. As an eclectic, he achieved powerful solutions. Among his main works were the Dresden opera house (1837–41, rebuilt 1871–78); the Zürich Polytechnikum (1858–64); and with Karl von Hasenauer, the Vienna Burgtheater (1874–88) and the two imperial museums in Vienna (1872–81). In his influential writings, principally *Der Stil in den technischen und tektonischen Kunsten* (1860–63), he stressed a rational interpretation of techniques as a source of style, recommending colour in decorative arts and architecture. He died in Rome on May 15, 1879.

His nephew KARL SEMPER (1832–1893) was a zoologist, noted for his studies of Philippine fauna. (A. K. P.)

SEMPILL OF BELTREES, a family of Scottish poets. SIR JAMES (1566–1626) was born in 1566 and brought up with the young King James VI, and like him was tutored by George Buchanan. He went to St. Andrews university, and became ambassador to England (1599) and to France (1601). He was knighted in 1600 and died at Paisley in Feb. 1626. His writings include a few controversial tracts, but he is remembered rather for his satirical poem *The Pack-mans Pater Noster*, an antipapal dialogue between a pedlar and a priest, written in English in rhyming couplets. This poem, when reprinted in 1669, contained additions by his son ROBERT (*c.* 1595–*c.* 1665), who was born *c.* 1595, was educated at Glasgow university, fought for the royalists during the Great Rebellion and wrote the elegy on Habbie Simson, "The Life and Death of the Piper of Kilbarchan." This humorous poem in Scots was included by James Watson in his *Choice Collection* (1706) and its fame was assured when Allan Ramsay called its metre "Standart *Habby*" and used it himself in several poems. "Standart *Habby*" was later known, after its greatest exponent, as the Burns stanza. Robert Sempill died *c.* 1665. Little is known of FRANCIS (*c.* 1616–1682), son of Robert, except that he was born *c.* 1616, was made sheriff-depute of Renfrewshire some time before 1677 and died at Paisley on March 12, 1682. He is reputedly the author of some popular and entertaining Scots poems, "The Blythsome Bridal," "Maggie Lauder" and "Hallow Fair."

An earlier ROBERT SEMPILL (*c.* 1530–1595), also a poet, has no proved connection with the Sempills of Beltrees. He was born *c.* 1530 and appears to have held some office at the Scottish court, being mentioned in the treasurer's accounts in 1568. From the evidence of his poems he was probably present at the sieges of Leith (1560) and of Edinburgh (1573). He probably died in 1595.

Sempill wrote many vigorous, bitter and often coarse satirical Scots poems on religious and social topics of the day; his work is violently Protestant and suited to his description of himself as "rakles [reckless] Robert."

BIBLIOGRAPHY.—For the earlier Robert Sempill see *The Sempill Ballates*, ed. by T. G. Stevenson (1872) and *Satirical Poems of the Time of the Reformation*, ed. by J. Cranstoun, Scottish Text Society, 2 vol. (1891–93). For the other Sempills see *The Poems of the Sempills of Beltrees*, ed. by J. Paterson (1849) and G. Eyre-Todd, *Scottish Poetry of the 17th Century* (1895). (E. G. M.)

SENA, an Indian dynasty ruling in Bengal in the 11th and 12th centuries. Their ancestors came from the south and became petty rulers in southwestern Bengal early in the 11th century, as tributaries of the Pala (*q.v.*) dynasty. Hemantasena, in the mid-11th century, became independent and assumed imperial titles. His successor, Vijayasena (reigned *c.* 1095–1158), built an empire on the ruins of that of the Palas, gaining control of all Bengal and north Bihar. The next ruler, Vallalasena, was the author of religious literature and is the reputed founder of the Bengali system of hypergamy (*i.e.*, socially "upward" marriage of women); he reestablished the caste system where it had become lax owing to Buddhist influence under the Palas. The last important Sena king, Lakshmanasena (*c.* 1178–*c.* 1205), after military successes, became a great patron of literature; the poets Jayadeva and Dhoyi wrote at his court at Nadia. In 1202 Lakshmanasena was expelled from Nadia by the Turkish chief Mohammed Bakhtiyar Khalji and died about three years later. Sena kings continued to rule in east Bengal for some decades, but the main political power in Bengal passed to the Muslims.

There was a marked revival of orthodox Hinduism under Sena rule.

See R. C. Majumdar (ed.), *The History of Bengal*, vol. i (1943). (A. L. BA.)

SÉNANCOUR, ÉTIENNE PIVERT DE (1770–1846), French author whose best-known work, *Oberman* (2 vol., 1804), marks him as a precursor of the Romantic movement. He was born in Paris in November 1770, the son of a *contrôleur des rentes*. His father intended him to be a priest, and in 1789, to avoid entering the seminary of Saint-Sulpice, he went to Switzerland, where in September 1790 he made a rash, and eventually unhappy, marriage with a Mlle Daguet. At the Revolution his name was included on the list of emigrants and he did not return permanently to Paris until 1803. In 1795 *Aldomen; ou, le bonheur dans l'obscurité*, an early draft of *Oberman*, was published. *Rêveries sur la nature primitive de l'homme* (1799) contains passages of impassioned description of the nature of man. *Oberman* (in later editions spelled *Obermann*) consists of a series of reveries in letter form written by a solitary in a valley in the Jura. The book, which was to a great extent inspired by Jean Jacques Rousseau (*q.v.*) was edited by Sainte-Beuve (1833) and George Sand (1852), and was praised for its poetic interest in nature and in man by Matthew Arnold (*Essays in Criticism*, Second Series, 1888; the essay, "Obermann," in the *Academy*, Oct. 9, 1896). It had considerable influence in France (*see* FRENCH LITERATURE: *The Revolution and the 19th Century*) and in England, and with Chateaubriand's Romantic novel *René* (published as part of the *Génie du christianisme*, 1802) it opened the way to a new type of writing, the main characteristic of which was a pervasive melancholy. Sénancour also wrote a treatise on love, *De l'Amour consideré dans les lois réelles et dans les formes sociales de l'union des deux sexes* (1805); and a meditative work, *Libres méditations d'un solitaire inconnu* (1819). In 1833 his other novel in letters, *Isabelle*, was published. He died at Saint-Cloud on Jan. 10, 1846.

BIBLIOGRAPHY.—*Rêveries sur la nature primitive de l'homme*, ed. by J. Merlant (1910); *Oberman*, reprinted from the 1st ed., ed. by A. Monglond, with a critical explanation entitled *Journal intime d'Oberman*, 3 vol. (1947). *See* also J. Merlant, *Sénancour* (1907); G. Michaut, *Sénancour, ses amis et ses ennemis* (1909); A. Monglond, *Vies préromantiques* (1925), *Jeunesses* (1933).

SENATE, ROMAN, the most permanent element in the Roman constitution. The Latin word *senatus* suggests an assembly of old men (from root *sen-*, as in *senex,* old), originally the heads of the chief families. The word corresponds with the Greek *gerousia* (*q.v.*), the name of the council of elders at Sparta. The Athenian Areopagus (*q.v.*) may in some ways be compared to the Roman Senate.

Origins.—The Roman authorities ascribe the origin of the Senate to Romulus, who chose 100 of the best of his subjects to form an advisory council. In 509 B.C. it contained 300 members and a distinction existed within it between *patres maiorum gentium* and *minorum gentium,* the heads of the greater and the lesser families. Throughout the monarchical period the Senate consisted entirely of patricians. Probably the rise in the number of the senators was due to the incorporation of fresh elements into the patrician community, the new clans comprising *gentes minores* (lesser families). The appointment of senators depended entirely upon the king, who possibly might change his advisers during his reign; a new king could certainly abstain from summoning some of those convened by his predecessors. The powers of the Senate at this time were very indefinite.

Republican Period.—With the abolition of monarchy the Senate became the advisory council of the consuls (*see* CONSUL), meeting only at their pleasure, and owing its appointment to them; it remained a power secondary to the magistrates. Few, if any, plebeians were introduced into the Senate at this time, but the existence of an elderly plebeian senator is attested in 401 B.C. In one respect, substitution of consuls for kings tended to the subordination of the chief magistrates to the Senate. The consuls held office only for one year, while the Senate was a permanent body; in experience and prestige, its individual members were often superior to the consuls of the year. The magistrate would seldom venture to disregard the advice of the Senate, especially as he himself, in accordance with steadily growing custom, would become a senator at the end of his year of office. It was probably in their capacity of former magistrates that plebeians first entered the Senate. Of the two powers which the Senate inherited from the monarchy, the interregnum and the *patrum auctoritas* (senatorial claim to ratify), the first came to be exercised only rarely; for if either consul existed, interregnum could not be resorted to. The *patrum auctoritas,* however, developed into a definite right claimed by the patrician senators to give or withhold their consent to any act of the *comitia* (*q.v.*). The influence which it had long exercised over foreign policy increased the importance of the Senate in a period of constant warfare with the peoples of Italy. But in the early Republic the Senate remained an advising body and assumed no definite executive powers.

In the last two centuries of the Republic, a great change took place. The Senate became a self-perpetuating, automatically constituted body, independent of the annual magistrates, a recognized factor in the constitution, with extensive powers. About 312 B.C. the selection of senators (*lectio senatus*) was transferred from the consuls to the censors (*q.v.*), who normally chose former magistrates. In 81 B.C. Sulla secured an automatic composition for the Senate by increasing the number of quaestors to 20, and enacting that all former quaestors should pass at once into the Senate (*see* QUAESTOR).

The Senate's powers had then extended far beyond its ancient prerogatives. The power to appoint an interrex had fallen into practical disuse, and the power to ratify decisions of the *comitia* had become a mere form by the last century of the republic. But the Senate had acquired more effective control through the observance of certain unwritten rules regulating the relation between Senate and magistrates, to whom it formally gave advice. It became the chief governing body in Rome and tendered advice on home and foreign policy, on legislation, and on financial and religious questions. It acquired the right to assign duties to the magistrates, to determine the two provinces to be entrusted to the consuls, to prolong a magistrate's period of office (*prorogatio imperii*), and to appoint senatorial commissions to help magistrates to organize conquered territory. Its earlier influence upon foreign policy developed into a definite claim to conduct all negotiations with a foreign power, although the formal declaration of war and ratification of treaties were referred to the people. It often acted as arbitrator in disputes among Italian communities, provincials, or client-states.

Though individual senators after 218 B.C. were debarred from trading, the control of finance was in the Senate's hands. Three circumstances had combined to bring this about. The censors, who were only occasional officials, were entrusted with the leasing of the public revenues; the Senate could order them to redraft contracts. The details of public expenditure were entrusted to the quaestors, young and inexperienced magistrates whom the Senate could guide. Third, the general control exercised by the Senate over provincial affairs implied its direction of the income derived from the provinces. It also claimed the right of granting occupation and decreeing alienation of public lands. Every branch of state finance was therefore in its hands: it controlled revenue and expenditure and supervised the treasury (*aerarium*). In matters of criminal jurisdiction, the Senate claimed the right, in times of crisis, to set free by its decree the full powers of coercion that were contained in the *imperium* (supreme administrative power) of a magistrate but that were limited normally in capital cases by laws of appeal. The first certain use of this right, which amounted to a declaration of martial law, concerned Gaius Gracchus; the last was in 40 B.C.

This ever-widening influence and power of the Senate was challenged by tribunes (*see* TRIBUNE) from the time of Tiberius Gracchus onward (133 B.C.); and more particularly by the military leaders, from Marius onward, who pitted their *imperium* against the *auctoritas* of the senate. Despite the short-lived attempt of Sulla to reinstate the Senate's ascendancy, the Republic collapsed under these repeated blows at the authority of the Senate. As a result of the civil war the number of senators, which Sulla had raised to 500 or 600, was seriously depleted. Julius Caesar revised the list and increased the Senate to 900, naturally filling it with his own supporters. This reflects the practice by which the early magistrates had chosen their own body of councilors. The composition of the Senate thus underwent a considerable change: few of the senators who had opposed Caesar survived; the new senators included many *equites* (*q.v.*) and municipal Italians and even a few provincials from Gaul.

The Empire.—Since Augustus officially "restored the Republic" (27 B.C.), it was essential to preserve the prestige of the Senate, which came to be regarded in the early principate as the representative of republican institutions. With this object, and also in order to provide an adequate body of administrators, Augustus established a *senatorius ordo* (senatorial order), limited to men of personal integrity, who had completed a term of military service and who possessed a minimum property qualification. From this order, men who had served in the army and had then held one of the minor magistracies could enter the Senate at the age of 25 through the quaestorship. The *senatorius ordo* tended to become a hereditary body because normally only sons of senators could become senators.

Since the election of magistrates was transferred from the people to the Senate, to a large measure under Augustus and completely under Tiberius, the Senate became a self-recruiting body, whose numbers were reduced to 600. The emperor could influence the quaestorian elections; he also had the right of appointing as new members of the Senate men who had not been magistrates. Most of the important offices in the state were filled by senators, whose careers depended largely on the good will and confidence of the emperor.

Augustus did not share his basic power with the Senate, but he did allow it to co-operate with him in most of the spheres of government. It was left at the head of the ordinary administration of Rome and Italy, together with those provinces which did not require any military force or present special administrative difficulties. It continued to administer the *aerarium* but was soon overshadowed by the emperor, who allowed it to supervise the copper coinage alone. The Senate received judicial functions and for the first time became a court of law, competent to try cases of extortion in the senatorial provinces (for these *see* PROVINCE,

ROMAN: *The Principate*) and with criminal jurisdiction over offenses by its own members. The legislative powers of the *comitia* became very gradually extinct, and *senatus consulta* (decrees of the Senate) came to take the place of *leges* in ordinary matters although they did not at first acquire full recognition as laws. On the other hand, the Senate lost all its control of foreign policy, and though it was occasionally consulted by the emperor it was entirely subordinate to him in this department. The emperor could convene and preside over the Senate, his report (*relatio*) taking precedence; he was also *princeps senatus* (*i.e.*, his name headed the list of senators). Relations varied with the different emperors, but although recognition by the Senate provided formal acceptance of an emperor's claim to rule, the Senate in practice lost independence of action and freedom of discussion declined.

The number of Italian and provincial senators increased (especially under Vespasian), but the Italians were not outnumbered by the provincials until after the reign of Septimius Severus. At first the provincials came predominantly from Spain and Narbonese Gaul, but later there were more Asians and Africans. Under Gallienus senators lost the right to command legions and much of their part in provincial administration. Under Constantine they were virtually amalgamated with the *equites*, who had benefited from these changes. The number of the new senators rose in the 4th century to about 2,000. That the Senate was still regarded as a representative and necessary part of the constitution is shown by Constantine's creation of a duplicate Senate in Constantinople.

The most important senators were the great landowners throughout the empire, whose position became almost feudal. A great number of them failed to leave their estates to attend meetings, and the Senate often acted—as it had in the early days of the Republic—merely as a town council for Rome, under the chairmanship of the prefect of the city. Senators were free from municipal burdens but not from all taxation. Many of the great senatorial landowners were men of culture who represented Roman civilization amid increasing barbarism and tried to uphold paganism in Italy. In the 5th century, however, some of them helped the barbarian leaders against the imperial authority. In the 6th century the Roman Senate disappears from view: it is last mentioned in A.D. 580. *See* also ROMAN HISTORY.

BIBLIOGRAPHY. —P. Willems, *Le Sénat de la république romaine* (1883); T. Mommsen, *Staatsrecht* (1907); A. H. J. Greenidge, *Roman Public Life* (1901); G. W. Botsford, *Roman Assemblies* (1919); O'Brien Moore in Pauly-Wissowa, *Realencyclopädie*, suppl. vi, pp. 660–800 (1935); M. Hammond, *The Augustan Principate in Theory and Practice during the Julio-Claudian Period* (1933). On the senatorial class: F. Münzer, *Römische Adelsparteien* (1920); R. Syme, *The Roman Revolution* (1939); S. J. De Laet, *De Samenstelling van den Romeinschen Senaat* (1941) for 1st century A.D.; B. Stech, *Senatores Romani* (*Klio*, Beiheft x, 1908) for Vespasian to Trajan; P. Lambrechts, *La Composition du sénat romain* (1936–37) for Hadrian to Commodus and for Septimius Severus to Diocletian; C. Lécrivain, *Le Sénat romain depuis Dioclétien* (1888). (X.; H. H. SD.)

SENDAI, the capital and largest city (pop. [1960] 452,272) of Miyagi Prefecture, Japan. Site of a great castle, Sendai enjoyed prosperity during the feudal period as headquarters of the powerful Date Clan. The largest city and commercial centre for northern Honshu, it is also the regional seat of federal administrative agencies. It is a rail centre but depends upon an outport, Shiogama, in the southwestern corner of Matsushima Bay, for ocean contacts. The relatively few Sendai manufactures are primarily for local consumption; a small industrial zone is developing around the Shiogama port facilities. Tohoku University is outstanding among Sendai's many schools. (J. D. EE.)

SENEBIER, JEAN (1742–1809), Swiss pastor remembered for his contributions to the knowledge of photosynthesis, was born at Geneva on May 6, 1742. The observation by C. Bonnet of bubbles of oxygen on the leaves submerged in aerated water induced Joseph Priestley, Jan Ingenhousz, and Senebier to carry out a series of experiments showing the influence of light on vegetation. Senebier reported his conclusions in his most important work, *Physiologie végétale*, 1800.

Senebier proved that the presence of carbonic acid is a deciding factor in the development of oxygen by green plants in sunlight and that with decomposition of carbonic acid oxygen is liberated. Thus he was first to give a connected view of the whole process of vegetable nutrition. He died at Geneva on July 22, 1809. (V. C. AS.)

SENECA, ANNAEUS (*c.* 55 B.C.–*c.* A.D. 39), of Corduba (Córdoba) in Spain, commonly known as Seneca the Elder to distinguish him from his more famous son Lucius Annaeus Seneca, wrote an account of certain Roman rhetoricians of the age of the emperor Augustus. Little is known of his life except that he visited Rome at least twice, probably for long periods.

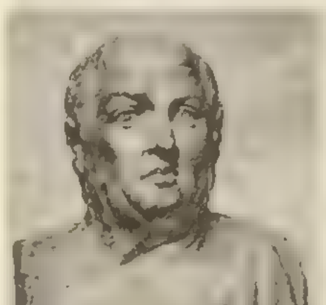
His book is concerned with the practitioners of declamation. "Declamations," originally speeches delivered privately for practice or rehearsal, had been transformed by Seneca's time into a public entertainment. They were of two kinds: those based on themes drawn from mythology or history, such as "Should the 300 Spartans at Thermopylae run away?," known as *suasoriae* because they purported to offer advice, and *controversiae*, fictitious legal cases based on some stated "law," itself either fictitious or an extremely simplified version of some existing enactment.

Seneca, who shows a fine and balanced literary judgment, disapproved of the artificial cleverness, often degenerating into absurdity, of many declaimers, but nevertheless produced a record of their work, with abundant quotation. Only about half of his book, which is entitled *Oratorum sententiae divisiones colores*, is extant; an epitome, written probably in the 4th century, covers most of the remainder. *Sententiae* were brief pointed utterances, *divisiones* the subdivisions of the subject matter, and *colores* the "slants" which the declaimers gave to the prescribed facts of their themes, so as, for instance, to make apparently unfavourable circumstances tell in favour of an imaginary defendant.

The perverse ingenuity of most of Seneca's quotations soon palls. The literary fashion which they record is interesting only from the harm it did to some major writers such as Ovid. Seneca's work is, however, valuable for the introductions to the books into which it is divided and for the comments interspersed throughout, which give much important and entertaining information about literary circles in Augustan Rome and also preserve several interesting passages which have not otherwise survived, for instance, various accounts, including Livy's, of the death of Cicero.

BIBLIOGRAPHY.—Complete editions by A. Kiessling (1872) in the Teubner series and by H. Bornecque with French trans., new ed., 2 vol. (1932); *Suasoriae*, ed. by W. A. Edward with Eng trans (1928). *See* also H. Bornecque, *Les Déclamations et les déclamateurs d'après Sénèque le père* (1902); S. F. Bonner, *Roman Declamation in the Late Republic and Early Empire* (1949). (A. E. DO.)

SENECA, LUCIUS ANNAEUS (*c.* 4 B.C.–A.D. 65), Roman statesman, philosopher, satirist, and tragedian, one of the most important Latin writers of the 1st century A.D., was born at Corduba (Córdoba), second son of Seneca the Elder, the rhetorician, of Italian or near-Italian stock. Taken to Rome while young and educated by men who combined rhetoric with philosophy, he became famous as a pleader and obtained a quaestorship. The emperor Caligula, jealous of his success, refrained from executing him only because he believed that he was about to die of disease. At the accession of Claudius I (A.D. 41) he was exiled to Corsica through the machinations of Messallina, Claudius' wife, on a charge of adultery with the emperor's niece Julia Livilla, and in spite of lavish imperial flattery was not recalled until A.D. 49; at the request of Agrippina, Claudius' new wife, he was appointed tutor to her son Nero and was immediately made praetor. When Nero became emperor (A.D. 54) Seneca with the collaboration of the praetorian prefect Burrus gradually broke Agrippina's power, turned the maladjusted adolescent emperor to dilettante



BY COURTESY OF THE STAATLICHE MUSEEN, BERLIN

LUCIUS ANNAEUS SENECA, A 3RD CENTURY COPY OF A 1ST CENTURY MARBLE BUST

pursuits and organized efficient government for five years. His power under Nero was based on personal ascendancy, not legal office. In A.D. 59 he was compelled to be privy to Nero's plan to murder his mother and is said to have written the letter to the Senate in which the emperor accused her of conspiracy. Seneca became increasingly irksome to Nero, and the death of Burrus in A.D. 62 finally broke his power. He withdrew from public life and offered to resign the great wealth amassed through political positions, which had brought him discredit; Dio Cassius alleges that his agents practised usury in Britain immediately before the rebellion of Boudicca (Boadicea) in A.D. 60. In A.D. 65 Nero, on detecting the Pisonian conspiracy, had an excuse to be rid of him and ordered his suicide. Seneca met his death unafraid and talked philosophy until the end.

Seneca belonged to the third stage of the Stoic movement, Stoics of imperial times, who placed spiritual counseling higher than first principles (*see* STOICS). His most important philosophical work is the collection of *Epistulae morales* (A.D. 63–65). These letters to Lucilius written in retirement blend Epicurean elements with Stoic exhortation and resignation; they also contain vivid illustrations from personal observation, reflections on literature and satirical presentation of vices similar to those portrayed by his contemporary Petronius. The essay *De clementia* (A.D. 55–56) on the mercy of the just ruler, dedicated to Nero, is a guarded warning against tyranny. His prose works also include the essay *De brevitate vitae*, a masterly analysis of the frivolous occupations of a corrupt society, *De tranquillitate animi*, on participation in public life, three works of consolation addressed to those who had suffered loss, and other moral essays. *Naturales quaestiones* is an exposition of atmospheric and terrestrial phenomena. The so-called correspondence of Seneca with St. Paul is a forgery of the 4th century A.D.

Seneca's prose style, an extreme example of the declamatory rhetorical manner of the 1st century A.D. with its short splintered sentences, epigrammatic concluding phrases (*sententiae*) and wide range of metaphor, was admired in his own day but attacked by Quintilian and other followers of Cicero.

Seneca's *Apocolocyntosis Divi Claudii* ("pumpkinification," a degradation of apotheosis) is one of the masterpieces of Roman satire. It is the only complete surviving example of the Menippean convention (*i.e.*, satire written in a mixture of prose and verse). Composed immediately after the death and deification of Claudius (A.D. 54) it describes the refusal of the gods in their assembly to accept him, and his trial and condemnation in the underworld. Seneca combines savage ridicule of a dead enemy with serious criticism of his political absolutism. The prose parts are urbanely colloquial; the verse insertions include parodies of funeral lament and tragic speech.

Seneca's nine tragedies on mythological themes are different in kind from Greek tragedies, being concerned not with character and motive but with the rhetorical display of intense emotion (*see* LATIN LITERATURE). They were intended for recitation, not for stage production, and in language are influenced by Ovid's elegant artificiality. As speeches, episodes and choruses are independent declamatory entities, the plays lack unity. In *Phaedra* Seneca as Stoic moralist demonstrates the evils of uncontrollable passion, in *Troades* he interprets human suffering and political interrogation in contemporary terms, and in *Thyestes* combines Stoic theory with experience of imperial tyrants. Significant content is, however, frequently vitiated by rhetorical overstatement. The authorship of *Hercules Oetaeus* is problematical. *Octavia praetexta*, a play on a contemporary historical theme, cannot on grounds of subject matter and style be by Seneca, but was probably written by an admirer in the next generation.

Seneca came from a provincial origin to the highest eminence in Rome and was one of the most successfully versatile men in antiquity. It is impossible to resolve the contradictions in his life and work between opportunism and enlightenment, financial unscrupulousness and philosophical resignation. His influence was great: to the Latin Church Fathers he was a salutary castigator of vice and to Renaissance writers a model for drama and a supreme moral teacher.

BIBLIOGRAPHY.—Works ed. by E. Hermes *et al.* in the Teubner series (1902–14); and by F. J. Miller *et al.* with Eng. trans. in the Loeb series (1917–35); epistles also ed. by L. D. Reynolds (1965); *Apocolocyntosis* ed. by F. Buecheler and W. Heraeus, *Petronii saturae*, 6th ed. (1922). *See* also J. W. Duff, *A Literary History of Rome in the Silver Age* (1927); M. Schanz and C. Hosius, *Geschichte der römischen Literatur*, vol. 2, 4th ed. (1935); on tragedies *see* M. Coffey, *Lustrum*, vol. 2, pp. 113–186 (1957); on their influence, W. H. Clemen, *English Tragedy Before Shakespeare* (1961). On *Apocolocyntosis*, *see* M. Coffey, *Lustrum*, vol. 6, pp. 239–271 (1961). *See* also M. Pohlenz, *Die Stoa*, 2nd ed. (1959); S. Dill, *Roman Society from Nero to Marcus Aurelius* (1904); E. Albertini, *La Composition dans les ouvrages philosophiques de Sénèque* (1923). (ML. C.)

SENECA (from *Sinnekens*, a Mahican term for Oneida, applied to the upper Iroquois tribes by the Dutch after 1635), a tribe of North American Indians of Iroquoian stock in western New York State. In the 1960s their population exceeded 4,100. To themselves and their confederates they were "great hill people," named for the site of their original village at Canandaigua Lake, where it is said "they sprang out of the ground." As the fifth fire of the Iroquois "Longhouse" (*see* IROQUOIS) they were "keepers of the western door," and the first stop for delegations of "Far Indians" visiting the great council fire at Onondaga (*q.v.*). Their original territory between Seneca Lake and the Genesee River expanded during the 17th century when they destroyed the Neuter and Erie nations to occupy western New York State entirely from Niagara south along the Allegheny River into Pennsylvania. Remote from white contact, secure in game and corn (maize), they could afford to field 1,000 warriors who equaled the combined strength of the rest of the Iroquois League. Never bowing entirely to French or English power, they would remain the League's most important and conservative tribe.

Starting with two prehistoric communities that moved about twice each generation as soil and firewood were exhausted, they expanded into satellite settlements. Throughout history the Seneca have had a leading chief in each moiety (*see* DUAL ORGANIZATION); these offices have descended from the leaders of the two village bands that predate the League. The "Longhouse" was later extended to shelter two more reluctant village chiefs who were made doorkeepers. In New York two great towns (*Kandagaro*, St. Jacques, near Victor; *Sonnontouan* or *Tiotihakton*, Conception, Rochester Junction) and two lesser communities of Huron and Neuter captives (*Gandougarae*, St. Michel, East Bloomfield; and *Gandachiragon*, St. Jean, Lima) persisted during the period of Jesuit missions (1655–87) when J. R. de Brésay, the marquis of Denonville's, expedition destroyed them. For eight years Seneca warriors had composed more than half the war parties that successively destroyed Huron (*q.v.*), Tobacco, Neuter, and Erie nations, making up their own losses by adopting whole towns, until by 1657 they had incorporated 11 different tribes. A century later they were dispersed in 30 villages as far as Ohio. The number of real Seneca dwindled in a population of 5,000 that filled about 300 houses, comprising 900 fireside families of possibly six persons each. These families were linked by maternal kinship as longhouse families and ultimately as eight clans, four in each moiety. One married outside his clan and preferably in the other moiety; both moieties had mainly ceremonial functions in games and funerals.

This social system survived the French and Indian War, Pontiac's conspiracy (*see* PONTIAC), and the American Revolution (in which the Seneca were British allies, resulting in the destruction of their villages by American General John Sullivan [*q.v.*] in 1779). A losing struggle to retain their lands followed the 1784 Treaty of Ft. Stanwix, when their chief Cornplanter (John O'Bail) compromised western Seneca territory. Relief from such humiliation finally came in the treaty of "peace and friendship" signed at Canandaigua (1794), long regarded as basic to land titles and tribal rights. Twelve parcels of land were reserved three years later at Big Tree; four of these reservations (Tonawanda, Cattaraugus, Oil Springs, and Allegany) existed in the 1960s.

At that time chiefs still governed about 700 at Tonawanda, but the Seneca Nation itself was a republic. Councilors were elected by adult males in a population of 3,300 (two-thirds registered at Cattaraugus, and one-third at Allegany, many actually living and

working scattered over the U.S.). In northeastern Oklahoma lived 900 so-called Seneca-Cayuga, descended from the Seneca of Sandusky, Ohio. Though most Seneca were nominally Christian, the followers of Handsome Lake (the Seneca prophet) numbered 800 and maintained the old ceremonies at three longhouses in western New York. The Seneca language was expected to survive for another generation.

Completion of the Kinzua Dam in December 1965, with resulting inundation of Allegany reservation bottomlands along the Allegheny River, forced relocation of the 482 bottomland Seneca. The U.S. appropriated $15,000,000 for this purpose and for rehabilitation, including the building of modern homes for the tribe at Steamburg, N.Y., and elsewhere. *See* also CAYUGA.

BIBLIOGRAPHY.—J. Curtin and J. N. B. Hewitt, "Seneca Fiction, Legends, and Myths," *32nd Ann. Report, Bur. Am. Ethnol.* (1910–11); W. N. Fenton, "Toward the Gradual Civilization of the Indian Natives . . . ," *Proc. Am. Phil. Soc.*, vol. 100 (1956), "The Iroquois Confederacy in the Twentieth Century . . . ," *Ethnology*, vol. IV (1965); A. C. Parker, "An Analytical History of the Seneca Indians," *Res. and Trans. New York State Archaeological Association*, vol. 6 (1926); Edmund Wilson, *Apologies to the Iroquois* (1960); L. H. Morgan, *League of the Iroquois* (1851; new ed., 1962); W. L. Chafe, *Handbook of the Seneca Language, Bull. N.Y. St. Mus. Sci. Serv.* 388 (1963).

(W. N. F.)

SENECIO, the groundsel or ragwort genus of plants, of the composite family (Compositae); one of the most numerous of all flowering-plant genera, it includes about 2,000 species, among them many coarse herbs and shrubs that occur as weeds round the world and several that are cultivated for their flowers, foliage, or climbing growth habit. The leaves are alternate and sometimes all basal. The flower heads are single or clustered, variously coloured, and usually with fertile rays, but sometimes rayless. The surrounding bracts (involucre) are usually arranged in a single series, reinforced at the base with calyxlike shorter bracts (bracteoles). The dry fruits (achenes) are crowned with a pappus of soft, whitish, often very numerous, bristles.

More than 70 species occur in North America, most numerously in the southern and western parts of the United States. Representative American species are the strong-growing *S. aureus* (golden ragwort or swamp squaw-weed), a conspicuous, golden-yellow wild flower, a one- to two-foot-high perennial inhabiting swamps and wet meadows; *S. pseudo-arnica* (seabeach groundsel); *S. obovatus* (round-leaved squaw-weed); *S. pauperculus* (balsam groundsel); and *S. lobatus* (cress-leaved groundsel). Particularly noxious weed species, both in North America and in Europe, are *S. vulgaris* (common ragwort or common groundsel) and *S. jacobaea* (stinking Willie or tansy ragwort), which move readily into fields, pastures, and roadsides. Among the more widely cultivated species are *S. cineraria* (dusty miller); *S. cruentus* (cineraria, *q.v.*); *S. petasitis* (velvet groundsel); and *S. mikanioides* (German ivy), a native of South Africa, naturalized in California.

See also GROUNDSEL.

SENEFELDER, ALOYS (1771–1834), German inventor of lithography, was born at Prague on Nov. 6, 1771, the son of an actor. An author of plays, for the engraving of which he was unable to pay, Senefelder set about engraving them himself. His work on copper plates, however, was not very successful. Further, they were so expensive that he was forced to reuse them, and the necessary grinding and polishing were extremely tedious. His first work on stones began with a fine piece of Kelheim stone he had bought for grinding ink. Since it could be ground and polished easily, he began to use it for practice in writing backward. His discovery of its possibilities for etching came about almost by accident (*see* LITHOGRAPH; LITHOGRAPHY). Senefelder continued his experiments with lithography until the art reached a high degree of excellence. He died at Munich on Feb. 26, 1834, having lived to see his invention brought to comparative perfection.

SENEGAL, REPUBLIC OF (RÉPUBLIQUE DU SÉNÉGAL), since 1960 an independent state, formerly a territory of French West Africa, extends between latitude 12°18′ and 16°41′ N and longitude 11°21′ and 17°32′ E. Area 76,124 sq.mi. (197,161 sq.km.). It is bounded west by the Atlantic Ocean, north by the Senegal River separating it from Mauritania, east by the Falémé River (an affluent of the Senegal) separating it from Mali, and south by Portuguese Guinea and the Republic of Guinea. The state of Gambia forms an enclave in Senegal along the Gambia River, thus isolating the Casamance region, which is the southern part of Senegal. The capital is Dakar.

Physical Geography.—Crystalline bedrock outcrops in the Bondu (Boundou) and Bambouk (respectively in the lower valley of the Falémé and to the west of it, and in the middle valley of the same river and to the east of it) toward the Guinea frontier, where it is surmounted by knolls of Primary sandstones and dolerites of the Fouta Djallon. Most of the country corresponds

MAJOR CITIES AND PHYSICAL FEATURES OF SENEGAL

with the sea gulf of the Cretaceous and early Eocene and consists of sandy clays more or less decomposed into laterite, which gives light soils in all regions. During the Quaternary period, climatic changes brought about the formation of terraces, alluvium estuaries, and Continental and maritime dunes. The sandy clay plains are gently undulating, surmounted by sandstone knolls and dissected by valleys formed by rivers drying up (as in the Sine region between the Dakar Peninsula and the Saloum River) or dried out (as in the Ferlo in the north). The coast, swept by the trade wind and by the prevalent swell from the northwest, is in places bordered by cliffs and by belts of Continental dunes isolating marshy depressions known as *niayes*. Such belts have linked volcanic peaks with the coast to form the peninsula of Cape Verde.

The climate over most of the country is Sudanese (transitional between dry desert and moist tropics); north of a line joining Thiès with the lower Falémé it is Sahelian (peculiar to the edge of the southern Sahara). There are two main seasons: dry and wet, separated by two transition periods. The rainy season lasts from June to October in the Sahelian zone, from May to October in the Sudanese zone, and from May to December in Casamance, with annual rainfalls of about 14 in. (355 mm.) in the north and about 60 in. (1,524 mm.) in the south. It is preceded by a series of tornadoes. Dakar has maximum and minimum average temperatures of 27° and 18° C (81° and 64° F) in January and 33° and 25° (91° and 77°) in August, but the temperature rises rapidly inland, where the trade wind and sea breezes are replaced by the harmattan, a dry warm wind that blows from the east-northeast from January to May.

The vegetation varies with the latitude and climate. In the north the Sahelian zone is characterized by scattered shrubs and spiny acacias and by pasture land of various gramineous plants. South of Cape Verde the savanna grows steadily thicker (African mahogany, silk-cotton trees, palmyras, and baobabs). On the coast are true oases formed by the *niayes* and relict forests of Guinean type, which are also found in Casamance where the tropical forest persists locally. In central and south Senegal the soils are generally leached, by rainfall or because of permeability. Elsewhere ferruginous or sandy belts create infertile soils. In the north the soils are less leached and include fine ferruginous gravels or, in the *tannes* ("marshy depressions"), are saline.

The animal life is Sudanic in type: there are several species of monkey and antelope, warthogs, lions, panthers, hyenas, wild hunting dogs, and hippopotamuses. Most of the large game has been seriously reduced and exists only in the reserves and in the national park of Niokolo-Koba in the neighbourhood of Tambacounda. The numerous birds include partridges, guinea fowl, spur-winged geese, and also the tiny *mangemil*, which does great damage to crops. Among reptiles are crocodiles, turtles and tortoises, monitors, chameleons, huge boas and pythons, and the hooded cobra. Great centipedes, millipedes, scorpions, and numerous kinds of insects are common. (J. D.)

The People.—True Senegalese peoples include the Wolof, Serer, Lébou (Lebu), Fulani (Peul), and Tukulor (Toucouleur). The Wolof, numbering about 1,200,000, inhabit the more populous west. They are very tall, slender, and deep black. They seem to have come from the Sahara and they established formal kingdoms with a well-developed hierarchy. Their society is patrilinear but shows traces of matrilinear origins. They grow millet and peanuts and some are traders. The Serer (*q.v.*), numbering about 500,000, resemble the Wolof but are smaller in stature; they are renowned as farmers. The Lébou (40,000) also resemble the Wolof; they form a group of fishermen in the Dakar region. The Fulani (*q.v.*) number about 400,000. They are light skinned. In the Ferlo they are nomadic herdsmen, but in Casamance they are settled and semiagricultural, and have much intermarried with Negro peoples. The Tukulor (*q.v.*) number about 350,000; they occupy the Senegal River district and have spread southward.

Sudanese peoples in Senegal include about 200,000 Malinke (*see* MANDINGO) in the east, and about 50,000 Sarakolé (Soninke) on the Senegal River around Bakel. These two groups, who cultivate millet, often find work in Dakar and neighbouring towns. The Diola (150,000) of lower Casamance are related to the coastal paleonegritic folk of Portuguese Guinea. They are industrious farmers of the swamps, jealous of their independence and without extensive political organization. The residential Moors, apart from many who come to trade in Senegal, comprise about 20,000, mostly along the river.

The Wolof, Serer, and Lébou languages are distinct but doubtless related. The Fulani and the Tukulor both speak Peul (Poular). The Malinke and Sarakolé speak languages of the Mandingo group; the Diola have separate dialects. French is spoken by the ruling elite, but Wolof is spreading as a commercial language.

The dominant religion is Islam, with a predominance of the Tijaniya (Tidjanist) sect, and around Diourbel, of the Mouride sect. The Diola and a large section of the Serer are pagans, with a few Christians. The houses are generally round, of clay in the east and of straw in the west. The Diola have huge family houses, rectangular in shape and built of clay. They wear few clothes, whereas the other peoples wear ample *boubous* (long robes), with jewels and ornamented leather objects.

Population.—At the 1960 census Senegal had a population of 3,109,840 including 61,700 foreigners (more than half of whom were French). The average density was 40.9 per sq.mi. (15.8 per sq.km.) but distribution varied considerably, being 10 or below in the river area, about 1 in the Ferlo, reaching 50 in the hinterland of Dakar and Thiès, 400 in the Cape Verde Peninsula, about 2 in the east, and between 5 and 15 in Casamance.

Apart from Dakar (*q.v.;* pop. [1958] 194,605, commune) the chief towns are Rufisque (47,000, commune), Thiès (36,246, commune), and Saint-Louis (*q.v.*), the former capital with 37,104 inhabitants. Kaolack (*q.v.*) is the port of the Saloum River and Ziguinchor (28,483) is the chief town in Casamance.

History.—Man has occupied Senegal since very ancient times. Paleolithic axes, Neolithic two-headed axes and arrows have been found near Dakar, and stone erections as well as copper and iron objects in the Sine-Saloum region.

The first Negroes of whom there is historical mention are the Tukulor, who occupied the Senegal Valley in the 11th century. The name Senegal appears to be derived from that of the Zenaga Berbers of Mauritania. The Sarakolé empire of Ghana lay to the east, between the Senegal and the Niger. Toward 1040 the Zenaga established, perhaps on an island of the river, the Muslim monastery in which the Murabti sect was to have its origin. This sect converted the Tukulor, destroyed Ghana, and conquered Morocco. Pagan invasions began in the 13th century. About 1400 some Fulani founded a dynasty on the middle Senegal, in the region thereafter known as Fouta-Toro. In the 18th century the Muslim Tukulor revolted and set up a feudal theocratic republic.

On the coast and in the Cayor the Wolof founded an independent kingdom; this was later divided into four states: Dyolof, Walo, Cayor, and Baol, which were often at war with each other and with their neighbours. As well as the warrior aristocracy, the Muslim marabouts also had some influence there. The Serers in the Sine and in the Saloum were organized into states by Mandingo chiefs, the Guellawar. The other peoples did not form any large states.

Portuguese navigators reached Cape Verde about 1444. They set up factories at the mouth of the Senegal, at Gorée (opposite the modern Dakar), at Rufisque (Rio Fresco), at Joal, at Portudal, in the Saloum, and in Casamance (Casa Mansa).

In 1638 a French factory was set up at the mouth of the Senegal and was rebuilt, in 1659, on N'Dar, an island in the river, which was to become Saint-Louis du Sénégal. The Dutch meanwhile had set themselves up on the island of Gorée (*q.v.*), which the French took over in 1677, after which the French agent J. B. Ducasse managed factories on the coast and did much to promote trade. Dealers from Saint-Louis began business on the lower reaches of the river, where they bought gums and slaves. André Brue spent 25 years trying to penetrate the hinterland by journeys up the Senegal. In 1700 he built Fort Saint-Joseph near the Falémé confluence and began to buy gold from the Bambouk. After his death this fort was destroyed by the Negroes. The French botanist Michel Adanson (1727–1806) spent five years exploring the interior before publishing his *Histoire naturelle du Sénégal* (1757).

The English occupied the factories during the Seven Years' War, and gave only Gorée back at the conclusion of peace in 1763. The duc de Lauzun recovered Saint-Louis in 1779; one of its first governors was the chevalier de Boufflers (1738–1815), whose love letters to Mme de Sabran contained descriptions of the country. François Blanchot de Verly, who became governor in 1787, had to endure a blockade and great privations but nevertheless was able to repel several English assaults. He died in 1807, and Saint-Louis surrendered two years later. The English governor had a mausoleum built for the gallant Blanchot.

In 1816 Saint-Louis and Gorée were returned to France. In the same year, on an expedition to Senegal, "La Méduse" was shipwrecked; this incident gave rise to Géricault's famous painting "Raft of the Medusa." Julien Schmaltz, as governor, founded a station up the river at Bakel. Schmaltz also tried to grow cotton in the neighbourhood of Saint-Louis, and his successor Baron Roger continued this experiment, but it came to nothing; after which trade was confined to gum. In 1818 Gaspard Mollien crossed the Ferlo and reached the Fouta Djallon, discovering the sources of the Senegal and Gambia rivers. Between 1835 and 1837 two stations, Carabane and Sedhiou, were acquired in Casamance.

The precarious position of the French, local wars, and the tribute paid to the Negro kinglets, all combined to reduce trade to very little. Under the Second Republic the slaves were emancipated and given French citizenship. In 1854, at the request of the businessmen of Saint-Louis, Comdt. Louis Léon César Faidherbe (*q.v.*) was appointed governor. To put an end to insecurity Faidherbe pacified the Walo east of Saint-Louis and then drove back the Moors on the right bank of the river. But the Tukulor marabout Omar el-Hadj was threatening the river. Faidherbe established a station at Médine (near Kayes) and left there a small garrison under Paul Holle, a mulatto from Saint-Louis. In 1857 Holle strenuously resisted attacks of the marabout's thousands until the rising of the high water enabled Faidherbe to reach him with reinforcements. In 1857 Captain Protet, of the French Navy, founded a station at Dakar. Faidherbe occupied the Cayor, thus uniting parts of the colony. In 1865 he went back to France, having organized Senegal and laid the foundations of native policy based on respect for existing usages.

His successor, Col. J. M. E. Pinet-Laprade, consolidated his work, occupying the Petite Côte and the Saloum and developing the town of Dakar. Peanuts had been introduced into the Cayor and were being exported from Rufisque. The Gambian fort of Albreda had been ceded to Great Britain in 1857. In 1886 the last Wolof king, the damel of Cayor, Lat Dior, had been killed during an insurrection, and Col. J. S. Galliéni was destroying the power of the marabout Mamadou Lamine in the east. Faidherbe's creation of the Tirailleurs Sénégalais (Senegalese riflemen) marks the beginning of France's use of Senegal as a base for expansion into the Sudan and toward the Rivières du Sud. Dahomey was conquered (1892–94) by Col. (later Gen.) Alfred Amédée Dodds (*q.v.*), another mulatto soldier from Saint-Louis who had subdued the Dyolof and the territory of the Tukulor. The pacification of the traditionally anarchic Lower Casamance was not achieved until 1903. The Gouvernement Général de l'Afrique Occidentale, established by decrees of 1895 and 1904, was at first the responsibility of the governor of Senegal, resident at Saint-Louis, but was later raised to the status of a separate authority based on Dakar. The latter town grew rapidly, to become the main port for Senegal and for the French Sudan.

In World War I the Negro deputy for Senegal, Blaise Diagne, helped to enlist large numbers of riflemen. In World War II the governor-general, Pierre François Boisson, remaining loyal to the Vichy government, held Dakar against a British attack in 1940; but he transferred his allegiance to the Allied side in November 1942. In 1946 all Senegalese became French citizens, whereas previously this right had been restricted to natives of the four ancient *communes*, Saint-Louis, Gorée, Dakar, and Rufisque; and the colony became an overseas territory of France. The two Senegalese deputies, Lamine Gueye and Léopold Sédar Senghor, played a prominent role in the drawing up of the French Constitution of 1958. In 1958 Senegal became an autonomous republic within the French Community; in September 1960, after a short period of federation with the Sudanese Republic in the Mali federation, Senegal was recognized as an independent state, remaining within the French Community. Senghor was elected president of the Republic. In December 1962 a tentative *coup d'état* by the president of the Council of Ministers, Mamadou Dia, failed and Senghor took over the office. (HU. DE.)

Administration and Social Conditions.—Under the 1963 constitution Senegal is a presidential republic, in which a National Assembly of 80 members elected by universal suffrage exercises legislative power. The assembly, together with the delegates of the provincial and communal councils, elects the president. The president appoints the prime minister. The ministry is responsible to the assembly.

Administratively the republic is divided into seven regions (Kaolack, Sine-Saloum, Thiès, Fleuve, Sénégal Orientale, Cap-Vert, and Casamance), each with a governor and local assembly and each subdivided into *départements*.

Elementary and secondary education are provided by schools and colleges, some of which are conducted by missionary bodies, with an increasing enrollment. Further education is provided by *lycées* at Dakar and Saint-Louis, technical and training colleges, and by the University of Dakar established in 1957.

Economy.—The traditional agricultural economy varies in the different regions. Cattle raising is practised by the Fulani in Bondu, and also in the Ferlo where a scattered bushy vegetation allows widespread seasonal migration of stock. Agricultural techniques vary; thus the soils along the Senegal River are cultivated by using the floodplain after the waters have subsided; by building dikes; or by dry farming. Millet cultivation is associated with cattle raising. The savanna regions are cultivated by the slash and burn method and with long fallow periods, although the Serer practise crop rotation (millet, peanuts, fallow). Irrigated cultivation is used in the *tannes* of the Sine district and in Casamance by the Diola, who have managed to polder the estuarine mud by arrangements of canals, sluices, and dikes. Fishing is carried on by the Lébou of the Cape Verde Peninsula and by the Guet N'Dar fishermen at Saint-Louis, and also in the rivers.

The economy introduced by Europeans is one of the oldest in Africa. Trade, initially concerned with slaves, developed chiefly in gum and peanuts. The soil, the population density, and the existence of maritime outlets encouraged peanut cultivation to an extent that transformed the country. First practised in the north, peanut cultivation spread along the routes of communication and toward the wetter south, to such an extent that labourers (*navétanes*) had to be brought from neighbouring countries. At one time, indeed, the cultivation of staple foodstuffs was neglected, the vegetation stripped, and the soil exhausted. But the use of mineral fertilizers and the distribution of selected seed and of agricultural implements resulted in increased yields, especially after World War II. By the mid-1960s annual harvests averaged 900,000 metric tons. Traditional methods have been improved by horse-drawn plows, and the establishment of regional (cooperative) development centres and some mechanized agricultural experiments. Harvesting activity reaches a peak from November onward.

The risks inherent in a single-crop economy and the growing needs of the population compelled the authorities to develop other food crops and to utilize new acreage. Work was undertaken to drain the *tannes* of the Saloum River and to increase the productivity of the Senegal Valley. Also, the cultivation of rice in the Senegal Delta was extended by pumping water from a dam on the Tauey effluent of Lake Guiers. Increased peanut production and the development of the port of Dakar (*q.v.*) encouraged small consumer and construction industries in that city and its environs. But lack of mineral resources precludes the development of heavy industries; the only minerals extracted are salt (Kaolack), ilmenite, and zircon near the coast (Rufisque) and phosphates of lime and of alumina near Thiès.

A four-year development plan (1961–64) was launched to encourage capital investment, especially in industrial development.

The chief imports of Senegal are wheat, rice, sugar, petroleum products, textiles, and machinery. The chief exports are peanuts, peanut oil, oilcake, phosphates, and titanium concentrates. Chief trading partners are France and countries of the franc zone, and the Federal Republic of Germany. As a member of the Association of African States, Senegal signed a convention of association with the European Economic Community in July 1963. Senegal is also a member of the West African Monetary Union (established in May 1962). Its currency is based on the franc CFA (1 fr. CFA = 0.02 French francs).

Communications.—Initially, links with the interior were maintained by the Senegal River, along which ports of call were established and on the estuary of which lies Saint-Louis. But the river is navigable only during the summer flood season, when there is little to transport. Therefore between 1880 and 1923 the metre-gauge Dakar-Niger Railway was built in stages, and extended in 1922–31 by branch lines to Kaolack, Linguère, and Touba. This redirected the flow of trade to Dakar, which has a well-equipped port and an airport used by international and local services. But the river ports of Kaolack and Ziguinchor were little developed. Road links between Dakar and the countries adjoining Senegal were completed after World War II. Radio diffusion du Sénégal, in Dakar, broadcasts in French and a number of vernacular languages and the Dakar area has a television service.

See also references under "Senegal, Republic of" in the Index.

(J. D.)

BIBLIOGRAPHY.—J. Richard-Molard, *Afrique Occidentale Française* (1952); R. J. H. Church, *West Africa* (1962); V. M. Thompson and R. Adloff, *French West Africa* (1958); C. R. Niven, *Land and People of West Africa* (1958); R. Theobald (ed.), *New Nations of West Africa* (1960); J. D. Fage, *Introduction to the History of West Africa* (1960); W. Fitzgerald, *West Africa* (1961); M. Crowder, *Senegal: a Study in French Assimilation Policy* (1962); H. Deschamps, *Le Sénégal et la Gambie* (1963); current history and statistics are summarized annually in *Britannica Book of the Year*.

SENEGAL RIVER in West Africa flows into the Atlantic Ocean. It is 1,015 mi. (1,633 km.) long and is formed by the union of three rivers of which the Bafing (Black River) and the Bakoy (White River) rise in the Fouta Djallon Plateau and drain north, joining at Bafoulabé; tributary to the Bakoy is the Baoulé. The principal tributary, the Bafing, flows north from the southern slopes of the Fouta Djallon in a generally enclosed valley,

its course including a series of rapids and falls, notably the Gouina and Félou. Below Kayes, 559 mi. (900 km.) from its mouth and only 197 ft. (60 m.) above sea level, the river (known as the Senegal below Bafoulabé) flows through a wide valley named Fouta Toro between Bakel and Dagana. It forms the frontier between the republics of Mauritania and Senegal; the right-bank flood plain in Mauritania is known as Chemama. The only major left-bank tributary is the Falémé; right-bank tributaries are of no significance in this part of its course as the valleys of Mauritania are dry most of the year. At low water the river flows in *marigots* (lateral channels) between islands of sand, *e.g.*, Île à Morphil, and alluvial banks which isolate the main river bed from the flood plain. Downstream from Kaédi the river passes through a region of northeast-southwest trending dunes. Below Dagana lies the delta with its many distributaries; the Senegal estuary is blocked by a sandbar and a 15-mi.-long sandspit, Langue de Barbarie. Near the head of the delta two minor tributaries of the Senegal drain the lakes Rkiz (Cayar) to the north and Guiers to the south.

The headstreams of the Senegal drain the well-watered plateau region where annual rainfall varies between 40 and 80 in. (1,000–2,000 mm.), but below Bakel rainfall is less than 20 in. (500 mm.). Percolation and evaporation contribute in reducing the total discharge of the river. The regime is seasonal; summer floods begin upstream in July but are not felt at Dagana until the end of August. The floods vary considerably in volume; in September they usually exceed 105,900 cu.ft. (3,000 cu.m.) per sec. but many reach as much as 247,200 cu.ft. (7,000 cu.m.) per sec. at Bakel, and the water there rises 30 ft. (9 m.) above low-water mark. The river is navigable throughout the year as far as Podor but during the floods (for 2½ months) Kayes can be reached. Because of the sand obstructions at the mouth and its seasonal nature the river is little used for transport except by local craft. Water level drops considerably during the dry season; the rate of flow at Bakel drops to 106 cu.ft. (3 cu.m.) per sec. at the close of the dry season. At these times the tide is felt above Dagana and sometimes above Podor.

In 1947 under the direction of the *Mission d'Aménagement du Sénégal* (MAS) a scheme to irrigate the flat land to the southwest of Richard Toll was begun. A barrage on the Taouey River prevents salt water from penetrating Lake Guiers in the dry season. Further work has been undertaken to make the valley more productive, and a project was prepared to use the falls of Félou and Gouina to regulate the water level and produce energy. (J. D.)

SENESCHAL (Lat. *senescalcus*, from Old Teutonic *seni*, "old," *skalko*, "servant," the oldest and thus the senior servant) was, from Merovingian times, the official who had charge, under the major domus or mayor of the palace, of the personnel of the royal household. Under the Carolingians he became responsible for organizing the journeys of king or emperor, and by 1071 was steward of the household, head of the army, and administrator of the royal demesne. With the title *dapifer* he headed the names of those witnessing royal diplomas. In 1107 the office passed from the powerful family of Rochefort to that of Garlande. But Étienne Garlande, already chancellor when he became seneschal in 1120, was deprived in 1127, and Louis VI gave the office (1131–52) to his own cousin Raoul de Vermandois. When Thibaut V, count of Blois, was seneschal (1152–91) the office seems to have been only honorary, and on his death no new seneschal was appointed.

Most of the great French feudatories, the dukes of Normandy and Aquitaine, the counts of Anjou, Poitiers, and Toulouse, had their seneschals. In Anjou and Poitou the title was given to subordinate officials on the counts' demesne. When these provinces became united to the crown after 1203, these officials were retained to perform the same duties as the bailiffs of the royal demesne elsewhere. This was the basis of the later classic administrative division of France into *bailliages* and *sénéchaussées*, although in Normandy, Maine, Anjou, and Auvergne the seneschals were replaced by bailiffs.

BIBLIOGRAPHY.—*Gallia regia, ou, état des officiers royaux des bailliages et sénéchaussées de 1328 à 1515*, 6 vol. (1942–61); J. Boussard, *Le Gouvernement d'Henri II Plantagenêt* (1956); F. Lot and R. Fawtier, *Histoire des institutions françaises au Moyen Âge*, vol. i; ii (1957–58). (MI. F.)

SENIGALLIA (SINIGAGLIA; anc. SENA GALLICA), a seaside resort and episcopal see of the Marches, central Italy, in Ancona Province, lies on the Adriatic coast, 17 mi. (27 km.) NW of Ancona town. Pop. (1961) 34,884 (commune). Principal buildings include the castle (1480); the Convento Delle Grazie (15th century) with a Madonna by Perugino; the 17th-century Palazzo Comunale and the Chiesa della Croce with a Deposition by Barocci; and Portici Ercolani (18th century). The town is a market centre of a predominantly agricultural area, and manufactures cement, bricks, agricultural implements, paper bags, and furniture. It is on the Adriatic state road and on the Bologna-Bari railway, and is served by the airfield at Falconara, 11 mi. (16 km.) SE. It has a small port. Sena Gallica was founded by the Senonian Gauls in the 6th century B.C., and became a Roman colony *c.* 290 B.C. In the 6th century A.D. it was one of the five cities of the Pentapolis. It was destroyed by the Arabs in 1264 and rebuilt by Sigismondo Malatesta of Rimini in 1450. Its lordship was assigned to the Della Rovere family by Pope Sixtus IV (late 15th century). It became part of the papal states from 1631 to 1860, and in this period was famous for its fair. Pope Pius IX was born at Senigallia in 1792. (R. CI.)

SENLIS, a small town of northern France, headquarters of an *arrondissement* in the *département* of Oise, 27 mi. (43 km.) NNE of Paris by road. Pop. (1962) 8,717. It is situated on the right bank of the Nonette River, a tributary of the Oise, among the extensive forests that occupy the sand cappings overlying the Tertiary limestone of the Île-de-France. Senlis was part of the royal domain and a royal residence from the time of Clovis (d. 511), and Hugh Capet was proclaimed king there in 987. The town has an inner perimeter of massive Gallo-Roman walls; and the remains of an outer ring of medieval walling. Senlis is rich in medieval buildings, including the ruins of the royal castle and the former cathedral church of Notre Dame, one of the finest achievements of Gothic architecture, which was begun on a grand scale in 1155, but not completed until the 16th century. Senlis is now a much frequented, convenient resort for Parisians. (AR. E. S.)

SENNA, a popular purgative, consisting of the leaves of *Cassia acutifolia* or *C. angustifolia*, plants of the pea family, Leguminosae. These are small shrubs about 2 ft. high, with numerous lanceolate leaflets arranged pinnately on a main stalk with no terminal leaflet; the yellow flowers are borne in long-stalked racemes in the leaf axils and are succeeded by broad flattish pods about 2 in. long. *C. acutifolia*, native to tropical Africa, is the Alexandria senna of commerce. *C. angustifolia* is the Bombay, East Indian Arabian, or Mecca senna of commerce. This plant grows wild in Arabia, Somaliland, Pakistan, and the Punjab, India. It is also cultivated in the extreme south of India, and there grows larger leaves, known commercially as Tinnevelly senna.

The laxative principles are apparently two glycosides, sennoside A and sennoside B, and the anthraquinone, emodin. Senna was introduced by the Arabs, who described its cathartic properties as early as the 9th century.

SENNACHERIB (Akkadian SIN-AKHKHE-ERIBA) (d. 681 B.C.), king of Assyria from 705 to 681 B.C., was the son and successor of Sargon II (*q.v.*). His most enduring work was the rebuilding of Nineveh and its adoption as the capital, which it remained until the fall of the Assyrian Empire. He carried out a great scheme of town planning, laying out streets, widening the squares, and providing the palace area with a wall of limestone slabs for flood protection. The palace itself, with paneling and doors of aromatic woods, was of great splendour, showing some features of Syrian construction. By the palace a botanical garden and orchards were laid out, a canal being cut for six miles to bring water to them. Sennacherib expressly mentions that this splendour was created through the slave labour of conquered peoples. (*See* further NINEVEH.) This king was keenly interested in technology, and he mentions a new method of metal casting devised in his reign and the discovery of new mineral resources.

The peace bequeathed by Sargon was broken after two years by the insurrection in Babylonia of Merodach-baladan, who had

united the Chaldean and Aramaean tribes behind him, secured the support of Elam, and worked for a widespread anti-Assyrian coalition involving, among others, Hezekiah of Judah (Isa. 39:1; II Kings 20:12). Merodach-baladan occupied Babylon and held other cities in northern Babylonia with Elamite troops. By clever generalship Sennacherib cut Elamite communications, whereupon northern Babylonia fell into Assyrian hands, Babylon receiving Sennacherib with enthusiasm (703 B.C.). The Assyrian army then occupied and defortified the Chaldean area in south Babylonia, which was left under Assyrian officials, a native Babylonian, Bel-ibni, being made king in north Babylonia. A rebellion broke out in 701 B.C. in Palestine, Hezekiah of Judah, in alliance with Egypt, being involved (*see* HEZEKIAH). A strong Assyrian force overran Palestine, defeated the Egyptian troops and took the rebel cities, with the exception of Jerusalem, which was spared on payment of a heavy indemnity (*cf.* II Kings 18–19; it was 19:35 that inspired Byron's well-known poem, "The Destruction of Sennacherib").

Meanwhile Merodach-baladan began fresh intrigues in Babylonia, so that in 700 B.C. another Assyrian foray into Chaldean territory was necessary. Bel-ibni, who had proved unable to maintain order, was replaced by Ashur-nadin-shum, a younger son of Sennacherib, who reigned for six years. Merodach-baladan died shortly after, but his Elamite allies, in whose territory disaffected Chaldean tribesmen were able to find sanctuary, remained a threat to Babylonian security. Sennacherib sent a fleet of ships down the Euphrates and attacked Elam by sea across the Persian Gulf (694 B.C.), looting the Elamite cities and taking prisoner survivors of Merodach-baladan's tribe. With the Assyrian forces still in the south, Elam reacted by a raid into metropolitan Babylonia, capturing Ashur-nadin-shum and replacing him with a Babylonian puppet. The returning Assyrian army dealt with the Elamite nominee, but a new Chaldean leader, Mushezib-Marduk, had emerged and held Babylon. After the Assyrian army had returned home, Mushezib-Marduk offered a considerable bribe and obtained massive military support from Elam. The Elamite and Assyrian armies met at the Diyala and a bloody carnage ensued, leaving the Assyrian army incapable of following up immediately the victory it claimed. In the following year, 689, the Assyrians were able to deal with Mushezib-Marduk, whose forces, after a nine months' siege in Babylon, were finally defeated. Sennacherib, departing from previous Assyrian policy, savagely looted and sacked Babylon.

Sennacherib died at Nineveh in January 681 at the hands of his sons. *See* also BABYLONIA AND ASSYRIA.

See Sidney Smith, *The First Campaign of Sennacherib* (1921); D. D. Luckenbill, *The Annals of Sennacherib* (1924). (H. W. F. S.)

SENNAR, a town in the Republic of the Sudan and the name of the area between Nubia (*q.v.*) and Ethiopia (Abyssinia). The town lies at the western end of the Sennar Dam on the Blue Nile, 110 mi. (177 km.) S of Khartoum. Pop. (1955–56) 8,093. Sennar as the name of the area was not known before the 16th century. It derived from Sennar, the capital of the Fung (Funj) kingdom, a Muslim dynasty known as the Black Sultanate (*sultana zerqa*), which with an 'Abdallab Arab viceroy ruled (1504–1821) most of what is now the Republic of the Sudan. The old capital town of Sennar, now uninhabited, was situated on the west bank of the Blue Nile a mile or so downstream of the site of the Sennar Dam and near Sennar railway junction, from where lines run west to El Obeid and Nyala, north to Khartoum, east to Kassala, and south to Ar Rusayris (Er Roseires).

The Sennar Dam, 9,922 ft. (3,025 m.) long with a maximum height of 130 ft. (40 m.), was completed in 1925 and irrigates the cotton and other crops in the Gezira (*q.v.*). It is sited on a rocky sill which provided a ford across the Blue Nile in Napatan and Meroitic times, and remains of both these periods have been found there. The ford was also presumably the reason for the siting of the Fung capital on an important route between the kingdoms of Darfur and Abyssinia.

The Fung kingdom replaced the southern Christian kingdom of 'Alwa, which had gradually collapsed after the Muslim conquest of the Christian kingdom of Mukurra (Dongola) in the 14th century. Who the Fung were is not yet certain, but it is believed that the ruling family were a cadet branch of the Muslim Bulala royal family of Bornu, who were expelled *c.* 1500 after a civil war and who, on arrival on the Nile, defeated the Arabs under the 'Abdallab at Arbagi. The name Fung is probably a corruption of the Shilluk word for strangers which was given to them on their arrival on the White Nile, and the name Hameg (Hamaj), which they gave to their indigenous subjects, came with them from Lake Chad, where it is used for the subject population. The Fung Chronicle, composed *c.* 1800, gives the names of the sultans and details of the later history. An indeterminate war was fought with Abyssinia (1618–19). In 1704, Jacques Le Noir du Roule, a French envoy to Abyssinia, was murdered at Sennar, and in 1744 an Abyssinian attack on Sennar was defeated (largely by the strategy of an exiled Darfur prince). In 1821 the degenerate Fung kingdom surrendered to Ismail Pasha without a struggle. Sennar then became the capital of a Turko-Egyptian province of the same name, until it was destroyed by the Mahdists after a siege in 1883. For some time it was the headquarters of a province under the Anglo-Egyptian condominium but was later reduced to a district. *See* also SUDAN, REPUBLIC OF THE: *History*.

See A. J. Arkell, *History of the Sudan* (1961). (A. J. AL.)

SENNETT, MACK (MICHAEL SINNOTT) (1880–1960), creator of the Keystone Kops and father of American slapstick comedy, was a dominant figure in the early era of U.S. film production. He was born at Richmond, Que., Jan. 17, 1880. Joining Biograph studios in 1909 after unsuccessful sorties into circuses and burlesque, he learned movie technique from D. W. Griffith (*q.v.*), then left to form his own independent Keystone Company late in 1911. At first he both directed and acted in his short comedies, many of which were improvised on the spot to take advantage of parades, fires, and similar events too costly to stage. Although his name is generally associated with bathing beauties and custard pies, the vintage Sennett films (1914–24) were often biting parodies or incisive satires that mocked the foibles of an increasingly mechanized society. He developed a coterie of clowns and comediennes that made the Keystone trademark world famous —Mabel Normand, Marie Dressler, Gloria Swanson, Roscoe ("Fatty") Arbuckle, Harry Langdon, Ben Turpin, and Charlie Chaplin among them. In 1914 Sennett's company produced the first American feature-length comedy, *Tillie's Punctured Romance*, and in 22 years he turned out more than 1,000 short subjects. The coming of sound and the advent of double features, combined with the Wall Street crash that wiped out his large personal fortune, severely crippled Sennett's style, and he retired in 1933. In 1937 the Academy of Motion Picture Arts and Sciences voted him a special award "for his lasting contribution to the comedy technique of the screen." He died in Hollywood on Nov. 5, 1960.

See Gene Fowler, *Father Goose* (1934); Mack Sennett, as told to Cameron Shipp, *King of Comedy* (1954). (A. KT.)

SENOI, a Veddoid people (*see* VEDDA) found in the Malay Peninsula and in small groups along the coastal plains of Siak in Sumatra. In the 1960s they were estimated to number between 20,000 and 27,000. Traces of such a people also appear in the eastern islands of Indonesia. Apparently they are remnants of a once widespread population that followed the Negrito into this area. They are sometimes known as Sakai, a derogatory term applied by the Malays. In the peninsula they speak a language apparently related to or influenced by Mon-Khmer, but in Sumatra they have adopted the language and matrilineal institutions of the Menangkabau (*q.v.*). There are indications that the Senoi entered the area with a well-developed dry land agriculture; and that they lived in communal houses and had some degree of political development. Despite their cultivation of manioc and rice, they are also active in hunting, fishing, and food gathering. In the hunt they employ the blowgun and poison darts, like the neighbouring Malay and Negrito (*see* SEMANG) of the peninsula. Their highly developed religious beliefs and practices stand in sharp contrast to their rather primitive material culture. Where there has been little intermixture the Senoi differ markedly from their neighbours. They are slight in build, short in stature, with long heads, and generally have lighter skin colour and narrower noses than the Malay or Negrito (*qq.v.*). Their long hair is allowed to grow until

it falls in curls over their shoulders. Although skilled in basketry, they do not weave or do metalwork. *See* also MALAYSIA; MALAY ARCHIPELAGO: *Anthropology*.

BIBLIOGRAPHY.—H. H. Bartlett, "The Problem of Negrito and Vedda Elements in the Population of Sumatra," *Proceedings*, Fifth Pacific Science Congress (1934); F.-C. Cole, *Peoples of Malaysia* (1945); F. M. LeBar, G. C. Hickey, and J. K. Musgrave, *Ethnic Groups of Mainland Southeast Asia* (1964). (F.-C. CE.; X.)

SENONES, the name of two ancient Celtic tribes, or perhaps rather of two divisions of the same people, one living in Gaul, the other in Italy. The Gallic Senones lived in the area which includes the modern French *départements* of Seine-et-Marne, Loiret, and Yonne. They fought against Caesar (53–51 B.C.); in 51 a Senonian named Drappes threatened the province of Gallia Narbonensis, but was captured and starved himself to death. In later times they were included in Gallia Lugdunensis. Their chief town was Agedincum (later Senonus, whence Sens). The Senones in Italy crossed the Alps into Italy, perhaps *c.* 400 B.C., and settled on the east coast between Ariminum (Rimini) and Ancona, in the so-called *ager Gallicus*, driving out the Umbrians there. In 391 they invaded Etruria and besieged Clusium. Roman intervention, in reply to an appeal from Clusium, led to the Gallic capture of Rome (390). Livy and Diodorus, but not Polybius, record that the Senones led the Gauls that captured Rome. For the next century the Senones were hostile to Rome, but they were finally defeated and expelled by P. Cornelius Dolabella in 283. Their territory was used either for colonies—Sena (Senigallia) and Ariminum—or for the allotments for individual Roman citizens that were proposed by Flaminius in 232 B.C.

BIBLIOGRAPHY.—A. Holder, *Altceltischer Sprachschatz*, ii, 1485–98 (1904). *Gallic Senones:* T. Rice Holmes, *Caesar's Conquest of Gaul*, 2nd ed. (1911). *Archaeological Evidence:* Keune in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft*, 2nd series, vol. ii, 1488–94 (1923). (H. H. SD.)

SENS, a town of north-central France, centre of an *arrondissement* in the *département* of Yonne, Burgundy region, and an archiepiscopal see, lies on both banks of, and on an island in, the River Yonne, 70 mi. (113 km.) SE of Paris by rail. Pop. (1962) 19,692. The central part of the town is on the right bank of the river, and is enclosed by broad, wooded boulevards on the site of the old Roman fortifications, of which there are some remains. The chief monument is the Cathedral of Saint-Étienne (*c.* 1130–16th century), one of the earliest examples of the Gothic style. The architect of its nave and choir (completed *c.* 1168) was William of Sens, who later rebuilt the choir of Canterbury Cathedral. The west front has three portals with 12th-13th-century carvings, those above the central portal depicting the life of St. Stephen (Étienne). Only the southern tower was ever completed; it was reconstructed during the 14th-16th centuries. The cathedral has magnificent stained glass of the 12th–17th centuries, and its treasury, one of the richest in France, contains a collection of ancient fabrics and liturgical vestments, including those of St. Thomas Becket. It was in the cathedral of Sens that St. Louis, in 1234, married Marguerite of Provence. South of the cathedral are the 13th-century officiality (now a lapidarium), restored in 1849 by Viollet-le-Duc, and the mainly 16th-century archbishop's palace. The earliest sanctuary of Sens is the partly Romanesque Basilica of Saint-Savinien, the foundation of which dates from the 3rd century. The municipal museum and library have an important collection of Gallo-Roman sculpture, a 5th-century ivory diptych, and some precious manuscripts. There are several 16th-century houses, including the Jean Cousin museum of local history and the so-called "House of Abraham."

Sens is on the Paris-Lyons-Mediterranean railway and is also connected by bus with Montargis and Troyes. It is intersected by the Paris-Geneva and Nancy-Orléans trunk roads. The navigable Yonne links it to the French inland waterways system. Important industries are tanning, flour milling, metalworking, and the manufacture of dairy products, cutlery, electric cables, chemicals, brushes, and articles of plastic, horn, and whalebone. An agricultural and commercial fair is held annually.

Sens (Agedincum) was the capital of the Senones, a powerful Gallic tribe. The Romans first subdued it after the defeat of Vercingetorix. In 395 it became the metropolis of the 4th Lugdunensis, whose boundaries became those of the archiepiscopal see. The inhabitants, converted by the martyrs Savinien and Potentian repelled the Alamanni and the Franks in 356, the Saracens in 731 or 738, and the Normans in 886. Several ecclesiastical councils were held in Sens, notably that of 1140, at which St. Bernard and Abelard met. Granted a communal charter in 1146, it suffered in the Hundred Years' War, was an early member of the Holy League, and was deprived of its privileges by Henry IV. The archbishopric, dismembered in 1622 on the elevation of Paris (hitherto suffragan to Sens), was reduced to a bishopric in 1791. This was suppressed during 1801–17, but in 1817 the archiepiscopal see was restored.

SENSATION commonly denotes any concrete conscious experience resulting from stimulation of a specific sense organ, sensory nerve, or sensory area in the brain. The word is used in a more general sense to indicate the whole class of such experiences. In ordinary speech the word is apt to be ambiguous; it is frequently used so as to leave uncertain whether the speaker is referring to the process of sensing, or to whatever it is that is thereby sensed (the apparent painful stimulus, sound of a bell, or red glow of a fire). This double meaning has produced confusion about whether or not sensations are purely mental (as opposed to physical). Though the process of sensing is thought by some to be purely mental, some psychologists and philosophers hold that what is sensed is normally a physical quality existing independently of mind; *e.g.*, the grass is literally green whether or not any person is present to perceive it (*see* REALISM). To avoid this ambiguity Bertrand Russell introduced the term sense-datum to signify what is sensed or "given in sensation," and later writers proposed the term sensum; the word sensation is then reserved for a so-called mental process or activity (*see* KNOWLEDGE, THEORY OF: *Epistemology in the 20th Century*).

More empirically inclined psychologists and physiologists prefer to regard sensation as a concept (not a datum) defined in terms of dependent relationships between discriminatory responses of organisms and properties of physical stimuli. Characteristics of sensory functions may be ascertained by training a laboratory animal or asking a human being to respond differentially to various aspects of the stimulus. This approach regards sensation much as sensing is regarded in modern automated devices (*see* AUTOMATION: *Elements in an Automated System*). Sensing elements (sensors) in automated systems indicate characteristics (presence, absence, intensity or degree) of some form of energy impinging on them. These sensors are called transducers; they convert their input energy into electrical currents that can be used as signals. For example, the electrical current generated by a photocell can be used as a signal indicating the intensity of a light falling on the photocell (*e.g., see* SELENIUM CELL). When the signal reaches a critical strength, a relay may be activated to start a motor. The motor starts (responds) when the photocell senses light of a certain intensity. The definition of sensation in terms of discriminative responses in living organisms is analogous. When a stimulus impinges on a sense organ and the organism responds appropriately it is said that the stimulus has been sensed. Nonetheless many people (including some psychologists) regard a mentalistic definition of sensation as basic to the psychology of sensation. Historically, conceptual formulations regarding sensation arose out of mentalistic philosophies (*see* IDEALISM) that deny efforts to equate the terms mental and physical (*see* MATERIALISM; FECHNER, GUSTAV THEODOR; EPIPHENOMENALISM).

Toward the close of the 18th century, Thomas Reid (*q.v.*) argued that, although sensation furnishes purely mental experiences (*e.g.*, those called pain, colour, sound and smell), perception yields direct awareness of physical things that exist independently of a percipient (*see* PERCEPTION). The associationists, however, held that objective reference and meaning are given to a sensation by the addition to it of other sensations or images (memories of sensations), as when the smell of a rose is identified by the associated visual image of a rose (*see* ASSOCIATION, MENTAL). Indeed this was the commonly accepted doctrine of the 19th century: perception is more complex than sensation, involving both sensations and images; perception refers, as sensation does not, to

an object and may also have other meaningful implications; and the complexity of perception adds meaningful reference to it.

But toward the close of the 19th century critics of the associationistic school strongly denied that percepts are compounded by adding mental images to actual sensations. They held that objective reference is present in every form of sensory awareness from the very first, and that the percept is a synthetic unity given directly in ordinary experience. When a child first sees a jumping, barking dog and screams with fear, he does not begin by noticing a brown sensory patch moving about, then connect it with noise, and finally associate both with the mental image of a possible bite to perceive a complex object in the external world. His cognitive response was held to be the immediate result of built-in mechanisms through which he at once perceives a dangerous-looking animal.

In the 1960s the differentiation of perception from sensation was of little systematic or theoretical significance in objective psychophysiology. While discrimination of fundamental stimulus dimensions (*e.g.*, wavelength or intensity of light) may be said to indicate sensation, and discrimination of more complex stimulus patterns (*e.g.*, speech sounds, visual space or shape) can be said to involve perception, the distinction is historical and conventional, and the terms are often interchanged.

Senses.—Traditionally, there are five classical senses (*see* HEARING; SMELL AND TASTE; TOUCH; VISION). This simple classification is helpful as far as it goes, but touch, for example, is known not to be a simple unitary modality; even as commonly used, the word includes cutaneous sensations of heat, cold, pain, and pressure. Thus, throughout the body are free nerve endings that uniquely mediate sensations of pain. The traditional list must also be expanded to include organic sensations sometimes grouped under the term somesthesia, and subdivided as kinesthetic, vestibular (static), and visceral. Kinesthesia, so named because it mediates awareness of bodily movement, posture, and perception of weight, arises from sense organs in muscles, tendons, and joints; it furnishes sensory clues that permit precise movements. The vestibular sense mediates balance, bodily posture, and sensations of rotation and dizziness; its sense organs are part of the inner ear (*see* EAR, ANATOMY OF; EQUILIBRIUM, ANIMAL; VERTIGO AND DIZZINESS). The kinesthetic and vestibular senses are often classified as proprioceptive; they are stimulated primarily by the movement and posture of the body itself. Visceral sensations underlie such experiences as hunger, nausea, and sexual feelings. Clearly, there are more than five human senses (*see* FEELING, PSYCHOLOGY OF: *Theory of Feeling;* HUNGER AND THIRST).

H. L. F. von Helmholtz (*q.v.*) suggested that sensory experience be divided into modalities according to qualitative criteria. Sensory qualities within a single modality were to be those that fall in an uninterrupted continuum: for example, colour, where oranges merge into reds and yellows, and grays form a continuum between black and white. Discrete qualities for which no intermediates exist were to be placed in separate modalities.

The number of modalities by Helmholtz's definition cannot be specified with precision. Visual sensations form one such modality; all hues (as well as blacks, whites, and grays), brightnesses, and saturations can be represented by a continuous solid figure (*see* COLOUR). Audible tones constitute a continuum for pitch (*see* PITCH, MUSICAL). Smell (olfactory sensation) is more complex. Hans Henning's researches (1915) led him to believe that all olfactory qualities could be represented on the surface of a triangular prism, with fragrant, spicy, ethereal, resinous, putrid, and burnt odours at the six corners. Thus in his system smell is a single modality. Henning also placed the principal gustatory or taste qualities (bitter, sour, salt, and sweet) at the corners of a tetrahedron, and unified them as a single modality, asserting the existence of intermediate gustatory sensations. What are popularly called flavours depend heavily on odours perceived when food vapours ascend to the nose from the mouth (*see* FLAVOUR). It is doubtful whether intermediate qualities can be found among the principal cutaneous qualities (pressure, pain, warmth, and cold). Some have maintained that all deep sensibility (from muscles and joints, viscera, inner ear) reflects pressure-pain patterns and adds nothing new to the organism's repertoire of sense qualities.

Modern classification tends to be based less on verbally described experiences than on discriminability of stimuli as related to physiological mechanisms of separate sense organs and their associated neural systems.

Attributes.—A sensation is characterized by attributes, conscious dimensions along which it can vary independently; the chief attributes are quality, intensity, extensity, and duration.

Sensations are generally identified and named by their qualities; for example, red, blue, fragrant, putrid, sweet, bitter, tickle, ache. Any of these qualities can vary in intensity. Tones can be soft or loud, odours faint or strong, pains mild or excruciating. It would seem that visual sensation has two intensive attributes, brightness and saturation. A red or gray sensation may be light or dark; that variation is brightness. Red may also vary in saturation, the intensity of redness as opposed to weakness or grayness, without becoming either lighter or darker. When most people speak of bright red, they usually mean highly saturated red.

Visual and tactual sensations have extensity; space perception is based on this characteristic. There is also convincing evidence that tones are sensed to vary along a kind of extensity dimension called volume. Low tones tend to seem large, high tones small; loud tones tend to appear large, faint tones small. Hence a faint low tone may be matched in size to a loud high tone.

All sensations have duration (protensity) and can be compared in terms of how long they appear to last.

In the 19th century it was supposed that attributes of sensation must correspond to dimensions of the stimulus. However, most of the attributes depend on joint variation of two or more dimensions of the stimulus. For instance, while the loudness of a tone varies mainly with the energy dissipated in the vibration of the source, it also depends on frequency of vibration. A very low-pitched (or high-pitched) tone requires more energy than does a tone in the middle range for the two to appear equally loud. Apparent pitch depends primarily on frequency, yet apparent pitch will alter slightly when the energy of vibration is changed and frequency is kept constant. Thus attributes of tone (pitch, loudness, and volume) can be altered by varying the frequency and energy of the stimulus (*see* SOUND: *Acoustic Reception*).

Sensitivity.—The accuracy with which differences in sensation attributes can be discriminated is measured by determining absolute and differential thresholds. An absolute threshold marks the limit of some sensory continuum; *e.g.*, the lowest or highest audible tone, the limits of the visible spectrum in terms of wavelength, and the least noticeable intensity of any sensation. A differential threshold is the just noticeable difference (j.n.d.) between two stimuli. Since sensitivity varies from moment to moment, every threshold is statistically defined as the average value of a stimulus difference resulting from a number of judgments; *i.e*, the difference that is sensed just as often as it is not.

Differential threshold for intensity is often expressed as a Weber fraction (the ratio of j.n.d. to stimulus intensity). If the light from 51 candles is reported five times out of ten by a given person as perceptibly brighter than 50 candles when intensity is increased step by step from well below 50 candles to well above, the observer's differential threshold for this intensity is 1 candle, and his Weber fraction is 1/50 or 0.02. Weber's law assumes this fraction to be constant at all intensities; *i.e.*, the j.n.d. would be 10 candles for a stimulus of 500 candles, and 0.1 candle at 5 candles. The law fails to hold exactly, however, especially at lower intensities. Different people have different thresholds. Typical average determinations of the minimal Weber fraction for trained adults are as follows: subcutaneous pressure 1/77; visual brightness 1/62; weights lifted by hand 1/53; tones 1/11; smell of rubber 1/7; taste of salt 1/5 (*see* PSYCHOPHYSICAL METHODS).

Determinations of absolute intensity thresholds show that sense organs are extremely sensitive under optimal conditions. A barely visible star delivers to the eye about 10^{-8} microwatt (1/100,000,000 millionths watt). Tone can be heard at some frequencies when the eardrum moves somewhat less than 10^{-9} cm. If the ear were a bit more sensitive, people probably could hear the random movements of air molecules.

Physiology.—Sensory physiology investigates mechanisms that translate stimuli into nerve impulses and brain mechanisms that mediate discriminative responses.

Just as a microphone (*q.v.*) translates sound waves into electrical waves, sense organs are transducers that translate stimuli into nerve impulses. Each sense organ is selectively most sensitive to a specific form of energy called its adequate stimulus. Mechanoreceptors initiate sensory nerve impulses when they are mechanically deformed; these include pressure, kinesthetic, and vestibular receptors, as well as the receptor cells in the cochlea of the inner ear. The skin also contains thermoreceptors (not unequivocally identified in the 1960s) that respond to changes in temperature. Sense organs of taste and smell are chemoreceptors; their adequate stimuli are chemical properties of sapid and odorous substances. The receptor cells (rods and cones) of the retina are photoreceptors. It has been estimated that absorption of one quantum of light by the photochemical substance in a rod is sufficient to discharge it. Many more impulses, however, are required to produce a sensation of light; *i.e.*, discrimination of the presence or absence of a flash of light. Other sensory structures do not mediate awareness, but instead elicit important reflexes. For example, mechanoreceptors in the large artery (the aorta) leading from the heart detect, signal, and act to regulate the level of blood pressure (*see* CIRCULATION OF THE BLOOD: *The Nervous Regulation of the Heart Beat*).

Neural structure and activity are used to account for differences in perceived sensory quality. The primary quality distinctions are those among modalities; *e.g.*, colour, tone, odour, and taste. Ancient Greeks believed that objects give off faint copies of themselves that are conducted to the mind (*see* DEMOCRITUS: *Perception and Knowledge*). Some such view was still prevalent enough in the early 19th century for Johannes Peter Müller (*q.v.*) to react against it with his theory of specific nerve energies. Müller held that there are five kinds of sensory nerve, each of which carries its own specific kind of energy. He argued that the mind does not perceive objects themselves, but the quality of neural impulses the objects excite. He noted that any sensory nerve always gives rise to its own peculiar quality; thus pressure on the eyeball (clearly not the adequate stimulus for visual sensation) yields a perceived pattern of light. Later, when all neural impulses were found to have identical electrical qualities and sensory centres in the brain had been discovered, the modern view of sensory quality developed. In the 1960s there seemed to be no better theory of modal quality than the gross statement that there is a specific brain area for each sensory modality (*see* PSYCHOLOGY, PHYSIOLOGICAL).

The use of microelectrodes to record electrical potentials from individual neurons permits direct study of how the nervous system encodes sensory information. The technique has revealed ganglion cells in the retina that discharge only to stimulation from a relatively narrow band of wavelengths of light (the colour sensors); others show a much broader response sensitivity to wavelength (light sensors). Some fibres in the optic nerve discharge during stimulation; others activate only briefly at the beginning and end of a flash of light; a third type responds briefly only at the cessation of the stimulus. Microelectrode studies of other modalities indicate a similarly complex situation.

Recordings from single cells in the visual system of the brain suggest even more complicated encoding principles. Some populations of neurons in the cat's visual brain cortex are activated by retinal excitation from a stimulus object in the form of a line held at a given angle from the vertical. If the angle of the stimulus line is changed, new groups of brain cells are activated. Apparently some of the cat's cortical cells "see" only straight lines across the retina, and only when the lines are at specific angles. In the frog's optic tectum (part of the midbrain) there are cells that discharge only in response to a small moving spot on the retina. Apparently the way in which the nervous system encodes sensory information involves much more than simply the place or region to which the nerve fibres go; encoding may be in terms other than basic attributes derived from discriminative responses.

Sensed intensity seems definitely to depend upon amount of neural excitation. Since a single nerve fibre operates under the All-or-None Law (*q.v.*), giving rise either to its maximal excitation or to none at all, neural mediation of intensity differences requires explanation. When an increase in stimulus intensity at the sense organ is transmitted to the brain, sometimes the number of transmitting fibres is increased, and sometimes the rate at which successive impulses are sent along each fibre is increased; often both principles work together.

In considering extensity, note that in vision a point-to-point correspondence of the retina with the visual brain is well established. This correspondence is topological; *i.e.*, the order of points in the retinal field matches a corresponding order in the brain, but shapes and sizes need not be the same in the two (*see* GEOMETRY: *Topology*). The spatial order of retinal excitation thus is preserved in the brain; this is one way in which extensity values of sensation are coded.

It is assumed that tactual perception of space is mediated in a similar manner. Like the retina, the body surface is projected topologically on the cerebral cortex. Nerve fibres in the spinal cord are sorted into separate bundles according to modality (*e.g.*, pressure, pain, and temperature from the skin). When the projection fibres from the skin reach the brain cortex, however, they are arranged, not according to modality, but spatially in terms of a projective map of the surface of the body. In general, spatial difference in perception suggests the existence of spatial difference in cerebral excitation. Since the different modalities of the skin senses seem not to be separated on the cortex, this indicates that modality-specific information is also coded in other than spatial terms.

How the brain mediates sensations of duration was not known in the 1960s.

See also EYE, HUMAN; OLFACTORY SYSTEM; SKIN, SENSORY FUNCTIONS OF; VISION; and references under "Sensation" in the Index.

BIBLIOGRAPHY.—S. S. Stevens (ed.), *Handbook of Experimental Psychology* (1951); C. T. Morgan, *Physiological Psychology*, 3rd ed. (1965); F. A. Geldard, *The Human Senses* (1953); E. G. Boring *Sensation and Perception in the History of Experimental Psychology* (1942); M. H. L. Pirenne, *Vision and the Eye* (1948); R. M. Evans, *An Introduction to Color* (1948); S. S. Stevens and H. Davis, *Hearing: Its Psychology and Physiology* (1938); R. W. Moncrieff, *The Chemical Senses* (1944); J. Field (ed.), *Handbook of Physiology*, vol. I (1959)

(G. J. TH.)

SENSITIVE PLANT: *see* MIMOSA.

SENTA, a town in the autonomous region of Vojvodina, in the Socialist Republic of Serbia, Yugoslavia, lies on the Tisa River, 23 mi. (37 km.) SE of Subotica. Pop. (1961) 25,062. The town has a law court, a museum, and several schools. Agriculture and the manufacture of consumer goods (soap, candles, and furniture) are the main occupations. The town is connected by rail and road with Novi Sad and Subotica, and with Hungary and Rumania. Senta was mentioned in 1216 and from the mid-16th century it was in Turkish hands. In 1697 Prince Eugene of Savoy defeated the Turks there. In 1918 it became part of Yugoslavia. *See* also SERBIA. (V. DE.)

SENTINUM, an ancient town of Umbria, Italy, lying to the south of the modern town of Sassoferrato, in the low ground. The foundations of the city walls are preserved, and a road and remains of houses have been discovered, including several mosaic pavements. In the neighbourhood the battle took place in which the Romans defeated the combined forces of the Samnites and Gauls in 295 B.C. It was taken and destroyed in 41 B.C. by the troops of Octavian, but revived under the empire. The people of Sentinum had many of the rights of Roman citizenship.

SENUFO, a group of closely related indigenous tribes of northern Republic of Ivory Coast and southeastern Republic of Mali (*qq.v.*), numbering about 750,000 in the 1960s. They speak at least four distinct languages (Palaka, Dyimini, and Senari in Ivory Coast; Suppire in Mali) which belong to the Gur branch of the Niger-Congo language family (*see* AFRICAN LANGUAGES). Within each group, numerous subdivisions use their own names for the people and language; the name Senufo is of external origin. Palaka separated from the main Senufo stock well before the 14th century A.D.; at about that time, with the founding of the town of

Kong as a Bambara (*q.v.*) trade-route station, the rest of the tribe began its migrations to the south, west, and north, resulting in the present divisions.

Senufo peoples are agricultural, major crops including corn and millet. Their farms cluster around villages of small mud-brick houses, thatched in the south but with flat roofs in the drier north. The domestic unit is the extended family: a patriarch with his sons and their wives and children. Marriage is by parental arrangement; polygyny (*q.v.*) is fairly common. Inheritance goes to a brother by the same mother, or to other members of the mother's family. Initiation rites for adolescents are the introduction to adult tribal responsibilities. Burial is in a village plot; the body is wound in cloth with one hand exposed, and buried in a niche in the wall of a grave, which is then sealed with bricks.

Senufo are among the outstanding musicians of Africa. They use marimbas, tuned iron gongs, a variety of drums, horns, and flutes; marimba and drum orchestras are common.

See W. E. Welmers, "Notes on Two Languages of the Senufo Group . . . ," *Language*, vol. 26 (1950); R. Goldwater, *Senufo Sculpture from West Africa* (1964). (W. E. WE.)

SENUSI (SENUSSI; Arab. SANUSI), in a strict and narrow sense a member of the Sanusiya, an Islamic order or fraternity established in Arabia and North Africa in the 19th century by Sayyid Mohammed bin 'Ali al-Sanusi, a sharif (descendant of the Prophet) from Mostaganem, Alg. It is one of a large number of Sufi or dervish orders (*see* SUFISM), all of which require from their members observances of one sort or another, such as recitation of litanies, in addition to the duties incumbent on every Muslim. There has always been a tendency for new orders to form out of older ones as eminent teachers make additions to earlier teachings and persuade others to follow the new rites and rules. Each has claimed orthodoxy and authority by citing the chain of his spiritual succession from some admired theologian of the past whose orthodoxy was not questioned. The Sanusiya was on the whole conservative, conventional, rational (not ecstatic), and puritanical. It was a reformist movement aimed at a return to the simple faith and life of the earliest days of Islam. It was also a missionary order which sought to instruct the Bedouin and persuade them to live a life more in accordance with the commands of the Koran and the teachings and example of the Prophet Mohammed than they were wont to do. It aimed further at converting the pagans or semipagans of the Sahara and central Africa.

The Sanusiya.—Sayyid Mohammed bin 'Ali al-Sanusi (1791–1859), generally spoken of as al-Sanusi al-Kabir, the Grand Senusi, to distinguish him from other notable members of the Senusi family, was a pupil during his residence at Mecca of another well-known Sufi teacher, Sayyid Ahmad bin Idris al-Fasi (d. 1837), the head of the Moroccan Khadiriya order, a branch of the Shadhiliya order, and later the founder of a new suborder of his own, the Idrisiya. On Sayyid Ahmad's death his two chief disciples founded two new orders, the Mirghaniya, which became prominent in the Sudan, and the Sanusiya, the head of which, Sayyid Mohammed Idris, grandson of the Grand Senusi, became king of Libya in 1951 (*see* IDRIS I). The Grand Senusi first established his headquarters at Mt. Abu Qubais, near Mecca, in 1837, and this year is regarded as the official date of foundation of the order.

Henceforth the word "Senusi" ("Sanusi") is used in a wider sense. Members of the Sanusiya in the narrower and strictly religious sense, *i.e.*, those who practised its rites, were few in number. The vast majority of people called Senusi were personal followers of the Grand Senusi and his family, an attachment which in North Africa was eventually to develop into a political relationship. It does not appear that the Grand Senusi originally intended to direct his activities toward North Africa. The new order was making rapid progress among the Bedouin of the Hejaz when it aroused the enmity of various authorities in Mecca, and the Grand Senusi felt it expedient to depart, accompanied by a number of his disciples, for his native land; but, hearing of French advances in Algeria, he turned back to Tripolitania and then to Cyrenaica, where in 1843 he founded the mother lodge, *al-zawiya al-Baida* (*zawiya*, "lodge"). Eventually he moved the centre of the order to the distant oasis of Jaghbub (Giarabub), possibly to avoid too much attention from the Turkish administration then in control of Cyrenaica, and there he founded an Islamic university which became famous for its library. From Jaghbub he intensified his propaganda in the Sahara and the Sudan. He died at the oasis on Sept. 7, 1859. The Sanusiya order was by that time the dominant order, outside the towns, in western Arabia and in North Africa from the Nile Valley to the borders of Tunisia and from the Mediterranean to the Sahara.

Under his son and successor, Sayyid Mohammed al-Mahdi (1845–1902), the order expanded considerably, particularly into the Sahara and the Sudan, for which reason he moved its seat to the remote and isolated oasis of Kufra in 1895. This expansion of his theocratic empire clashed with French interests, for at the time the French were making rapid military and political advances into the Sahara. Sayyid al-Mahdi died in 1902. At the time of his death almost all the Bedouin and the oasis dwellers under their influence in Cyrenaica and the Sirtica and most of those in the Western Desert of Egypt were adherents of the order, and it had also a considerable following in southern Tripolitania, Fezzan, central Sahara, and the Hejaz. In all these regions there had been planted lodges of the order, over 140 in all. Then its fortunes declined, mainly, as will be seen, because its empire was attacked by the French in the Sahara and by the Italians in Cyrenaica and Tripolitania, while Wahhabi rule in Arabia curtailed its influence and prevented further expansion there. Sayyid Ahmad al-Sharif, the third head of the order and Sayyid al-Mahdi's nephew, was too occupied in meeting these threats in Africa as leader of the Sanusiya militant to be able to devote himself to anything else. From 1902 to 1912 he was resisting the French in the Sahara, from 1912 to 1918 he was directing the Bedouin resistance to the Italians, and later to the British, in Cyrenaica, and from 1918 to 1933, the year of his death at Medina in Arabia, he was an exile. From 1918 therefore, though Sayyid Ahmad never formally renounced headship of the order, effective control of it was in the hands of his cousin Sayyid Mohammed Idris, son of Sayyid al-Mahdi, who later as King Idris I of Libya had the dual position of head of the order and head of the state.

Organization of the Order.—The creation of an empire of the size it was in the time of Sayyid al-Mahdi and among a people so intractable as the Bedouin was a remarkable achievement. It was done through an organization which, however simple, served the purpose of unifying a number of different tribes and peoples through their common acknowledgment of loyalty to the Senusi family. The head of the order was advised by a small inner circle known as the *Khawass*. Other sheikhs of lodges and other persons who lived in them and practised the rule of the order were called *Ikhwan* ("brothers"). Apart from this very small category of persons, the adherents to the order were *muntasabin*, illiterate people who identified themselves with its interests. In every region where the Grand Senusi and his followers gained influence it was consolidated by the building of *zawiyas* at the request of the local population. They were placed at points of political or commercial significance, often in Cyrenaica where there had been settlements in Greco-Roman times. They may have been rather simple structures but they were imposing in regions where a tented people roamed. The local population, having asked for a *zawiya*, endowed it with land, wells, springs, or date palms, and the inmates lived by exploiting these donations, by their own labour, and on tithes and other contributions of money, goods, transport, and labour. Each lodge lived on its own revenues and also sent a surplus to maintain the centre of the order at Jaghbub or Kufra. The lodges were more or less autonomous in day-by-day running of affairs, but the sheikh of each was appointed by the head of the order and was responsible to him, though the post tended to become hereditary. Through its lodges the order rendered services to the people who maintained them. Each was a small centre for propagation of the faith, for simple schooling, and for settlement of disputes, and provided shelter and hospitality for travelers and asylum for those in jeopardy. They were small islands of civilization in what was then a barbaric sea.

Establishment in Cyrenaica.—It was in Cyrenaica that the order became most firmly established and for longest retained its

influence, and also where its organization and history have been most closely studied. Several circumstances favoured it there. Cyrenaica was isolated by deserts to south, east, and west. The country was not dominated by the towns, and the Turkish administration, which operated from such small towns as there were, exercised very little control over the interior. The Sanusiya therefore did not have to compete with any rival power, and the services it rendered were not being supplied by any other organization. Also the Bedouin were used to holy men, marabouts, coming from the west on pilgrimage to Mecca. Many settled in Cyrenaica where their tombs, familiar landmarks in the countryside, are venerated by the Bedouin; and it was further greatly to the advantage of the order that the sheikhs of its lodges were usually foreigners from the Maghreb (Maghrib), for this meant that they were not involved by sectional loyalties in the feuds and disputes endemic to Bedouin society and hence were suitable arbitrators. Also, whether that was his purpose or not, the Grand Senusi, by making Jaghbub the headquarters of the order and the residence of its head, ensured that it could not, by territorial location, be identified with any one tribe or tribal section and thereby arouse the jealousy of others.

But perhaps the main reason for the success of the order in Cyrenaica was that the Grand Senusi was able to integrate the *zawiya* organization with an already existing tribal system. There were nine noble, what are locally called *Sa'adi*, tribes and several client, or *Marabtin*, tribes, each divided into sections. In every main tribal section the order established one of its lodges, so that the more fractionized a tribe was the more lodges it tended to have. Altogether there were 45 *zawiyas* in Cyrenaica proper. In this manner the order was planted in every political segment of the country, and each segment, through its *zawiya*, participated in the order. The tribal system and the Sanusiya organization interpenetrated. This situation among the Bedouin and attached oasis folk may be contrasted with that obtaining in the towns. In the small towns of Cyrenaica eight other dervish orders were represented, only two of which had any representation among the Bedouin, and that slight. The lodges of these orders—and this applies also to the urban lodges of the Sanusiya—were little more than clubs for religious exercises and social insurance. They had no political significance because they were both small in membership and lacked cohesion and unity of direction. Their political ineffectiveness may be said to have reflected the individualism of urban communities, people's interests being confined to small quarters and even single streets, as contrasted with the solidarity and wider loyalties of tribal life.

Political Functions.—The Sanusiya had some proto-statal functions during Turkish times. So long as there was a reasonable amount of security and taxes were, at least sometimes, paid, the Turks left the Bedouin to their own devices. This meant that the order performed some services which are usually a governmental responsibility and also that it acted as intermediary between the Turkish administration and the tribes. The lodges in Cyrenaica were separated by Egypt from those in Arabia, and the Sanusiya's Saharan and Sudanese empire was crumbling year by year before fresh French advances. Yet the effect of the struggle that followed the Italian landing on the Cyrenaican coast in October 1911 was to accentuate the political functions of the Sanusiya.

Wars With Italy.—From the beginning of the Italo-Turkish War (*q.v.*), sheikhs of the lodges organized the tribes to aid the Turkish forces stationed in the country in their resistance. It was primarily a war between Italians and Turks, but the Bedouin gave support to their fellow Muslims. Being unable to reinforce their garrisons in Libya and facing serious troubles in the Balkans and in Yemen, the Turks made peace in October 1912 and promised to withdraw all armed forces from Libya. In the event, however, they did not withdraw their troops from Cyrenaica. Meanwhile Sayyid Ahmad had arrived at Jaghbub from Kufra and there met the Turkish commander, Enver Bey, who asked him to continue resistance in the name of the sultan; in the following year Sayyid Ahmad took charge of operations. At this point it became an Italo-Senusi war, the Senusi fighting in the name of the Sanusiya order and under the direction of Sayyid Ahmad and the sheikhs and brothers of the order. Most of the Turkish forces eventually departed for Egypt in late 1913 or very early in 1914. When Italy entered World War I in 1915 on the side of the Allies the Senusi thereby became involved with British forces in Egypt and were finally routed by them in February 1916. Sayyid Ahmad fled to the oasis of al Dakhla (ad Dakhilah) in Egypt and finally to Turkey in September 1918.

After prolonged negotiations a *modus vivendi* was agreed upon, Sayyid Idris representing the people of Cyrenaica in the discussions with the British and Italians which led to it. The political significance of these negotiations lay in the fact that, since they could not be carried on with an acephalous body, the only person who in the circumstances could speak for the people of Cyrenaica was the head of the Sanusiya, who throughout the discussions claimed sovereignty for the order over the whole country, over most of which it exercised control, both political and military, at the time. The truce resulted in an uneasy Italo-Sanusiya condominium, during which the Sayyid was accorded by the Italians (the British had vanished from the scene) the somewhat ambiguous title of amir and various other recognitions of political status as distinct from that which he had as leader of a religious fraternity.

When it became clear that the Italians intended to end the truce as soon as it suited their convenience, the amir fled in December 1922 to Egypt, where he continued, as far as he was able, to represent the Senusi resistance when the Italians began a further attempt to subjugate them in March 1923. This second colonial war lasted till 1932, Senusi-Bedouin forces in it being led by Sidi 'Umar al-Mukhtar, a Sanusiya sheikh who throughout conducted operations as representative of the order and of its head, the amir, thereby further enhancing the political role of the order and marking a further step in its secular development. The first war was fought in the name of the sultan and caliph, the second in the name of the Sanusiya. In the bitter and long-drawn-out suppression of the Bedouin and in the measures they took afterward, during the period of colonization, the Italians aimed at the complete destruction of the order: *delenda est Sanusiya* (*see* LIBYA). The *zawiyas* were demolished or fell into ruins, their sheikhs were scattered or exiled, their estates were confiscated, and the head of the order was in asylum in Egypt. The order was to all intents and purposes eliminated.

King Idris I.—When Italy entered World War II in June 1940 the amir Idris offered the services of his people to the British cause, and they played an ancillary part in the operations which took place in the Western Desert. In 1943, when the fighting was over, the amir reentered Cyrenaica, just a century after his grandfather had founded his first *zawiya* there; but he did not return as head of a religious fraternity, though he was that of course, but rather as a head of state. The Italians, by destroying the Sanusiya organization, had made yet more evident the political status of its head. When in 1951 the amir became King Idris I of Libya (Tripolitania, Cyrenaica, and Fezzan), the order, which had begun as a religious movement and developed into a political one could not be reconstituted to exercise political functions in the new state.

BIBLIOGRAPHY.—Octave Depont and Xavier Coppolani, *Les Confréries religieuses musulmanes* (1897); E. E. Evans-Pritchard, *The Sanusi of Cyrenaica* (1949); Carlo Giglio, *La Confraternità senussita dalle sue origini ad oggi* (1932); Fabrizio Serra, *Italia e Senussia* (1933).

(E. E. E.-P.)

SEONI, a town in Madhya Pradesh, India, and the administrative headquarters of the district of the same name, lies on the Nagpur-Jabalpur road, about 80 mi. (129 km.) N of Nagpur and on the branch line of the Satpura Railway from Nainpur to Chhindwara. Pop. (1961) 30,274. The town has a government high school and a girls' school. Seoni provides a market for a large area of the Satpura Hills region; it is also a weaving centre and is noted for the manufacture of lac bangles.

SEONI DISTRICT has an area of 3,376 sq.mi. (8,744 sq.km.) and a population (1961) of 523,741. A large part of the district is forested, yielding teak, bamboo, lac, and myrobalans. The chief crop is wheat, but gram, rice, cotton, and oilseeds are also important. In the south, Kurai is the centre of wholesale trade in

teak, which is exported to Kamptee in Nagpur district, Maharashtra. (D. G. Na.)

SEOUL (Kyongsong; Jap. Keijo), the capital of the Republic of Korea, at 37° 34′ N. and 127° 6′ E. and an altitude of 120 ft., is 25 mi. from Chemulpo (*q.v.*), its seaport. Pop. urban area (1960) 2,445,402. It lies in a basin among granite hills of 2,000 to 3,000 ft., remarkable for their craggy peaks and barren slopes. Seoul was the capital of Yi dynasty (1392–1910). Shortly after its founding, a stone wall pierced by eight gates, 20 to 30 ft. in height and about 11 mi. in circuit, was built upon the natural base of the hills. The city is 3 mi. N. of the Han river, upon which it depended for its major water supply in early days. The northern part of the old city against the highest hills was the site of the wall-enclosed palace grounds of the Yi family. It was laid out in wards after the pattern of Peking. Above the flat skyline of single story houses rose the city gates and palace walls.

With the opening of Korea to the outside world in 1876 and particularly after the annexation of Korea to Japan in 1910, many changes came to the city. The city walls were not preserved, and the gates became isolated symbols of the past. A railroad centre was developed between the south gate and the Han river; residential suburbs sprang up, some of them across the Han. An imposing granite Japanese government-general building was built on one of the palace grounds; other brick and stone buildings for bank and commercial concerns modified the appearance of the city. A streetcar system was developed. A university was established on another palace ground in the northeast of the city, and other educational establishments were opened. The Christian missionary influence is demonstrated by the many churches, including the imposing Roman Catholic cathedral.

After World War II, with Korea divided by the 38th parallel, Seoul became the centre of the U.S. military government and on Aug. 15, 1948, the capital of the Republic of Korea. Refugees from Communist-dominated North Korea flooded into the city. New buildings were constructed and the population spread into the suburbs.

Seoul was an immediate target for attack at the outbreak of the Korean war and was occupied by Communist forces on June 28, 1950, three days after they crossed the 38th parallel. The bridge across the Han was dynamited by Republic of Korea (R.O.K.) forces as a delaying action. The UN forces re-entered Seoul on Sept. 29, 1950, but it was occupied again by Communist troops on Jan. 4, 1951, though it was retaken on March 14, 1951. Great damage was inflicted, especially to the major buildings. After 1951 reconstruction progressed, broad streets were laid out and new buildings were erected. Seoul again became the political, industrial, financial and cultural centre of the republic. It forms a special governmental unit, equivalent to a province. (S. McC.)

SEPARATION OF POWERS. The separation of powers, an essential feature of constitutional government, was classically expounded by Montesquieu in a celebrated chapter of the *Esprit des lois* (1748). In his discussion of the English constitution, he drew a sharp, logical distinction between the legislative, the executive and the judicial powers and insisted that they should be handled by separate and distinct bodies. A variant of this doctrine had been previously stated by the English philosopher John Locke. It would appear ultimately to have derived from ancient and medieval concepts of mixed government (*see* below).

Locke was concerned with making sure that only one power, the legislative power, be divided between the king and parliament. Before him, the 17th-century political theorist James Harrington had advocated a more abstract notion of a necessary balance of power, which he elaborated in his *Oceana* (1656). Harrington's work shows more clearly than that of the later theorists how the separation of powers derived from the theory of mixed government. The advantages of combining monarchy, aristocracy and democracy had been a commonplace of political thought ever since Aristotle and Polybius. Indeed, it was a doctrine which the most diverse writers agreed upon, including St. Thomas Aquinas and Marsilius of Padua, Machiavelli and the Machiavellians and Monarchomachs. But it lacked precision until it was transformed into the doctrine of the separation of powers.

Following Montesquieu, the separation of powers was given institutional expression in many constitutions, notably the United States constitution of 1787, the French constitutions of the Revolutionary period and various monarchical constitutions of western Europe in the post-Napoleonic period. It was most fully elaborated in the United States by John Adams, who refined it by a rather flexible doctrine of "checks and balances." Kant gave it its most abstract philosophical form, by interpreting it as the institutionalization of a logical syllogism.

In the light of subsequent constitutional experience, it became clear that the triple separation of powers was only one means of solving a broader constitutional problem: how to insure the restraint of governmental power by dividing such power without carrying division to an extreme incompatible with effective government. Federalism (*see* Federal Government), the division of power between coexisting territorial jurisdictions, was another means of achieving this purpose. When seen as part of this broader design, the classical functional division of power loses some of its rigidity and dogmatic purity. Any comparative study of modern constitutional systems discloses a great variety of arrangements, as illustrated by the very divergent treatment of the judicial power in the U.S., British, French and Swiss systems, and the organizing and structuring of the legislative power is a problem whose solution likewise varies according to circumstances.

The concept of the separation of powers was sharply attacked, in the name of both democracy (*q.v.*) and efficiency, by fascists and communists who radically rejected the idea of dividing governmental power and insisted upon the need of concentrating it in the hands of ruling party groups dedicated to the revolutionary transformation of society. Other critics, from Jeremy Bentham onward, objected to the inhibiting effect of the separation of powers, insisting that it does not actually operate, that it cannot be effectuated. Thus J. Allen Smith, in *The Growth and Decadence of Constitutional Government* (1930), a penetrating critique focused upon the serious disequilibrium resulting from large-scale modern industry, announced the impending demise of the time-honoured scheme. However, experience with the totalitarian dictatorships of the interwar and post-World War II era persuaded many to reconsider. One conclusion was clear: constitutional democracy presupposed a balanced system of divided powers, for only within such a system could the citizen hope to enjoy a measure of independence and freedom through a guarantee of civil liberties. (C. J. Fr.)

SEPARATOR, MAGNETIC, a device for the separation of iron or steel and of feebly magnetic ores, for the purification of materials and the prevention of damage to machines. There are a great many designs of machines, including numerous continuous feed types for handling large quantities of material. One common form has a rotating drum carrying magnets on which the material falls; the magnetic portions cling to the drum and are then brushed off. Another type has a belt passing over two pulleys, the belt being fitted with magnetized feelers. A larger type with long conveyor belt is employed in mining practice for the separation of weakly magnetic minerals from nonmagnetic, such as wolfram from tin ores. A series of magnets above the belt deal with the material in its passage below them. Some machines are portable and may be made at dumps or tips, while for treating house refuse a rather large fixed separator supplied by a conveyor belt lifts tins and other iron pieces out of the waste and places them automatically into a chute. Used foundry sand can be dealt with by a machine which recovers small splashes of iron, brads, chippings, etc., and also sieves and grades the sand ready for use again. Mixed iron and brass turnings are separated in the rotary drum class of apparatus, while some separation, as for the pottery and china trade, is effected in a wet trough device. Many kinds of powders and seeds, cocoa beans, tea, tobacco, etc., are treated by a magnetic separator. Substances that have to go through a crusher are treated magnetically to avoid the risk of damage by "tramp" iron.

SEPHARDIM, ASHKENAZIM, AND ORIENTAL JEWS are the three major divisions of the Jewish people.

The Sephardim are the Jews whose ancestors lived in the Mid-

dle Ages in Spain (*Sepharad* in Hebrew). After their expulsion from Spain (1492) they settled in France, Holland, England, Italy, the Balkans, Turkey, Palestine, and North Africa as well as overseas, where they have continued to live. They preserved everywhere their own customs, religious ritual, and the Ladino (Judeo-Spaniolic) language, a form of medieval Spanish with some Hebrew terms, written in Hebrew characters. The Sephardim numbered about 500,000 by the 1960s.

The Ashkenazim (from the Hebrew *Ashkenaz*, Germany) are the Jews whose ancestors lived in the Middle Ages in German lands and migrated thence to east and west Europe, and, in the 19th and 20th centuries, overseas. Their customs and religious ritual differ markedly from those of the Sephardim. Up to the end of the 19th century most Ashkenazim everywhere spoke Yiddish or Judeo-German, a form of medieval German with a certain percentage of Hebrew expressions, written, like Ladino, in Hebrew characters (*see* YIDDISH LANGUAGE). The total number of Ashkenazim in the 1960s was estimated at 11,000,000, including almost all the 5,600,000 Jews in the U.S. and some 3,000,000 in the Soviet orbit.

The Oriental Jews, more varied in their ethnic characteristics than either the Sephardim or the Ashkenazim, are the descendants of Jews who, following the Assyrian, or Babylonian, or Roman exile from Palestine, settled in countries of the Middle East and North Africa. In Arab lands (Lebanon, Syria, Palestine, Iraq, South Arabia, Egypt, Libya, Tunisia, Algeria, and Morocco) they acquired, and have retained, Arabic as their mother tongue; in Persian-speaking lands (Iran, Afghanistan, Bukhara), Persian; in Kurdistan, Neo-Aramaic. From Iraq Oriental Jews moved on to India, and from Afghanistan and Bukhara into other parts of Central Asia and even into China. In some Oriental Jewish communities (notably those of Yemen and Iran) polygyny has been practised. Following the establishment of Israel in 1948 practically all the Yemenite, Iraqi, and Libyan Jews, and major parts of the Turkish, Tunisian, Algerian, Moroccan, and Syrian Jewish communities were resettled in Israel. The Oriental Jews numbered in the 1960s about 1,500,000.

Some minor Jewish groups do not fit into the above threefold division. These comprise the 2,000-year-old Italian-speaking Jewish communities; the even older Greek-speaking Jewish communities (mostly destroyed during the Nazi occupation); the Falashas (*q.v.*) or black Jews of Ethiopia; and the Bene-Israel (*q.v.*) or black Jews of India, of obscure origin. (RA. P.)

SEPIA, a semitransparent dark-brown pigment obtained from the ink sacs of cuttlefishes and squids. The sacs (which are speedily removed on capture and dried to prevent putrefaction) are dissolved in dilute alkali; the solution is filtered; and the pigment, precipitated with dilute hydrochloric acid, is then washed, filtered, and dried. The chemically inert pigment is fairly permanent and is used as a drawing ink and as an artist's watercolour, particularly in monochrome. (J. KA.)

SEPIK RIVER, one of the largest rivers on the island of New Guinea (*q.v.*), rises in the Victor Emmanuel Range, part of the great central cordillera, and for much of its course flows eastward following a great structural depression between the cordillera and the coastal ranges, emptying into the Bismarck Sea. The river is about 600 mi. (966 km.) long and is navigable for more than 300 mi. (483 km.) for vessels drawing 13 ft. (4 m.) of water, but much of this part of the course consists of a wilderness of sago and nipa palm swamp covering hundreds of square miles, while extensive mangrove swamps occur at the delta. As with all the great rivers of New Guinea, there are no large settlements along the course of the Sepik, and the whole of the lower basin is very sparsely populated. (D. W. F.)

SEPSIS means, literally, putrefaction or decay and implies illness from the absorption of noxious substances. It is too inexact a word to convey the specific meaning required by modern medical nomenclature and is used usually as a descriptive adjective (septic) to denote a special appearance or attitude. The word toxic is frequently used synonymously but it is not altogether correct. A toxin (*q.v.*) is a noxious protein that may be a plant extract, animal venom, or bacterial product. Toxins are introduced from without, whereas the victim of sepsis is carrying the responsible focus within his body. An example of this is a patient whose bowel is obstructed; he is absorbing back into his system products which normally would be eliminated.

Before bacteria were discovered and before the communicable nature of disease was known, putrefaction and fermentation were considered the basis of the septic state. Theodor Schwann in 1836 found that putrefaction was caused by living organisms and Louis Pasteur finally laid to rest the old belief.

Sapremia.—When dead protein material lies static within a hollow organ such as the intestine or in a body cavity, certain bacteria gain access to it. Access may be by direct inoculation, as with a needle or instrument used to evacuate the cavity, or by direct extension through membranes which are in contact with the outside. These bacteria are saprophytic; *i.e.*, they live and multiply in a necrotic environment and are incapable of invading healthy tissue. They are not pathogenic (capable of originating disease) in the human body. These bacteria multiply, use the dead protein for food, and elaborate waste products. The body absorbs these waste products and a state of sapremia, a form or subgroup of sepsis, is thereby set up. The condition is marked by loss of appetite and weight, depletion of energy, and slight to moderate fever—the usual clinical picture of chronic illness. If sapremia continues for a period of years, irreversible cellular changes such as amyloid degeneration may take place in the liver and kidneys particularly, thus making complete recovery, even with proper care, virtually impossible. Early diagnosis, prompt evacuation of the putrid contents under strict aseptic conditions, the prevention of secondary infection by pathogenic organisms during the application of dressings, and the building of an adequate protein and vitamin reserve will usually bring about complete recovery from the sapremic state.

Septicemia.—It is not always possible to know definitely by the symptoms that saprophytic bacteria are entirely responsible: other subgroups of sepsis must then be considered. Pathogenic bacteria may also invade the putrid cavity and from it gain access to the body. They are capable of living and multiplying in otherwise healthy tissue and originating a disease known as septicemia (blood poisoning). Exact diagnosis depends on identifying the organism; there is streptococcal septicemia, pneumococcal septicemia, typhoid septicemia, etc. Identification is usually made from cultures of the blood or of the discharge from the wound through which the bacteria entered the body. Unlike sapremia, septicemia is caused by bacteria that do not need to inhabit a large putrefying cavity. The originating site in septicemia is therefore difficult to locate. The highly virulent streptococcus, for instance, may gain entrance through a scratch or prick of which the person was barely aware. Occasionally, sudden "lighting up" of a focus inside the body, as an abscess at the root of a tooth or infected tonsils, sinuses, gall bladder, prostate, etc., is indicative.

Pyemia.—Another subgroup under the broad heading of sepsis is pyemia. As in septicemia, the bacteria causing pyemia are pathogenic but they are also pyogenic; *i.e.*, capable of local tissue destruction and pus formation. The most common pyemic organism is the staphylococcus. On gaining access to the bloodstream the bacteria are carried throughout the body, to become lodged finally in the small end arteries where they set up local abscesses in the bones, joints, lungs, brain, liver, and other tissues. Pyemia is a frequent disease of children and is the forerunner of an affliction of bone in childhood, acute osteomyelitis (blood-borne infection of bone).

Effect on the Body.—Bacteria, then, form the background of sepsis. Their effect on the body is both local and general and the intensity of the effect is determined by the virulence of the organism and the body's resistance. The local reaction is known as inflammation (*q.v.*). It is characterized by redness, swelling, increase in regional heat, and pain. These phenomena are the result of the stimulation and activation of nature's defenses. The rate of blood flow to the part is quickened and the white cells of the blood (defenders) are mobilized. The small blood vessels and capillaries become more permeable and permit an outpouring of cells and plasma into the involved tissues. If the invaders are

pyogenic bacteria, pus is produced. This is a combination of destroyed and liquefied tissue, blood serum, dead bacteria, and white blood cells. If the suppurating area has a sufficiently large adsorptive surface, the systemic effects are then produced, and these are those of the septic state. The symptoms of fever, chills, malaise, etc., are the result of the absorption of the broken-down tissue at the point of infection, or the result of toxins stored in the bacterial bodies and released after their death and destruction by the white blood cells.

Treatment.—Exact diagnosis of the subgroup of sepsis must first be made. The sapremic state usually can be traced to a cavity containing putrescible material under tension. When the tension is relieved and drainage is instituted, improvement will begin; but not until complete evacuation, sterilization, and obliteration of the cavity have taken place will a cure be effected. Evacuation is accomplished by incision and drainage. This allows the local defense forces to phagocytize (neutralize and destroy) the bacteria and render the cavity sterile. The reparative process then obliterates the cavity with scar tissue. Aside from general supportive measures no specific treatment need be used other than surgical drainage.

If a culture of the blood or discharge shows that pathogenic bacteria are present, surgery is not indicated until the infection is definitely localized. To cut into an area of diffuse infection is to risk opening more channels for invasion, and of breaking down the barriers nature has already set up to keep the infection localized. The involved part is rested and moist warm dressings are applied to increase the blood flow and prevent discharges from crusting and thus interfering with free drainage. Sulfa drugs and antibiotics are useful supplements.

Pyemia is treated similarly. Diagnosis of abscesses cannot be made nor treatment instituted until clinical signs are evident. Each abscess must be treated as a separate entity, and treatment will depend in great measure on the type of tissue in which the abscess is located. The best weapons for sterilization of the blood stream and prevention of abscesses are the antibiotics.

See also BACTERIA. (J. K. SK.)

SEPTARIUM, a subspherical concretionary body, an inch to a foot or more in diameter, found in some shales and characterized by an internal network of veins. *See* CONCRETION.

SEPTEMBER, the ninth month of the modern calendar, with 30 days. As is shown by its name (from Lat. *septem*), it was the seventh month in the early Roman calendar, which began with March. In the United States the first Monday of September is celebrated as Labour Day (*q.v.*), and the weekend in which it falls has become a major holiday, conventionally regarded as the end of the summer season. Michaelmas, the feast of the archangel Michael, is Sept. 29. An old saying runs, "If you eat goose on Michaelmas Day, you will never want money all the year round." The autumnal equinox occurs in this month. (F. R. WN.)

SEPTUAGINT, the earliest extant Greek translation of the Old Testament out of the original Hebrew, made during the 3rd and 2nd centuries B.C. for the use of the Jewish community in Egypt.

Legend and History.—The name Septuagint is derived from the legend that there were 70 (Lat. *septuaginta*, abbreviated as LXX), or rather 72, translators (six from each of the 12 tribes of Israel), who worked in separate cells, translating the whole, and in the end all their versions were found to be identical. The tradition that translators were sent to Alexandria by Eleazar, the chief priest at Jerusalem, at the request of Ptolemy II Philadelphus (285–246 B.C.), a patron of literature, first appears in the letter of Aristeas to Philocrates, probably written toward the end of the 2nd century B.C. but possibly making use of older material (*see* ARISTEAS, LETTER OF). This letter advocates merely a revised version of the Pentateuch, which must therefore have already existed in an earlier translation. The Septuagint does show traces of Egyptian origin, and it is probable that from various contemporary Greek translations made in several places those of the largest centres became the most important. Then, probably, the version used in Alexandria, the greatest centre of all, eventually superseded the others.

Analysis of the language of the Septuagint shows that the Pentateuch was the first part of the Old Testament to be translated, at about the beginning of the 3rd century B.C. This accords with the important place held by the Pentateuch in Jewish religion. The prophetic books of the Bible were translated into Greek in the 2nd century B.C., and the prologue to Ecclesiasticus, written *c.* 132 B.C., may imply that the third division of the Hebrew canon, the Writings or Hagiographa, already existed by then in Greek translation. It is not known how far translations of biblical books outside the Pentateuch obtained the authority of the synagogues.

The early Christian Church, whose language was Greek, used the Septuagint, not the Hebrew original, as its Bible, so it found in the Septuagint text the prophecies that it claimed were fulfilled by Christ. To Jewish eyes this was a misuse of the Holy Scriptures, and Jews therefore tended to cease using the Septuagint, particularly after the fall of Jerusalem to the Romans in A.D. 70, when normative Judaism restricted itself more and more to the Hebrew text of the Bible. A Greek translation was still needed, however, and the Jews attempted to replace the Septuagint by a new translation more in harmony with the stricter ideas of the period. The version by Aquila, as well as those of Symmachus and Theodotion (*q.v.*), fulfilled this need for a time.

The later history of the Septuagint lies entirely within the Christian Church. The authors of the New Testament drew their Old Testament quotations chiefly from the Septuagint, not the Hebrew, and the Septuagint was the Old Testament used by the Greek Church Fathers in constructing their theology. Origen (*q.v.*) in the 3rd century A.D. attempted in his *Hexapla* to check the degeneration of its text, which by then varied widely from copy to copy. Lucian of Antioch (*c.* A.D. 300) also consulted the Hebrew in order to make the Septuagint text more accurate. The Septuagint, not the original Hebrew, was the main basis for the Old Latin, Coptic, Ethiopic, Armenian, Georgian, Slavonic, and part of the Arabic translations of the Old Testament. The Septuagint has never ceased to be the standard version of the Old Testament in the Greek Church, whose liturgical texts are imbued with its language.

Contents.—The Palestinian collection of Old Testament books was fixed by Jewish authority in the 1st century A.D. The Septuagint contains all the books in this Jewish canon together with the extra books usually known as the Apocrypha (*see* APOCRYPHA, OLD TESTAMENT). The order of the books in the Septuagint differs from that of the Hebrew canon. Instead of the Jewish division into three—Law, Prophets (including history), and Writings or Hagiographa—the Septuagint has four divisions—Law, history, poetry, Prophets. The books of the Apocrypha are inserted where appropriate and do not form a separate group. This new division was perpetuated in the Western Church by way of the Old Latin version and the Vulgate and lives on in most modern translations of the Bible, save that in Protestant versions the Apocrypha is either omitted or grouped separately.

Language, Text, and Editions.—In vocabulary and grammar there is little to distinguish the Greek of the Septuagint from the Hellenistic Greek, known as Koine, of its period (*see* GREEK LANGUAGE: *Koine and Byzantine*). The syntax however betrays the Semitic character of the original in many ways, though this is not equally true of all the books: IV Maccabees, for example, which is not translated from the Hebrew, is of high literary quality. In a number of cases the method of translation and the vocabulary make it possible to distinguish more than one translator in the longer books such as Jeremiah and Ezekiel.

The importance of the Septuagint, apart from its interest as the oldest biblical translation of any scope, lies in the fact that it is the earliest witness to the Hebrew text, before the latter was standardized 1,000 years later by the Masoretes (*q.v.*). It does not follow, however, that the Septuagint reading is preferable in all cases where it diverges from the Hebrew. The translators sometimes made mistakes because they did not know enough Hebrew, occasionally they changed the meaning in order to make it suit the local Egyptian situation, and often they interpreted instead of translating. So the Septuagint can be used to clarify the

Hebrew text only in a limited way and only after examining the methods of translation book by book.

Further, the use of the Septuagint is hindered by the unsatisfactory nature of its own text. It is contained in a few early, but not therefore necessarily reliable, manuscripts, written in uncials (*see* BIBLE: *Canon and Text*), the best known being the Codex Vaticanus (B) of the 4th century A.D., the Codex Alexandrinus (A) of the 5th century A.D., and the Codex Sinaiticus (ℵ or S) of the 4th century. There are numerous papyrus fragments written earlier than the uncials, and many later manuscripts.

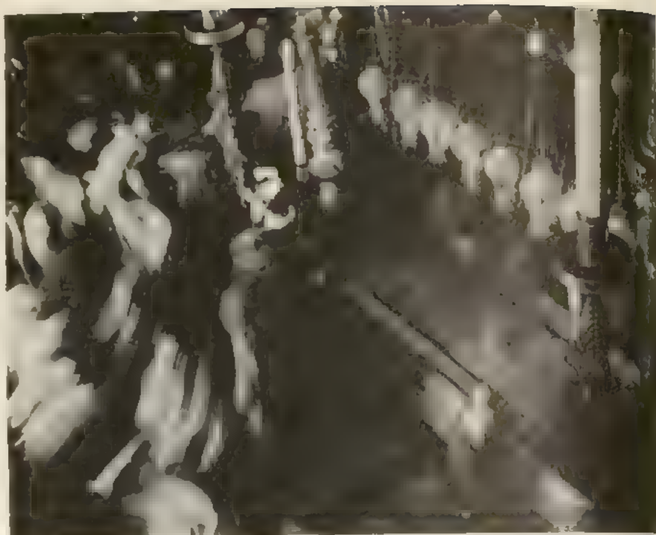
The first copy of the Septuagint to be printed was in the Complutensian Polyglot in 1514–17 (*see* POLYGLOT BIBLES), but it was not put in circulation until 1522, by which time the actual *editio princeps* had appeared from the Aldine Press at Venice in 1518. H. B. Swete's edition in three volumes (1887–94, revised 3rd and 4th ed. 1907–12) gives the text of B and of A or ℵ where B is lacking, with variants from a limited number of manuscripts. It is thus not a critical edition but a selection based on the uncial manuscripts and on manuscripts available in facsimile or in print. E. A. Brooke and N. McLean's edition (1906–   ) has practically the same text as Swete, but the critical apparatus mentions every known uncial manuscript together with a selection from the later manuscripts and from some early translations based on the Septuagint as well as from patristic quotations. An edition on a similar scale by A. Rahlfs under the auspices of the Göttingen Academy of Science (1931–   ) gives a critically compiled text and, in the critical apparatus, the readings of groups of manuscripts in families corresponding to the different recensions which the Septuagint text unmistakably exhibits. (Rahlf's two-volume edition of 1935 gives the same text but omits most of the critical apparatus.)

As a result of the research of P. E. Kahle, emphasis in biblical studies falls on the gradual development of the different parts of the Old Testament translation. It is seen that official control which led to an authoritative text came rather at the end than at the beginning of its development. It is possible that the Christian Church—and modern knowledge of the Septuagint depends almost entirely on the material used by the church—took over one of several Jewish texts in circulation. The lucky finds of fragments (Papyrus Fouad; John Rylands fragments) from the Septuagint of Jewish origin in Egypt and near the Dead Sea show how dangerous, if not impossible, is the attempt made by Paul Antoine de Lagarde and his followers to reconstruct the "original" Septuagint.

BIBLIOGRAPHY.—Editions are mentioned above; manuscripts listed in A. Rahlfs, *Verzeichnis der griechischen Handschriften des Alten Testaments* (1914); concordance by E. Hatch and H. A. Redpath, 2 vol. with supplement (1897–1906; reprinted 1954). Eng. trans. by L. C. L. Brenton (1844, frequently reprinted). *See* also H. B. Swete, *Introduction to the Old Testament in Greek* (1900; revised ed. by R. R. Ottley, 1914); R. R. Ottley, *A Handbook to the Septuagint* (1920); B. J. Roberts, *The Old Testament Text and Versions* (1951); P. E. Kahle, *The Cairo Geniza*, 2nd ed. (1959); H. St. J. Thackeray, *A Grammar of the Old Testament in Greek*, vol. 1 (1909); S. R. Driver, *Notes on the Hebrew Text and the Topography of the Books of Samuel*, 2nd ed., introduction (1913); D. Barthélemy, *Les Devanciers d'Aquila; première publication intégrale du texte des fragments du Dodécaprophéton, trouvés dans le désert de Juda, précédée d'une étude sur les traductions et recensions grecques de la Bible réalisées au premier siècle de notre ère sous l'influence du rabbinat palestinien* (1963); J. Ziegler, *Die Septuaginta. Erbe und Auftrag* (1962). (P. A. H. DE B.)

SEPULCHRE, HOLY, the tomb in which Jesus was buried, and so the name of the church built on the traditional site of Jesus' crucifixion and burial. Since, as John 19:42 says, "the tomb was close at hand," near the cross, the church has been planned to enclose the site of both cross and tomb (*see* CALVARY).

The Church of the Holy Sepulchre (more correctly, Anastasis, or "Resurrection") lies in the northwest quarter of modern walled Jerusalem. Constantine the Great first built a church on the site. It was dedicated about A.D. 336, burned by the Persians in 614, restored by Modestus, destroyed by the caliph al-Hakim about 1009, and restored by the Byzantine emperor Constantine Monomachus. In the 12th century the crusaders carried out a general rebuilding of the church. Since then, because of fire, war, earthquake, and deterioration, frequent repair, restoration, and re-



SABINE WEISS—RAPHO GUILLUMETTE

PILGRIMS MOVE PAST THE MARBLE SLAB THAT COVERS THE SUPPOSED TOMB OF CHRIST IN THE CHURCH OF THE HOLY SEPULCHRE, JERUSALEM

modeling have been necessary. The present church dates mainly from 1810. Though structurally weak and aesthetically unsatisfying, the church is the focus of great Christian interest. Various Christian groups, including the Greek, Roman, Armenian, and Coptic churches, control parts of the structure and conduct services regularly. The main features of the church are the *Martyrion*, on the supposed site of Jesus' crucifixion, and the rotunda over the tomb shown as the place of his burial and resurrection; auxiliary chapels commemorate events of Jesus' passion and resurrection.

This site, thus, has been continuously recognized since the 4th century as the place where Jesus died, was buried, and rose from the dead. But whether it is the actual place of his crucifixion and burial has been hotly debated. Two main questions are involved. The first of these is whether Christians of the first three centuries could and did preserve a continuous authentic tradition as to just where these events occurred. The Jerusalem church fled to Pella about A.D. 66. Titus destroyed Jerusalem in A.D. 70. In A.D. 135 Hadrian again destroyed the city and in its place built the pagan city Aelia Capitolina. It is said that over the place where the tomb of Jesus is now shown he built a shrine of Venus, possibly to stop Christians from visiting the spot in memory of Jesus' death. But it is not certain that Christians in the apostolic age continually focused attention on this spot as the exact place where Jesus died and was buried, and times of war, destruction, and confusion could have prevented the preservation of such exact information.

The other main question involves the course of the second north wall of ancient Jerusalem. The first wall ran eastward from near the present Jaffa Gate to the west wall of the Temple area. The second wall, according to Josephus (*Wars*, v, 4, 2), started from the Gennath Gate, near the western end of the first north wall, circled north and east to enclose an area north of the first wall, and ended at Antonia, at the northwest corner of the Temple area. The third wall was built later than the time of Jesus, and so throws no light on where Jesus died and was buried. Some archaeological remains on the east and south sides of the Church of the Holy Sepulchre are widely interpreted to mark the course of the second wall; if so, the site of the church lay just outside the city wall in the time of Jesus, and could be the actual place of his crucifixion and burial. This site probably did lie outside the city wall at that time; it may well be the very place where Jesus died and was buried; no rival site is supported by any real evidence; but it is not certain that knowledge of the exact spot was continuously preserved in the early, turbulent days of the church, so complete certainty is now unattainable. *See* also PALESTINE: *The Holy Places*.

BIBLIOGRAPHY.—Eusebius, *Life of Constantine*, iii, 25–40; C. W. Wilson, *Golgotha and the Holy Sepulchre* (1906); L. H. Vincent and F. M. Abel, *Jérusalem*, vol. ii, *Jérusalem Nouvelle*, pp. 89–300 (1914).

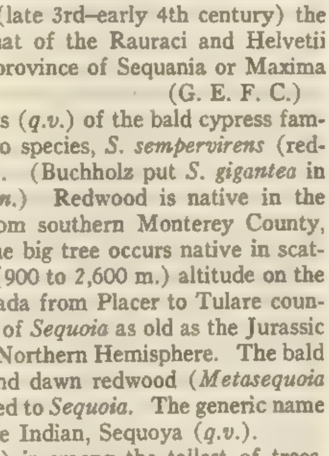
J. Jeremias, *Golgotha* (1926); G. H. Dalman, *Sacred Sites and Ways*, pp. 346–381, Eng. trans. (1935); L. H. Vincent and M. A. Steve, *Jérusalem de l'Ancien Testament*, pp. 51–174 (1954); A. Parrot, *Golgotha and the Church of the Holy Sepulchre*, Eng. trans. (1957). (F. V. F.)

SEQUANI, a Celtic people who by Caesar's day were occupying the territory between the Saône, Rhone, and Rhine rivers, with their centre at Vesontio (Besançon). Quarrels with the Aedui (*q.v.*) over tolls on the Saône led them to call in the German Ariovistus, who defeated the Aedui (perhaps in *c.* 61 B.C.), but occupied Sequanian territory in modern Alsace and gradually raised his demands. Sequanian nobles supported the Aedui in appealing to Caesar in 58, but after Caesar had expelled the Germans he compelled the Sequani to restore all Aeduan land they had seized. The Sequani joined the Gallic revolt led by Vercingetorix in 52, but after its suppression by Caesar they were apparently not penalized. Under the empire they were part of Gallia Belgica and ruled by magistrates on the Roman model; Vesontio (possibly a Latin colony) had walls and important public buildings, though the well-known arch (Porte Noire) which survives was probably built after A.D. 150 (*see* BESANÇON). Like the Aedui the tribe supported Julius Vindex in 68, and next year opposed the Treveri in their attempt at an independent Gaul.

In Diocletian's reorganization (late 3rd–early 4th century) the territory of the Sequani, with that of the Rauraci and Helvetii to the east, became the separate province of Sequania or Maxima Sequanorum. (G. E. F. C.)

SEQUOIA, a genus of conifers (*q.v.*) of the bald cypress family (Taxodiaceae), comprising two species, *S. sempervirens* (redwood) and *S. gigantea* (big tree). (Buchholz put *S. gigantea* in a separate genus, *Sequoiadendron.*) Redwood is native in the fog belt of the Coast Ranges from southern Monterey County, Calif., to southern Oregon, and the big tree occurs native in scattered groves at 3,000 to 8,500 ft. (900 to 2,600 m.) altitude on the westerly slopes of the Sierra Nevada from Placer to Tulare counties in California. Fossil remains of *Sequoia* as old as the Jurassic period are widely dispersed in the Northern Hemisphere. The bald cypress (*Taxodium distichum*) and dawn redwood (*Metasequoia glyptostroboides*) are closely related to *Sequoia.* The generic name commemorates the great Cherokee Indian, Sequoya (*q.v.*).

The redwood (*S. sempervirens*) is among the tallest of trees, often exceeding 300 ft. (91 m.), and the trunk diameters range to 15 ft. (5 m.), or exceptionally to 20 ft. (6 m.), measured above the swollen bases. The insect-, fungus-, and fire-resistant bark is reddish-brown, fibrous, deeply furrowed, and may be 12 in. (30 cm.) or more thick on old trees. The bases of the trees form massive buttresses, and hemispherical burls occur on some trunks. Beautifully grained veneer, bowls, trays, and turned articles are made from large burls; small burls, sawed from the tree and placed in a tray of water, send forth decorative green sprays. Young trees are conical in form and clothed to the ground with densely leaved, slightly drooping branches; in age the lower limbs fall away, leaving a columnar clear trunk to a height of 100 ft. (30 m.) or more. When a tree is cut, numerous shoots arise from the sapwood below the cut surface and grow into trees in 30 to 60 years if protected from fire; vigorous leaders grow as much as eight feet in length during a year. The needlelike leaves are dark green above, whitish beneath, one-third to one inch long, arranged in two-ranked sprays, and remain functional three to seven years. Cones are one to two inches long and produce tremendous seed crops, but germination rarely exceeds 25%. Thus seedlings are rare in the forests and reproduction is mainly vegetative.



JOHN MARKHAM

FOLIAGE AND CONES OF REDWOOD (SEQUOIA SEMPERVIRENS)

The sapwood is thin, white, and decays readily, but heartwood is reddish-brown, evenly fine grained, strong, and durable. It resists attack by fungi, termites, and other insects. Some redwood fence posts have lasted at least 50 years.

Heavy cutting beginning in the 1860s has drastically reduced the original 1,454,000 ac. (588,434 ha.) of redwood forests, but extensive stands still exist. Logging continues, but the annual cut dropped after World War II. Yields run about 150,000 bd.ft. per acre in normal stands, but cuts as high as 1,500,000 bd.ft. per acre are on record. Second-growth timber on favourable sites produces nearly 140,000 bd.ft. per acre in 60 years, the greatest volume growth of any known softwood tree in the world.

The big tree (*S. gigantea*) is the largest of all trees in bulk and long was reputed to be the oldest living thing. Many reports about the great age of the trees were exaggerations, for the largest stumps accurately examined yielded counts of about 4,000 years. The largest specimen is the General Sherman tree in Sequoia National Park, measuring 101.5 ft. (31 m.) in circumference at the base; a mean base diameter of 32.2 ft. (10 m.); diameter 27 ft. (8 m.), 8 ft. (2 m.) above the ground; diameter 18.5 ft. (6 m.), 100 ft. (30 m.) above ground; 272.4 ft. (83 m.) tall; diameter of largest branch (130 ft. [40 m.] above ground) 6.5 ft. (2 m.); total estimated weight 6,167 tons. A few specimens are over 300 ft. (91 m.) high, but have less bulk than the General Sherman tree.

AUTHENTICATED NEWS

BIG TREE (SEQUOIA GIGANTEA) IN KINGS CANYON NATIONAL PARK, CALIF.

Although several big tree groves were cut, the lumber is more brittle than that of the redwood; since the lumber is less desirable, it has been easier to save the big trees from destruction. Most of the 70 distinct groves now are under the protection of state or national forests and parks. The groves contain as few as 4 and as many as 3,500 mature trees each, and seedlings and young trees are very numerous in some of the groves.

The earliest-known mention of the big trees is a narrative of Capt. Joseph Walker's expedition of 1833 published in 1839. Though other persons must have seen the big trees in the years following, the Calaveras grove did not become known generally till 1852.

BIBLIOGRAPHY.—C. S. Sargent, *The Silva of North America* (1896); W. L. Jepson, *The Trees of California* (1923); W. Fry and J. R. White, *Big Trees* (1938); J. C. Shirley, *The Redwoods of Coast and Sierra* (1938); N. A. Bowers, *Cone-Bearing Trees of the Pacific Coast* (1942); Save-the-Redwoods League reports (1920 *et seq.*); National Park Service pamphlets of information. (I. L. W.)

SEQUOIA NATIONAL PARK, a reservation of 386,863 ac. (156,680 ha.) in the Sierra Nevada Mountains of California, was established by act of the U.S. Congress on Sept. 25, 1890, in order to protect groves of *Sequoia gigantea*, giant sequoia or big tree, among the largest members of the vegetable kingdom. Until specimens of bristlecone pine were discovered to be older, *Sequoia gigantea* was also believed to have the oldest living members of any species on earth. *Sequoia gigantea* is not known to exist anywhere outside the Sierra Nevadas. It does not reproduce rapidly, very few of the seeds taking root. The great girth and lack of taper of mature specimens of *Sequoia gigantea* account for its

achieving a greater bulk than any other species, although other species grow taller. The General Sherman tree is the largest in the park (*see* Sequoia).

Repeated efforts have been made to save the redwoods from the lumberman's saw. The Mariposa Grove, adjacent to Yosemite Valley, was set aside by Congress in 1864. The General Grant National Park was created in 1890 and assimilated in Kings Canyon National Park when the latter was created in 1940. Other groves are in California state parks. Sequoia National Park is accessible by automobile and bus from Fresno, Visalia, or Tulare. Accommodations are available at Giant Forest throughout the year. Lodges and camping areas are open elsewhere from June to September. Fishing, hiking, and horseback riding are among the recreations. (J. E. Cl.)

SEQUOYA (Sequoyah) (?–1843), a half-blooded Cherokee Indian who created the Cherokee alphabet, the only man in history known to have developed an entire alphabet or syllabary. Born at a Tuskegee village on the Tennessee River in what later became the state of Tennessee, Sequoya was probably the son of a British trader named Nathaniel Gist. Reared by his Indian mother, he never learned to speak, read, or write English. He was, however, an accomplished silversmith, painter, and warrior, and served with the U.S. Army in the Creek War in 1813–14. Perhaps as a result of wounds received in the Creek campaigns, he became crippled in one leg, which enhanced his inclination toward a contemplative life.

Like many Indians of his era, Sequoya wished to learn the secret of the white men's superior power. He became convinced that their secret was written language, which enabled them to accumulate and transmit more knowledge than was possible for a people dependent on memory and word of mouth. About 1809 Sequoya began working systematically to develop a written language for the Cherokees, experimenting first with pictographs and then with symbols representing the syllables of the spoken language. His young daughter helped him to identify the Cherokee syllables. For symbols, he adapted letters from an English speller and, probably, from printed words in Greek and Hebrew obtained from missionaries. By 1821 he had perfected 86 letters, representing all the syllables of the Cherokee language.

Sequoya convinced his tribesmen of the utility of his syllabary by transmitting messages between the Cherokees of Arkansas (with whom he went to live) and those of the east, and by teaching his daughter and selected youths of the tribe to write. The simplicity of his system enabled pupils to learn it rapidly, and soon Cherokees throughout the nation were teaching it to one another. The tribe established schools where Sequoya's alphabet was taught, and published books and newspapers in the Cherokee language.

Sequoya's name (spelled Sequoia) was given to the giant redwoods of California and to the Sequoia National Park in California. In recognition of Sequoya's contribution to the early development of Oklahoma, that state chose him as one of its two representatives in Statuary Hall in the national capitol at Washington, D.C.

See Grant Foreman, *Sequoyah* (1959). (M. E. Yo.)

SERAING, an industrial town in the province of Liège, Belgium, lies on the Meuse River, 6 mi. (10 km.) upstream from Liège town. Pop. (1961) 41,239. In the town are an engineering and technical institute, the royal lyceum, the royal athenaeum, and the statue of John Cockerill (1790–1840). The château of the prince bishops of Liège houses the amalgamated Cockerill-Ougrée Company (1955). In 1835 an engineering enterprise founded earlier by John Cockerill, in partnership with King William I of the Netherlands, constructed at Seraing the first steam locomotive on the continent. Seraing is a principal centre of Belgium's metallurgical and machine-building industry, and there are also glassworks.

SERAMPUR, a town in Hooghly District, West Bengal, India, on the right bank of the Hooghly River, 4 mi. (6 km.) N of Calcutta. Pop. (1961) 91,521. Originally a Danish settlement, the town was acquired by the East India Company in 1845. A Baptist mission, which began work in Serampur in 1793, did much to spread knowledge of Christianity in Bengal. Serampur College (1818) owes its establishment to three Baptist missionaries who were the first people in India to cast types in the Indian alphabets, and the earliest newspapers in Bengali were issued from the press at Serampur in 1818. The first paper mill in India was built at Serampur in the 1870s. The town, which has a college of textile technology and two other colleges, both affiliated with Burdwan University, is an important centre of the cotton and jute industry. (S. P. C.)

SERAO, MATILDE (1856–1927), Italian novelist, one of the best of the southern Italian realist school of writers, was born at Patras, Greece, on March 7, 1856. She spent most of her life in Naples, where she founded and directed a daily paper, *Il Giorno*. She died in Naples on July 25, 1927. Many of her innumerable novels and short stories were very popular, and the best of them, which include *La virtù di Checchina*, *La conquista di Roma*, and *Il paese di Cuccagna*, depict with minute accuracy and feminine tenderness the life of the humble people.

See B. Croce, *La letteratura della nuova Italia*, vol. III (1929); M. Sandri, *M. Serao* (1922).

SERAPHIM, SAINT, of Sarov (Prokhor Isidorivich Moshnin) (1759–1833), is, together with St. Sergius (*q.v.*) of Radonezh, perhaps the saint who most clearly embodies all the features of Russian Orthodox monastic spirituality. He was born on July 19 (old style), 1759, in Kursk, the son of a merchant. Unusually devout even as a boy, on Nov. 20 (O.S.), 1779, he entered the Sarov monastery (in the Oka region of East Russia). On Aug. 13 (O.S.), 1786, he was consecrated as a monk and took Seraphim as his name in religion. He received priestly orders in 1793 and, until he left the monastery, celebrated the Eucharist daily, which was rare in 18th-century Russia. From 1794 to 1810 he lived as a hermit in a hut in the woods, for some years spending the night in prayer as a stylite on a great stone and from 1807 in complete silence. He constantly carried the Gospels in a sack on his back, in order to read them daily and to be reminded of the weight of the cross on Christ's back. Following Matt. 5:39, he did not defend himself against an attack by robbers and as a result was permanently crippled. He attributed his recovery to the Virgin Mary, who (as in earlier illnesses) appeared to him in a vision.

From 1810 to 1825 he lived as a recluse in the monastery of Sarov. In the latter year he left his cell on the command of the Virgin Mary and devoted himself to pastoral care and spiritual healing. He told his flock, most of them laity, that the most important aim in life was to obtain the "gift of the Holy Ghost," through a spiritual life in the old monastic tradition, which Seraphim was the first to extend also to laymen. He died at Sarov on Jan. 13, 1833 (O.S.), and was canonized on July 19 (O.S.), 1903.

Bibliography.—V. Zander, *Saint Seraphim de Sarov* (undated); I. Smolitsch, *Leben und Lehre der Starzen*, 2nd ed. (1952); I. Kologrivof, *Essai sur la sainteté en Russie* (1953); G. P. Fedotov (ed.), *A Treasury of Russian Spirituality* (1952). (F. v. L.)

SERAPHIM, six-winged beings seen by Isaiah in his vision (Isa. 6:2–6) standing by the throne of the Lord and singing his praises. The Hebrew term is essentially an adjective meaning "winged" or "fiery ones." The singular saraph occurs in Num. 21:6–8 and in Deut. 8:15 and Isa. 30:6, mentioned in the last case in association with scorpions. The concept apparently derives from ancient Near Eastern folklore, and the seraphim have been identified with the six-winged figure, holding a serpent in each hand, portrayed on a relief at Tell Halaf, one of the oldest sites in north Mesopotamia.

In early Christian times the seraphim came to be considered as a category of angels, comparable to the cherubim, and in the nine "choirs" of angels they held the highest rank. In Christian liturgy, the best-known use of the word is in the *Te Deum:* "To thee Cherubim and Seraphim continually do cry, Holy, Holy, Holy, Lord God of Sabaoth." *See* Angel.

SERAPIS: *see* Sarapis.

SERBIA (Srbija), a country of southeastern Europe, in the north of the Balkan Peninsula. After a long history of great vicissitudes, which constitute the subject of the present article, Serbia, with an area expanded to 49,528 sq.mi. (128,278 sq.km.)

as a final result of World War I, was in December 1918 merged in the Kingdom of the Serbs, Croats, and Slovenes, which in 1929 changed its name to Yugoslavia (*q.v.*). When Yugoslavia was reorganized in 1946, Serbia became one of the component republics of the federation, with an area of 34,116 sq.mi. (88,361 sq.km.), including the autonomous region of Vojvodina (*q.v.*) in the north and that of Kosovo-Metohija (*q.v.*) in the south. For physiography *see* YUGOSLAVIA.

The Serbs use a Cyrillic alphabet, rather than the Latin one with diacritical marks, as used by the Croats, whose language is scarcely different from Serbian (*see* SERBO-CROATIAN LANGUAGE). In this article the latter alphabet is used for Serbian names, instead of the English system of transliteration from the Cyrillic: for instance *c* instead of *ts; č* or *ć* instead of *ch; ja* instead of *ya; š* instead of *sh; ž* instead of *zh*.

The Serbs (*Srbi*, singular *Srbin*), who belong to the South Slav group of the Slavonic peoples (*see* SLAVS), arrived in the Balkans during the 7th century A.D. The etymology of this name, which first occurs, in the form Sorabi, in the *Frankish Annals* for the year 822, is obscure; it must be remarked that the Slavs of Lusatia (*q.v.*) similarly call themselves Sorbs.

In their Balkan settlements, along the Piva, Tara, upper Drina, Ibar, Lim, and Western Morava rivers, the Serbs did not at once form a united state. Clans more or less related to each other occupied a certain territory under the political and military leadership of a *župan*. Attempts by one *župan* after another to subjugate his neighbours led to much bloodshed; and the custom of the ancient Slavs, whereby a *župan* was succeeded by the eldest surviving member of his family, not necessarily by his own son as he might naturally wish, gave rise to further conflict.

The House of Višeslav.—The first *župan* whose name is recorded was Višeslav (Visheslav), who was living *c.* A.D. 780. He had a son, Radoslav, whose grandson Vlastimir fought a war against the Bulgars (*c.* 850). Vlastimir's son Mutimir accepted Orthodox Christianity (*c.* 879); and the work of evangelization was carried forward by disciples of SS. Cyril and Methodius (*see* CYRIL AND METHODIUS, SAINTS), in particular by SS. Clement and Naum, who founded Serbia's first monasteries on the Prespa and Ohrid lakes.

After Mutimir's death (*c.* 890) the Byzantines and the Bulgars began a struggle for Serbia. Peter Gojniković, who ruled from 892 as the successor of his cousin Prvoslav (Mutimir's son), adopted a pro-Byzantine policy, but was lured under the promise of safe conduct to an interview with the Bulgarian tsar Simeon, who then arrested him and took him in shackles to die a prisoner in Bulgaria. The Bulgars set up Mutimir's grandson Paul as ruler of Serbia (917–920). When Paul initiated a *rapprochement* with Constantinople, they replaced him with Zacharias, another of Mutimir's grandsons, who had originally been the Byzantines' candidate against Paul but had fallen into Simeon's hands. When Zacharias likewise made new approaches to the Byzantines, Simeon drove him into exile in Croatia (924).

On Simeon's death, in 927, Časlav (Chaslav), a great-grandson of Vlastimir, escaped from captivity in Bulgaria and soon made himself ruler over most of the Serbs and was able to annex Bosnia, but had to acknowledge the Byzantine emperor as his suzerain. He was killed in battle against the Magyars (Hungarians) in 960. Bosnia soon detached itself from the rest of the Serbian territory.

Zeta and Raška.—One of the Serbian territories was Duklja, the ancient Dioclea, later called Zeta (*see* MONTENEGRO). Its prince, Vladimir, was captured by the Bulgarian tsar Samuel in 989 but was allowed to return to his country after marrying Samuel's daughter Kosara. On May 22, 1016, however, he was treacherously murdered at the behest of Samuel's successor Vladislav.

The subsequent course of events is not well known. Prince Vojislav of Travunija (Trebinje), a cousin of Vladimir, led an insurrection against Byzantine overlordship, but was defeated (1036) and taken to Constantinople. Escaping to the mountains of Zeta, he led a second rising, which was successful. Vojislav's son Michael (Mihajlo), who succeeded him *c.* 1050 and was later styled king, sent his son Bodin to help an anti-Byzantine insurrection in the region of Skopje (Skoplje) in 1073, whereupon the insurgents proclaimed Bodin tsar of the Bulgars. Though Bodin fell eventually into Byzantine hands, he was set free by the Venetians and returned to Zeta to succeed his father as king (*c.* 1081). He added Raška (Rashka, on the Ibar River) and Bosnia to his dominions. In another war against the Byzantines he was again taken prisoner (1091). Freed again, he reinstated himself as king, and died in 1101.

Raška, detached from Zeta before the end of the 11th century, produced its own *župan*, Vukan, against whom the Byzantine emperor Alexius I conducted three campaigns. The *župan* Uroš (Urosh) I, recorded *c.* 1114, may have been a nephew of Vukan and was probably the father of the *župan* Uroš II, against whom the emperor Manuel I fought a war in 1149. Manuel destroyed Uroš II's capital, drove him to refuge in the mountains, and, in 1150, attacked him again, forcing him this time to swear allegiance. By *c.* 1155, however, a certain Desa was ruling Raška in the place of Uroš II, whose cause Manuel then supported.

The Nemanja Dynasty.—About 1167 Stephen (Stefan) Nemanja (Nemanya) became grand *župan* of Raška, under Byzantine suzerainty. When he joined the Venetians against Manuel in 1171, a Byzantine army was sent to install his brother Tihomir in his place, but Nemanja won a battle near Kosovo, and Tihomir perished. Defeated by the Byzantines in 1172, Nemanja obtained pardon from Manuel; but after Manuel's death (1180) he began war again, conquering several Byzantine places on the Adriatic coast. He allied himself with Ivan Asen I, founder of the Second Bulgarian Empire; and in 1189–90 he took Skopje, Prizren, and Tetovo. Later in 1190 he was defeated by the emperor Isaac II on the Morava River; but even so he retained Kosovo, Peć, Prizren, Bar, Skadar (Shkodër), and Kotor, while his son, Stephen, was married to Isaac's niece, Eudocia. After Isaac was deposed, Nemanja abdicated (March 25, 1196) and became a monk. He died in 1200, on Mount Athos.

Nemanja left his throne, not to his eldest son, Vukan, but to his second son, Stephen, probably because Stephen's father-in-law had become Byzantine emperor as Alexius III. Apart from Vukan's resentment, Stephen had to face danger from the Bulgars, from the Hungarians, and from the Latin Empire, which the leaders of the Fourth Crusade set up in place of the Greek in 1204. He received invaluable support, however, from his youngest brother, Sava, the future saint, one of the greatest personalities in Serbian history. Sava returned to Serbia from Mount Athos, where he had been living as a monk before Nemanja's retirement; and in 1219 he was consecrated as the first archbishop of an autocephalous Serbian Orthodox Church. While it is generally accepted that Stephen received a royal crown from Pope Honorius III in 1217, Serbian sources state that he was crowned king of the Serbs by Sava in the Church of the Seven Doors at Žiča in 1220. Known as Stephen the First Crowned, he died *c.* 1228.

The succeeding kings were Stephen's sons Stephen Radoslav (reign 1228–34), Stephen Vladislav (1234-43), and Stephen Uroš I (1243–76). Uroš in 1252 made war against the Adriatic city of Dubrovnik, which soon came to terms with him; but in 1253 Dubrovnik allied itself with the Bulgars, whose forces penetrated Serbia as far as the Lim River. For that reason the seat of the Serbian archiepiscopate was transferred from Žiča to Peć. Peace, however, was made in 1254. In a war against the Nicean emperor Theodore II Lascaris, Uroš occupied Skopje, Prilep, and Kičevo, but not for long. Taken prisoner in war against the Hungarians (1268), he obtained his freedom by promising to accept as coruler his son Stephen Dragutin, who was married to Elizabeth, granddaughter of the Hungarian king Béla IV. His violation of this promise caused Dragutin to revolt; and Uroš, defeated in battle in 1276, retired as a monk to Hum (Hercegovina), where he died.

Having maintained good relations with his neighbours for most of his reign (1276–82), Dragutin joined Charles I of Naples and Sicily and the Venetians in their anti-Byzantine alliance of 1281. Then he broke his leg in a fall from his horse. Taking this a judgment from God, he abdicated in 1282, in favour of his brother Milutin (also reckoned as Stephen Uroš II). Later, however, he claimed that he had merely made Milutin regent and had reserved

FIG. 1.—TERRITORIAL EXPANSION DURING THE NEMANJA DYNASTY

the eventual succession to his own descendants (*see* below). Though Dragutin was granted Belgrade, Mačva, and northeastern Bosnia by his brother-in-law László IV of Hungary, he finally became a monk, dying in 1316.

Milutin's reign (1282-1321) marks the rise of Serbia in the international field. A few months after his accession he conquered Skopje, Tetovo, Ovče Pole, and Pijanec from the Byzantines. An army sent against him by the emperor Andronicus II was defeated, and the Serbs advanced to the walls of Strumica, Prilep, and Ohrid. Peace was not concluded till 1299, when Milutin married the emperor's daughter Simonis, and the Byzantines formally ceded the conquered lands as her dowry. Planning a new war against the Byzantines, Milutin in 1308 made an alliance with the ambitious French prince Charles (*q.v.*) of Valois and entered into relations with Pope Clement V, but this project came to nothing when Charles gave up his designs on the Eastern Empire. A Hungarian attack on Serbia in 1319 was without effect.

Internally, Serbia's mineral resources were exploited, commerce thrived, and about 40 churches were built during Milutin's reign. There were two political crises, however. The first, arising from Dragutin's dynastic claims for his sons, was overcome in 1312. The second was the rebellion (1314) of Milutin's own son and heir apparent, Stephen (Uroš III, otherwise called Stephen of Dečani, or Stefan Dečanski), who was captured, condemned to be blinded, and exiled but was later allowed to return to Serbia.

Milutin died of apoplexy on Oct. 29, 1321. His younger son Constantine and Dragutin's son Vladislav then contested the throne. Suddenly, however, the rumour spread that the blinded Stephen had miraculously recovered his sight. Popular sympathy was with him as a martyr, and at Epiphany 1322 he was crowned king, while his son Stephen Dušan (Dushan) was also crowned as "young king" or heir apparent. Constantine was killed soon afterward, and Vladislav gave up the struggle. The Byzantines and the Bulgars allied themselves against Serbia in 1330; but the Bulgars were decisively defeated at Velbuzhd (Kyustendil) on July 28, when their tsar Michael Shishman was killed. The Byzantine emperor Andronicus III thereupon abandoned his campaign.

There followed a quarrel between the Serbian king and his heir. In the spring of 1331 they seemed to be reconciled, but in the following summer Dušan dethroned his father, who died soon afterward (Nov. 11, 1331). Dušan's sole rule now began. A man of magnificent physique, energetic, resolute, and ambitious, he fascinated his subjects and could lead them wherever he wished. Continuing the war against the Byzantines, he forced them to cede some Macedonian territory in 1334. The Byzantine civil war, which broke out in 1341, gave him a new opportunity. Having conquered all Macedonia as far eastward as Philippi, with the exception of Salonika, he assumed the title of "tsar [emperor] of the Serbs and Greeks" in 1345. He then procured the establishment of the Serbian Church as an independent patriarchate (March 1346); and on Easter Day, 1346, he had himself crowned as tsar at Skopje by his new patriarch. Dušan then turned his attention westward and southward, occupying Albania, Epirus, Aetolia (which gave him a foothold on the Gulf of Corinth), and finally Thessaly. He was dreaming of taking Constantinople when he died on Dec. 20, 1355.

The tsar Uroš, Dušan's son and successor, was a weak man. His reign (1355–71) saw the decline of Serbian power. The process of disruption was begun by Simeon, Dušan's half brother: he proclaimed himself "tsar of the Greeks, Serbs, and all Albania" and ruled Epirus and Thessaly in independence from 1359. Other provincial rulers asserted themselves in the 1360s: the Balšić family in northern Albania and along the Adriatic coast; Nikola Altomanić at Rudnik and southwestward from there; and the brothers Vukašin and Uglješa in Macedonia. Vukašin, ruling Skopje, Prizren, Prilep, and Tetovo, assumed the title of king in January 1366. Campaigning against the Ottoman Turks, who had first appeared in the Balkans in the 1340s, Vukašin and his brother were killed in the Battle of the Maritsa River, near Adrianople, in the stormy night of Sept. 25–26, 1371, when the Serbs sustained a terrible defeat still proverbially remembered as a national disaster. Two months later, Tsar Uroš died.

Prince Lazar and the Battle of Kosovo.—The foremost of the princes who took over the succession of Uroš in 1371 was Lazar, a member of the Hrebeljanović family and, by marriage, a relative of the House of Nemanja. The Balšić family, however, remained powerful; Vukašin's son Marko (the Marko Kraljević of Serbian legend) also had some territory; the Dejanović family were strong in eastern Macedonia; and Vuk Branković dominated the Kosovo region. With help from Tvrtko of Bosnia, Lazar overthrew Nikola Altomanić in 1373 and partitioned his territory with Tvrtko.

Though Tvrtko's subsequent assumption of a royal title (*see* BOSNIA-HERCEGOVINA) implied some diminution of Lazar's status

the two were in agreement on the need to cooperate against the Turks. Having reduced most of Bulgaria and Macedonia to vassality, Sultan Murad I was demanding tribute from the Serbs. Lazar resisted, and Murad, who attacked Serbia and captured Niš in 1386, was halted at Pločnik. In 1389, however, Murad led another invasion, and on June 15 the Battle of Kosovo was fought: Murad himself was killed, but the Serbs were catastrophically defeated. Lazar was taken prisoner and beheaded. The Battle of Kosovo, symbolizing the choice of death rather than compromise with the enemy, is one of the permanent memories of the Serbian national consciousness.

The Despotate.—Lazar's son Stephen succeeded to the Serbian principality in 1389 as a vassal of Sultan Bayazid I. He thus took part in several Turkish campaigns, fighting at the Battle of Rovine against the Walachians (1395), at the Battle of Nicopolis against the crusaders from Hungary (1396), and at the Battle of Ankara, where Bayazid's army was defeated by Timur (1402). Stephen, returning home via Constantinople, accepted the title of despot (that is, lord) from the Byzantine regent John VIII Palaeologus. The discomfiture of the Turks gave him a freer hand in Serbia. After nominating his nephew George Branković as his successor, he died on July 19, 1427.

George Branković was a highly cultured man and an experienced soldier. The Turks, whose power was growing again, resented George's friendly relations with Hungary, and in 1439 Sultan Murad II took the new Serbian capital, Smederevo, and captured two of George's sons, who were later blinded. George went to Hungary for help; and in autumn 1443 he and János Hunyadi (*q.v.*) led a large army against the Turks. Having advanced as far as Sofia, the allies might have taken Adrianople if the cold weather and the shortage of food had not induced Hunyadi to retreat to Hungary for the winter. George then made a separate peace with Murad, who recognized him as despot of the restored Serbia.

When Hunyadi returned to the Balkans with the Hungarian king Ulászló I (Władysław III of Poland), in 1444, he was completely defeated at Varna; and his expedition of 1448 ended in disaster at Kosovo. These two Hungarian reverses were due in part to the attitude of George Branković, who was no longer willing to commit himself hastily against the Turks. Even so, Sultan Mohammed II, after his conquest of Constantinople in 1453, decided to attack Serbia. After a first campaign in 1454, he captured Novo Brdo in 1455, while George was in Hungary seeking help again. George might have received it more easily if he had become a Roman Catholic, but this he steadfastly refused to do, lest he should forfeit the respect of his Orthodox subjects. He died on Christmas Eve 1456.

After the short reign of George's son Lazar (1456–58), there was an attempt to save the remnant of Serbia by marrying Lazar's daughter Jelena to the Bosnian prince Stephen Tomašević. But the sultan attacked Serbia again and took Smederevo on June 20, 1459. That was the end of Serbia as a territorial power.

From 1389, however, the Serbs had been migrating to Hungarian territory, and the numbers of such migrants increased after 1459. The Hungarians welcomed them as fighters against the Turks. Matthias (*q.v.*) Corvinus granted the nominal title of despot in 1471 to George Branković's grandson Vuk (d. 1485) and in 1486 to the latter's first cousin George (d. 1496 or 1497), from whom it passed to his brother John (d. 1502). From 1504, it was borne by members of other families.

Sultan Suleiman I the Magnificent conquered Šabac in 1521 and Belgrade a few months later. Crossing the Sava River, the Turks devastated Srem. The Serbs Radić Božić and Pavle Bakić took a leading part in the resistance, but Suleiman's great victory over the Hungarians at Mohács in 1526 heralded the subjection of all the Serbs to Turkish rule. John (Jovan) Nenad, known as "the Black Man," who won a large following among the Serbs of Hungary and assumed the title of tsar, was killed in 1527; and ten years later Pavle Bakić, the last titular despot, was killed in battle.

Turkish Rule.—All land in Serbia became the property of the sultan, who entrusted it to *spahis*, military tenants who held it from him in return for their services. The Serbs themselves were bondslaves of the land, most of the revenue being the perquisite of the *spahis*. In the early period of the Turkish occupation, however, there took place an event of great significance for the Serbs, namely the restoration of the Serbian patriarchate at Peć (1557). The first of the new patriarchs was Makarije, brother of the Turkish grand vizir Mohammed Pasha Sököllü (Sokolović). With a large territorial jurisdiction, the patriarch was the personification of national survival as well as the spiritual head of his people.

The defeat of the Turks outside Vienna (1683) was followed by Austrian counteroffensives in the course of which Prince Louis of Baden advanced as far south as Kosovo. The Serbs then rose against the Turks; but a new Turkish effort, under the grand vizir Mustafa Köprülü, not only won the country back but also spread such terror that thousands of Serbs, with their patriarch Arsenije III Crnojević migrated to Hungary. For these refugees the Holy Roman emperor Leopold I promulgated a charter (1690), promising to respect their Orthodox religion and guaranteeing them the rights to elect their patriarch. These undertakings, however, were not long observed.

The Banat (*q.v.*), which had a large population of Serbs, was left under Turkish rule, with the rest of Serbia and Bosnia, by the Peace of Carlowitz (Karlovci) in 1699, though other lands of the Hungarian crown were then ceded by the Turks to Austria. The Austro-Turkish war of 1716–18, however, led to the Peace of Passarowitz (Požarevac), whereby Austria obtained both the Banat and also the adjacent part of Serbia (known as Šumadija) south of the Sava and Danube rivers, with Belgrade. After Austria's unfortunate invasion of Turkish Serbia in 1737, the Peace of Belgrade (1739) restored the Turkish frontier on the Sava and Danube.

In 1766 the Turkish government abolished the Serbian patriarchate; but by the Treaty of Kuchuk Kainarji, in 1774 (*see* RUSSO-TURKISH WARS), Turkey formally acknowledged Russia's claims to champion Orthodox and Slav interests in the Ottoman Empire. When Russia and Austria together began another war against Turkey in 1787, the Serbs under Mihajlo Mihaljević formed a free corps in the Austrian service: the Austrians captured Belgrade in 1789, but retroceded it to the Turks at the Peace of Sistova in 1791. The Serbs' disillusionment at Austria's conduct was expressed in the remark of one of Mihaljević's officers, Aleksa Nenadović: "The Serbs should not trust the Germans."

The reforms initiated by the sultan Selim III in the last decade of the 18th century provoked a number of revolts by the reactionary janizaries in the outlying provinces of the Ottoman Empire. At the end of 1801 Haji Mustafa, the Ottoman governor of the pashalik of Smederevo (or of Belgrade; *i.e.*, Serbia), was assassinated by a faction of janizaries, four of whom then assumed power themselves as *dahije*, or deys. Their regime was so tyrannical that the Serbs began to plan insurrection for the spring of 1804. Aleksa Nenadović, unmindful of the strictures that he had made after the Peace of Sistova, wrote a letter appealing for Austrian help, but his letter fell into the hands of the *dahije*, who put him and four other patriots to death (Jan. 23, 1804). (M. A. P.)

Karageorge and the First Rising (1804–13).—One of the Serbian patriots, Karageorge (*q.v.*; *i.e.*, Karadjordje), fled to the forest to escape death at the hands of the janizaries. Within ten days 2,000 men joined him, and they began attacking Turkish positions. When Karageorge was proposed for the supreme command of the incipient insurrection (Feb. 14, 1804), he declared: "I know myself too well, I am violent . . . ; if I were disobeyed I should not know how to use fair words. I should strike." "All the better," came the reply; "we want a leader whom everyone will fear." The *dahije* tried to negotiate with Karageorge, but after two ineffectual meetings the Serbs seized Rudnik, Valjevo, Jagodina (now Svetozarevo), and Šabac (March–April). They unsuccessfully asked for Austrian and Russian protection. Austrian mediation, at the request of the *dahije*, came to nothing, and in May the Serbs addressed a petition to Sultan Selim III, took Požarevac, and sent a deputation to Russia. A meeting between Karageorge and the pasha of Bosnia, at the latter's request, took place in July, and the Serbs were allowed to capture and behead the four *dahije*. The Serbs then brought their demands before the pasha, but he rejected them.

The Serbian representatives were well received in St. Petersburg; but, for fear of Napoleonic France, the Russian emperor Alexander I decided to uphold Turkey's integrity, and advised the Serbs to send a deputation to Turkey, which they did in June 1805. On learning from the Austrians what the Serbs intended, the Turkish government decided to crush the rising. The Serbs, however, conquered Karanovac and Užice (now Titovo Užice) (June–August). Until then the rising had been directed against the outlawed janizaries, but at Ivankovac, on Aug. 18, 1805, the Serbs for the first time defeated Turkish regular troops. Intoxicated by their successes and aiming now at independence, they organized the Serbian State Council, a supreme administrative and judicial body. They conquered Smederevo in November and sent a deputation to Austria. Though the Serbs were again advised to settle their differences directly with Turkey (February 1806), both Austria and Russia promised their good offices, fearing that Napoleon I's recent victory over Austria (*see* NAPOLEONIC WARS) might induce the Serbs to turn toward France. Napoleon, however, wrote in June 1806 to Sultan Selim that he should use the strongest means to bring the Serbian rebels into subjection. Even so, the Serbs continued their successes; they took Poreč, Paraćin, Ražanj, Aleksinac, Kruševac, the Stari Vlah district, and half of the Novi Pazar district; routed the Turkish regulars again at Mišar and at Deligrad; and finally in December won Belgrade. They had liberated Serbia without outside help.

War broke out between Turkey and Russia at the end of the year. Asked by Turkey for help against Russia, the Serbs replied: "Serbia considers herself an independent state . . . and refuses to take up arms against her brothers and allies" (March 31, 1807). The Russians, however, instead of helping the Serbs, at first asked them to make a diversion; but when Karageorge protested they sent a token force to Serbia. The allies routed the Turks at Štubik, a Serbo-Russian convention was signed on July 10, 1807, and K. K. Rodofinikin arrived as the first Russian diplomatic agent in Belgrade. The Serbs of Serbia made contact with their fellow nationals in Bosnian, Austrian, and Montenegrin territory.

Defeated by Napoleon at Friedland (modern Pravdinsk), Russia concluded the armistice of Slobozia with Turkey (Aug. 24, 1807), in which Serbia was not even mentioned. Karageorge resented such behaviour, and his attitude aroused suspicion in some of his commanders, who wanted more freedom of action for themselves. The Serbian State Council, however, seeing that Russia was attempting to gain exclusive influence, passed the first Serbian constitution (Dec. 26, 1808): this proclaimed Karageorge "first and supreme hereditary leader," who was to command the loyalty of all other Serbian leaders; but it stipulated that his laws and decrees were to be issued through the council, with the common consent of its members, for transmission by it to the military commanders of the provinces. Schools were opened everywhere, including the Belgrade *lyceum* (later university). Karageorge, after informing Russia, rejected an Austrian offer of protection, as well as a Turkish offer of peace.

The Russo-Turkish War restarted in March 1809, and the Russian commander in chief informed the Serbs in May that Alexander I aimed "at the liberation of Serbia from any kind of dependence on Turkey." The Serbs took Sjenica, Suhodol, and Novi Pazar (May–June), but their commanders were torn between Karageorge and his opponents; and the Turks, concentrating a large army in the south, won a victory at Kamenica in June and recovered Soko Banja, Deligrad, Stalać, and Požarevac by the end of August. As the Russian armies were idle, the exasperated Serbs sent a delegate to Napoleon and appealed again to Austria. The Russians, however, invaded Bulgaria (at that time still Turkish territory) in September and the Turks withdrew their armies from Serbia. With renewed confidence in Russia, an assembly of Serbian leaders challenged Karageorge by sending a delegation to the Russian commander in chief, P. I. Bagration, to disavow Karageorge's fruitless approaches to France and Austria. Russia sent 2,500 men to Serbia and in the autumn of 1810 the Turks were defeated at Varvarin and Loznica. Finally, at Karageorge's request, Russia agreed to establish four garrisons in Serbia. To consolidate his position and to silence the opposition, Karageorge convened a leaders' assembly, which passed a new constitution; and before the Russian garrisons reached Belgrade he appointed his first government (Jan. 20, 1811).

On the eve of attack by Napoleon, Russia concluded the Treaty of Bucharest with Turkey (May 1812). Its Art. VIII stipulated an amnesty for all Serbs, the demolition of newly built fortresses, the surrender of old ones, and a limited autonomy, but details were to be negotiated directly between Serbs and Turks. Forsaken again, the Serbs were bitterly disappointed. The opposition held Karageorge responsible, and dissension grew among the leaders. Rejecting a Turkish summons to surrender (July 1813), the Serbs were attacked from three sides by powerful armies and were defeated. Karageorge took refuge in Austria, and on Oct. 7, 1813, the Turks reentered Belgrade. For a fortnight, every Turk was permitted to kill any Serb over 15 years old and to enslave women and children. In a single day 1,800 slaves were sold in Belgrade. After Napoleon I's defeat at Leipzig, the Turks granted a general amnesty.

The Second Rising and Miloš Obrenović I (1814–39).—Miloš Obrenović I (*see* MILOSH), one of the leaders who had not followed Karageorge into exile, was in August 1814 appointed head of three Serbian central districts by the new vizir, Suleiman Pasha Skopljak. When the vizir, to enrich himself, levied new taxes, Serbian leaders sent a delegate to the Congress of Vienna, but Russia was the only power there to feel any interest in Serbia's fate. Russia indeed tried to raise the Serbian question, but failed because of Napoleon's return to France from Elba (March 1815). Then Miloš, who in September 1814 had even helped the Turks to suppress a rebellion, decided to act against them. He summoned all the Serbian leaders to the Takovo church on Palm Sunday and a second rising began.

The Serbs routed the Turkish regulars at Ljubić, Palež (Obrenovac), and Dublje and, in August 1815, took Požarevac. Controlling an area of about 24,440 sq.km., Miloš styled himself "supreme prince and ruler of the Serbian nation" and was recognized as such by the *Skupština* (*see* below) and, in November, by Marashli Ali Pasha, the new vizir; but when the Turks mustered two large armies he realized his weakness and tried to compromise.

The *Skupština* was the National Assembly of the Serbs. Under Miloš, it comprised the 22 district heads (*nahijske starešine*), who were nominated by the prince; and representatives of the town and parish councils, whose members—prominent merchants, priests, and other local notables—were nominated by the district heads. A purely consultative body, it was convened at the prince's pleasure, usually once a year. It was distinct from the leaders' assembly which met regularly twice a year for the discussion of taxes between the prince, the pasha, and the district heads.

In September 1815, Russia threatened military intervention, whereupon the Turks lent Miloš a helping hand. At the beginning of 1816, Serbian demands were partly accepted. Miloš agreed to the supremacy of the Turkish authorities, but soon secured parallelism between Serbs and Turks and finally eliminated the latter. This settlement was guaranteed only by the vizir's word, and Miloš wanted it confirmed by Sultan Mahmud II. He approached the Turks as leader of the Serbs while impressing the Serbs as being a friend of the Turks. Simultaneously, he extended Serbian autonomy, disposed of all rivals, and even had Karageorge assassinated (July 1817)—an act which started the feud between the Obrenović and Karadjordjević (Karageorgevich) dynasties. He then convened the *Skupština* and was recognized as hereditary prince (Nov. 18, 1817).

At last the sultan delivered a decree, but when his representative reached Belgrade and read it to the *Skupština*, the Serbs rejected it (September 1820). As the sultan consistently deferred recognizing Miloš's agreement with the vizir and any solution of the Walachian and Greek problems (*see* EASTERN QUESTION), Russia in March 1826 sent an ultimatum demanding a settlement and fulfillment of Turkey's obligations toward Serbia under the Bucharest treaty. A new Russo-Turkish convention was signed at Akkerman in October, and Turkey undertook to settle the Serbian question within 18 months and to grant Serbia "independence in its interior administration." Turkish negotiations with the Serbs

FIG. 2.—SERBIA 1817–1913

were interrupted by the Russo-Turkish War of 1828–29, which ended with the Treaty of Adrianople (Sept. 14, 1829). As regards Serbia, this treaty reiterated the stipulations of the Akkerman convention, and Turkey undertook to restore the "Six Districts" (Jadar, Crna Reka, Paraćin, Kruševac, and Stari Vlah in the south, and Krajina in the east) which had been liberated during the first rising but later recovered by Turkey. A *hatti-sherif* (imperial edict) was promulgated in Belgrade on Dec. 12, 1830, with a brevet recognizing Miloš as hereditary prince. Serbia became an internationally recognized autonomous principality under Turkish suzerainty and Russian protection; Turkish garrisons were retained only in Belgrade and five other fortresses; and Turkish civilians had to sell their properties in the towns and leave the country or settle inside the walls of the citadels. Turkey, however, still kept the Six Districts. Taking advantage of Mohammed Ali's rebellion in Egypt (1831), Miloš fomented sedition in the Six Districts, gathered troops along their border, and informed the Porte that he was ready to help the rebels if the Turks tried to crush them. Turkey at last gave up, and the union of the Six Districts with Serbia was proclaimed (June 10, 1833), bringing the area of the principality to about 37,740 sq.km. Miloš then paid a state visit to the sultan.

Before achieving his goal, Miloš had solved the agrarian problem by enacting that the land, which the *spahi* tenants of the sultan (*see* above, *Turkish Rule*) had gradually vacated, should belong to those who worked it. In this way he put Serbia's economy on a sound basis and also unwittingly prepared the way for the future peasant democracy of Serbia. He instituted a judicature, formed police forces, and opened new schools. The Serbs agreed with his policy toward the Turks, but in home affairs they were against his despotic tendencies and wanted more representative institutions. Having suppressed five rebellions, Miloš eventually had to let the *Skupština* pass a constitution limiting his powers (1835). Austria, Russia, and Turkey immediately protested against this constitution as too liberal, and Miloš gladly abolished it. Even so, paradoxically, autocratic Russia tried to restrain the prince's Oriental despotism, while liberal Great Britain, in order to build up Serbia into a bulwark against Russian advance in the Balkans, supported the extension of his authority. Finally a new constitution, negotiated in Istanbul between Turkey and Miloš's representative Avram Petronijević (1791–1852), who was under Russian influence, was promulgated by *hatti-sherif* in 1838. A State Council of 17 life members and a cabinet of four were duly instituted. As soon as the constitution became operative (Feb. 24, 1839), the councilors, all of whom belonged to the opposition, were appointed.

The Serbs had sent countless envoys to Istanbul; the Turkish officials had ignored some of them, negotiated with others, and imprisoned others for years or kept them as hostages. Bribery had served the Serbian cause well enough; but the achievement of autonomy was due above all to the political skill, instinctive realism, and indomitable peasant energy of Miloš Obrenović. Temperamentally, however, Miloš could not tolerate having to submit his decisions to a State Council whose members did not depend on him. After trying unsuccessfully to abolish the new constitution by force, he was compelled to end his first reign by abdication (June 13, 1839).

The Period of Oligarchy (1839–58).—Milan Obrenović II, the elder son of Miloš, died after a nominal reign of only 25 days and was succeeded by his infant brother Michael (*q.v.*) Obrenović III. The sultan therefore appointed a regency. Its members represented the so-called "Defenders of Constitutionalism," and it was the embryo of the oligarchy which ruled Serbia for the next two decades. Conflict soon grew up between the Defenders and Prince Michael. The regents, having been sent into exile and then pardoned, finally organized an armed rebellion in complicity with the vizir. The young prince then fled to Austria (Sept. 7, 1842).

On Sept. 14, 1842, the *Skupština* elected Karageorge's son Alexander (*q.v.*) to the throne. Turkey recognized the election, but Russia contested it on the ground that it followed a rebellion instigated by the vizir. At a second election (June 27, 1843), Alexander was again chosen. Two rebellions fomented by Miloš were suppressed in 1844; but the rivalry between the Obrenović and the Karadjordjević dynasty was to continue for 60 years.

Serbia's external relations for the next 12 years were marked by the "Plan" of Ilija Garašanin for foreign policy (*see* GARAŠANIN, ILIYA), by the Hungarian revolution of 1848–49, and by the Crimean War. On the basis of a memorandum by the Polish patriot Prince Adam Jerzy Czartoryski (*q.v.*), Garašanin, the Serbian home secretary, in 1844, developed a plan for a general rising of all Christians under the Turkish yoke and for the creation of a large South Slav state under Serbia's leadership. Originally the plan refrained from anticipating the liberation of the South Slavs of the Austrian Empire; but when the Hungarian revolution broke out Garašanin wanted Serbia to help the Serbs of the Vojvodina, who rose against the Hungarians. The government decided not to intervene, but numerous volunteers flocked to help their kinsmen, and Stefan Petrović-Knićanin (1807–55), a former member of the State Council, became the insurgents' commander in chief. Serbia's choice of neutrality during the Crimean War and Garašanin's connections with the West were rewarded by the Treaty of Paris at the end of that war (March 30, 1856), since it placed Serbia's autonomy under the collective guarantee of the Powers and restricted Turkey's sovereign rights by forbidding armed interference in Serbia without the Powers' consent.

At home, the two main objects of the Defenders were to secure the civic rights of the Serbs and to limit the prince's power. The first was achieved by introducing modern institutions, a competent bureaucracy, and a series of new laws. Whereas the old *Skupština* had never been able to limit Miloš Obrenović I's power, the newly instituted State Council succeeded in limiting Alexander Karageorgević's. Later, however, people realized that their political rights were not necessarily guaranteed by the securing of their

civic rights and that the prince's power was limited not by them, but only by an oligarchy. Furthermore, the stipulations of the constitution were not precise, and the organic laws passed after the constitution were sometimes inconsistent with it. Since the prince's power was more limited by the organic laws than by the constitution, a struggle was inevitable between him and the State Council, which despite its restricted composition was really a kind of representative body of true national leaders. When Alexander asked the council, in 1848, to make the laws conform to the constitution, it made no reply; and his relationship with it became intolerable after 1853, when, under Russian pressure, he dismissed the pro-Western Garašanin (his prime minister and foreign secretary from 1852). In the earlier phases of the Crimean War the prince and the council, which was pro-Russian, observed a truce, but when it became evident that Russia could not win the war dissension was revived. In March 1855, when Alexander informed the council that he was still waiting for an answer to his demand of 1848, he received a flat refusal, and a life-and-death struggle ensued.

Incited by the long-deposed Miloš, Stefan Stefanović-Tenka (1797–1865), the president of the State Council, became the ringleader of a conspiracy to overthrow Alexander; and with a few of his fellow conspirators he organized a new plot against the prince's life (1857). When this plot was exposed, 4 of the 17 members of the council were imprisoned, 6 were compelled to resign, 1 resigned voluntarily, and 2 fell under suspicion. Alexander appointed new councilors; but Russia and France, resentful of the pro-Austrian policy which he had recently adopted, took his action against the old council as a pretext for asking Turkey to send a commissary to Belgrade to settle the conflict. The imprisoned councilors were reprieved and handed over to Turkey, those who had resigned were reinstated and a new government was sworn in. The discovery of the plot had enabled Alexander to strike a major blow at the power of the council, but his victory was short-lived. On Dec. 23, 1858, the *Skupština* deposed him.

Miloš Restored (1858–60).—The *Skupština* which dethroned Prince Alexander recalled Prince Miloš Obrenović I, whose election was recognized by Turkey. The council and the *Skupština* had overthrown Alexander together, but the latter elected Miloš without consulting the former. This created a rift which provided the elements of the first two political groups, based on personal allegiances rather than on any nationwide party or class structure: namely the Conservative supporters of the council, and the Liberal followers of the *Skupština*. The Conservatives, who were called *kajmakanci* (from *kajmakan*, regent) because their leaders had wanted to be elected as regents on the overthrow of Alexander, desired only a change of dynasty, with the legislative power still shared between the council and the prince and with the civil service retaining its entirely conservative character. The Liberals, on the other hand, who were called *dukatovci* because their leaders' anonymous articles appeared in the press over a symbol meaning ducat, wanted to make the *Skupština* the third partner in the legislature and to reform the civil service.

As soon as Miloš reached Belgrade (Feb. 6, 1859), he appointed new councilors and a Liberal government and began dismissing the civil servants. The difference between his populism and that of the Liberals was evident: both wanted a *Skupština*, but Miloš expected it to be dependent on him, while the Liberals wanted it dependent on a strong political party. Miloš, disliking new ideas, asked Jevrem Grujić (1840–1913), his permanent home undersecretary and one of the two Conservative leaders: "If you are responsible, then who am I? I prefer power with responsibility to irresponsibility without power." He estranged the Liberals without attracting the Conservatives, ignored the constitution and the council, and reduced the *Skupština* to a body of illiterate peasants. At the beginning, however, he made himself popular with the peasantry by flattering them. Whereas during his first reign he had been a despot feared by the masses, during his second he was a potentate afraid of them. When he died (Sept. 26, 1860), everyone was tired of his tyranny and demagogy.

Michael Obrenović and Enlightened Absolutism (1860–68).—His son Michael, returning to the throne after 18 years, formulated his program as follows: the liberation of the Balkans; the creation of a large South Slav state led by Serbia; a strict and just application of the law; and economic development. To achieve this he wanted the Serbs to acquire the habit of submitting to law. "As long as a Prince Michael is on the throne," he said, "it is his wish that law be the highest will in Serbia." By law he meant the prince. An era of enlightened absolutism began.

Prince Michael stopped all political persecution and reinstated the dismissed civil servants or gave them a pension, but was unable to reconcile the Conservatives and the Liberals. His first government therefore was a caretaker one, and it passed the laws that he desired. The Council's authority was limited to legislation. The irremovability of the councilors was abolished. The ministers were made responsible only to the prince, not as before both to the Council and to the prince. The government was granted discretionary powers to dismiss and to pension off members of the civil service. The *Skupština* remained as a consultative representative body though its powers were somewhat increased. There was no freedom of the press.

No longer needing to cultivate the Liberals, Michael appointed a Conservative government with Garašanin as prime minister. The latter was thus able to develop and enforce his "Plan," the main object of which was war with Turkey. The strength of the Serbian army was raised from 4,000 men to 50,000 regulars and 40,000 reserves, with a strong artillery; its organization was entrusted to Hippolyte Florentin Mondin, a French army officer. The bill was met by the prince out of his own wealth.

Before starting war, Serbia had to settle the question of the Turkish garrisons. This question was in fact raised by the action of the Turks themselves in June 1862: a brawl between Turks and Serbs queuing for water at a fountain in Belgrade led to firing from the Turkish barracks; and when fighting broke out again later in the day the Turks bombarded the town from the citadel. The Serbian government would have declared war on Turkey but for the intervention of the powers: instead, it negotiated through Jovan Ristić (*q.v.*), its diplomatic representative in Istanbul and he secured the withdrawal of the last Turks. After Prince Michael had paid a state visit to the sultan, the fortresses were surrendered in April 1867. In further preparation for war, meanwhile, Serbia concluded a series of alliances: with Montenegro (1866), with the Bulgarian Revolutionary Committee (1867), with Greece (1867), and with Rumania (1868). The first Balkan League was thus created.

Prince Michael's reign was clouded by two unfortunate incidents. The first arose from the discovery of a plot to overthrow the dynasty. After the court which first tried the accused had found some of them not guilty and had sentenced others for attempting to organize a plot and for *lèse-majesté*, the high court decided that talking in private was not a crime and released them all. The government thereupon put the judges of the high court on trial, and they were all convicted of passing a wrong verdict (1864). The second incident arose from Michael's desire to marry his cousin, Katarina Konstantinović, with whom he was deeply in love. Garašanin, who was against this marriage, was dismissed (1867), and the prince formed a highly unpopular government. Michael now found himself in a most difficult position: the youth of Serbia wanted war with Turkey, but he could not declare war because Russia withdrew support after Garašanin's dismissal; the Liberals, who were in favour of his projected marriage, wanted more political freedom, which he was not willing to concede; and the Conservatives, who supported his strong-handed rule, were against his marrying.

On June 10, 1868, Michael was assassinated. The government suspected Alexander Karageorgević of being behind the plot, but the man who really organized it was the brother of a criminal whom the prince had refused to reprieve.

The Regency for Milan Obrenović IV (1868–72).—Prince Michael having died without leaving a son and heir, the government appointed a provisional regency council until the *Skupština* could elect a new prince. However, Milivoje Petrović-Blaznavac (1824–73), secretary for war, challenged both the regency and the government. He announced by a military *pronunciamiento* that

Milan Obrenović IV, a boy of 14 and the only Obrenović male, had been elected prince. When Milan reached Belgrade, he was greeted by Petrović-Blaznavac and the Army, but the regency and the government were not represented. Eventually they compromised, accepting Milan as prince-elect until his election could be "proclaimed" by the *Skupština*. This was accomplished on July 2, 1868, and Petrović-Blaznavac, Ristić, and Jovan Gavrilović (1796–1877) were appointed regents for Milan. Although elected second regent, Ristić soon took the lead.

The *Skupština* then demanded a widening of its powers and more liberal institutions. The regents drafted a new constitution without consulting either Turkey or the council, with which they still shared the legislative power. The draft was submitted to a Constituent *Skupština* and was passed with a few alterations (July 11, 1869). Its main features were: recognition of the hereditary rights of the Obrenović dynasty; the legislative power to be shared by the prince and the *Skupština* and not, as before, by him and the council; the right to table bills to be reserved to the government, while the *Skupština* could only propose them to the government; the council to remain solely as a consultative body; the government to be responsible only to the prince; the principle of the freedom of the press to be acknowledged; and, finally, judges to be independent but not irremovable. Turkey recognized *de jure* both the hereditary right of the Obrenović dynasty and Serbia's right to pass a new constitution.

Before adjourning, the Constituent *Skupština* in 1870 passed an electoral law, regulations of parliamentary procedure, the law of ministerial responsibility (whereby ministers were not made politically responsible to the *Skupština* but could be required by it to answer certain criminal charges before a special court), the statute of the Council of State (whereby the political State Council was transformed into an administrative Council of State), and an act giving limited freedom of the press. The reforms were not a victory of the people over the bureaucracy, but rather a compromise, and they produced different reactions: the peasants were satisfied that the *Skupština* should have wider powers; the Conservatives disapproved, preferring to continue to support enlightened absolutism; the Liberals, while accepting the new policy, looked forward to further progressive reforms; and a new militant Socialist group led by Svetozar Marković (*q.v.*) was scornfully opposed to such "reforms." Ristić, responsible for the new constitution, avoided stating anything in it which might give the slightest suggestion that the government was dependent on the *Skupština:* Serbia was a constitutional, not a parliamentary, monarchy. Ristić felt it his first duty to protect the Obrenović dynasty and the state, on the one hand, from supporters of the Karageorgević dynasty and, on the other, from Marković's Socialists: he persecuted both sides of the opposition, which considered that he deliberately exaggerated the danger so as to strengthen himself. Meanwhile he paid great attention to education and to the economic development of the country: he balanced the budget and continued to improve the Army (though he slowed down its equipment).

Before being elected regent, Petrović-Blaznavac had promised Austria to "respect Turkey's integrity and to liberate Serbia totally from Russian influence." The relationship with Austria was therefore good, though Ristić managed to prevent Serbia from falling under an exclusively Austrian influence. Since the constitution of 1838 and the sharing of legislative power between the prince and the council had been of Russia's making, the Russian government regarded the constitutional changes in Serbia as a defeat of its own policy and distrusted the new regime. Nevertheless it wanted to draw the young prince closer—and definitely before he came of age. He was invited to visit the emperor Alexander II at Livadia in 1871. This visit resulted in a deterioration of Serbia's relationship both with Turkey and with Austria, and the regency inclined more toward Russia.

As the Serbs of Montenegro adopted a separatist attitude and as Bulgaria became antagonistic to Serbia, the Balkan League which Prince Michael and Garašanin had brought into being gradually fell apart. Ristić in any case had no faith in such a league, believing that the Balkan states, even if united, were too weak to defeat Turkey without outside help, and that any success won with the help of the Great Powers would make Serbia too dependent on them.

Milan's First Governments (1872–75).—When Milan attained his majority (Aug. 22, 1872), the regency ended, but Milan retained Petrović-Blaznavac in office as prime minister and Ristić as minister of foreign affairs. Petrović-Blaznavac died, however, on April 5, 1873, and Ristić then formed a short-lived nonparty government which alienated both Conservatives and Liberals. For various reasons, moreover, Serbia's relationship with Germany, France, Austria, and Turkey became disturbed. To improve the situation, Milan paid visits to Vienna and to Paris. He was well received, but his visits worsened his relationship with Turkey. Ristić, feeling that he was losing Milan's confidence, submitted his resignation (Nov. 3, 1873) in the belief that Milan would not dare to accept it; but Milan, who could not forgive his former guardian for having tied his hands with the constitution of 1869, in fact accepted it. Backed by the League of Three Emperors (Austria, Germany, and Russia), Milan then appointed a Conservative government; in May 1874 he paid a state visit to the sultan. After a general election in October had returned a Liberal majority, discontent grew in the *Skupština*, and, for the first time, it was dissolved. A fresh election, however, in August 1875, gave the same result, and Milan reluctantly appointed a Liberal government.

The Wars of 1876–78 and Independence.—In July 1875 the Serbs of Hercegovina had risen in rebellion against the Turks. The Serbian Liberal government decided to help the insurgents even at the risk of war; but Milan, advised by the League of Three Emperors against intervention, in October 1875 appointed a new government comprising both younger Conservatives and Liberals. As the political groups were divided on the question of war, the decision rested with Milan, who could not curb the growing nationalism incited by the Russian Slavophils (*q.v.*), though officially Russia tried to restrain it. In order to wage war, the prince had to appoint a Liberal government again (May 1876).

Ristić, who became foreign minister, preferred to avoid war, but Turkey rejected his proposal to entrust Serbia with the administration of Bosnia-Hercegovina. The Serbia government then declared war (June 30, 1876), mistakenly believing that the first shot would produce a general Balkan rising and that the Slavophils would force Russia to join. No such rising occurred, and Russia not only abstained from war against Turkey but also agreed, on July 8, that Austria-Hungary, instead of Serbia, should occupy Bosnia-Hercegovina. Even so, the Serbs chose a Russian general, M. G. Chernyaev, to be their commander in chief. His troops, however, in which Russian adventurers outnumbered Serbian officers, were defeated by the Turks at Veliki Izvor on July 18. After an interval of confusion, during which the Russian foreign minister, A. M. Gorchakov, advised a cease-fire and the Russian ambassador to Turkey, N. P. Ignatiev, advised continuation of the war, the Serbs won a victory at Šumatovac. The Great Powers then tried mediation, but Turkey's terms proved unacceptable. Finally the Serbs, after a series of reverses (September–October) appealed to Russia, and on March 1, 1877, peace was concluded on the basis of the status quo.

When Russia in April 1877 declared war on Turkey (*see* RUSSO-TURKISH WARS), Serbia again wanted to fight, but at first received evasive answers from Russia. The Turks, however, halted the Russian advance at Pleven, and Russia then appealed for help. Serbia declared war against Turkey on Dec. 13, and proceeded to liberate large tracts of territory on the southern frontier. Then Russia imposed the Treaty of San Stefano on Turkey (March 3, 1878) without consulting Serbia. This treaty gave Serbia independence and possession of Niš, but only 150 sq.km. of new territory, while a Greater Bulgaria, which Russia needed for further expansion, was created. The opposition of the Great Powers frustrated this plan; and the Congress of Berlin (*q.v.*) in July besides granting Serbia's independence increased the amount of new territory to 3,860 sq.mi. (10,000 sq.km.), adding Pirot, Toplica, and Vranje to Niš. This was achieved through Gyula Andrássy, the Austro-Hungarian foreign minister, after Ristić had come to terms

with him. Disillusioned with Russia, the Serbs now looked to Austria-Hungary. Serbia's unrestricted independence was proclaimed on August 22.

Ristić, appointed prime minister at the head of a new government in October 1878, had to negotiate a railway convention and a trade agreement with Austria-Hungary. The railway convention was signed on April 9, 1880, but the trade agreement raised questions. Even at the risk of a tariff war, Ristić could not sign an instrument which would make Serbia economically dependent on Austria-Hungary. The government was hampered in the *Skupština*, and Milan promised Vienna in September to settle the matter with or without Ristić. When Ristić resigned (Oct. 26, 1880), Milan appointed a pro-Austrian Progressist (younger Conservative) government headed by Milan Piroćanac (1837–97); the Progressists and the Radicals led by Nikola Pašić (*q.v.*) fought an election against the Liberals in December and won nearly all the seats; and on May 5, 1881, a trade agreement which satisfied Serbian peasant needs was concluded with Austria-Hungary.

When Austria-Hungary, Russia, and Germany renewed the League of Three Emperors on June 18, 1881, they divided the Balkans into spheres of influence: Austria-Hungary agreed to the eventual absorption of Eastern Rumelia by Bulgaria, Russia's protégé; and Russia agreed to the annexation of Bosnia-Hercegovina by Austria-Hungary. The Serbo-Austrian secret convention of June 28 was a consequence of these agreements. Negotiated by Milan, its terms, before it was signed, were known only to his foreign secretary, Čedomilj Mijatović (1842–1932). By it, Serbia undertook neither to conclude any political treaty without Vienna's consent nor to tolerate any agitation against Austria-Hungary, while Vienna in return promised to support the Obrenović dynasty, to recognize Serbia as a kingdom, and to sanction Serbian expansion to the south. Later, when its existence was first revealed (1893), this convention was considered humiliating, but it is better regarded as an act of political expediency. Estranged from Russia, Serbia could preserve independence only if Austria-Hungary regarded it as inoffensive.

Milan's Kingdom (1882–89).—Though political groups had existed since 1858, it was only in 1881 that formally organized political parties came into being in Serbia: the Radical Party, the Progressist Party, and the Liberal Party. Acts were passed abolishing censorship, guaranteeing the freedom of the press, the right of public meeting and association, and the irremovability of judges. The Army was reorganized and the number of schools increased, but higher taxation made the government unpopular. The bankruptcy of the Union Générale (Jan. 30, 1882), Serbia's bankers, aggravated a crisis.

Serbia was proclaimed a kingdom on March 6, 1882. The Progressists' policy, however, alienated the Radicals, and a long political fight between the latter and King Milan ensued. The king moreover undermined the prestige of the dynasty by his personal policy and by his notorious private affairs. A general election in September 1883 returned a Radical majority, and the government resigned; but the king appointed first a nonparty government and then again a Progressist one (Feb. 19, 1884), headed by Milutin Garašanin.

The union of Eastern Rumelia with Bulgaria, precipitated by a *coup d'état* in Plovdiv on Sept. 18, 1885, and accepted by Bulgaria, disturbed the Balkan balance of power. Serbia decided to prevent it. Though the war was planned to be a *Blitzkrieg*, Serbia postponed action for more than three weeks because Vienna was officially against war, though Budapest and the Austro-Hungarian minister in Belgrade, Graf Rudolf von Khevenhüller, were in favour of it. When Milan launched his attack (Nov. 14), the Bulgars had already concentrated troops west of Sofia. The Serbs were defeated at Slivnitsa (Nov. 17–19) and driven back across the frontier before Austria-Hungary checked the Bulgars by threatening intervention. The Treaty of Bucharest (March 3, 1886) made peace between Serbia and Bulgaria without any gain to the former. (*See* SERBO-BULGARIAN WAR.)

Milan and the Progressists became unpopular. An election (May 1886) returned a one-seat Progressist majority. In 1887, after Milan and his queen, Natalie, had separated (April 18), the government resigned (June 13). Ristić, who wanted a *rapprochement* with Russia and a new constitution, became prime minister at the head of a Liberal-Radical coalition, which opposed the Progressists in a new election (September). When the Radicals won a majority, they withdrew their support, and Ristić resigned (Dec. 31). For the first time, a homogeneous Radical government was appointed, and a general election in March 1888 resulted in an overwhelming Radical victory. The Radicals, however, passed new laws which the king refused to sanction, and the government resigned (April 26). The king appointed a caretaker government.

With the German chancellor Bismarck's complicity, King Milan had his young son and only child, the crown prince Alexander, abducted from Wiesbaden (July 14, 1888), where he had been in the custody of the king's estranged wife. This act and the king's subsequent divorce (Oct. 24) were serious blows to the prestige of the dynasty. To rehabilitate it, the king initiated a liberal constitution based on a compromise between Liberal, Progressist, and Radical views. He appointed a tripartite Constituent Commission (Oct. 27), over which he successfully presided. An elected Constituent Assembly passed the new constitution (Jan. 3, 1889). This guaranteed civil and political rights, including the freedom of the press; abolished summary courts and the king's power to suspend civil rights in case of emergency; gave the *Skupština* the right to table bills, to receive petitions, to put down questions, to hold inquiries, to pass the budget, and to supervise the government; introduced direct voting by secret ballot; and admitted the principle of local government. Thus Serbia became both a constitutional and parliamentary monarchy. Milan then abdicated (March 6, 1889) in favour of his son Alexander (*q.v.*) for whom he appointed Ristić, Gen. Jovan Belimarković and Gen. Kosta Protić as regents.

The Regency for Alexander Obrenović (1889–93).—Milan having left Serbia, the government resigned, and the regents appointed a Radical government under Sava Grujić (1840–1913). The government passed new laws qualifying the balance between the executive and the legislature to the benefit of the latter; freed political parties from coercion; promoted economic development, and reduced the deficit.

Pressure from Austria-Hungary caused the regents to give an undertaking to comply with the secret convention of 1881. This had been renewed by Milan before he abdicated; but the government was unaware of its existence, and mutual distrust existed between the Radicals and Austria-Hungary. The government was reconstructed in February 1891, and for the first time Pašić became prime minister. Both Ristić and Pašić tried to improve relations with Russia, but while Pašić could do this freely, Ristić was hampered by his knowledge of the secret convention. Even so the young Alexander paid a state visit to Russia. As the radicals became too powerful, the opposition objected that neither person nor property was adequately secured, and that party interest was put before everything else; but the Radicals were more embarrassed by the interference of the king's parents, the divorced Milan and Natalie, and by the public discussion of their quarrel.

Protić, the third regent, died on June 16, 1892. His successor had to be elected by the *Skupština*. Ristić knew that the new regent would be a Radical, but he wanted to avoid having Pašić. To assure Pašić's election, the Radicals proposed a fresh poll and a new *Skupština* to elect the regent. Ristić opposed dissolution and the government resigned (Aug. 21). Despite a Radical majority, a Liberal government under Jovan Avakumović (1841–1928) was appointed. Only then was the *Skupština* dissolved. A political fight between the Liberals and Radicals ensued. A fresh election (March 1893) returned a manipulated four-seat Liberal majority. Then by an unexpected *coup d'état* on April 13, 1893, the young king, although still a minor, proclaimed himself of age, dismissed the two regents and the Liberals, and appointed a Radical government under Lazar Dokić (1845–93). Except for the Liberals, public opinion accepted the change. The Radicals, who had helped the king to make his *coup*, were jubilant.

The Last Decade of Obrenović Rule (1893–1903).—In the first eight months following Alexander's *coup* there were three successive Radical governments. These ratified a new trade agre

ment with Austria-Hungary, reorganized the National Bank, negotiated a new state loan, and brought the former Liberal government to trial. In January 1894, however, the king recalled his father, Milan, a bitter anti-Radical, and brought the first of three nonparty governments to power. On Milan's prompting, he suspended the constitution of 1889 (May 21, 1894) and reestablished that of 1869. While the Liberals and the Progressists were overjoyed, the Radicals were indignant at this retrogressive step, which did irreparable damage to the dynasty.

New elections in April 1895 returned a Progressist majority, and a Progressist government took office in July. It passed several financial and economic laws and secured the appointment of Serbian instead of Greek bishops in Macedonia; but its pro-Russian policy exasperated Austria-Hungary, and Hungary closed its frontier to the import of Serbian live pigs. The "Pig War" worsened after the princes of Bulgaria and of Montenegro paid state visits to Serbia (1896). Replaced in the government by the Radicals, the Progressists decided to dissolve their party (1897) in order to disguise their failure. A new election returned a Radical majority, which tried to revive the Balkan League and signed a political agreement with Bulgaria. The king paid state visits to Sofia, Cetinje, and Athens.

From 1898 to 1900 Serbia was ruled by the king's father through a "strong" nonparty government under Vladan Djordjević (1844–1930). "General H.M. King Milan" was appointed commander in chief of the Army, and became co-ruler with Alexander. Elections in June 1898 returned a docile Liberal majority. Djordjević, blaming parliamentarism for party warfare, passed retrogressive political laws, but promoted economic development. He expanded the army, increased the number of agricultural and technical schools (but decreased the number of elementary and secondary schools); and reduced the budget deficit. He tried unsuccessfully to promote good relations with Turkey and kept up friendship with Austria-Hungary. Russia was against the regime because of King Milan. The government used an attempt on Milan's life (1899) as a pretext for persecuting the Radicals.

While his father and the prime minister were abroad seeking a bride for him, King Alexander announced his betrothal to Draga Mašin, a commoner and widow of questionable reputation. The government resigned (Oct. 11, 1900), and King Milan remained abroad. In the last phase of Alexander's reign government succeeded government: eight in three years. Serbia was flung into Russia's arms, and the emperor Nicholas II acted as best man at the king's wedding (Aug. 5, 1900). The king promulgated a new constitution (April 19, 1901), established a bicameral legislature, and guaranteed freedom of the press and of association; but in 1902 he reverted to reaction. He became the object of widespread aversion, and on June 11, 1903, he and Queen Draga were assassinated.

The Karageorgević Restoration.—An all-party provisional government was formed. The *Skupština* restored the constitution of 1889 and then, on June 15, 1903, elected Peter Karageorgević (*see* PETER I) to the throne. Russia and Austria-Hungary, the former being against any German candidate and the latter against any Russian one, recognized the change. Other powers followed suit, but Great Britain withheld recognition till 1906. When the provisional government resigned, Serbia became a parliamentary democracy, with a predominantly two-party system, being ruled mainly by Radicals, once by Independent Radicals (a party formed in 1901) and, in emergency, by coalition governments. Elections were fought in 1905, 1906, 1908, and 1912. The new regime inherited a budget deficit of 11,500,000 dinars, but in 1904 there was a surplus of 6,500,000. The Army was provided with modern weapons. Confidence revived at home and abroad.

The regenerated Serbia constituted an obstacle to Austria-Hungary's plans for the Balkans. A Serbo-Bulgarian customs union (Aug. 4, 1905) led to a breach in Serbia's trade negotiations with Austria-Hungary and eventually to a tariff war (from Jan. 25, 1906), which compelled Serbia to seek fresh markets, with the result that foreign trade increased by 10,000,000 dinars. On Austria-Hungary's annexation of Bosnia-Hercegovina (*q.v.*) in October 1908, Serbia was strongly inclined to declare war, incited by Russia; but pressure from Berlin and Vienna caused the government to make the humiliating statement that the annexation "does not affect Serbia's rights" (March 1909). The experience, however, strengthened Serbia's determination to form a Balkan alliance. King Peter I paid state visits to the principal European capitals, with the exception of Vienna.

The Balkan Wars (1912–13).—Serbia concluded a treaty of alliance (March 13, 1912) and a military convention (May 12) with Bulgaria, and a political and military convention with Montenegro (Sept. 27); and in October 1912 the three states and Greece declared war on Turkey. (*See* BALKAN WARS.) The Serbs won notable victories at Kumanovo and at Monastir (Bitola) and advanced across Albania to the Adriatic coast. These successes strengthened Serbia's sense of a vocation to play for the South Slavs a role corresponding to that of Piedmont in the Italian Risorgimento. To check the Serbs, Austria-Hungary mobilized, but European public opinion had little sympathy for the Turks, who were compelled to conclude an armistice (Dec. 3). This was followed by a peace conference in London, where all the belligerents found themselves under pressure from the Great Powers to reach an agreement. After six weeks the conference broke up, and war restarted on Feb. 3, 1913. Thanks to 32 Serbian battalions and 92 siege guns, Adrianople (Edirne), which Bulgaria wanted from Turkey, surrendered on March 26. Negotiations were resumed in London on May 20, and the Treaty of London (May 30) gave to the Balkan allies all Turkish territories west of the Enos-Midia (Enez-Midye) line.

While Bulgaria had failed to meet some of its obligations under the alliance, Serbia had fulfilled more than had been stipulated; but in the event Serbia was deprived of an outlet to the Adriatic by Austria-Hungary's veto and by the frontier assigned to the newly created Albania. Serbia therefore demanded that Bulgaria should agree to revise the partition of Macedonia in Serbia's favour and, in accordance with the Treaty of London, proposed that the Russian emperor Nicholas II should act as arbiter. Bulgaria, however, being also in dispute with Greece over Macedonia, on June 28, 1913, decided to attack both Serbia and Greece, without declaration of war. The Serbs defeated the Bulgars on the Bregalnica River (July 9), and Austria-Hungary was held back from intervention against Serbia by Germany and Italy. Bulgaria had to sign an armistice, followed by the Treaty of Bucharest (Aug. 10), which gave to Serbia the contested territories. With Novi Pazar, Kosovo, Štip, Skoplje, and Bitola, Serbia was thus expanded to 87,788 sq.km. Austria-Hungary, dismayed at the Bulgarian defeat, concentrated attention on Albania and, by an ultimatum (Oct. 18), compelled the Serbs to withdraw the troops that they still had there.

World War I.—Incapacitated by ill-health, King Peter I appointed his heir, Prince Alexander (*q.v.*), to be regent (June 24, 1914). The *Skupština* was dissolved. Pašić was canvassing his constituency and the commander in chief, Field Marshal Radomir Putnik (1847–1917), was at Gleichenberg, when Gavrilo Princip (1894–1918), a Bosnian Serb but an Austro-Hungarian citizen, assassinated the Austrian archduke Francis Ferdinand (*q.v.*) at Sarajevo on June 28, 1914. Though Col. Dragutin Dimitrijević (*q.v.*) was one of the inspirers of the plot, the complicity of the Serbian government was never proved: Serbia was exhausted, and the government had every reason to avoid giving offense. Austria-Hungary, however, delivered a formidable ultimatum (July 23). Serbia consented (July 25) to "such collaboration [of Austro-Hungarian officials] as agrees with the principle of international law," and rejected only the demand for the participation of Austro-Hungarian officials in the actual inquiry into the crime, while admitting that even this demand could be met by "communications in concrete cases." Finally, Serbia offered to submit the case to the International Tribunal at The Hague. Austria-Hungary declared war on July 28. (*See* WORLD WAR I.)

The dissolved *Skupština* was recalled, and it unanimously endorsed the government's action. The Austro-Hungarian invaders from Bosnia were routed on Cer Mountain on Aug. 19, 1914, but in September a Serbian offensive across the Sava had to be withdrawn, and further attacks forced the Serbs to retreat eastward

and southward (Belgrade fell on Dec. 1). The Austro-Hungarians, however, were again routed on the Kolubara River; and on Dec. 15 Serbia was cleared of invaders.

On Dec. 7, 1914, the *Skupština* had unanimously resolved that "Serbia's only aim is the liberation and union of all our Serb, Croat, and Slovene brethren." But the policy of the Western Allies, based on expediency, ran counter to Serbia's aim and even to Serbia's integrity as established in 1913. Without the knowledge of Serbia, the Entente concluded the secret Treaty of London with Italy in April 1915, made concrete proposals to Bulgaria in May, and forbade a Serbian declaration of war on Bulgaria in September, when that state was on the verge of attacking Serbia. Though the Allies undertook to provide 150,000 men to support Serbia's ally Greece at Salonika, and though Sir Edward Grey, the British foreign secretary, promised help "without reserve or qualification," there were only 35,000 French and 13,000 British troops (the latter moreover under strict injunctions not to enter Serbia) at Salonika on Oct. 6, when the rupture with Bulgaria occurred. Bulgaria attacked on Oct. 11; and by Nov. 25, Serbia, overcome by combined German, Austro-Hungarian, and Bulgarian forces, was partly occupied by Austria-Hungary and partly annexed by Bulgaria. The retreat of king, government, *Skupština*, and masses of the people with the Army across Albania stands out as one of the great exploits of any war.

Corfu became the seat of the Serbian government. Reequipped, the Army joined the Allies at Salonika. The Serbs conquered Kajmakčalan Mountain on Oct. 4, 1916, and Bitola was liberated on Nov. 19. A formidable rising of the Serbs in the Toplica and Jablanica districts (February–March 1917) was ferociously repressed by the occupying forces. The trial of Dimitrijević and others, at Salonika, on charges of plotting against the regent and the government resulted in political unrest.

The government, the opposition, and the Yugoslav committee on July 20, 1917, signed the Corfu declaration, affirming that Serbs, Croats, and Slovenes constitute one nation and demanding union in a constitutional, democratic, and parliamentary monarchy. In 1918, when the Allies launched the Salonika offensive, the Serbs routed the Bulgars at Dobro Polje (Sept. 16) and liberated Prilep; the French entered Skopje, and Bulgaria capitulated (Sept. 29). Niš, Prizren, and Mitrovica were liberated, and the victorious Serbian Army entered Belgrade on Oct. 31. The war had cost Serbia 23% of its population.

Serbia and Yugoslavia from 1918.—On Nov. 24, 1918, at Zagreb in Croatia, the National Council of Slovenes, Croats, and Serbs, representing the Slavs of Dalmatia, Croatia, Slavonia, Slovenia, and Bosnia, proclaimed the union of their territories, hitherto all subject to Austria-Hungary, with Serbia and Montenegro in one State of the Serbs, Croats, and Slovenes. It then sent a delegation to invite Serbia to adhere to this union. On Nov. 25, however, a National Assembly of the Serbs of Vojvodina, at Novi Sad, voted for the immediate union of their country with Serbia; and on Nov. 26 a Montenegrin Grand National Assembly, at Podgorica, voted for the union of Montenegro with Serbia. Serbia's area was increased to 128,278 sq.km. in the few days before Alexander, as regent for Peter I, on Dec. 1, 1918, in Belgrade, accepted the invitation from Zagreb and so brought into being the Kingdom of Serbs, Croats, and Slovenes (*see* YUGOSLAVIA).

The Treaty of Neuilly (*q.v.*), in 1919, made a slight adjustment to the Bulgarian frontier whereby Serbia obtained (1) a strip of territory between Negotin and Zaječar, (2) the districts of Caribrod and Bosiljgrad, and (3) in the extreme southeast, Strumica. The Conference of Ambassadors, moreover, in 1925, adjusted the Albanian frontier so as to give the monastery of St. Naum to Serbia. These accessions increased the area of Serbia to 130,741 sq.km.; and Serbian territory constituted 20 of the 33 provinces into which the Yugoslav state was divided administratively.

For later history *see* YUGOSLAVIA. In 1929, when the kingdom officially took the name Yugoslavia, the ancient component regions were dissolved into nine *banovine* (administrative provinces), Serbia being then split among five of them. The Yugoslav constitution of 1946 gave to the federal republic of Serbia a smaller area than the Serbia of 1918–29, since Montenegro and the formerly Serbian Macedonia became separate federal republics at the same time.

See also references under "Serbia" in the Index. (K. ST. P.)

BIBLIOGRAPHY.—*General:* F. P. Kanitz, *Das Konigreich Serbien und das Serbenvolk, von der Römerzeit bis zur gegenwart*, 3 vol. (1904–14); J. K. Jireček, *Geschichte der Serben*, 2 vol. (1911–18; 2nd Serbian ed., augmented, 1952); H. W. V. Temperley, *History of Serbia* (1917); A. Mousset, *La Serbie et son église* (1938); M. A. Purković, *Srpski vladari* (1958); L. Stavrianos, *The Balkans Since 1453* (1958).

19th Century and After: Among Serbian works, those of Slobodan Jovanović (*q.v.*; collected ed., 17 vol., 1932–40) are of major importance; *see* also J. Ristić, *Spoljašnji odnošaji Srbije novijega vremena*, 3 vol. (1887–1901). For the non-Serbian reader, G. Yakchitch, *L'Europe et la résurrection de la Serbie* (1907); R. W. Seton-Watson, *The Southern Slav Question* (1911); L. Léger, *Serbes, Croates et Bulgares* (1913); E. Denis, *La Grande Serbie* (1915); J. Buchan (ed.), *Yugoslavia* (1923); A. Mousset, *Le Royaume serbe-croate-slovene* (1926); E. Haumant, *La Formation de la Yougoslavie* (1930); R. J. Kerner (ed.), *Yugoslavia* (1949).

SERBIAN ORTHODOX CHURCH, the autocephalous national representative in Yugoslavia of the Orthodox Eastern Church (*q.v.*). The Orthodox, who comprise more than a third of the Yugoslav population, are found mostly in Serbia and Macedonia.

The conversion of the Serbs to Christianity is officially placed in the reign of Mutimir (d. *c.* 890), but only after disciples of St. Methodius began their work, with Slavonic translations of the Scriptures and liturgical books, did the Christianization of the Serbs become effective (*see* CYRIL AND METHODIUS, SAINTS). For long the southern Slavs wavered in their ecclesiastical allegiance between Rome and Constantinople, their final adherence to the latter being largely the work of St. Sava (*q.v.*) who became the first archbishop of an independent Serbian Orthodox Church in 1219. The territorial expansion and the prestige of medieval Serbia reached their zenith in the reign of Stephen (*q.v.*) Dushan (1331–55), who assumed the imperial title of tsar; in 1346 the Serbian archbishopric of Peć was correspondingly raised to the rank of patriarchate. Stephen's policy was on the whole to strengthen the church, and his celebrated code (*zakonik*), promulgated at a council of representatives of church and state in 1349, confirmed its privileges.

Stephen's empire fell soon after his death, and in 1389 the overwhelming Serbian defeat at the Battle of Kosovo brought Serbia under Turkish suzerainty. In 1459 the last relics of Serbian independence were extinguished; Serbia became a Turkish pashalik and the patriarchate was abolished. Thereafter the church, administered from Ohrid, was the sole channel for expression of Serbian national spirit and aspiration. The patriarchate of Peć was restored in 1557, and among notable occupants of the see were Jovan, who inspired the "Rebellion of St. Sava" at the end of the 16th century, and Arsenye III, who led a migration of Serbian families into Austria (1690). The patriarchate was abolished again in 1766 and the church remained under the jurisdiction of the ecumenical patriarch of Constantinople until it became autocephalous in 1879, the year after the recognition of Serbia as an independent state. After World War I all the Serbs were united under one ecclesiastical authority and the patriarchate was reestablished in 1920, the patriarch's full title being "archbishop of Peć, metropolitan of Belgrade and Karlovtsi, and patriarch of the Serbs." The Communist government after World War II separated the church from the state and prohibited the teaching of religion except in church buildings.

The supreme authority in the Serbian Church is the Holy Synod, composed of all the bishops, which meets once a year. The patriarch presides also over a standing synod of four members which administers the day-to-day affairs of the church. There are two theological seminaries. The patriarchal journal *Glasnik* is published monthly. There are 22 dioceses, including the extraterritorial one in North America.

The Serbian Church was introduced to the United States as a result of the Serbian immigration that began in the 1890s; the first church was St. Sava's in Jackson, Calif., founded in 1894. The diocese was formally created in 1920 with the appointment of a deputy administrator by the patriarchate and embraced the whole

of Canada, the United States, and Mexico. In 1926 the diocese received its first bishop, and the see was established at the St. Sava's Serbian Monastery, Libertyville, Ill.

BIBLIOGRAPHY.—A. Hudal, *Die serbisch-orthodoxe Nationalkirche* (1922); R. M. French, *Serbian Church Life* (1942); M. Spinka, *A History of Christianity in the Balkans* (1933). (RE. M. F.)

SERBO-BULGARIAN WAR (1885), one of the stages in the long-protracted solution of the Eastern Question (*q.v.*). The Congress of Berlin (*q.v.*), in 1878, had given only Niš, Pirot, Leskovac, and Vranje to Serbia, and Serbia was dissatisfied with this award. Considerable fears were entertained lest Bulgaria, which was under Russian influence, should annex areas of mixed nationality (Serbs and Bulgars). Bulgaria claimed Pirot, while the Serbs wished to seize Vidin from Bulgaria. The ruler of Serbia, Milan Obrenovich IV, looked to Austria-Hungary for support and, on June 28, 1881, signed a secret convention with that power. By this convention Serbia undertook to renounce agitation among the Serbs of Austria-Hungary and in Bosnia-Hercegovina in return for a mutual defense agreement, the real meaning of which was that Austria-Hungary would support Serbian claims on territories in the Vardar Valley and elsewhere in the south. Serbia moreover promised to conclude no alliances without the approval of Vienna, which thus obtained virtual control of Serbian foreign policy. With Austria-Hungary's consent, Milan assumed the title of king of Serbia in March 1882.

On Sept. 18, 1885, a *coup d'état* was carried out in Plovdiv (Philippopolis), capital of Eastern Rumelia—the autonomous province constituted south of the Balkan Range under Turkish suzerainty by the Treaty of Berlin (*see* BULGARIA: *History*). The governor was expelled and union declared with Bulgaria. Prince Alexander (*q.v.*) of Bulgaria entered Plovdiv on Sept. 21 and formally accepted the union. Alexander's relations with Russia, however, had not been easy since his election in 1879: neither Conservative nor Liberal ministries had favoured too close a Russian tutelage; the prince in 1883 had been obliged to dismiss two Russian generals, A. V. Kaulbars and L. N. Sobolev, from the Bulgarian service; and the Russian emperor Alexander III blamed the prince for the embarrassments of Russian policy in Bulgaria. When therefore the union of "the two Bulgarias" was declared without Russian approval, the emperor gave orders that Russian officers serving with the Bulgarian Army should be withdrawn (Sept. 23) and informed a Bulgarian delegation sent to meet him in Copenhagen (Oct. 3) that Bulgaria could expect no Russian aid under its existing government.

In Serbia there was considerable agitation for a cession of territory by Bulgaria to compensate Serbia for Bulgaria's enlargement. The Great Powers considered a solution of this kind inevitable, and Russia was not reluctant to see a solution which would detract from Prince Alexander's prestige. G. S. Kalnóky, the Austro-Hungarian foreign minister, was himself unwilling to restrain King Milan, lest a popular rising should overthrow him and install the head of the rival house of Karadjordjević (Karageorgevich), who would almost certainly place Serbia under the influence of Russia. Kalnóky merely asked Milan to delay action until the ambassadors' conference in Istanbul had arrived at a formula to settle the crisis, but Milan was under such pressure at home that he could not wait. He issued a declaration of war against Bulgaria in the night of Nov. 13–14, 1885, and Serbian troops crossed the frontier.

The initial progress of the Serbs was rapid because the Bulgarian Army was not adequately deployed to meet their attack; but, contrary to all expectation, the Bulgars, commanded by relatively junior officers, met and defeated the Serbs decisively in the Battle of Slivnitsa on Nov. 17–19, 1885. On Nov. 26 the Bulgars entered Serbia, and within two days they took Pirot.

The danger now was that Serbia might collapse. The Austro-Hungarian ambassador in Belgrade, Graf Rudolf von Khevenhüller, hastened to Pirot and insisted on an armistice, to which Prince Alexander was unwilling to agree. Austria-Hungary would have been willing to take military action in support of Serbia, but the German chancellor, Bismarck, warned Kalnóky that there could be no modification of the *status quo* in the Balkans without the consent of the powers which had signed the Treaty of Berlin. Peace was concluded at Bucharest on March 3, 1886. Bulgaria surrendered no territory to Serbia.

The result of the crisis gave Russia good reason to be satisfied with Germany. There seemed to Russia to be positive advantages in maintaining the *Dreikaiserbund*, or League of the Three Emperors (Russia, Austria, and Germany), formed in 1881 and renewed in 1884. Austria-Hungary had less reason to be pleased. In 1887, since the *Dreikaiserbund* was not renewed, Russia had to be content with the so-called Reinsurance Treaty with Germany. (R. F. LE.)

SERBO-CROATIAN LANGUAGE, a member of the southern group of the Slavic languages (*q.v.*), is the principal language of Yugoslavia. It is native to the four republics of Serbia, Montenegro, Bosnia-Hercegovina, and Croatia, and serves as an important secondary language in Slovenia and Macedonia. The number of speakers of Serbo-Croatian in Yugoslavia in mid-20th century was over 12,000,000.

The Serbs and Croats are of the same racial stock and speak a single language, but because of nationalistic antagonisms deeply rooted in history they often exaggerate the minor regional differences in an attempt to prove that there are two languages. The Catholic Croats lived for centuries under Venetian or Austro-Hungarian rule, while the Orthodox Serbs enjoyed a brief period of independence within the sphere of Byzantium and then were subject to the Turks for five centuries. Consequently the Croat writes his language in Roman letters, calling it Croatian, while the Serb uses the Cyrillic alphabet and calls the language Serbian (for a table of the Cyrillic alphabet, *see* SLAVIC LANGUAGES). The Croats speak a western type of dialect, but not all Serbs speak eastern dialects. In vocabulary it is possible to point out some words which are specifically Croatian and others which can be called Serbian, yet nearly always where lexical doublets exist, both forms are known and understood all over the country.

History.—The earliest surviving Serbo-Croatian texts date from the 12th century. In Serbia, the local variant of Church Slavonic was in use until Russian Church Slavonic books were adopted at the end of the 17th century. During the following decades the Serbs worked out a compromise style called Slavono-Serbian. In Croatia the chief written language was Latin, but nonetheless in some areas the Croats wrote their own language with Glagolitic letters, using their own style of Church Slavonic for literature, and the local *ča*-dialects (*see* below) for records and legal documents. Dalmatian poets wrote in the *ča*-dialect in the 16th century, but the *što-(i)je* dialect of Dubrovnik soon became the literary language. Attempts in the 18th century to base a literary language on the *kaj*-dialects gave way to enthusiasm for an artificial "Illyrian" language which was to unite all south Slavs. At this time a Serb, Vuk Stefanović Karadžić (1787–1864) began to write in his native *što-(i)je* Hercegovinian dialect. He worked out an excellent phonemic orthography, adapting Cyrillic letters to make an alphabet truly suited to the language. Vuk's dialect was the one spoken by the largest number of people, it was central, in a sense a compromise among peripheral dialects, and moreover it was close to the language of Dubrovnik literature. After long controversy, a group of leading Croat and Serb writers formally accepted Vuk's language in 1850, but dissension about details continued. Since the end of the century, the *e*-dialects have been used in the Serbian east, while the rest of the country has retained Vuk's *(i)je* forms.

Classification.—Serbo-Croatian is joined to its South Slavic neighbours by transitional dialects, in the east and south to Bulgarian and Macedonian, and in the northwest to Slovene. It is closer to Slovene, however, in that both have retained full nominal declensions and vocalic systems which distinguish long from short and rising from falling vowels, whereas Macedonian and Bulgarian have virtually lost the case system and have neither length nor intonation as distinctive features of the vowels.

The three major Serbo-Croatian dialects are named *što*, *ča*, and *kaj*, according to their respective interrogative pronouns for "what?"

The *što* dialect covers the largest area and is the basis for the

standard language. The *ča* dialect, formerly spoken widely in the west, is rapidly disappearing, though it can still be heard on the Dalmatian islands, in much of Istria, and in a few spots on the mainland. It is important for historical linguistics because it has preserved with great fidelity the accent on the same syllable that bore it in Common Slavic. The *kaj* dialect is spoken in most of northern Croatia, including the Zagreb region. It seems once to have been a part of Slovene, but during the last millennium its character was changed because of strong influence first from the *ča* and then from the *što* dialects.

Što-Serbo-Croatian is characterized by the development of Common Slavic *ǫ* to *u*, *tj* and *dj* to the strongly affricated palatal stops *ć đ*, syllabic *l* > *u*, and, generally, post-vocalic *l* > *o*, and both *ŭ* and *ĭ* > *a*. There is a strong tendency to avoid difficult consonantal clusters, either by inserting a vowel or by dropping one of the consonants. The aorist and imperfect tenses are regularly used in the east but restricted to the written language in the west. In the east the infinitive is usually avoided, being replaced by *da* plus present tense.

The subdivisions of the *što* dialect are generally defined by a combination of criteria, chiefly the development of Common Slavic *ě* and the degree of innovation in the accentual system. In the east *ě* > *e* (**dětę* "child" > *dête*, **děca* "children" > *dèca*), in the west *i* (*dîte*, *dìca*), and in the intermediate area *ije* if long and *je* if short (*dijète*, *djèca*). The original geographic division has been obscured by mass migrations of *(i)je*-speakers to the west, and in some areas (especially Bosnia) Catholics and Muslims use the *i*-forms, and members of all three religions the *(i)je*-forms. Isolated dialects have a mixed development (*i/je* or *ē/je*) or have maintained a six-vowel system where *ě* has not fallen together with any other vowel. In accentuation, there is a general tendency to shift the stress one syllable toward the beginning of the word. Conservative dialects have admitted shifts rarely or not at all, but most dialects have carried the shift through in many or all possible cases. The standard language has the newer accentuation, and for *ě* either *e* (Serbia and the Vojvodina) or *(i)je* (Montenegro, Bosnia-Hercegovina, Croatia). The vocalic system here includes *i e a o u* and syllabic *r*, all of which may occur long or short, with rising or falling intonation. Rising intonation never occurs in a final syllable. There can be only one rising syllable in a word, and it automatically receives the stress, which is weak. If there is no rising intonation, the stress falls on the first syllable.

BIBLIOGRAPHY.—*Dictionaries:* V. Karadžić, *Lexicon serbico-germanico-latinum*, last ed. (1934), is the classical work; F. Iveković and I. Broz, *Rječnik hrvatskoga jezika* (1901); S. Ristić and J. Kangrga, *Enzyklopädisches serbokroatisches-deutsches Wörterbuch*, 2 vol. (1928), the best source for modern terms; M. Drvodelić, *Hrvatsko-engleski rječnik* (1953). The exhaustive dictionary begun by the Yugoslav Academy (Zagreb) in 1880 had progressed as far as Tro (Roman alphabet) by 1963. *Grammars:* A. Leskien, *Grammatik der serbokroatischen Sprache* (1914), with information on dialects and history; D. Subotić and N. Forbes, *Serbian Grammar*, 2nd ed. (1926); M. Stevanović, *Savremeni srpskohrvatski jezik*, I (1964); A. Belić, *Savremeni srpskohrvatski književni jezik*, 2 pt. (1948–49). *Dialects:* P. Ivić, *Die serbokroatischen Dialekte*, I (1958). *History:* A. Belić, *Istorija srpskohrvatskog jezika*, vol. 2, 2 pt. (1950–51). Detailed bibliography regularly in the journal *Južnoslovenski filolog*. (H. G. L.)

SERENA: *see* LA SERENA.

SERENADE (French, "evening music"), a musical composition which was originally a song of courtship. The serenade to Zerlina, "Deh! vieni alla finestra," in Mozart's *Don Giovanni*, is an outstanding example in opera. Losing its association of courtship, the term serenade was applied about 1770 to instrumental works for a small ensemble in several movements, similar in kind to works of the same period entitled cassation, divertimento and notturno. The sequence of movements, which was indefinite, included marches, minuets, and sonata movements, and resembled that of the suite (*q.v.*). Mozart's *Haffner* serenade is one of the finest examples. Later, in works by many composers from Beethoven to Sir Edward Elgar, Richard Strauss, and Arnold Schoenberg, the term was applied to works in several movements for small orchestra, a solo instrument, or wind or string instruments.

SERER, an industrious Negro people of Senegal, south of the railway between Thiès and Kaffrine, numbering about 500,000 in the 1960s. Traditions indicate that their ancestors moved westward under pressure from Wolof, Malinke, and Fulani (*qq.v.*). The dialects of Serer belong to the West Atlantic group of languages (*see* AFRICAN LANGUAGES: *The Niger-Congo Family*). Some similarities in vocabulary are found between Serer and Fulani.

The compact Serer villages have round or square thatched houses of reed or millet stalks. Feud was important in their former political organization, and law and order were maintained by supernatural sanctions; *e.g.*, trial by ordeal and formal cursing. They lack the caste system of the Wolof, and are organized in exogamous matrilineages, though a wife joins her husband's community. They have a mixed economy, keeping cattle, growing millet for food and peanuts for sale; crop rotation is practised, and fodder is gathered from *Acacia albida* trees. Property, such as land or cattle, is vested in the matrilineage head. Most Serer are pagan, with few converts to Islam and Christianity. Spirits of the dead are appeased or invoked; graves are covered with thatched roofs, a man's being marked with a bow and arrow, a woman's with mortar and pestle; special ceremonies are held for rainmaking.

See also SENEGAL, REPUBLIC OF.

See P. Pelissier, "Les Paysans sérères," *Les Cahiers d'Outre-Mer*, no. 22 (1953); D. P. Gamble, *The Wolof of Senegambia* (1957). (D. P. GA.)

SERFDOM AND VILLEINAGE, terms used to describe the status of peasants who possessed their own means of subsistence but lacked personal freedom. One of the problems facing the historian of serfdom is that of definition. For some authorities, serfdom is an institution of public law requiring for its existence certain legal definitions separating the serf from the free man. For others, serfdom is a social condition that may or may not be defined in law. Since social realities often exist for long periods of time without receiving legal recognition, it follows that legally defined serfdom would not necessarily coincide chronologically or geographically with the serfdom that is defined in sociological rather than legal terms. One would expect the social condition to appear before its legal definition and to decline in practice before it disappeared from the law. This happened in England, where in law there does not seem to have been an intermediate position between freedom and slavery until the 12th century, although servile social conditions existed centuries before that date. Some English lawyers were still claiming that villeinage existed in law in the 17th century although as a social reality it had by then disappeared. Some historians (*e.g.*, Alexander Eck, George Vernadsky) hold that serfdom did not appear in Russia until the end of the 16th century. Others (*e.g.*, B. D. Grekov) consider that the depression of the free peasantry in Kiev Russia into a servile condition began in the 9th century. In Muslim law there was no intermediate status between slave and free, but in the Seljuk dominions in the Middle East many peasants were economically and socially servile.

Economic and Social Aspects.—It is advisable, therefore, to describe the economic and social aspects of serfdom, without neglecting the legal definitions that were later found necessary to consolidate the privileges of those who benefited from the servility of others. The vast majority of serfs were peasants. This means that they obtained their subsistence from a plot of land. This, rather than their place in public law, was the essential feature differentiating serfs from slaves. The slave was an instrument of production owned by a master who provided him with food and clothing. The serf provided his own food and clothing from his own productive effort. Under fortunate circumstances he could produce more than his own subsistence requirements and accumulate reserves of his own. He was, however, a dependent peasant; this meant that a substantial proportion of the surplus product of his holding was taken by his overlord. Alternatively, or additionally, the lord used the serf's labour for the cultivation of the portion of his land that was not held by tenants (called demesne land). The payment of rent in money, in kind, or in labour was not, however, the only or even the essential sign of the serf's dependence, for rents are paid by free men as well as by serfs. The essential additional mark of serfdom was lack of freedom of movement or

the peasant family, and restrictions on the free disposal of its property. When the peasant could leave his holding, and if necessary his village, without a lord's permission, he was no longer a serf.

Occurrence.—Serfdom was a very ancient and widespread institution. But there has been no simple line of development in history from freedom to serfdom or vice versa. In ancient China there were periods of serfdom and periods of relative peasant freedom, the latter usually coinciding with periods when the central government was strong and the local power of the landlords was weak. In ancient Egypt, too, when the hereditary power of the landowners was great (from the 5th to the 12th dynasties) the peasants were bound to their holdings. Before and after that period, when the ruling dynasties established strong central control, freedom of movement and free alienation of land by the peasantry were permitted. Serfdom in the ancient civilizations of the Mediterranean is less well documented than slavery. It seems clear, however, that in Crete and Sparta during the era of Dorian domination of the indigenous populations, the serfs were bound to the hereditary domains of the aristocracy; they were deprived of freedom of movement and were obliged, in the case of Sparta, to pay half the product of their holdings to their overlords. In Crete in the 5th century serfdom rather than slavery was paramount in the economy. There, serfs had rights, guaranteed in law, that distinguished them from chattel slaves. But not until the late Roman Empire and the Middle Ages in Europe is there sufficient evidence for historians adequately to relate the expansion and decline of serfdom to changing economic and political circumstances.

SERFDOM IN EUROPE

The serfdom of medieval and modern Europe was a complex social institution that changed in form according to a variety of specific circumstances. The ratio of land to labour within any society was a basic factor that could operate in different ways according to political conditions. If abundant land was available for colonization, free conditions of tenure might be offered by some landlords to attract peasant settlers. This could reduce the general level of rents and services. On the other hand, the landlords, in order to combat this tendency, might bind the peasants by legal restrictions to prevent them from leaving their holdings. If there was pressure of population and a great demand by the peasants for landed holdings, then landlords would be in a position to exact onerous returns from tenants. But rents and services might be so heavy that enserfment would be the only guarantee that they would continue to be rendered. Another factor making for enserfment, especially in times of weak central governments, was the development of ties of personal dependence between the rich and strong and the poor and weak. This was the case in dynastic Egypt, in the late Roman Empire, in the barbarian states from the 6th to the 12th centuries (*see* FEUDALISM), and in the Byzantine Empire. The social framework within which these fluctuations of serfdom operated was, of course, one where the principal classes were landowners with established property rights and political power, and peasants occupying dependent holdings. The extent to which the state power acted on behalf of the landowners varied not only with its own strength but also according to the degree to which it represented landowning as against urban, commercial, and industrial interests. Only when the varying interaction of all these factors is examined can the strength of serfdom in any particular period of history be correctly understood.

Origins of Serfdom.—One of the sources of the servile population of medieval Europe was the body of slaves of the late Roman Empire and of the barbarian states that succeeded the Empire. As early as the 2nd century A.D. the slackening of production for the market caused changes in estate management. Large estates that had been worked by slave gangs housed and fed by the estate owner were divided into peasant holdings, some of which were granted to the slaves. It is to these slaves of the Roman period and later who were given their own holdings that we may trace the probable origin of the servile holdings of the Carolingian estates of the 9th century. When they were provided with peasant holdings the slaves were not necessarily freed. There were, however, numerous instances of the freeing of slaves both in the empire and in the barbarian successor states of the Goths, Franks, and Lombards. In some cases these enfranchisements put the former slave on a level with the free members of society, but for the most part the freedman continued to be under the protection of his former master and owed him loyalty and services.

The bulk of the peasants of the late empire and Dark Ages were neither slaves nor freedmen but *coloni*. These *coloni* might have taken up holdings granted by a proprietor, or they might have surrendered their own lands to him in return for protection, receiving back their land as a dependent tenement. This surrender and reissue of land frequently accompanied the act by which a person commended himself to a great man and swore fealty to him, though the act of commendation did not necessarily involve any land transaction. It was made in order to obtain protection (especially in the empire) from the state tax-gatherer and (especially in the disorders of the Dark Ages and early Middle Ages) from invaders and oppressive neighbours. Whatever the origin of the *coloni*, a problem concerning them, which was as old as the Roman Empire, was to prevent them from leaving their land. The solution was to bind them to their holdings. This was done, probably originally for fiscal reasons, by the emperor Constantine in A.D. 332. But it was so useful to the landlords that by the 6th century it was still firmly established, and *coloni*, though legally free, were treated as an inferior element in society.

A substantial proportion of the peasant population in the former territories of the western Roman Empire, in that part of Germany which had formed part of the Carolingian dominions, and in the Byzantine Empire was therefore deprived of freedom of movement, at any rate, in law. This was the case whether they were considered to be attached to the land or whether they were considered to be attached to the person of their lord, as were most dependents of the early Middle Ages. There is evidence of them in the law codes and in the Carolingian estate surveys of the 9th and 10th centuries, but it is uncertain what proportion of the total peasant population they constituted. What this proportion was depends on how much of the total cultivated area was included in the huge estates of the secular and ecclesiastical landlords. This amount cannot be known, though it is certain that in many parts of both the western and Byzantine Empires there was still a substantial survival of small free peasant property. However, on the big estates restrictions on movement were doubtless essential because the rents and services due from the peasant holdings were heavy. Among other things they included labour services on the landlords' demesne land (*see* MANOR). In theory labour services were no more symptoms of servile tenure than other forms of rent or service. But in practice the exaction of labour services required the exercise by the landlord's agents of labour discipline. This was recognized as one of the clearest symptoms of a man's personal subjection. It is not surprising that later on in England the obligation to perform labour services became one of the tests of servile status.

In the 9th and 10th centuries, the serfs in the strict sense of the word, those who in fact were called *servi*, were mainly those descended from the former slave population, though many on the big estates who were not called *servi* (*e.g.*, the *coloni*) had conditions of life that were virtually servile. In the next three centuries economic and political conditions in Europe changed considerably. Population increased, commerce and industry developed, towns grew up, and, both within the old territories of the Carolingian Empire and in sparsely occupied Slavonic lands east of the Elbe, forests and marshes were opened up for cultivation. Politically there was, first of all, a disintegration of the old forms of authority, both central and local, and from the 12th century a reconstitution of feudal monarchies, duchies, and counties. Statistical evidence about serfdom during this period is scanty, but by the 13th century it is clear that there was an immense increase in the number of dependent peasants, whether in the territories of present day France, England, or western Germany. It has been shown that this increase in the numbers of peasants whose social and economic condition was virtually equivalent to servility did not occur, as was once believed, as a result of the spread of the

legal status of the Carolingian *servi* to the *coloni,* freedmen, and other categories of dependent peasants on the big estates.

The *servi* were considered to be personally subjected to one master "to the very bones." Their condition was hereditary, usually through the mother. They had no legal rights in the public courts. They could not normally enter the clerical order. The only way in which their servile status could be ended was by an act of enfranchisement or manumission by their master. Since these serfs constituted a hereditary caste, the only way in which their numbers grew over the centuries was by natural increase. This was offset by manumissions and by escape. By the middle of the 13th century, the serfs in the stricter juridical sense of the word were relatively no more numerous than four centuries earlier. The increase was in the numbers of dependent peasants whose status was socially, if not juridically, servile. This was demonstrated by the fact that they shared with the serfs economic and social disadvantages which brought the two groups so close together that contemporaries as well as historians confused them. The chief disadvantages, which more and more peasants were sharing, were not even specifically of servile origin (in the juridical sense) but resulted from the restrictions of peasant freedom necessary to guarantee the rendering of rents and of labour services on the big estates. The increasing subjection of the English villein in the 12th and 13th centuries is to be understood largely in the context of this type of estate management. But in other parts of western Europe the connection between demesne cultivation and the labour services owed from the peasant holding was disappearing after the 9th century. And yet, as in England, peasant dependency was on the increase. This was because lords at all levels in the feudal hierarchy were strengthening their rights (or seizing rights) of private jurisdiction through the right of command, *bannum*, exercised not merely over their tenants but over all inhabitants within their area of jurisdiction. Even the juridically free tenants had no redress against the lord who had appropriated for himself a jurisdiction that had once been public. In the case of both types of lordship, whether personal or real, or in many instances a mixture of both, there were three principal obligations shared by serfs and by other dependent peasants.

Obligations of Serfs and Dependent Peasants.—The first obligation was the payment to the lord of a head payment or *capitagium* (French: *chevage*). This was common to serfs, to enfranchised men, to those who had commended themselves to a religious overlord, and to other dependent peasants. It was not a heavy payment but it was the sign that the payer accepted his dependence on the lord. Next was the obligation (*forismaritagium*) to obtain the lord's permission for marriage outside the social group or outside the lord's domain. This was particularly important in the case of juridical serfs, for if a male serf married a free woman, the offspring followed the mother's condition. But in the case of both juridical serfs and dependents, marriage outside the domain was (apart from flight) the most serious threat to the lord's future supply of tenants and labourers. The third principal obligation was that of surrendering part of the peasant inheritance to the lord at the death of the servile or subject tenant. This *mainmorte* of the serf was the whole or a substantial part of the inheritance, where there were no direct heirs to succeed. If there were direct heirs the lord would allow the inheritance to them, so that they should continue production and pay rent. Subject peasants who were not juridical serfs had a modified obligation, that of paying the best beast or chattel, but this was paid whether the direct heirs succeeded or not. These were the three most common symbols of dependence; other symbols included the payment of an arbitrary tallage, a seigneurial tax, which, while not specifically a mark of juridical servility, was a burden shared by juridical serfs and subject peasants alike.

This necessarily simplified description of serfdom and peasant dependence between the 9th and 13th centuries must be modified in a number of respects. From region to region there were varying degrees of dependence and of servile obligations. The German *hörigen,* descendants of demesne slaves, and other grades of dependents merged into a general class of *leibeigene* (bondmen) by the 13th century. The evolution is not fundamentally different in western Germany from that of the other parts of the former Carolingian Empire, except for the *ministeriales* or *Dienstmänner.* These were serfs who were originally employed by lords in military and administrative duties and were closely associated with the lord and his household. Their military role became emphasized, they were endowed with fiefs, and by the 13th century they were on their way into the nobility, in spite of such relics of servile status as restrictions on freedom of marriage and alienation of property. In England serfdom was much more related to the holding of land than to personal dependence on a lord. This was partly due to the fact that private jurisdiction and the private right of the *bannum* were there undeveloped, owing to the strength of the central government. It was also due to the maintenance of the close tie between demesne and peasant tenements on the big estates. Hence, 13th-century English lawyers, acting in the interests of estate owners, sometimes attempted to lay down that the performance of labour service was a criterion of unfree condition. This equation of unfreedom with the performance of labour services is also found in Bavaria and is obviously a local reaction to specific economic needs. It was to become important again in eastern Europe.

Factors Affecting the Extent of Serfdom.—Another important modification to the picture of serfdom arises from the fact that all over Europe factors tending to decrease serfdom existed contemporaneously with those that strengthened it. Most important of these was the attempt by landlords and political authorities to encourage the colonization of waste and forest land. Hence significant groups of free tenants who paid money rent established themselves in newly colonized areas. In England and France free peasant tenures were created in considerable numbers in the 12th century, especially in areas where virgin forest survived. Abbot Suger of St. Denis in the early part of the century took the initiative in inviting peasant *hospites* (*hôtes*) to settle on abandoned or uncultivated sites, on free conditions of tenure. At the end of the century the bishop of Worcester, also in a heavily wooded region, was granting holdings that had formerly been woodland to cultivators for a money rent with hereditary right. In Spain, particularly in Castile, there were other reasons for an expansion of free peasant tenures during this same period. It was again a call for colonizers, this time of land won in the reconquest from the retreating Moors by the Christian kingdoms of the north. But the most spectacular shift of the balance between peasant serfdom and freedom occurred in the newly colonized lands of eastern Germany and in the neighbouring Slav and Magyar states. The vast move of Flemings and Germans east of the Elbe is well known. The new villages they settled, often in former Slav territory, under the so-called "German Law" were essentially settlements of free men, for freedom was the bait. The Slavs themselves were promoting the same sort of free peasant settlements, numbering many hundreds before the middle of the 14th century in Greater and Lesser Poland, not to mention over a thousand in Silesia where the German element was also present.

Partly in order to offset the lure of freedom in new areas of colonization, partly in order to raise cash, landlords in the old settled districts of western Europe tipped the balance against peasant subjection still further. In France many charters were issued, particularly in the mid-13th century, to whole communities. These exempted their inhabitants from burdens such as *mainmorte*, *formariage*, and *chevage*, sometimes, but by no means always, including juridical serfs in the act of emancipation. In addition the French crown, for purely fiscal reasons, adopted a policy of individual manumission of its serfs particularly on crown lands east of Paris. This was systematized in the 1290s and *c.* 1310–20. The reasons given for emancipation are of some interest. They express one of a number of common medieval doctrines about serfdom, widespread in time and place . . . "As according to the law of nature everyone should be born free, but by certain usages and customs of great age preserved in our kingdom . . . and also perhaps because of the misdeeds of their predecessors, many persons of our common people have fallen into the bonds of servitude and into various conditions, which much displeases us, considering that our kingdom is called the Kingdom of the Franks . . . we have ordered . . . that these servitudes shall be brought to freedom and

to those who by birth or long standing or recently through marriage or residence have fallen into servile condition, or could so fall, freedom shall be given on good and convenient terms . . ." (Preamble to Letters of Commission of July 3, 1315 and Jan. 23, 1318, quoted in M. Bloch, *Rois et Serfs*, 1920). Evidently the attitude of those social classes of medieval society who profited from serfdom must have been at all times ambivalent. In Italy there was a similar movement for enfranchisement at the same period. The initiative came from the urban communes which were among the most powerful rulers of the peninsula. Their motives were more mixed than those of the French crown and included the wish to weaken the power of the landowning nobility; to increase the number of free taxpayers to the commune; and to increase the urban labour force. These real motives were also covered by moralizing. The enfranchisement of serfs by Bologna in 1256–57 was attributed to the realization that slavery was the consequence of the fall of man, that man's natural condition is freedom, and that Bologna was the home of freedom.

By the 14th century economic conditions in western Europe were favourable to the replacement of the servile by the free peasant tenant. The growth of the power of central and regional governments permitted the enforcement of peasant-landlord contracts (as in *métayage* or stock and land lease contracts) without the need for peasant servility. The final abandonment of labour services on demesnes removed the need for the direct exercise of labour discipline. The drastic population decline after 1350, as a result of the Black Death, altered the land-to-labour ratio in favour of tenants. And, not least, peasant risings which were endemic in England, France, Italy, and Spain in the 14th and 15th centuries, culminating in the German Peasant War of the early 16th century, also forced more favourable terms of peasant tenure. Peasant demands in some of these risings illustrate vividly both the character of serfdom at the time and the new conditions envisaged by the rebels. Local rebellions in England, frequent from the middle of the 13th century, were directed against servile status mainly because those of such status had no legal redress against increases in rents and services. By the time of the Peasants' Revolt of 1381 the rebels still aimed at free status and low rents, and among other things, free access to the commons and a free market in land. In 12th and 13th-century Italy peasant rebellions were partly, at any rate, the cause of grants of charters of liberties which included the fixing of rents and services and safeguards against imprisonment without legal process. The peasant risings in Catalonia in the late 15th century were against those personal obligations which had replaced land rent as the main burden on the peasant. Between 1493 and 1525 the west German peasant movement had, of course, many political and religious objectives. But here again demands for the abolition of serfdom were linked with requests for the reduction of rents and services and the restoration of access to the commons. Such late references to western serfdom can, however, be misleading. Whatever legal rights landlords might have over their tenants, or over peasants within their sphere of jurisdiction, for practical purposes most peasants had freedom of movement and terms of tenure that were contractual rather than customary. They were not necessarily better off economically than their servile forebears, but their status in society conformed to their economic position rather than to the legal classification of their holdings or their persons.

Modern Serfdom in Eastern Europe

This favourable evolution was not shared by the peasants of eastern Europe. In the later Middle Ages peasant status from eastern Germany to Muscovy deteriorated sharply; a second serfdom began which was prolonged well into modern times, in Russia until 1861. Peasant conditions in eastern European countries before the 14th century do not seem to have been worse than those of the west. In some ways they were better, for in eastern Germany, Poland, Bohemia, Moravia, and Hungary the colonization of forest land had led to the establishment of many free peasant communities. But a combination of political and economic circumstances reversed these developments.

Eastern Europe from the Baltic to the Balkans in the 14th and 15th centuries suffered wars even more devastating than those of western Europe. These not only diminished the peasant population, and increased the area of uncultivated land, as in the west, but also increased the power of the nobility at the expense of central governments. In eastern Germany, Prussia, and Poland this coincided with an increased demand for grain from western Europe. To profit from this demand landlords took back peasant holdings, expanded their own cultivation, and made heavy demands for peasant labour services. These were increased to as much as four days a week from the peasant holding in 15th-century Poland. This increase in services eventually led to pressure by landlord interests for means of legally binding the peasant to his holding. This was beginning in Prussia as early as 1412 and was achieved in Poland by legislation of the Diet early in the 16th century. In Bohemia, Moravia, and Hungary, the same background of devastation by dynastic, religious, and national wars created similar results. Despite the temporary upsurge of the peasantry of Bohemia during the Hussite Wars, the power of the nobility was restored by the end of the 15th century and legislation in 1497 and 1500 tied the peasants to their holdings. In Hungary, which was at war with the Turks, the same enserfment of the peasantry was achieved early in the 16th century. Not until the end of the 18th century were the peasants of the Austro-Hungarian Empire freed from a serfdom that was defined by the Emperor Joseph II himself as the absence of freedom of movement, of freedom of marriage, and of the right to learn a profession according to one's choice.

The intensification of serfdom in Russia was comparable with that of the other eastern European countries. To say that serfdom in Russia did not begin until the 16th century is to take too legalistic a view. As early at least as the 12th century there were in addition to slaves more and more peasants who, through indebtedness, were becoming dependents of their landlords. Even some of the free peasants were being drawn into dependent association with the landowner's household. But the opportunities that the vast uncultivated wastes and forests presented to peasants who were discontented with their conditions undoubtedly kept a considerable element of freedom in Russian peasant status. The expansion of the state of Muscovy in the 15th and 16th centuries led, however, to an increase in state taxation at the same time as landlords were demanding heavy rents and labour services under the stimulus of an increasing internal grain trade. This caused peasant discontent which expressed itself in flight. The government, therefore, in the interests of the new class of military fief-holders on whom it depended for its armed force, at various times between 1450 and 1550, first of all prevented peasants from leaving their holdings other than in the period near St. George's day (Nov. 26, in the East). At the end of the 16th century the privilege of moving on St. George's day was abolished, at first temporarily, though fully servile conditions were embodied in the code of 1649. Peasant conditions continued to deteriorate in the later 17th and 18th centuries, the requirements of the landlords for heavy rents and labour services and for proprietary rights in the serf analogous to slavery being fully supported by the state. (For the emancipation of the serfs under Alexander II *see* Russian History.)

See also Land Tenure: Economic and Agrarian Aspects; Manor; and references under "Serfdom and Villeinage" in the Index.

Bibliography.—"Le Servage," *Recueils de la société Jean Bodin*, vol. ii (1937); *The Cambridge Economic History of Europe*, vol. i (1941), ed. by J. H. Clapham and E. Power; M. Bloch, *Mélanges historiques*, vol. i (1963); G. von Below, *Geschichte der deutschen Landwirtschaft des Mittelalters* (1937); L. Verriest, *Institutions médiévales*, vol. i (1946); P. Vinogradoff, *Villainage in England* (1892); F. Pollock and F. W. Maitland, *History of English Law Before the Time of Edward I*, 2 vol. (1898); G. C. Homans, *English Villagers of the Thirteenth Century* (1942); R. H. Hilton, "Peasant Movements in England Before 1381," *Economic History Review*, 2nd series, vol. ii, no. 2 (1949); P. S. Leicht, *Operai artigiani agricolturi in Italia* (1946); Jerome Blum, *Lord and Peasant in Russia: from the Ninth to the Nineteenth Century* (1961); M. Malowist, "The Economic and Social Development of the Baltic Countries From the Fifteenth to the Seventeenth Centuries," *Economic History Review*, 2nd series, vol. xii, no. 2 (1959); B. D. Grekov, *Kiev Rus*, Eng. trans. by Y. Sdobnikov (1959); E. M. Link, *The Emancipation of the Austrian Peasant, 1740–1798* (1949); A. K. S. Lambton, *Landlord and Peasant in Persia* (1953); R. F. Wil-

letts, *Aristocratic Society in Ancient Crete* (1955); G. Duby, *L'Economie rurale et la vie des campagnes dans l'occident médiéval* (1962). (R. H. Hi.)

SERGE, a fabric, much used for military uniforms, made in an even-sided twill weave (*see* Weaving) and usually clear finished—that is, the fibre ends on the surface of the cloth are sheared or singed so that the twill weave is very prominent, producing a flat diagonal rib pattern that goes from the lower left to the upper right selvage on the right side. Serge was originally made of wool, but is now sometimes made of other fibres, or of wool blended with other fibres, in weights ranging from 5 to 31 oz. per sq. yd.

SERGEANT, JOHN (1622–1710), English Roman Catholic controversialist, especially well known for his criticism of Locke, was born in 1622 at Barrow-upon-Humber, Eng. He was educated privately and at St. John's College, Cambridge, and then became secretary to Thomas Morton, bishop of Durham. The result of his researches into early church history was his conversion to Roman Catholicism. After a period of theological study at the English College at Lisbon, he was ordained and was sent on the English mission in 1652 to defend the Catholic cause. In 1675 he was in France, where he lived for a time with Bossuet. Most of his life was spent in animated theological and philosophical controversy, for which he had an outstanding, if unrestrained, talent. Hardly any Anglican writer of note escaped his argumentative pen, Bishop Stillingfleet, Jeremy Taylor, and Bishop Tillotson being among his more celebrated adversaries. Sergeant was also one of the many critics of Locke, whom he attacked in his *Solid Philosophy Asserted Against the Fancies of the Ideists* (1697). In all he published 34 works, most of which were pamphlets. He was said to have died "with a pen in his hand."

Sergeant holds that we can extend and explain our knowledge of the world by the application of metaphysical principles and of the general principles of reason or "maxims," and that where empirical investigations themselves yield no new knowledge, we must resort to these principles. He therefore criticized Locke, who denied the importance of these principles in extending our knowledge, though he did not rule them out entirely. *See* Locke, John.

See for Sergeant's criticism of Locke, J. W. Yolton, *John Locke and the Way of Ideas* (1956).

SERGEANT AT ARMS. In early usage in England a sergeant at arms was generally an armed officer of a lord, often one of a special body required to be in immediate attendance on the king's person, to arrest traitors and other offenders. Thus the title now denotes certain court, parliamentary, and city officials with ceremonial (and ostensibly disciplinary) functions. It is perhaps most readily associated in modern times with an officer in each house of the British Parliament (where the word "sergeant" is spelled with a "j"), paralleled by two officers in the United States Congress. In the British House of Commons, the serjeant at arms is appointed by the sovereign, but thereupon becomes the servant of the House. His duties include attendance on the speaker, with the mace, and maintenance of order in the House and its precincts. In the House of Lords, though similarly appointed by the crown, he remains the officer of the lord chancellor (whom he attends with the mace) rather than of the House. The sergeants at arms in the United States Congress, on the other hand, are elected by the members of Congress. The sergeant at arms of the House of Representatives has similar duties with the mace and also exercises certain police functions about the legislative buildings. (E. C. Tn.)

SERGIPE (originally Sergipe d'el Rey), a state in northeastern Brazil bounded east by the Atlantic, south and west by the state of Bahia, and north by Alagôas, from which it is separated by the São Francisco River. It is the smallest of Brazilian states, with an area of 8,492 sq.mi. Its population (1960) 760,273, is about three-fourths of mixed racial origin and Negroes. The northern part of the state slopes downward to the São Francisco River, while the southern part drains into the Atlantic. Sergipe is traversed by a number of small rivers which are frequently dry in their upper reaches. The largest of these are the Vasa Barris, the Real, and the Cotinguiba, which are not navigable to ocean shipping. From east to west the state is divided into two major geographic regions, with a narrow, heavily forested coastal strip giving way to a higher zone of rough open country. The interior is devoted primarily to livestock raising. The lower, fertile lands are cultivated and produce sugar, coconuts, cotton, maize, tobacco, rice, beans, and cassava. Sergipe is Brazil's leading producer of coconuts, although sugar is its principal money crop. The only manufacturing industries of importance are cotton mills, sugar factories, and distilleries. One of the largest sugar *usinas* in Brazil is located at Riachuelo near Laranjeiras. However, a considerable expansion of vegetable and mineral processing industries is anticipated, since Sergipe has a surplus of electric power available from the Paulo Afonso hydroelectric project.

The capital of the state is Aracajú (*q.v.*), on the estuary of the Cotinguiba River, near the coast. It is connected by rail with the town of Propriá on the São Francisco River and with Bahia to the south. Highways link Aracajú with interior towns in Sergipe and with Bahia and southern Brazil. Domestic airlines provide regular service between Aracajú and all major Brazilian cities. Other important towns in Sergipe are Estância, Laranjeiras, Capela, São Cristóvão, formerly the capital, and Lagarto. Sergipe was settled in the 16th century by cattlemen and sugar planters from Bahia and it remained an administrative district of Bahia until 1821, when it became an independent captaincy. During the period of the Dutch conquest (1624–54) Sergipe was a focal point of the Brazilian resistance. It became a province of the empire in 1824 and a state in the republic in 1889. (R. E. P.)

SERGIUS, SAINT, of Radonezh (in Russian Sergii Radonezhski; baptismal name Varfolomei Kirillovich) (*c.* 1315–1392), Russian spiritual leader to whom the monastery of Troitse-Sergiyevo owed its lasting fame as a centre of the religious and social life of the nation. He was born in the city of Rostov, probably in 1315, of noble parents who soon after became impoverished and moved to the village of Radonezh, north of Moscow. At this time Tatar political oppression and social, economic, and moral decline within Russian society made "worldly" life tedious and distressing for high-minded men; therefore, after his parents' death the young Varfolomei Kirillovich withdrew into a wild forest and for several years lived as a hermit in a log cabin. Taking monastic vows in 1337, with the name Sergius, he became the abbot of the new monastery at Troitse (named Sergiyevo after him, but renamed Zagorsk in 1930).

Even when Sergius had won fame for his sanctity, he remained the poorest, humblest, and most hard-working member of the community and firmly refused the metropolitan Alexis' invitation to be his successor. In the second half of the 14th century great numbers of Russians, princes and poor peasants alike, began to make pilgrimages to the monastery. In 1380 the grand prince Dimitri (Donskoi), when he was preparing his fateful struggle against the Tatars, turned for advice and blessing to Sergius, whose message helped to inspire the Russian warriors before their victory at Kulikovo. Sergius died in 1392, revered throughout Russia. His feast day is Sept. 25.

See N. Zernov, *St. Sergius, Builder of Russia* (1939). (S. G. Pu.)

SERGIUS, the name of four popes.

Saint Sergius I, pope from 687 to 701, came of an Antiochene family which had settled at Palermo. He was elected successor to Conon after a fierce struggle between two other candidates, Paschal and Theodore, and was consecrated on Dec. 15, 687. In 689 he baptized King Ceadwalla of Wessex in Rome. For rejecting certain canons of the Trullan council of 692, Justinian II commanded his arrest and transportation to Constantinople, but the militia of Ravenna and the Roman citizenry forced the imperial officer to abandon the attempt. Sergius died *c.* Sept. 8, 701. His feast day is Sept. 8.

Sergius II, pope from 844 to 847, a Roman of noble birth, elected by the clergy and people to succeed Gregory IV in January 844, was forthwith consecrated without waiting for the sanction of the emperor Lothair I. Lothair accordingly sent his son Louis (later Louis II) with an army to punish the breach of the Roman constitution of 824, which had affirmed the emperor's sovereignty over the pope. A pacific arrangement was ultimately made, and

Louis was crowned king of the Lombards by Sergius. In this pontificate Rome was ravaged and the churches of St. Peter and St. Paul robbed by Saracens (August 846). He died Jan. 27, 847.

SERGIUS III, elected pope by one of the factions in Rome in 898, simultaneously with John IX, was expelled from the city by his adversaries. He reappeared early in 904, seized the two claimants, Leo V and Christopher, who were disputing the succession of Benedict IV, and had them strangled, being himself consecrated on Jan. 29. His adherents rallied round the master of the wardrobe Theophylact and his wife Theodora. Sergius is reputed to have been the lover of Theodora's daughter Marozia and the father of her son, the future Pope John XI. Sergius was very hostile to the memory of Pope Formosus and refused to recognize any of the ordinations celebrated by him, thus causing grave disorders. He also affected to consider as antipopes not only John IX but also Benedict IV and Leo V as well as Christopher. He restored the Lateran basilica, which had fallen down in 897. He died in the second quarter of the year 911.

SERGIUS IV, pope from 1009 to 1012, had been known as "Pigmouth" before he was consecrated successor to John XVIII in July 1009. He was powerless in the hands of John II Crescentius (*see* CRESCENTII) and the Roman nobles. He died in May 1012.

See also PAPACY.

SERGIUS (IVAN NIKOLAEVICH STRAGORODSKI) (1867–1944), patriarch of Moscow and all Russia, who achieved the recognition of the Russian Orthodox Church by the Soviet government after the Revolution. Born on Jan. 11, 1867 (old style), in Arzamas near Nizhni Novgorod (now Gorki), he came of a family with a tradition of entering the priesthood and attended the St. Petersburg academy. After a holiday in the famous monastery on Valaam Island, in Lake Ladoga, where he felt an inner call to the monastic life, he took his vows on Jan. 30, 1890 (O.S.), and chose Sergius as his name in religion. After a brilliant final examination he alternated between missionary and scholarly activity. He was a member of the Orthodox Japanese Mission (1890–93 and 1897–99), assistant inspector at the Moscow Theological Academy (1893), priest to the embassy church in Athens (1893), lecturer in theology (1896), rector of the priestly seminary in St. Petersburg and assistant inspector of the theological academy there (1899), and then rector of the same academy (1901). In the same year he was consecrated titular bishop of Yamburg. These years were filled with intensive theological work for lectures and publications, his special subject being dogmatic differences with the Western confessions. In 1905 Sergius became archbishop of Finland and Viborg and until 1907 he devoted all his strength to the erection of a bulwark there against Scandinavian Protestantism. Sergius played a large part in the calling of the regional synod of 1917 and the restoration of the patriarchate, himself becoming metropolitan of Nizhni Novgorod.

After the death of the patriarch Tikhon (*q.v.*) in 1925 Sergius soon became locum tenens of the patriarchal see. After some months' imprisonment in 1927 he sent out two epistles which called upon the faithful to refrain from anti-Soviet activity and to recognize the Soviet government as permitted by God. At the same time he achieved by negotiation with the government, in return for a declaration of loyalty, the recognition of the Russian Orthodox Church, its faith, its canonical organization and leadership, as compatible with the Soviet order, and also won the registration of the whole church as a religious association, which had hitherto been refused. Although difficult years still lay ahead of the Russian Church, Sergius thereby gave it a legal basis for its defense against the revivalist "Living Church" Movement, which had severely endangered its faith and unity, being at first supported by the Soviet government. During World War II, at the time of the German invasion, Sergius, who had again been arrested during Stalin's 1937 purge, at once called upon the faithful to defend their country and resist the enemy; he proved the church's patriotism, which had up to then been continually called in question by the Soviet government, by monetary and other collections, and by fitting out the famous tank division, "Dimitri Donskoi." On Sept. 4, 1943, he reaped the first fruits of long and patient work. Stalin received representatives of the Russian Orthodox Church and gave them permission to choose a patriarch once more. On Sept. 8 Sergius was elected to the position by a simplified procedure, but died on May 15, 1944, in Moscow. During the last months of his activity the church benefited by the ending of internal divisions and by contact with other Orthodox bodies and with the Anglican Church.

See J. Shelton-Curtiss, *The Russian Church and the Soviet State* (1953). (F. v. L.)

SERGIUS AND BACCHUS, SAINTS, two 4th-century martyrs, according to legend, officers of the Roman Army on the Syrian frontier. On their refusal to sacrifice to Jupiter, they were sent to Rosafa in Mesopotamia, where they were scourged so severely that Bacchus died. Sergius later was beheaded. The church over Sergius' grave was restored in 431, and shortly afterward Rosafa became the seat of a bishopric; it was renamed Sergiopolis. Sergius and Bacchus became protectors of the Byzantine Army. Their feast day is Oct. 7.

See H. Thurston and D. Attwater (eds.), *Butler's Lives of the Saints*, vol. iv, p. 50 (1956).

SERI, a tribe of Mexican Indians on Tiburón Island in the Gulf of California; numbering about 300–400 in the 1960s, some of them lived on the mainland in Sonora State (*see* SONORA), working for the most part as farm labourers. Many of their ancient ways of life continued on Tiburón. Traditionally nonagricultural, they lived by gathering, hunting, or fishing for such edibles as shellfish, turtles, pelicans, and cactus fruit. Their crude, weakly constructed dwellings of twigs and cacti, simple cane spears, hunting bows, and pebble tools earned them a reputation as being among the most primitive aboriginal groups of North America. However, they did produce serviceable pottery and fashioned boat-shaped rush rafts on which they successfully navigated to the mainland. Without chiefs, and lacking a real political structure, the matrilocal Seri practised polygamy and tree burial (replaced by inhumation in more recent years). Linguistically, they are classified among the Hokan speakers (*see* HOKAN).

BIBLIOGRAPHY.—W. J. McGee, "The Seri Indians," *17th Annual Report of the Bureau of American Ethnology* (1898); A. L. Kroeber, *Serian, Tequistlatecan, and Hokan* (1915), *The Seri* (1931); E. H. Davis and E. Y. Dawson, "The Savage Seris of Sonora," *Scientific Monthly*, vol. 60 (1945); W. B. Griffen, *Notes on Seri Indian Culture* (1959); B. Malkin, *Seri Ethnozoology* (1962); W. H. Lindig, "Tree Burial Among the Seri Indians," *Ethnology*, vol. 3 (July 1964).

SERIEMA (CARIAMA), a South American bird belonging to the family Cariamidae, allied to the cranes (*q.v.*) and trumpeters. There are two species, both restricted to the southern half of South America. The crested seriema (*Cariama cristata*), with long legs and neck, stands about two feet high. The beak and legs are red, the plumage brownish above and dull white beneath with bluish skin around the eyes. It inhabits the grasslands, where it runs swiftly. The nest is built in bushes or trees and contains two buff eggs blotched with red-brown. The young are hatched, after about 25 days' incubation, covered with gray down. The seriema feeds on insects, snails, reptiles, and berries.

Burmeister's, or the gray, seriema (*Chunga burmeisteri*), which inhabits wooded areas, is darker and grayer, with a shorter crest and shorter legs.

SERIES, in mathematics, is the indicated sum of a set of terms. The notion of a series has close connection with that of a sequence. (*See* NUMBER SEQUENCES.) The simplest example of a sequence is that of positive integers $1, 2, 3, \ldots, n, \ldots$; this particular example is important for the general definition. An infinite sequence of numbers exists if every positive integer $1, 2, 3, \ldots, n, \ldots$ is assigned a number; in what follows the expression "sequence" stands for "an infinite sequence of numbers." Some examples of sequences are as follows:

$$1^2, 2^2, 3^2, \ldots, n^2, \ldots \tag{1}$$

$$1, 0, 1, 0, 1, \ldots \tag{2}$$

$$1, -2, 3, -4, 5, -6, \ldots \tag{3}$$

$$1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots, \frac{1}{n}, \ldots \tag{4}$$

The numbers constituting a sequence are called the terms of the sequence (such as the first, the second, . . . , the nth, . . .). It is customary to denote a sequence by a single letter with a

subscript (index) indicating the position of the term in the sequence; for example, $a_1, a_2, a_3, \ldots, a_n, \ldots$, or $b_1, b_2, b_3, \ldots, b_n, \ldots$. The abbreviated notation $\{a_n\}$ for $a_1, a_2, a_3, \ldots, a_n, \ldots$ and similarly with other letters also is used. It will be assumed that the terms of a sequence are real numbers (*see* NUMBER: *The Real Number System*); the case of sequences of complex numbers is considered below.

If the terms of a sequence $\{a_n\}$ have the property that $a_n \leq a_{n+1}$ for each n, the sequence is said to be increasing (or, sometimes, nondecreasing); if $a_n < a_{n+1}$ for each n, the sequence is strictly increasing. If $a_n \geq a_{n+1}$ for each n, the sequence is decreasing (or nonincreasing); if $a_n > a_{n+1}$, it is strictly decreasing. Sequences that are either increasing or decreasing are called monotone (monotonically increasing or monotonically decreasing, as the case may be). Sequences (1) and (4) above are monotone, while (2) and (3) are not. If all the terms of the sequence $\{a_n\}$ are less than a given number M (in symbols, $a_n < M$ for some M and all n) the sequence is said to be bounded above. If all a_n exceed a certain number, $\{a_n\}$ is said to be bounded below. Sequences bounded both above and below are simply called bounded. For example, sequence (1) is bounded below but not above, (2) is bounded, and (3) is not bounded either above or below.

A sequence $\{a_n\}$ is said to converge (or tend) to the limit a if the terms a_n approach a indefinitely as n increases; in symbols

$$\lim_{n\to\infty} a_n = a \text{ or, } a_n \to a$$

A precise definition is as follows: a_n tends to the limit a if given any positive number ϵ (no matter how small) we have $|a_n - a| < \epsilon$ for all n large enough. If $\{a_n\}$ tends to a limit, the sequence is said to be convergent; otherwise it is divergent. The following type of divergence is important: $\{a_n\}$ is said to diverge to $+\infty$ or, in symbols

$$\lim_{n\to\infty} a_n = +\infty \text{ or, } a_n \to +\infty$$

if for any number M (no matter how large) a_n is greater than M for all n large enough; divergence of a sequence to $-\infty$ is defined correspondingly. Of the sequences mentioned above, (4) converges to the limit 0; (1) and (2) diverge (the former to $+\infty$). Whether a given sequence converges and, if so, toward what limit, is not always easy to decide. The following facts can be helpful here.

I. If $a_n \to a$ and $b_n \to b$, then

$$a_n + b_n \to a + b,\ a_n b_n \to ab,\ a_n/b_n \to a/b$$

the last relation requiring, however, the additional assumption that $b \neq 0$.

II. Every convergent sequence is necessarily bounded; example (2) shows that the converse is not true: a bounded sequence need not converge.

III. If a sequence is increasing and bounded above, or decreasing and bounded below, it converges.

IV. If a sequence converges, then its terms, coming arbitrarily close to the limit, must also come arbitrarily close to each other, provided the indices are large enough. The converse of this is also true, though less easy to prove, and the result is the following important theorem of Cauchy: a sequence $\{a_n\}$ converges if and only if it has the following property: for any positive number ϵ (no matter how small) $|a_m - a_n| < \epsilon$ for all m and n large enough.

Starting with a sequence $a_1, a_2, \ldots, a_n, \ldots$ a new sequence $s_1, s_2, \ldots, s_n, \ldots$ is formed by the following rule:

$$s_1 = a_1,\ s_2 = a_1 + a_2,\ s_3 = a_1 + a_2 + a_3, \ldots$$
$$s_n = a_1 + a_2 + \ldots + a_n, \ldots$$

If $\{s_n\}$ converges to limit s, then it is said that the infinite series

$$a_1 + a_2 + a_3 + \ldots + a_n + \ldots$$

converges to sum s. Instead of $a_1 + a_2 + \ldots + a_n + \ldots$ it is customary to write $\sum_{n=1}^{\infty} a_n$ or, simply, Σa_n; the Greek letter sigma is a symbol for summation. The numbers s_n are called the partial sums of the series Σa_n, and the relation $s_n \to s$ is also written $\Sigma a_n = s$. The numbers a_n are the terms of the series Σa_n.

Given any sequence $b_1, b_2, \ldots, b_n, \ldots$, there is always a series

$$b_1 + (b_2 - b_1) + (b_3 - b_2) + \ldots + (b_n - b_{n-1}) + \ldots$$

whose partial sums are $b_1, b_2, b_3, \ldots, b_n, \ldots$. Hence, every series is represented by a sequence (that of its partial sums) and, conversely, every sequence by a series. It follows that every statement about series can be given in a form bearing on sequences, and conversely (but usually, in individual cases, one form may be preferable to the other). For example, Cauchy's theorem stated above and pertaining to sequences assumes, in the case of series, the following form:

V. A series Σa_n converges if, and only if, given any positive number ϵ, $|a_{m+1} + a_{m+2} + \ldots + a_n| < \epsilon$ for all m large enough and any n greater than m. In particular (taking $n = m + 1$), the terms of a convergent series must necessarily tend to 0.

Examples.—If a is any number and q is any number of absolute value less than 1, the series $a + aq + aq^2 + \ldots + aq^{n-1} + \ldots = \Sigma aq^{n-1}$ (called a geometric series) converges, and its sum is $a/(1 - q)$.

The series $\Sigma 1/n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots$, called the harmonic series, diverges.

If the terms of a series are positive numbers, the series is called positive. The partial sums of a positive series form a strictly increasing sequence and, in view of proposition III, a positive series converges if its partial sums are bounded above. If the partial sums of a positive series are not bounded above, the series diverges to $+\infty$. Proposition III also shows that:

VI. If Σa_n and Σb_n are positive series, and if $a_n \geq b_n$ for all n, then the convergence of Σa_n implies convergence of Σb_n and, equivalently, the divergence of Σb_n implies that of Σa_n. This result, usually called the comparison test, is very useful since the knowledge of convergence or divergence of some positive series can be utilized to obtain information about other series. For example, from the fact that the geometric series Σaq^{n-1} converges if a is positive and q is positive and less than 1, it is possible to deduce the following corollaries for positive series:

VII. Suppose that $a_{n+1}/a_n \to l$. Then, if $l < 1$ the series Σa_n converges; if $l > 1$ the series diverges (ratio test).

VIII. Suppose that $\sqrt[n]{a_n} \to l$. If $l < 1$, the series Σa_n converges; if $l > 1$, the series diverges (root test).

If the limit l in VII or VIII exists but is equal to 1, the test fails (the case $l = 1$ can occur for both convergent and divergent series) and one must look for other tests. There are many of them but the most useful is the following one based on the notion of the integral:

IX. If $f(x)$ is a function defined for x positive, itself positive and steadily decreasing to 0 as x increases indefinitely, then the series $\Sigma f(n) = f(1) + f(2) + f(3) + \ldots$ converges or diverges according as the integral $\int_1^{\infty} f(x)dx$ is finite or not (the integral test).

When applied to the function $f(x) = x^{-k}$, where k is a fixed positive constant, the test shows that the series

$$1^{-k} + 2^{-k} + 3^{-k} + \ldots + n^{-k} + \ldots = \sum_{n=1}^{\infty} n^{-k}$$

converges if $k > 1$ and diverges in all other cases. The result is not deducible from either the ratio or root test, since for the series $\sum_{n=1}^{\infty} n^{-k}$, l is equal to 1 both in VII and VIII.

Of interest among series that are not positive are alternating series; this name is given to series $a_1 - a_2 + a_3 - a_4 + \ldots$ where the numbers a_n themselves are of constant sign. It can be shown that such a series converges if $\{a_n\}$ decreases monotonically to 0. For example, the series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$ converges.

A series Σa_n is said to converge absolutely if $\Sigma |a_n|$ converges. Cauchy's theorem for series shows that if a series converges absolutely, then it converges in the ordinary sense; but the example of the convergent alternating series $1 - \frac{1}{2} + \frac{1}{3} - \ldots$ shows that a series may converge without converging absolutely. It follows

that absolutely convergent series form a special class among all convergent series. Since series $\Sigma|a_k|$ have positive or zero terms, the problem of whether a given series converges absolutely can be solved by applying one of the tests for the convergence of positive series. In many cases the simplest way of proving that a certain series converges is by showing that the series converges absolutely, and this shows the importance of positive series for the general theory of series. Absolutely convergent series have many properties that are not shared by all convergent series. The most important of them is the fact that if a series converges absolutely, then arbitrary changes in the order of the terms of the series affect neither the convergence nor the sum of the series. The situation is completely different for series that converge but do not do so absolutely (such series are called conditionally convergent); by a suitable change of the order of the terms such a series can be made to converge to any prescribed sum, or can even be made divergent (theorem of Riemann). Hence, unlike the case of absolutely convergent series, a conditionally convergent series has no intrinsic sum and everything depends on the order of the terms.

The number of important series that occurs in mathematics is enormous. A few basic convergent series are listed below:

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\ldots=\ln 2$$

$$1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\frac{1}{9}-\ldots=\frac{1}{4}\pi \text{ (Leibniz' series)}$$

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\frac{1}{5^2}+\ldots=\frac{1}{6}\pi^2$$

$$1+\frac{1}{3^2}+\frac{1}{5^2}+\frac{1}{7^2}+\frac{1}{9^2}+\ldots=\frac{1}{8}\pi^2$$

$$1+\frac{1}{1}+\frac{1}{1\times 2}+\frac{1}{1\times 2\times 3}+\frac{1}{1\times 2\times 3\times 4}+\ldots=e=2.7182818\ldots$$

Series With Complex Terms.—The convergence of complex sequences and series (*i.e.*, sequences and series whose terms are complex numbers) is formally defined exactly in the same way as in the case of real numbers, the absolute value $|z|$ of a complex number $z = x + iy$ (where i is $\sqrt{-1}$) being $\sqrt{(x^2+y^2)}$. This definition leads to the following important conclusion: if $c_n = a_n + ib_n$, the convergence of Σc_n is equivalent to the simultaneous convergence of Σa_n and Σb_n; and if $\Sigma a_n = s$, $\Sigma b_n = t$, then $\Sigma c_n = s + it$. Hence, the study of complex series reduces to that of real series. As in the case of real series, the absolute convergence of Σc_n is defined as the convergence of $\Sigma|c_n| = \Sigma\sqrt{(a_n^2+b_n^2)}$. Since the convergence of $\Sigma\sqrt{(a_n^2+b_n^2)}$ means the same thing as the simultaneous convergence of $\Sigma|a_n|$ and $\Sigma|b_n|$, the absolute convergence of $\Sigma(a_n+ib_n)$ is equivalent to the absolute convergence of both $\Sigma|a_n|$ and $\Sigma|b_n|$; it therefore implies the ordinary convergence of Σa_n and Σb_n, and so also of $\Sigma(a_n+ib_n)$. Among complex series, absolutely convergent series play the same important role as among real series; in particular, the convergence and the sum of an absolutely convergent complex series are independent of the order of the terms.

Double Sequences and Series.—A double sequence is defined by assigning to every pair m,n of positive integers a number $\{s_{m,n}\}$. A double sequence can thus be written as a double array of numbers:

$$\begin{array}{l} s_{1,1},\ s_{1,2},\ \ldots,\ s_{1,n},\ \ldots \\ s_{2,1},\ s_{2,2},\ \ldots,\ s_{2,n},\ \ldots \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ s_{m,1},\ s_{m,2},\ \ldots,\ s_{m,n},\ \ldots \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \end{array}$$

The indices m,n of a term can be interpreted as the coordinates of a point in a plane. The notion of convergence of a double sequence is not so immediate and can be introduced in many ways (which are not equivalent). The following definition is most commonly used: $\{s_{m,n}\}$ is said to converge to limit l if for any positive ϵ we have $|s_{m,n} - l| < \epsilon$ provided both m and n are large enough. With this definition, changing the terms in any finite number of rows and columns in the table above (the number of changes may be infinite) does not affect the convergence or limit of $\{s_{m,n}\}$.

The definition of convergence of a double series

$$\begin{array}{l} a_{1,1}+a_{1,2}+\ldots+a_{1,n}+\ldots \\ a_{2,1}+a_{2,2}+\ldots+a_{2,n}+\ldots \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \\ a_{m,1}+a_{m,2}+\ldots+a_{m,n}+\ldots \\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots \end{array}$$

can also be introduced in various ways; the series itself will be written as $\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}a_{m,n}$ or simply $\Sigma\Sigma a_{m,n}$. Three definitions follow, only the first of which has close connection with the just defined notion of convergence of double sequences (at any rate, connection between double series and sequences is less intimate than in the case of single series and sequences):

(a) $\Sigma\Sigma a_{m,n}$ is said to converge to sum s, if the double sequence $\{s_{m,n}\}$ tends to s, in the sense just described, where $s_{m,n}$ denotes the sum of the terms $a_{j,k}$ with $j \leq m$, $k \leq n$; such terms $a_{j,k}$ fill a rectangle in the upper-left corner of the double series table above and the sums $s_{m,n}$ are correspondingly called rectangular partial sums of the series; the kind of convergence which has just been introduced is called the rectangular convergence of the series;

(b) the series $\Sigma\Sigma a_{m,n}$ is said to converge to sum s, if S_R tends to s as R tends to $+\infty$, where S_R is the sum of the terms $a_{m,n}$ with $m^2 + n^2 \leq R^2$; indices m,n satisfying this inequality are coordinates of points situated in the circle with the centre at the origin and radius R, so that it is natural to call the S_R the circular partial sums, and the kind of convergence they generate circular convergence;

(c) $\Sigma\Sigma a_{m,n}$ is said to converge to sum s if the terms in each column of the table form convergent series and if $A_1 + A_2 + \ldots + A_n + \ldots = s$, where A_n is the sum of terms in the nth column of the table; this kind of summation is called summation by columns, and by interchanging the roles of columns and rows leads to a parallel notion of convergence by rows.

These definitions of convergence are not equivalent and a series can converge according to one definition and not according to another. The problem of additional conditions under which the definitions are equivalent is not very difficult but the conditions themselves are not of a simple nature and most (though not all) important series $\Sigma\Sigma a_{m,n}$ in analysis are absolutely convergent (that is $\Sigma\Sigma|a_{m,n}|$ converges). For such series all reasonable definitions of convergence are equivalent and the sum of the series is independent of the definition. Definitions and results can be extended to triple, quadruple, etc. series.

Series of Functions.—Let $\Sigma u_n(x)$ be a series whose terms $u_1(x), u_2(x), \ldots, u_n(x), \ldots$ are functions defined in an interval of the real variable x. If $\Sigma u_n(x)$ converges at each point of the interval the series is said to converge pointwise in the interval. Denote by $f(x)$ the sum and by $s_n(x)$ the partial sums of the series. Pointwise convergence means that, for any positive ϵ, there is an index n_0, depending in general on ϵ and x, in symbols $n_0 = n_0(\epsilon,x)$, such that $|f(x) - s_n(x)| < \epsilon$ for $n > n_0$. If for each $\epsilon > 0$ an n_0 can be found independent of x the series is said to converge uniformly in the interval considered; geometrically this means that for each $n > n_0$ the graph of the curve $y = s_n(x)$ and that of the curve $y = f(x)$ differ by less than ϵ in the whole interval under consideration. Pointwise convergence in general does not preserve properties of the terms of the series. For example, if the $u_n(x)$ are continuous and $\Sigma u_n(x)$ converges pointwise, the sum of $\Sigma u_n(x)$ need not be continuous. A simple example is provided by the series $(1-x) + x(1-x) + x^2(1-x) + \ldots$, which converges at each point of the interval $0 \leq x \leq 1$ and whose sum is equal to 0 for $x = 1$ and equal to 1 at the remaining points of the interval. Similarly, if the $u_n(x)$ are integrable over an interval and the series $\Sigma u_n(x)$ converges pointwise, the sum of $\Sigma u_n(x)$ need not be integrable. The significance of the notion of uniform convergence is that the sum of a uniformly convergent series of functions inherits many important properties of the terms, and in particular continuity and integrability. Not all properties, however, are preserved by uniform convergence; for example, the sum of a uniformly convergent series of differentiable functions need not be differentiable (as a matter of fact, examples of continuous and nowhere differentiable functions are usually given as sums of uniformly convergent series of differentiable functions). The following three theorems about uniformly convergent series are particularly useful.

X. If the functions $u_n(x)$ are continuous in an interval, and the series $\Sigma u_n(x)$ converges uniformly in the interval, the sum of the series is continuous in the interval.

XI. If the $u_n(x)$ are integrable over a finite interval (a,b), and $\Sigma u_n(x)$ converges uniformly in (a,b), then the sum $f(x)$ of $\Sigma u_n(x)$ is integrable over (a,b) and

$$\int_a^b f(x)dx = \Sigma \int_a^b u_n(x)dx$$

XII. If each term of the series $\Sigma u_n(x)$ has a continuous derivative in a finite interval (a,b), and if the series $\Sigma u_n'(x)$ of derivatives converges uniformly in (a,b), then the series $\Sigma u_n(x)$ also converges uniformly in (a,b), its sum $f(x)$ is differentiable and $f'(x) = \Sigma u_n'(x)$.

There are no special tests for the pointwise convergence of series since pointwise convergence simply means ordinary convergence at each point x separately. There are tests for the uniform convergence of a series of functions, among which the following, usually called Weierstrass' M-test, is particularly useful:

XIII. If there are positive constants $M_1, M_2, \ldots, M_n, \ldots$ such that the series ΣM_n converges and $|u_n(x)| \leq M_n$ for each n and all x of an interval then the series $\Sigma u_n(x)$ converges uniformly in the interval.

Only the uniform convergence of series has been considered. There are parallel definitions and results for the uniform convergence of sequences. One can also consider the uniform convergence of double series and sequences. Special types of series of functions are particularly important for various branches of analysis (*see*, for example, FOURIER SERIES). The convergence of series $\Sigma u_n(z)$ whose terms depend on a complex variable $z = x + iy$ may be considered. The most important among them is the case of power series.

Power Series.—Power series are of the form

$$a_0 + a_1z + a_2z^2 + \ldots + a_nz^n + \ldots = \Sigma a_nz^n$$

where the a_n are constants (in general, complex) and $z = x + iy$ is a complex variable; in the symbol Σa_nz^n the index n ranges through the values $0, 1, 2, \ldots, n, \ldots$ Each power series has a circle of convergence; this is a circle (disk) with the centre at the point $z = 0$ and the following properties: the series converges at each point interior to the circle and diverges at each point exterior to the circle. As to the behaviour on the circumference of the circle of convergence, Σa_nz^n may converge at some points, diverge at others. The intersection of the circle of convergence with the real axis is an interval $(-R,R)$ and the series Σa_nx^n, a power series of the real variable x, converges in the interior of this interval, and diverges in the exterior. The radius R of the circle of convergence is called the radius of convergence of the power series. The circle of convergence can degenerate to a point; each power series Σa_nz^n necessarily converges at the point $z = 0$ but may converge at no other point: the series $\Sigma n!z^n$ is an example of this. The circle of convergence may also cover the whole complex plane, as in the series $\Sigma z^n/n!$ Finally the radius of convergence may have any prescribed positive value R since the radius of convergence of the geometric series $\Sigma(z/R)^n$ is precisely R.

At each point interior to the circle of convergence the power series not only converges but converges absolutely. What is very important is that in each circle concentric with the circle of convergence but of smaller radius the power series converges uniformly. Series of the form $a_0 + a_1(z-b) + a_2(z-b)^2 + \ldots = \Sigma a_n(z-b)^n$, where b is a fixed complex number are called power series with centre b. Their study does not require the introduction of new concepts since by replacing $z - b$ with a new variable Z we reduce the series to the previously discussed case Σa_nZ^n. It follows that $\Sigma a_n(z-b)^n$ has also its circle of convergence, but this time the circle has its centre at the point b. Functions representable by power series have very important properties. Listed below are a few basic power series, considering for simplicity only real values of x.

$(1+x)^\alpha = 1 + \alpha x + \dfrac{\alpha(\alpha-1)}{1\times 2}x^2 + \dfrac{\alpha(\alpha-1)(\alpha-2)}{1\times 2\times 3}x^3 + \ldots$ (Newton's binomial series; $R = 1$)

$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \ldots (R = 1)$

$\text{arc tan } x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \ldots (R = 1)$

$e^x = 1 + \dfrac{x}{1} + \dfrac{x^2}{1\times 2} + \dfrac{x^3}{1\times 2\times 3} + \ldots (R = \infty)$

$\cos x = 1 - \dfrac{x^2}{1\times 2} + \dfrac{x^4}{1\times 2\times 3\times 4} - \dfrac{x^6}{1\times 2\times 3\times 4\times 5\times 6} + \ldots$ $(R = \infty)$

$\sin x = x - \dfrac{x^3}{1\times 2\times 3} + \dfrac{x^5}{1\times 2\times 3\times 4\times 5} - \ldots (R = \infty)$

Divergent Series.—Consider the series $1 - 1 + 1 - 1 + \ldots$ Its partial sums $1, 0, 1, 0, \ldots$ do not tend to any limit; therefore, according to the definition, the series is divergent. On the other hand, the average value of the partial sums $1, 0, 1, 0, \ldots$ is $\frac{1}{2}$, which made some mathematicians as early as the 17th century adopt the point of view that though the series, strictly speaking is divergent it somehow has the sum $\frac{1}{2}$. Such ideas, though for a long time unclear and confused, were finally clarified and led to a logically coherent and fruitful theory of divergent series and sequences. A few facts are here presented. The theories of divergent series and sequences being parallel, sequences are considered first.

Consider a sequence $\{s_n\} = s_0, s_1, \ldots, s_n, \ldots$ (in this context it will be convenient to include 0 as an index). A definition of convergence of $\{s_n\}$ already has been formulated. This definition, which may be called classical, is not the only one possible; alternate definitions could be adopted. The following one is very intuitive. Consider the averages of the successive terms of $\{s_n\}$; these averages will be denoted by $\sigma_0, \sigma_1, \sigma_2, \ldots, \sigma_n, \ldots$ so that

$$\sigma_0 = \frac{s_0}{1}, \sigma_1 = \frac{s_0 + s_1}{2}, \sigma_2 = \frac{s_0 + s_1 + s_2}{3}, \ldots, \sigma_n = \frac{s_0 + s_1 + \ldots + s_n}{n+1}, \ldots$$

It may be said that $\{s_n\}$ converges on the average to limit s if the sequence $\{\sigma_n\}$ converges to limit s in the classical sense. It can be shown that any sequence $\{s_n\}$ which converges in the classical sense also converges on the average, and that the limits are the same in both cases. There are sequences that do not converge in the classical sense but converge on the average (the sequence $\{s_n\} = 1, 0, 1, 0, \ldots$ is an instance in point, since here $\{\sigma_n\} = 1, \frac{1}{2}, \frac{2}{3}, \frac{1}{2}, \frac{3}{5}, \frac{1}{2}, \ldots$ tends to $\frac{1}{2}$). Hence, in adopting convergence on the average as a definition of convergence of a sequence, conflict with the old definition is avoided on the one hand and, on the other, sequences previously considered divergent (and useless) are made convergent. Instead of saying that $\{s_n\}$ converges "on the average" it is now customary to say that $\{s_n\}$ is summable by the method of the first arithmetic mean (*q.v.*). Not every sequence is thus summable. For example, if $s_n = 1, -2, +3, -4, +5, \ldots$, then σ_n does not converge in the classical sense; but it is easy to prove that σ_n is summable by the method of the first arithmetic mean to the limit 0, so that we may say that $\{s_n\}$ is summable to limit 0 by the method of the second arithmetic mean. We could similarly introduce the third, fourth, etc., arithmetic means. It is, however, preferable to adopt a more general point of view.

The averages considered in the definition of $\{\sigma_n\}$ are not the only conceivable ones, since "weighted" averages may be introduced. The general scheme is as follows. Consider a fixed matrix (double array) of numbers denoted by M:

$$\begin{matrix} a_{0,0}, & a_{0,1}, & a_{0,2}, & \ldots, & a_{0,n}, & \ldots \\ a_{1,0}, & a_{1,1}, & a_{1,2}, & \ldots, & a_{1,n}, & \ldots \\ a_{2,0}, & a_{2,1}, & a_{2,2}, & \ldots, & a_{2,n}, & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m,0}, & a_{m,1}, & a_{m,2}, & \ldots, & a_{m,n}, & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{matrix}$$

Given any sequence $\{s_n\}$ a new sequence σ_m may be introduced defined by the formulas

$$\sigma_m = a_{m,0}s_0 + a_{m,1}s_1 + a_{m,2}s_2 + \ldots + a_{m,n}s_n + \ldots$$

(in which $m = 0, 1, 2, \ldots$).

Then σ_m may be considered as a weighted average of the num-

bers $s_0, s_1, \ldots$ formed with weights situated in the mth row of the matrix. If $\{\sigma_m\}$ converges in the classical sense to limit s, it may be said that $\{s_n\}$ is summable by the matrix M, or simply summable M, to limit s; s is then the generalized limit of $\{s_n\}$, associated with the matrix M. It is easily seen that the method of the first arithmetic mean is given by the matrix

$$\begin{matrix} 1, & 0, & 0, & 0, & \ldots \\ \frac{1}{2}, & \frac{1}{2}, & 0, & 0, & \ldots \\ \frac{1}{3}, & \frac{1}{3}, & \frac{1}{3}, & 0, & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{matrix}$$

Classical convergence corresponds to the matrix whose terms on the main diagonal are all $1 (a_{n,n} = 1$ for all $n)$ and the remaining terms are 0, since then $\sigma_n = s_n$ for all n.

This definition of summability of a sequence was introduced by the German mathematician Otto Toeplitz. It is extremely general, though it could be generalized still further, and raises a very large number of problems which are the subject of the general theory. A few such problems are considered below.

The first question concerns the conditions under which this definition is a genuine generalization of the classical convergence. The answer is that a necessary and sufficient condition that every convergent sequence should also be summable M, and to the same limit, is that (1) the terms of M should tend to 0 in each column separately; (2) the sum of the absolute values of the terms in each row should stay below some constant; (3) the algebraic sum of elements in each row should tend to 1 as the index of the row increases. Many results about summable sequences are corollaries of this theorem. In the special case when the elements of the matrix are nonnegative, condition (2) is essentially a consequence of (3) and can be dropped.

It is said that a method M' is stronger than M if every sequence summable M is summable M', but not conversely, and if the limits in both cases are the same. It may then be asked when a given method M' is stronger than M. Theorems answering this question are called Abelian, after Niels Henrik Abel. If M' is stronger than M we may ask under what additional conditions, for the sequence summability by M' implies summability by M; results of this type are called Tauberian theorems, after the Austrian mathematician Alfred Tauber. The most interesting theorems of Abelian and Tauberian character pertain to specific and important methods of summability.

Given a series $a_0 + a_1 + \ldots + a_n + \ldots$ it may be said that it is summable M to sum s if the sequence of the partial sums of the series is summable M to limit s. To distinguish between series and sequences some writers say that a series is summable while a sequence is limitable, but no confusion arises if the word summable is used in both cases. However, certain forms of summability are more adapted to series than sequences, and vice versa. This situation, in particular, applies to Abel's method of summation, which defines the sum of Σa_n as the limit of $\Sigma a_n r^n$ for r tending to 1 through values less than 1. Abel's method is stronger than convergence, and even stronger than the method of the first arithmetic mean (Abelian theorems). A series that is Abel summable need not converge; for example, the divergent series $1 - 1 + 1 - \ldots$ is Abel summable to sum $\frac{1}{2}$ since $1 - r + r^2 - r^3 + \ldots = 1/(1+r)$ tends to $\frac{1}{2}$ as $r \to 1$. If, however, $a_n \leq 1/n$ for all n, and Σa_n is Abel summable, then Σa_n converges (a Tauberian theorem).

The method of the first arithmetic mean and that of Abel are among the most important ones. Their main applications are to the study of the behaviour of power series on the circumference of their circles of convergence and to the theory of Fourier series. There is a great variety of methods of considerable intrinsic interest and importance in applications. The method of arithmetic means of various orders (hinted at above) has developed into a wide and elegant theory with far-reaching applications. The method of Borel assigns to a sequence $s_0, s_1, \ldots, s_n, \ldots$ the sum

$$\lim_{x \to +\infty} \left\{ e^{-x} \sum_{n=0}^{\infty} \frac{s_n}{n!} x^n \right\}$$

and can sum power series outside their circles of convergence. There is also an extension of the theory to divergent integrals. The importance of the theory of divergent series lies, however, not only in individual results but in the general point of view upon the notion of convergence, a point of view which has had a great impact upon the development of mathematical analysis.

History of Series.—The origin of the modern theory of series precedes that of the calculus and can be associated in the 17th century with the names of Bonaventura Cavalieri, John Wallis, James Gregory and others. In the same century, the discovery of the calculus by Leibniz and Newton gave a very strong impulse to the theory; and the 18th century witnessed considerable progress in the work of Jakob and Daniel Bernoulli, D'Alembert, Leonhard Euler and others. Much of this work was of purely formal character, and mathematical rigour was not the primary concern of the investigators. Though Newton and Leibniz avoided divergent series, formal and uninhibited use of the latter was bringing such mathematical rewards that the temptation was too much to withstand, and divergent series became—mostly through the work of Euler—an accepted tool of investigation. Mathematical instinct of the great mathematicians of the 18th century prevented them, however, from making mistakes and their results, properly interpreted, can now be proved rigorously. Rigorous foundations for the theory of infinite series were laid in the first half of the 19th century by Augustin Louis Cauchy and Niels Henrik Abel; definitions they gave and results they proved can be found in all textbooks of calculus. Their attitude toward divergent series was purely negative, however. (Abel wrote in 1828: "Divergent series are the invention of the devil, and it is shameful to base on them any demonstration whatsoever.") The respect commanded by the work of Cauchy and Abel banished divergent series from mathematics for half a century. Their work on the foundations of the theory of series was continued by others in the 19th century, in particular by Peter Gustav Lejeune-Dirichlet, Georg Friedrich Riemann and Karl Weierstrass. By the last decade of the 19th century the effects of Abel's anathema began to disappear, and the work of Ernesto Cesàro, Émile Borel, Leopold Fejér and others showed that the theory of divergent series can be put on a rigorous basis and is a source of great progress in mathematics. In the 20th century the work of G. H. Hardy and J. E. Littlewood brought major breakthroughs and extended the field in depth. Important progress in Tauberian theorems was made by Norbert Wiener.

See also ANALYSIS: *Functions;* BINOMIAL THEOREM; FUNCTION; FUNCTIONS, ANALYTIC; FUNCTIONS, SPECIAL; INTEGRATION AND MEASURE; and references under "Series" in the Index.

BIBLIOGRAPHY.—Elements of the theory of convergent series will be found in any book of calculus; a good source is G. H. Hardy, *A Course of Pure Mathematics* (many editions). On a more advanced level are T. Fort, *Infinite Series* (1930); T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series,* 2nd ed. (1926); K. Knopp, *Theory and Application of Infinite Series,* 2nd ed. (1948), elegantly written and containing a wealth of material; I. I. Hirschman, *Infinite Series* (1962). *See* also G. H. Hardy, *Divergent Series* (1949); H. R. Pitt, *Tauberian Theorems* (1958); K. Zeller, *Theorie der Limitierungsverfahren* (1958); and, for summability of integrals and related topics, K. Chandrasekharan and S. Minakshisundaram, *Typical Means* (1953).

(A. ZD.)

SERIGRAPHY: *see* SILK SCREEN PRINTING.

SERJEANTS-AT-LAW (SERVIENTES AD LEGEM): *see* LEGAL PROFESSION.

SERJEANTY, in the feudal age, was a form of land tenure granted in return for the performance of some specific service to the lord, whether the king or another. In England, immediately after the Norman Conquest, the king created many serjeanties because he had much land at his free disposal and needed many services—the provision of bows, arrows, and knives, men to keep down vermin in his forests, to serve in his buttery and winecellar and as ushers. It was easier at that time to reward any service with land than with money. But it was uneconomic, and fresh grants of land in serjeanty became rare by the late 12th century. The tenure was widespread, and the Domesday account of every county ends with a list of the king's serjeants. Tenants by serjeanty were subject to the feudal dues of wardship, marriage, and relief, but they paid no scutage (*q.v.*) since they performed no

knight service (*see* KNIGHT SERVICE). Land held by serjeanty could not by law be sold or divided among heirs. Since by the 13th century much alienation (*i.e.*, transfer by deed) and subdivision had, however, actually taken place, an inquiry into the tenure was made in 1250 under Robert Passelewe, deputy treasurer. As a result, the holders of the alienated portions were mostly required to pay rent or do a quota of knight service, and the unalienated portion remained charged with the original duty. A statute of 1324 (17 Edw. II c. vii) provided that a "reasonable" fine was owed to the crown by serjeants who alienated their serjeanties "without the king's licence."

The miscellaneous nature of the services due to the king by his serjeants resulted in the inaccurate conception of serjeanty as falling into two divisions: grand serjeanty, a tenure so noble that it ranked socially above knight service, and petty serjeanty, a tenure so meagre that it ranked with the peasants' tenure, socage (*q.v.*). In origin there was no distinction between serjeanties, but inevitably those which brought their holders into immediate contact with the sovereign acquired prestige and became known as grand serjeanties. *See* also ENGLISH LAW: *Norman Feudal Land Law.*

BIBLIOGRAPHY.—J. H. Round, *The King's Serjeants and Officers of State* (1911); E. G. Kimball, *Serjeanty Tenure in Medieval England* (1936); A. L. Poole, *Obligations of Society in the XII and XIII Centuries* (1946). (D. M. S.)

SERLIO, SEBASTIANO (1475–1554), Italian architect and theorist, the founder of the classical school of architecture in France, was born at Bologna on Sept. 6, 1475. He worked in Rome with Baldassare Peruzzi from 1514 until the sack of the city in 1527, when he fled to Venice. Peruzzi bequeathed him all his own drawings, and these were put to good use by Serlio in his *Treatise*, his greatest achievement. The first part of it, published in 1537, was really Book iv of the complete *Treatise* but was soon reprinted and translated as an independent work (Eng. trans., 1611).

Book iii followed in 1540, and in 1541 Serlio moved to France, where he was supported for six years by Francis I. Books i and ii came out in 1545, Book v in 1547, Book vi exists only in manuscript, Book vii was published posthumously in 1575, but there is a manuscript of (apparently) an eighth book in Munich. Most important of all, Serlio published an *Extraordinario Libro* in 1551 which contains 50 fanciful designs for doorways. These were much copied in northern Europe.

Serlio's *Treatise* as a whole was very influential because it was essentially a practical handbook of the antique style and presented a number of models for copying; it was fundamentally a set of illustrations linked by commentary rather than an essay on archaeology or aesthetics. The only surviving buildings which can be connected with Serlio are one doorway at Fontainebleau and the château of Ancy-le-Franc in Burgundy. He died at Fontainebleau in 1554.

See W. B. Dinsmoor in *Art Bulletin*, xxiv (1942), and A. F. Blunt, *Art and Architecture in France, 1500 to 1700* (1953). (P. J. MY.)

SERMON: *see* PREACHING.

SERMON ON THE MOUNT, the popular name for Matt. 5–7, the longest discourse of Jesus in the Synoptic Gospels. In it are concentrated the most familiar ethical teachings of Jesus; however, they are so thoroughly founded upon the insuperable demand of God as to be sharply distinct from any merely humanistic ethic. In brief outline the sermon contains: (1) nine blessings, 5:3–12 (*see* BEATITUDES); (2) salt and light, 5:13–16; (3) validity of the Law, 5:17–20; (4) ". . . but I say to you . . . ," interpretations of six demands of the Law, 5:21–48; (5) three acts of piety (almsgiving, prayer, fasting) with (6) the Lord's Prayer (*q.v.*) appended to the second, 6:1–18; (7) true and false treasure, 6:19–34; (8) miscellaneous ending, mostly warnings (against self-righteous judgment, against profaning the holy, God's response to prayer, the Golden Rule, against any easy way to Life, against false prophets, against serving God by word only), 7:1–23, illustrated by the simile of a house with and without foundation, 7:24–27, and an editorial conclusion, 7:28 ff.

Most of these sayings have parallels in Luke, of which about half occur in Luke's much shorter Sermon on the Plain (6:20–49, named from Luke 6:17) and in nearly the same order; Luke's remaining parallels are scattered through his chapters 11–16. This parallelism suggests either that the author of Luke (*see* LUKE, GOSPEL ACCORDING TO SAINT) dispersed what had come down to him as a sermon or that the author of Matthew (*see* MATTHEW, GOSPEL ACCORDING TO SAINT) editorially composed a sermon out of mostly scattered sayings offered him by tradition. From the clearly discernible literary habit of the Gospel of Matthew to organize the words of Jesus into connected complexes concluding with "And when Jesus had finished these sayings . . . ," the latter possibility is the probable one. The plan and some basic material of the sermon were furnished by what is common to the Sermon on the Mount and the Sermon on the Plain; *i.e.*, by a definite section of the lost source common to Matthew and Luke, whether this was a written document or only crystallized and memorized tradition.

Neither the "mount" nor the "plain" is given a name and it is idle to attempt to identify either. The source evidently did not localize the discourse, but apparently did suggest that it was of one piece. Whether the latter is true is neither determinable nor centrally important. What is important is that the sermon contains some of the most certain and characteristic utterances of the historical Jesus.

See M. Dibelius, *The Sermon on the Mount* (1940). (KE. G.)

SEROV, a town in Sverdlovskaya (Sverdlovsk) Oblast' of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the left bank of the Kakva, a tributary of the Sos'va. Pop. (1959) 97,882. Originally known as Nadezhdinsk, the town developed as a centre of the Urals ironworking area. Although its development in the 1890s was late for the Urals, it soon became the largest pre-Revolutionary ironworking town in the area, making rails for the Trans-Siberian Railway. Today it is the main centre of the northern Urals mining and metallurgical region, with iron ore, bauxite, manganese, gold, and lignite obtained nearby. It produces high-grade steel. The lignite is used in the thermal power station. (R. A. F.)

SEROWE, town in Ngwato district, Bechuanaland (Botswana) Africa, lies 3,800 ft. (1,158 m.) above sea level, 30 mi. (48 km.) WNW of Palapye Road on the Cape to Congo railway, 200 mi. SW of Bulawayo, Rhodesia, and 245 mi. (394 km.) NNE of Mafeking, S.Af. Pop. (1964) 33,592. The chief town of the Ngwato tribe and the headquarters of the tribal administration, it has several government offices, a hospital, and an airfield. The European population consists largely of traders. Serowe was first occupied by the Ngwato in 1902 after they had abandoned their earlier home in Palapye. The statue of a duiker stands in memory of chief Kgama III, tribal leader until 1923. (AY. SY.)

SERPENTINE, a group of hydrous magnesium silicates with iron, nickel, and manganese commonly substituting for magnesium. The iron gives it a variety of green (although sometimes yellowish and brownish) colours; the pure material is grayish to white. When mottled with red the colour is due to a physical mixture of excess iron oxide. The usually massive serpentine rock may be composed of a mixture, the fine fibrous variety known as chrysotile and antigorite, both of the same composition but differing in crystal structure. Other varieties are identified by X-ray diffraction techniques.

Chrysotile serpentine, when found unmixed with antigorite and with long fibres, is used for most of the asbestos (*q.v.*) of commerce although some asbestos does come from the amphibole group of minerals.

Serpentine is a secondary mineral generally formed by the hydration of the magnesium silicates in peridotite, an olivine and pyroxene (*q.v.*) rich rock. The composition of serpentine is $Mg_3Si_2O_5(OH)_4$, so commonly both H_2O and SiO_2 must be added to an existing composition to form serpentine, especially in the magnesium carbonate rocks. This is a process that generates heat and requires an increase in volume for a large mass to develop. However, occasional distinct crystals show the form of olivine, pyroxene, or amphibole that has been replaced forming a pseudomorph which may grow by some leaching process. The density

of serpentine is about 2.5 to 3.0. It cleaves easily parallel to the plates in the variety antigorite or to the fibres in the variety chrysotile. It has a waxy feel and appearance and takes a high polish after carving or cutting as a gem or ornamental stone (verd antique).

(*See* also MARBLE; STONE.)

Serpentine occurrence is extremely widespread. It is characteristically found along the crests and axes, the low-pressure areas, of great folds whether island arcs or alpine mountain chains. These massive serpentines are commonly produced by metamorphism (*q.v.*) of peridotite, and indicate a development in these major structures of the earth's crust that progressed simultaneously with the growth of the fold system and may have continued after the mountain building ceased.

The serpentines along island arcs outcrop over areas covering from hundreds to thousands of square miles, as in New Caledonia and Cuba, and may be exposed by erosion or perhaps a solid extrusion.

See also references under "Serpentine" in the Index.

(K. O. B.)

SERPENT MOUND, GREAT, a prehistoric American Indian earthwork in Ohio on a narrow spur between Brush Creek and East Creek, Adams County. Feature attraction of Serpent Mound State Park, the effigy is 1,330 ft. (405 m.) long with seven serpentine convolutions, 15 to 20 ft. (4 to 6 m.) wide in the body, and 3 to 5 ft. (1 to 2 m.) high. The tail section has three coils, while the head is reconstructed as a serpent's open mouth. An oval embankment, 125 ft. long and 60 ft. wide, protrudes from the open mouth. Carefully excavated sections of the body and the oval disclosed a foundation of stone, ashes, and clay. There was no definite information as to the meaning of this effigy in the 1960s; apparently it had considerable importance as a symbol of religious beliefs. Serpents have had a prominent place in the mythology and folklore of Indians of the Mississippi and Ohio valleys. Materials excavated from conical mounds and a village site to the south of the serpent indicate that this Adena Culture effigy was constructed about 2,000 years ago.

See also MOUND BUILDERS; OHIO: *History*. (J. B. GN.)

SERPUKHOV, a town in Moscow Oblast' of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on both banks of the Nara River at its confluence with the Oka, 62 mi. (100 km.) S of Moscow. Pop. (1959) 106,387. The town was founded in 1374, although a settlement existed there earlier. It was one of the major strongholds protecting Moscow from attack from the south by the Tatars, and was sacked in 1382 and 1408. Ivan IV, the Terrible, built stone fortifications in 1556. Lying on the main route south from the capital (now on the trunk railway and highway from Moscow to the Crimea), the town soon became an important centre of trade and crafts, especially sailmaking. In the 19th century textile manufacture was developed and modern Serpukhov is one of the chief textile towns of the Moscow region, concentrating on cotton spinning and weaving and the production of cotton prints. Artificial silk is also made. The engineering industry is important, producing motor invalid carriages, machine tools, hydraulic presses, gasoline pumps, metal files, and parts for textile machinery. Concrete pipes and foodstuffs are also produced. (R. A. F.)

SERRA, JUNÍPERO (1713–1784), Spanish Franciscan missionary in America, known as the Apostle of California, was born on the island of Majorca, Nov. 24, 1713. After entering the Franciscan order in 1730, he became a university professor and distinguished preacher. In 1750 he arrived in Mexico City for missionary work among the Indians. Serra laboured in the Sierra Gorda missions for eight years, then became a circuit missionary in south-central Mexico until 1767. He was president of the Lower California missions, 1768–69.

When José Gálvez ordered the occupation of Upper California, Serra joined the expedition of Gaspar de Portolá (*q.v.*). On July 16, 1769, he founded Mission San Diego, the first within the present state of California. Between 1770 and 1782 he founded eight additional missions: San Carlos at Monterey (1770), changed to Carmel in 1771, which became Serra's headquarters; San Antonio and San Gabriel, the latter near Los Angeles (1771); San Luis Obispo (1772); San Francisco (Mission Dolores) and San Juan Capistrano (1776); Santa Clara (1777); and San Buenaventura (1782). He introduced cattle, sheep, grains, and fruits from Mexico. A strenuous defender of the Indians, Serra died at San Carlos, Aug. 28, 1784, and is there buried. The cause for his beatification has been introduced. A statue of Junípero Serra represents California in the National Statuary Hall in Washington. *See* also CALIFORNIA: *History: Missions.*

See Francisco Palóu's *Life of Fray Junípero Serra*, Eng. trans. by Maynard Geiger (1955); C. E. Chapman, *The Founding of Spanish California* (1916). (M. J. GE.)

SERRAI (SERRES), the chief town of the Serrai *nomos* (department) of Greek Macedonia, 43 mi. (69 km.) NE of Salonika. Pop. (1961): 40,063 (town), 248,041 (*nomos*). The town stands at the foot of the hills on the eastern side of the valley of the Struma (Strymon) River, and commands the route through the Rupel Pass into Bulgaria. It is the seat of the metropolitan bishop of Serrai and Nigrita and consists of the old quarter, Varosh, situated at the foot and on the slope of the hill crowned by the old castle, and of the new town, built on the plain and forming the commercial centre. The district is remarkably fertile, and there is a large trade in rice and cereals; other exports include tobacco, cotton, and hides.

Serrai is the ancient Siris in Paeonia, later called Sirrhae. It was fortified by the Byzantine emperors to guard the road between Thessalonica (Salonika) and Constantinople (Istanbul), and to serve as a stronghold on the Bulgarian frontier. It was unsuccessfully attacked by the Bulgars in the 10th century and again in 1195–96. The Serbians conquered it in 1345, but it was retrieved by the Byzantines in 1371, before falling to the Turks in 1383. It remained Turkish until 1913, when it was looted and burned by Bulgarian troops during the Greek advance up the Struma valley. It later passed into Greek territory, but was occupied by the Bulgarian army from 1916 to 1918, and suffered severely from bombardment. After 1922 there was a great influx of Greek immigrants. During World War II it was again occupied by Bulgaria, from 1941 to 1944.

See P. Lemerle, *Philippes et la Macédoine orientale à l'époque chrétienne et byzantine*, vol. i, pt. 1 (1945). (D. M. N.)

SERRANO Y DOMINGUEZ, FRANCISCO, DUQUE DE LA TORRE (1810–1885), Spanish general and statesman, a leader of the Revolution of 1868 and regent of Spain from 1869 to 1870, was born at San Fernando, Cádiz, on Oct. 17, 1810. He joined the army at the age of 12 and during the First Carlist War (1833–39) received rapid promotion. In 1839 he became brigadier and was elected deputy to the *Cortes* for Málaga. Supporting Baldomero Espartero, Serrano voted for him to take over the regency and Espartero made him field marshal (1840). However he later became Espartero's opponent, conspiring with Gen. Juan Prim and Luis González Bravo to drive him into exile in 1843. Serrano was captain general of Granada from 1847 to 1848. In 1854 he supported the *pronunciamiento* of Leopoldo O'Donnell, and was captain general of Cuba from 1859 to 1862, when he was created duque de la Torre.

On his return to Spain in 1863 he became minister of state in O'Donnell's government. He helped to suppress the rising at San Gil, in Madrid, in June 1866 and was rewarded with the Order of the Golden Fleece. In 1867 he succeeded O'Donnell as leader of the Union Liberal Party.

Serrano was deported to the Canary Islands in July 1868 for plotting against Isabella II, but in September he returned to take part with Adm. Juan Topete y Carballo and Prim in the rising at Cádiz, defeating Gen. Manuel Pavía y Lacy at Alcolea later in the month. Isabella was exiled and Serrano formed a provisional revolutionary government. In 1869 the *Cortes* elected him regent. After Amadeo of Savoy was proclaimed king of Spain (November 1870), he entrusted Serrano with the task of forming a new government.

On Amadeo's abduction (February 1873) and the establishment of the First Republic, Serrano went into exile at Biarritz, returning to Spain shortly before the *coup d'état* of Gen. Manuel Pavía

y Rodríguez de Alburquerque in January 1874. Serrano was elected president of the executive and concerned himself with the pacification of the north of Spain. The *pronunciamiento* of December 1874, which restored the Bourbons in the person of Alfonso XII, took Serrano by surprise. Excluded from the government, he formed the left-wing Liberal group. He died in Madrid on Nov. 26, 1885.

See The Marqués de Villa Urrutia, *El General Serrano* (1929).
(R. S. LL.)

SERTORIUS, QUINTUS (*c.* 123–72 B.C.), Roman statesman and military commander of genius, leader of an effective revolt against the Roman Senate and for eight years supremely powerful in Spain, was born at Nursia (modern Norcia), in central Italy. After acquiring some reputation in Rome as a jurist and orator, he fought in Gaul against the invading Cimbri and Teutones (105 and 102), and in 97 he served in Spain. In 90 he was quaestor in Cisalpine Gaul.

In the struggle between L. Cornelius Sulla (*q.v.*), and Gaius Marius (*q.v.*), Sertorius supported Marius (*see* also ROMAN HISTORY: *The Social War and Sulla's Dictatorship*). He was praetor in 83, and in 83 or early 82 escaped to his appointed province of Nearer Spain. When Sulla sent two legions against him he retreated to Africa. In 80 he returned to Spain, where he was joined by many Roman refugees and deserters and by Spanish volunteers. With this army he successfully fought the Sullan governor of Farther Spain, Q. Metellus Pius.

For some years Sertorius ruled Spain as in effect an independent state. He was in league with the Mediterranean pirates and negotiated with Mithradates VI of Pontus, as fellow enemies of the Roman government. In 77 he was joined by M. Perperna and other supporters of the unsuccessful rebel Lepidus (*q.v.*), but in the same year Pompey (*see* POMPEIUS: *Gnaeus Pompeius Magnus*) was sent to Spain with an army. During the next two years Sertorius demonstrated his skill as a strategist and tactician against both Pompey and Metellus. In 74, however, the tide turned against Sertorius. The Iberians lost heart, he was driven to take harsh measures to maintain order, and his formerly great popularity declined. In 72 he was murdered in a conspiracy headed by Perperna.

Sertorius was in revolt perhaps less against Rome than against the constitution Sulla had imposed on Rome. His object in Spain was to build up a stable government with the cooperation of the people, to whom he wished to bring Roman civilization. He established a Senate of 300 members, drawn mainly from Roman emigrants but probably with a few Spaniards, and his bodyguard was composed of Spaniards. Strict and severe with his soldiers, he was considerate to the people. He was accompanied everywhere by a white fawn, which was supposed to communicate to him the advice of the goddess Diana and which promoted his popularity among the superstitious tribesmen.

BIBLIOGRAPHY.—*Ancient Sources* (chiefly Plutarch): A. H. J. Greenidge and A. M. Clay, *Sources for Roman History 133–70 B.C.*, 2nd ed. rev. by E. W. Gray (1960). *Modern Works:* A. Schulten, *Sertorius* (1926); R. Gardner in *Cambridge Ancient History*, vol. ix, ch. 7 (1932); P. Treves, "Sertorio," *Athenaeum*, nuova serie, vol. x, 127–147 (1932); E. Gabba, "Le origini della Guerra Sociale e la vita politica romana dopo l'89 a.C.," *ibid.*, vol. xxxii, 41 ff., esp. 293 ff. (1954).

SERUM THERAPY. Serum, the liquid cell-free fraction of the blood, contains among other chemicals proteins called globulins. Linked to the globulins are antibodies, which are chemical fractions elaborated by the body in response to the presence of foreign proteins (antigens), some of which are toxic. The purpose of antibodies is to neutralize agents which are harmful to the organism.

The chemical structure of any toxic material is well defined and the corresponding antibody produced by the organism is highly specific, chemically exact, and limited in its action to the substance against which it is formed. Serum therapy, broadly defined, is the administration of serum to a sick person or animal to provide one or more fractions that are lacking or deficient; in the limited, classical definition it is the injection of serum for the beneficial effect of the antibodies on infection in the patient. In this sense serum therapy is also called passive immunization, since antibodies formed in the blood of another person or animal are merely transferred. It differs from vaccine therapy (*q.v.*), or active immunization, in which benign doses of toxins or of disease organisms (alive or dead) are introduced into the bloodstream to stimulate antibody production.

History.—Paul Ehrlich (*q.v.;* 1854–1915) developed the active-immunization theory. In 1890 Emil von Behring (*q.v.*) applied and extended this basic principle of immunology to prepare diphtheria antitoxin. This was done by repeated injections of diphtheria toxin (a poison elaborated by the diphtheria organism) into horses. The vaccination provoked formation of antibodies against the diphtheria toxin. Vaccinated horses were then bled and serum was prepared. Von Behring's theory that the administration of antibodies in the serum should neutralize the toxin in the tissues of diphtheria patients was quickly confirmed by the results. Treatment with the antitoxin had a dramatic effect on persons who contracted this severe disease: mortality declined from more than 33 to less than 5%. Also, children exposed to diphtheria infection but treated with serum escaped the disease. The serum thus proved to have prophylactic value, although the protection thus conferred was only temporary and each separate exposure required an injection.

The success of diphtheria antitoxin prompted scientific workers to attempt to produce antibody serums (antiserums) against other diseases. Most of the early efforts ended in failure. Over the years, however, antiserums were developed and used with moderate success in treating tetanus, epidemic meningitis, gas gangrene, scarlet fever, pneumonia, botulism, and especially poisoning from snakes and insects. Some serums, prior to the development and extensive use of good vaccines, were widely used for prophylaxis. This was especially true of diphtheria and tetanus antitoxin. In general, antiserums against bacterial toxins (antitoxic serums) were the most effective, whereas antiserums to combat bacteria themselves (antibacterial serums) were relatively ineffective.

Effective vaccines and chemotherapy have served to reduce the need for therapeutic or prophylactic antiserums. Widespread immunization against diphtheria, tetanus, and pertussis (whooping cough) so greatly reduced the incidence of these diseases that the need for serum therapy became very limited. Even the effective and historic diphtheria antitoxin was, in the 1960s, required only occasionally.

Types of Serum and Their Uses.—Serum prepared from the blood of healthy adult human beings is called normal human serum and is used principally in shock or in cases of protein deficiency. Serum is also prepared from blood of persons lately recovered from certain acute infections; *e.g.*, scarlet fever, measles, mumps and poliomyelitis. This is called human convalescent serum and is used to a variable extent for prophylaxis or treatment of the specific disease. The effectiveness of human convalescent serum depends on the fact that antibodies resulting from infection are at a peak level in the circulation soon after recovery from the illness.

Serum is also prepared from adults who have been intensively treated with vaccines; this human hyperimmune serum finds principal use in the prevention and treatment of pertussis and mumps. Human serums are generally preferred to animal serums because their injection into patients is usually free from the undesirable reactions (serum sickness) sometimes encountered with the use of animal serums. However, for obvious reasons, supplies of human serum are definitely limited, especially in the case of convalescent or hyperimmune serums. One of the fortuitous developments with the use of human serum was a method, evolved by Edwin J. Cohn, of fractionating the blood proteins in pure, concentrated, and stable form. One of the protein fractions, gamma globulin, found valuable use as a preventive of measles, in the same manner as measles convalescent serum. Since the supply of human gamma globulin is much greater than that of convalescent serum, it is widely used to prevent the disease or modify its severity.

Animal serums, prepared from the blood of horses or rabbits, are of the hyperimmune variety. The animals are treated with an intensive and repeated course of vaccination of the specific agent

such as diphtheria toxin, tetanus toxin, cobra venom, etc. When laboratory tests indicate that a satisfactory level of antibodies has been elaborated, blood is collected, from which serum is prepared. Such serum is further chemically processed, concentrated to make it more effective, and purified to reduce those alien substances in the animal serum which cause undesirable reactions.

A broader use of serum as an agent in combating shock from hemorrhage, trauma, burns, etc., was developed in 1940. Human serum and plasma (serum being the liquid fraction of clotted blood while plasma is the liquid fraction from unclotted blood and contains the clotting elements, prothrombin and fibrinogen) found extensive use for this purpose and was considered one of the three or four great medical weapons of World War II. The value of serum or plasma in shock or hemorrhage is that it provides the fluid and proteins to the circulation so that the vital tissues can be adequately oxygenated by the oxygen-bearing red blood cells. Illnesses accompanied by or resulting in a deficiency of circulating blood protein are also treated with human serum or plasma.

BIBLIOGRAPHY.—J. A. Kolmer and Louis Tuft, *Clinical Immunology, Biotherapy and Chemotherapy* (1941); S. O. Levinson and A. M. Wolf, "Human Serum and Plasma: Their Application in Medicine," *Med. Clns N.Am.* (Jan. 1943); Bret Ratner, *Allergy, Anaphylaxis and Immunotherapy* (1943); S. O. Levinson and H. J. Shaughnessy, "Immune Serum in the Prophylaxis and Treatment of Virus Diseases," *Med. Clns N.Am.* (Jan. 1946); E. A. Kabat and M. M. Mayer, *Experimental Immunochemistry* 2nd. ed. (1961); E. L. De Gowin, R. C. Hardin, and J. B. Alsever, *Blood Transfusion* (1949).

SERVETUS, MICHAEL (MIGUEL SERVETO) (1511?–1553), Spanish theologian and physician, whose execution as a heretic at Geneva led to severe criticism of John Calvin, was born at Tudela or Villanueva in 1511 (or, less likely, 1509). He studied law in Toulouse and on his own also the Bible, with special reference to the Trinity. In Feb. 1530 he accompanied his patron, the Franciscan John de Quintana, to the coronation of the emperor Charles V at Bologna; but he was so distressed by the papal pomp and by the emperor's deference to the worldly pope that he left Quintana and visited Lyons, Geneva and Basel. At Basel he met John Oecolampadius and discussed the trinitarian problem with him. He moved on to Strasbourg, where he met Martin Bucer, exchanged views with Kaspar von Schwenckfeld on the celestial flesh of Christ, and became acquainted with the local Anabaptist community. He published his new ideas on the Trinity in *De Trinitatis erroribus libri vii* (1531), attacking the orthodox teaching and attempting to reconstruct a view of his own, namely, that the Word is eternal, a mode of God's self-expression, whereas the Spirit is God's motion or power within the hearts of men. The Son is the union of the eternal Word with the man Jesus. Servetus' bold and scarcely lucid speculative effort made him odious to both Catholics and Protestants, who found themselves better informed on what he opposed than on what he was trying to formulate. He thereupon published a revised formulation, *Dialogorum de Trinitate libri ii* (1532).

Moving to Lyons under the name Villanovanus, Servetus edited scientific works and published a translation of Ptolemy's *Geography*. Around 1534 a rendezvous was arranged with John Calvin in Paris for the purpose of discussing several theological questions, but Servetus failed to arrive; the two were not to meet until nearly a score of years later. Servetus' writings in this period were a defense (1536) of his friend Leonard Fuchs, the botanist, a book significant because it shows Servetus' growing interest in medicine and physiology; and a treatise, *Syruporum universa ratio*, published in Paris in 1537, the year he moved there. He also published (1538) a work on astrology, became interested in the influence of the stars on health and was subjected to an attack by the Paris medical faculty on account of it. He matriculated in medicine at Paris in 1538 and was associated with Vesalius.

After brief residences in Louvain, Avignon and Charlieu, Servetus became physician to the archbishop in Vienne, where he worked also as an editor; while outwardly a conforming Catholic, he pursued his private theological studies and was presumably rebaptized in 1541. He later asserted that, in imitation of Christ, every believer should be immersed at the age of 30. About this time he wrote *Declarationis Jesu Christi filii Dei libri v* (until recently ascribed to an otherwise unidentified Alphonsus Lyncurius Terraconensis). This comprehensive work anticipates his magnum opus. In 1542 his *Biblia sacra ex Santis Pagnini tra[ns]latione* appeared at Lyons, remarkable for its theory of prophecy in the notes.

In 1546 Servetus opened a fatal correspondence with Calvin, forwarding the manuscript of an enlarged revision of his ideas, the *Christianismi Restitutio*, and expressing a desire to confer with the reformer in Geneva. After the first few letters, Calvin would have nothing more to do with him, kept the manuscript, and said to Guillaume Farel that if Servetus should ever come to Geneva he would not permit him to leave alive. Servetus rewrote the manuscript and had an edition of 1,000 copies secretly printed at Vienne (1553). It was in this comprehensive, systematic theology, in connection with the problem of the relationship of the Spirit and regeneration, that he almost incidentally made known his discovery of the pulmonary transit of the blood. (*See* HARVEY, WILLIAM: *Harvey's Work on the Circulation.*) The assumption of the book is that both God the Father and Christ his Son had been dishonoured by the Constantinian promulgation of the Nicene formula (*see* CREED: *Nicene Creed*), which obscured the redemptive role of Christ and brought about the fall of the church. Servetus, not without a streak of vainglory, felt that he could restore the church by separating it from the state and employing in theological formulations and church practice only that which could be proved from Scripture and the pre-Constantinian Fathers.

A former citizen of Lyons then living in Geneva (Guillaume de Trye) secured some of the letters written by Servetus to Calvin and exposed him to the inquisitor-general at Lyons. Servetus and his printers were seized. Servetus was able to contrive his escape during the trial, and the Catholic authorities had to be content to burn him in effigy. He quixotically appeared in Geneva, was recognized and arrested. His trial, in which Calvin took a prominent part, lasted from Aug. 14 to Oct. 25, 1553, and involved correspondence with the other Reformed Swiss churches in search of judicial and theological counsel. Their spokesmen agreed that Servetus had to be punished as a heretic, but they did not mention the death penalty. Servetus was found guilty mainly on two theological charges (the Trinity and baptism) and was sentenced and burned alive at Champel, Oct. 27, 1553. Calvin had pressed for execution, although by beheading rather than by fire.

The denial by Servetus of three equal persons in the Godhead, along with his belief in the celestial flesh of Christ, in Anabaptism and in psychopannychism (the sleep or even the death of the soul with the body), resulted in his condemnation by both the Catholics in Vienne and the Protestants in Geneva, despite his intense Biblicism, his passionate devotion to the person and work of Christ, and indeed his entirely Christocentric scheme of the universe. What Calvin regarded as pantheism can be best understood as a combination of revealed and natural theology in which the incarnation of the Word was paradigmatic of the process by which creative Light was ever penetrating matter to form created things. Among the few immediate followers was Peter Gonesius in Poland, who seems to have introduced the baptismal theology of Servetus, with its stress on immersion and regeneration, into Polish Antitrinitarianism. His execution produced a Protestant controversy on the death penalty for heresy, in which, among others, Theodore Beza (for) and Sébastien Castellio (against) took part.

BIBLIOGRAPHY.—E. M. Wilbur, *A History of Unitarianism*, vol. i, ch. v, ix–xii (1945); R. H. Bainton, *Hunted Heretic* (1953); B. Becker (ed.), *Autour de Michel Servet et de Sébastien Castellion* (1953); W. Emde, "Michael Servetus als Renaissancephilosoph und Restitutionstheologe," *Zeitschrift für Kirchengeschichte*, vol. l, 96–131 (1941); G. H. Williams, *The Radical Reformation* (1962); "Le procès de Michel Servet" in A. Dufour and H. Kingdon (eds.), *Les Registres de la Vénérable-Compagnie* (1962); his two early works on the Trinity are available in translation by E. M. Wilbur, *Two Treatises on the Trinity*, Harvard Theological Studies, vol. xvi (1932), and his geographical, medical and astrological writings by C. D. O'Malley (1953). (G. H. Ws.)

SERVICE, ROBERT WILLIAM (1874–1958), Canadian poet and novelist known for his ballads of the Yukon, was born at Preston, Eng., on Jan. 16, 1874, and educated at Hillhead Public School, Glasgow. He went to Canada in 1894 and settled

for a short time on Vancouver Island. He entered the Canadian Bank of Commerce in Victoria, B.C., and was afterward transferred first to Whitehorse in the Yukon and then to Dawson. In all he spent eight years in the Yukon and traveled widely. During his last years with the bank, he wrote verse describing life in the north, notably *Songs of a Sourdough* (1907) and *Ballads of a Cheechako* (1909), both of which were enormously popular. In 1910 appeared a novel, *The Trail of '98*, giving a vivid description of men and conditions in the Klondike. During the Balkan War of 1912–13, Service was war correspondent to the *Toronto Star*. He served this paper in the same capacity during World War I, in which he spent two years as an ambulance driver in the Canadian Army medical corps. He described his war experiences in *Rhymes of a Red Cross Man* (1916). His other early works include *Rhymes of a Rolling Stone* (1912); *Ballads of a Bohemian* (1920); and *The Roughneck* (1923). He returned to Victoria for a time during World War II but later lived in retirement on the French Riviera, where he died on Sept. 11, 1958. Two autobiographical volumes, *Ploughman of the Moon* and *Harper of Heaven*, appeared in 1945 and 1948 respectively. His later collections of ballads include *Songs of a Sun Lover* (1949); *Lyrics of a Low Brow* (1951); *Rhymes of a Rebel* (1952); *Songs for My Supper* (1953); *Carols of an Old Codger* (1954); and *Rhymes for My Rags* (1956). (A. J. M. S.)

SERVICE CLUB, an organization, usually composed of business and professional men or women, that promotes friendliness among its members and is devoted to the principle of service to society. The typical service club combines fellowship and volunteer community service with emphasis on good citizenship and high ideals. All have high-principled mottoes and creeds such as Rotary's "Service Above Self." About half of these clubs are in semirural communities and have memberships of less than 50. Most hold a luncheon or dinner meeting each week. The larger service organizations publish magazines that report their activities and also carry articles of general interest by well-known writers. Many of the service organizations have clubs in more than one country and endeavour to promote international goodwill. They take pride in the international spread of their influence and practices, and they show an awareness and an understanding of conditions in countries in which they have common fellowship.

The term has an entirely different meaning in the U.S. armed forces. A service club in military terminology is a social and recreational club for enlisted personnel.

The civilian service club idea originated in 1905 with Paul P. Harris, a young attorney in Chicago, Ill. His plan of organization envisioned all the essential features of the present-day service club and included the classification principle that restricts membership in a given club to a quota from each business or profession. Because meetings were to be held in rotation in members' offices, Harris proposed the name Rotary. Growth was rapid and in 1907 community service was made a part of the club activities. In 1908 the Rotary movement spread to California and a club was formed in San Francisco. With the formation of clubs in Winnipeg, Can., Dublin, Ire., and London, Eng., the name International Association of Rotary clubs was adopted but was replaced in 1922 by the name Rotary International. By 1912 there were 50 Rotary clubs with 5,000 members, and the service-club movement was well under way. Fifty years later there were more than 8,000 Rotary clubs with a total membership exceeding 400,000.

Other clubs of somewhat similar design followed in rapid succession. In 1911 in Detroit, Mich., a group of businessmen who for many years had been meeting informally at luncheon organized the first Exchange club. Others followed this example and in 1917 the National Exchange club came into being, adopting as its motto "Unity for Service."

In 1912 in Kansas City, Mo., a physician, George W. Smith, finding his classification in the Rotary club already filled, founded Sertoma (International), a name derived from the slogan "Service to Mankind." Gyro (International) came into being in the same year in Cleveland, O., taking as its emblem the gyroscope, signifying stability.

In 1915 Kiwanis (International) was organized in Detroit, the name being derived from Indian lore suggesting self-expression. Its purpose is to promote "the adoption and application of higher social, business, and professional standards" and the development of "intelligent, aggressive and serviceable citizenship." Kiwanis clubs may select two members from each business or profession. The organization has expanded its influence by sponsoring Key club (International), a service club for high school boys, and Circle K (International), a similar organization for college men.

The International Association of Lions clubs was organized in Dallas, Tex., in 1917. Lions clubs adopted more lenient rules for admission and did not follow the classification principle rigidly. As a result, their membership increased rapidly and the Lions club soon became the largest of the service club organizations. By the 1960s there were more than 16,000 Lions clubs with approximately 660,000 members. The organization emphasizes good citizenship, patriotism, education, safety, aid to the blind and support of the United Nations.

In 1919 a group of independent clubs met in Louisville, Ky., and formed Optimist (International), pledged to developing optimism as a philosophy of life. In 1923 the Optimists decided to direct their charitable work toward helping boys.

Civitan (International), with its motto "Builders of Good Citizenship," started in Birmingham, Ala., in 1920 and was dedicated to displaying the Golden Rule in action.

The story of King Arthur inspired the naming in Oakland, Calif., in 1922 of the Loyal Knights of the Round Table (now Round Table International), which has for its motto: "He who seeks to serve another, best serves himself."

In 1928 Ruritan (National) was founded in Holland, Va., with the specification that its membership be one-third rural, one-third business and professional men, and one-third chosen from either category. It uses the service-club technique to encourage a closer union between farmers and men in business and the professions. Most of its clubs are located in towns of 5,000 population or less.

Cosmopolitan (International) was organized in Oklahoma City, Okla., in 1933.

The year 1922 marked the founding in the United States of two young men's service clubs—Active (International) and 20–30 (International), both of which restricted membership to approximately the same (20–40) age group. Round Table (International), founded in 1926 in Norwich, Eng., grew rapidly and formed clubs in Europe, Asia, Africa and New Zealand. These three international organizations together with the Association of Kin(smen) clubs (Canada) and the Association of Apex clubs (Australia), grouped themselves in 1945 into the World Council of Young Men's Service clubs.

In 1960 Active and 20–30 combined to form Active 20–30 International; at the 1960 World council meeting Round Table International was disbanded, and each country entered the council as a national association.

The service-club movement is not limited to men's organizations; there are also several women's service clubs. Among these the pioneer is Altrusa (International), organized in Nashville, Tenn., in 1917 "to help solve community and world problems and promote international understanding." Following Rotary's example, it admits to each club only one member of a specific business or profession. Quota club (International) owes its origin in 1919 to the Kiwanis club of Buffalo, N.Y. It is a civic service club of women holding executive positions or owning their own businesses. Zonta (International), a classified civic service organization of women in business and the professions, was founded in 1919. In the same year the National Federation of Business and Professional Women's clubs was founded. It grew out of the work women performed during World War I and is the largest of the women's service organizations. The name of the American, European and British federations of Soroptimist clubs is compounded from the Latin words *soror* meaning "sister" and *optima* meaning "best." Pilot club (International) was formed in 1921 with the aid and counsel of the Rotary club of Macon, Ga.

It is estimated that each year service clubs carry on several hundred thousand local community projects, ranging in complexity from sponsoring Easter-egg hunts to building and equipping hos-

pitals or summer camps for underprivileged children. Most service organizations support youth activities. Exchange clubs conduct a nationwide model-airplane competition for teen-agers. Lions clubs operate a school for guide dogs for the blind. Rotary International has one of the most ambitious international programs; it administers a multimillion-dollar fund to provide scholarships for graduate students to study in countries other than their own and to become better acquainted with people of other lands. A Kiwanis publication visualizes the service club as "only the first step toward what may well be complete understanding and cooperation among men—the millennium of permanent peace and universal good will."

BIBLIOGRAPHY.—P. P. Harris, *My Road to Rotary* (1948); C. R. Hewitt, *Towards My Neighbour* (1950); John Creasey, *Round Table: the First Twenty-five Years of the Round Table Movement* (1953); C. F. Marden, *Rotary and Its Brothers* (1935). (E. B. R.)

SERVITES (SERVANTS OF MARY; ORDO FRATRUM SERVORUM SANCTAE MARIAE; O.S.M.), an order of mendicant friars of the Roman Catholic Church founded by St. Bonfilius of Florence and six fellow merchants in 1233. These men were canonized in 1888 and are invoked collectively as the "Seven Holy Founders" (feast Feb. 12). The first approbation of the new order was given by the cardinal legate in Tuscany in 1249; final papal approval was granted by Benedict XI in 1304.

The seven founders of the order first withdrew outside the gates of Florence to live a life of witness based on a literal interpretation of the gospel: a life of poverty and fellowship. After 1240 they departed for Monte Senario about 12 mi. from the city where they continued their penitential life and laid the eremitical foundations of the order. The Rule of St. Augustine was adopted (*see* AUGUSTINE OF HIPPO, RULE OF ST.), and the earliest extant redaction of the Servite constitutions shows heavy Dominican influence.

The order was saved from suppression and greatly propagated by its fifth prior general, St. Philip Benizi (1233–85). The eremitical life was established on Monte Senario about 1410 and later developed into a distinct congregation. Soon after this a vigorous reform movement began in northern Italy; in 1614 a far-reaching Germanic reform was begun at Innsbruck by the Venerable Anna Juliana, archduchess of Austria.

The characteristics of the order can be summed up in the two main religious currents that preceded its birth, the monastic and eremitical, while to these the founders added the life of a friar under the patronage of St. Mary, the Mother of God. The service of the order to the church is carried out both within the monastery and by apostolic works on six continents. There are also sisters who share in the active works of the order and nuns devoted entirely to prayer within the monastic enclosure. The friars' garb consists of black tunic, scapular, and cowl, with a rosary of the seven sorrows of Mary. The general motherhouse is at Rome.

See also ORDERS AND CONGREGATIONS, RELIGIOUS.

BIBLIOGRAPHY.—*Sources:* A. Morini *et al.* (ed.), *Monumenta Ordinis Fratrum Servorum Sanctae Mariae,* 20 vol. (1897–1930). *History:* A. Giani and L. Garbi, *Annales Ordinis Servorum,* 2nd ed., 3 vol. (1719–25); A. Rossi, *Manuale di Storia dell'Ordine dei Servi di Maria* (1956). *Summary of Actual State:* J. Besutti (ed.), *Ordo Servorum Mariae,* (1956); *Acta Ordinis Servorum B. V. M.* (1916 *et seq.*). (J. M. RY.)

SERVIUS (fl. *c.* A.D. 400), whose other names sometimes appear in manuscripts as Marius or Maurus and Honoratus, was a Latin grammarian, commentator, and teacher and author of a valuable commentary on Virgil. As an *adulescens* he was one of the speakers in the *Saturnalia* of Ambrosius Theodosius Macrobius (*q.v.*) and at least the greater part of his life was spent in Rome.

His commentary on Virgil is extant in two versions, a longer and a shorter. The longer and anonymous version, first printed in 1600 in an edition by Pierre Daniel, consists of Servius' own work—somewhat altered—in which he sought to meet the needs of schools and paid special, but not exclusive, attention to grammatical and stylistic points. With it are incorporated some valuable additions, in the main from a commentary—perhaps those parts of the commentary by Aelius Donatus which were not used by Servius—which mostly concern Virgil's rhetoric, mythology, and subject matter. These are a precious source of knowledge about Roman antiquities. They presuppose that Virgil had an exact knowledge of ancient Roman customs and institutions.

Servius was a pagan, convinced that Virgil represented the highest truth. He was a learned man and a capable verbal expositor; but he was not a literary critic. His work is interesting as an example of 4th-century exegesis for schools.

An *Explanatio in artem Donati* and three unimportant works—*De centum metris, De finalibus* and *De metris Horatii*—also go under Servius' name.

BIBLIOGRAPHY.—Modern editions of the Virgilian commentary by G. Thilo (1881–1887); E. K. Rand *et al.*, vol. 2 (1946). For minor works *see* H. Keil, *Grammatici Latini,* vol. iv (1864). *See* also E. Thomas, *Essai sur Servius et son commentaire sur Virgile* (1880); Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft,* ii, 2, col. 1834–1848 (1923). (G. B. A. F.)

SERVIUS TULLIUS, sixth king of Rome (traditionally 578–534 B.C.). In legend he was slave-born in the household of Tarquinius Priscus. He married Tarquin's daughter and succeeded him by the contrivance of his mother-in-law, Tanaquil, who was prophetic and foresaw his greatness (this story seems obviously designed to account for his name, which suggests a servile origin). He was eventually killed by his daughter and her husband Tarquinius Superbus. He is undoubtedly a historical personage, but authorities are divided on his nationality. According to one tradition he is Etruscan, a logical inference from his reign between the two Etruscan Tarquins, Priscus and Superbus. This is reinforced by another legend (late, however) identifying him with the Etruscan Mastarna who is associated in a 4th-century B.C. wall painting from Vulci with the killing of a Roman Tarquin. There is perhaps even stronger evidence, however, in the events of his career to suggest he was Latin. Many of these are fictitious, but several can be accepted as historically accurate. He founded the earliest and most important shrine of Diana, a Latin deity, on the Aventine Hill, and to his reign is assigned a crucial treaty between Rome and the Latin League. These both seem authenticated by the preservation in this temple of the text of the treaty down to the Augustan era.

Servius is also traditionally credited with a revision of the constitution which reclassified the population. The significance of this "Servian Constitution" is much disputed, but it is probably a reading back into the legendary past of reforms which were not effected until a much later date. The Servian Wall around the city of Rome ascribed to this period is likewise much later, dating from the 4th century B.C.

BIBLIOGRAPHY.—Livy, i, 39–48; Dionysius of Halicarnassus, iv, 1–40; Cicero, *De republica,* ii, 21–22; E. Pais, *Ancient Legends of Roman History,* ch. 7 (1906); H. Last in *Cambridge Ancient History,* vol. vii, pp. 390 ff. (1928); W. Hoffmann in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft,* 2nd series, vii, 804–820 (1939). (R. B. LD.)

SERVOMECHANISM, a device which tends to position an object in accordance with the command given by an arbitrarily varying position indicator capable of supplying only a small amount of power. The operation of a servomechanism is dependent upon the difference between the actual position of the object and the desired position. The servomechanism acts to reduce this difference, called the error, despite disturbing influences. A common example is the automatic pilot in which a signal indicating the actual direction and height of the aircraft is compared with a signal indicating the desired direction and height; any difference actuates ailerons and rudder to position the aircraft onto the desired course.

Most servomechanisms do not rely on human beings to carry out their function. A simple servomechanism in which a human being is involved, however, serves to illustrate all the essential elements. When an operator drives a car he becomes part of a servomechanism whose object is to keep the actual path of the automobile following the desired path, the road. As the road changes direction and the desired path of the car changes, the operator senses any difference between the actual path of the car and the desired path. His eyes detect any existing errors, his brain signals his muscles to act and by means of the steering mechanism of the automobile he corrects the actual path to coincide with the desired path. He is able to do this despite disturb-

ing influences in the form of wind or ruts that might try to force the car's direction from the desired path. Even this simple system does not always respond in the same way. For example, drivers have different reaction rates in the detection of errors and apply different magnitudes of correction for the same deviation from the desired path. Automobiles differ in their responses, too, depending upon the steering ratio and size of the load in the car. Actual malfunction of the apparatus, such as that caused by backlash in the steering mechanism or by soft tires, influences the accuracy with which the path is maintained. These analogies also find their counterparts in completely automatic systems.

The need for servomechanisms arises because man is limited in his ability to perceive errors swiftly and precisely enough, to use his muscles to correct them fast enough and to exert enough power to control the enormous masses often involved. Servomechanisms are also used to relieve man of the drudgery of many tasks; they operate tirelessly and, when functioning properly, with great accuracy and smoothness. They are capable of amplifying signals of no more power than the light from a star to the hundreds of horsepower necessary to control large masses.

Applications.—There are many practical applications of servomechanisms: one is the automatic pilot mentioned earlier, another is in the field of machine tools, particularly those used to duplicate the contours of a piece with rounded surfaces of different radii, such as an aircraft propeller. In this application, the master pattern is clamped beside the piece of raw metal from which a replica is to be made. The servomechanism positions a metal-cutting tool in accordance with the movement of a stylus that is passed over every part of the surface of the master pattern while the tool cuts away unnecessary metal. The master pattern can be made of some material more easily worked by hand tools than the metal from which the final replica is made.

Many applications of servomechanisms involve military devices. Numerous interconnected servomechanisms are utilized in tracking and firing on an enemy airplane that is invisible to observers at a missile emplacement below. A radar antenna is pointed toward the airplane guided by servomechanisms that detect any slight angular difference between the direction the antenna is pointing and the location of the airplane. Given the location and rate of travel of the aircraft, as shown by the angular speed of the radar antenna, automatic computers can predict the probable path of the aircraft and the proper direction for a missile to be fired to intercept this path; other servomechanisms are then used to position the missile launcher properly. All of this is accomplished without any human intervention.

Automatic Feedback Control.—Servomechanism systems or follow-ups, as they are sometimes called, are important tools in the field of engineering known as automatic feedback control or automatic regulation, serving that portion of the field concerned with controlling the position of an object. The systems have wide industrial uses in such tasks as the control of pressure, flow, liquid level, temperature, and many combinations of these in such industries as oil refining, chemical production and food processing.

Automatic feedback control of the more general sort is older in application; indeed, nature has utilized this phenomenon for the control of many functions of the human body. There is a body-temperature regulator; if there were none, the heat produced in 20 min. of muscular effort would cause the albuminous substances in the blood to harden like the white of a hard-boiled egg. At the slightest rise in body temperature, a series of operations goes into effect: the blood vessels dilate and constrict, diverting blood from the internal organs to the surface; blood volume is increased by dilution with fluid drawn from the tissues, thus producing further cooling; sweat glands produce more perspiration; the flow of blood is increased by more rapid pumping of the heart; and the respiratory rate increases to provide more air to the lungs. In addition, if there were no acidity regulator, the lactic acid produced from such muscular effort would upset the acidity of the blood and cause convulsions and death.

Common feedback control devices include the household thermostat, the speed governor on a household mixer, the temperature control in an electric iron and the pressure regulator in a household pressure cooker. Such devices are unified with servomechanisms by a common theory of operation even though they are concerned with the control of temperature or pressure rather than position. They all involve a closed loop of causal action; *i.e.*, the existence of an error causes the device to attempt a correction of the error. The result of this correction changes the error which in turn causes the device to act again. In actual operation these actions take place continuously and instantaneously. The only reason an error exists at all in a servomechanism is because the desired state or output is constantly changing and external influences are also tending to change the actual state of the object.

The principal measure of the excellence of performance of a servomechanism is the accuracy with which the output is positioned relative to its desired value; in other words, whether or not the error is within tolerable bounds. A consequence of this requirement is that the system be stable; *i.e.*, that the output not oscillate or "hunt" about its desired value.

These requirements can be illustrated by the example of a large ship at sea being steered in accordance with a desired heading. Assume first that a helmsman is in control whose job is to detect the difference between the actual heading of the ship, as given by a compass reading, and the desired heading, as given by orders from the captain. He is required to do this despite the disturbing influences of wind and waves, the hydrodynamic effect of the water through which the ship is passing and ocean currents that might tend to force the ship from the desired direction. If an error exists, the helmsman rotates his wheel; the ship slowly responds in accordance with his signal, reducing the error. A considerable amount of turning momentum is built up in this process and an inexperienced helmsman may not realize this until the error is reducing rapidly to almost zero. Seeing that the ship is turning through the proper heading, he applies the opposite rudder in an attempt to once again bring the error to zero. The ship slowly responds to his signal again and eventually builds up a considerable amount of momentum to rotate it in the correct direction and reduce the error. The course of the ship might again overshoot its desired course and this process could continue indefinitely unless the skill of the helmsman improved.

Helmsmen learn with experience, however, that small errors should be met with small corrections. They also learn that after they have introduced a rudder angle to correct an error, they must anticipate the tendency to overshoot and bring the rudder back to zero more swiftly than otherwise, or even apply a reverse rudder as the error reduces. By skillful application of this proportionate control and realization of the fact that the error is rapidly reducing and that something must be done before it reaches zero, a pilot may become very skilled, the oscillations practically nonexistent and the accuracy of the heading maintained despite external influences. An automatic pilot, which supplants the human pilot in a ship or aircraft, has the same problems and if properly designed can apply corrections in the same manner as the skillful human pilot.

Basic Components.—Every servomechanism consists of certain basic elements that combine to perform the functions described in the given examples. First of all, there must be an error detector, some device for detecting and measuring the difference between the desired position of an airplane, for example, and its actual position. There must also be a means of controlling the source of power, involving some method of amplifying what may be a tiny signal from the error detector to sufficient power to operate the next element in the chain, the prime mover. In the case of the airplane, the prime mover might be an electrical hydraulic system acting on ailerons and rudder to correct the position of the airplane and bring it into coincidence with the desired position. There must also be some means of modifying the performance of these devices in order to achieve the desired stability and accuracy.

These, then, are the essential elements of any servomechanism. When a human agency is involved, the methods of detection of the error are often the eyes and the brain. A man might also provide some or all of the muscular power needed to move the object to its desired position. His skill, experience and the fine

ness and swiftness with which he can move his hands cause him to supply corrections of the proper amount with proper anticipation in order to prevent oscillations. He may thus achieve the desired performance of stability and accuracy.

In the design of inanimate servomechanisms, components are carefully chosen for their desired application. The error is often detected in electrical form by means of potentiometers or rheostats which can give a voltage signal proportional to the angular difference between their shaft positions. In the case of the automatic pilot aboard ship, the shaft of one of the potentiometers is connected to the ship's compass, the shaft of the other is attached to a dial upon which the desired course can be indicated. The difference between these angular positions of the two shafts appears as an electrical voltage that can be used to control the rudder. Other error-detecting devices rely on magnetic coupling of their parts to produce a voltage proportional to the error.

The means used to control the source of power involves amplification which can be electrical, electronic, hydraulic, mechanical or a combination of these.

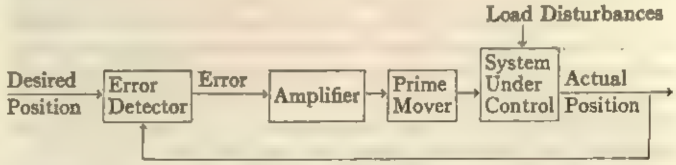
A prime mover supplies the muscles which actually move the object to its desired position. If this object has a small mass, such as in computers or other type instrument applications, the power of the prime mover need not be large, and is commonly a small electric motor. On the other hand, if the mass involved is several hundred tons, a large electric motor or large hydraulic pump and motor might be used to position the mass.

To make the servomechanism satisfactorily stable, some kind of process that damps any oscillations must be employed. Damping can be achieved electrically by the introduction of networks in the electronic amplifier or by feeding back into the amplifier a signal from a generator or other device that measures the speed of the object being positioned. Knowing the speed with which an object is approaching the zero error position, the amplifier can combine this information with a measurement of the size of the error and achieve stability in the same manner that the skilled helmsman anticipates the motion of the ship in response to his correcting forces.

Performance in a servomechanism is measured by comparing its accuracy to the accuracy required of the system in which it is to be used. Some systems do not require a high order of accuracy and may be required to follow desired angular motions only within an accuracy of 1°. However, other applications such as the positioning of optical mirrors in gunfire control may require the mirrors to be aimed within one minute of arc, $\frac{1}{21,600}$ part of a circle. For such an application, the components of the system have strict requirements.

For one thing, the error detector must be extremely accurate because the servomechanism is relying upon the error detector as its only means of knowing of an existing error in the position of the mirrors. The amplifier must likewise be capable of amplifying very minute error signals without distortion. These signals are very small because the servomechanism, if it is performing its job well, has a very small error and hence the signal from the error detector which is proportional to the error is also small. The error signals must be amplified many thousand-fold to be large enough to control a source of power such as an electric motor; the motor must be extremely responsive to signals from the amplifier. Some of the gearing used to connect the high-speed shaft of the motor to the relatively low-speed shaft of the mirror must be of a high order of accuracy so as not to introduce error into the system. Ingenious stabilizing devices must be used if the mirror is to follow its desired angle accurately and without oscillation.

Analysis and Design.—The analysis and design of servomechanisms is often carried out by the use of a block diagram which allows the designer to think more readily in terms of operational characteristics of the system rather than of the wires, vacuum tubes, hydraulic components and other elements involved in it. The block diagram is a graphical representation of the flow of information and the functions performed in the system. The designer can draw the block diagram by studying the operation of each element in order to recognize its function and by analyzing the relationship of the elements to learn how the information flows between them.



Block diagrams are also useful in learning about the fundamentals of servomechanisms. In the accompanying diagram, there are two bits of information of concern: the desired position of the object under control and its actual position. These two kinds of information flow into the error detector whose function is to measure their difference and convert it into a usable form. The signal emanating from the error detector is the error; this is fed into the amplifier and then to the prime mover to effect the proper position of the system under control.

From such a block diagram, it is easy to see the closed loop of causal action in which continuous correction is taking place to reduce the error. It is also seen that any modification of the characteristics of any element in the chain can cause modification of the operation of the entire system. More complicated servomechanisms might be represented by block diagrams having many interconnected loops in which information is fed from various portions of the system to other elements.

In addition to experimental techniques, a number of mathematical methods are used to analyze and design servomechanisms. The most commonly employed are the applications of differential equations and the use of operational calculus and frequency-response studies. While mathematics does not give the final answer and adjustments must always be made experimentally, the use of mathematical analysis can greatly shorten the time required to design a servomechanism and indicate the characteristics of components necessary to achieve the required accuracy.

To solve a problem in servomechanism design or analysis, simultaneous differential equations are written relating the various signals in the servomechanism loop. These can then be solved by any of the standard techniques. Often the servomechanism is a linear one or can be considered so for the purpose of analysis. Under these conditions, the solution of the differential equations involves the elementary techniques for solving linear differential equations with constant coefficients (*see* DIFFERENTIAL EQUATIONS, ORDINARY).

A more flexible approach employs the use of operational calculus, the most popular form of which is the Laplace transformation. This enables derivatives, integrals and trigonometric functions to be expressed algebraically in terms of functions of a complex variable. Such a technique permits the dynamics of elements of a servomechanism to be described in terms of algebraic quantities called transfer functions. The characteristics of the whole system are, therefore, simply the combination of the transfer functions of each of the various elements. The designer can then consider the characteristics of each element separately and how each affects the performance of the entire servomechanism loop.

The most flexible technique of design analysis is the use of frequency-response studies. In such a method of analysis, a study is made of the response of each element of the system to sinusoidal signals whose frequencies vary from zero to infinity. When the response to these signals is known, the performance of the servomechanism can be predicted for any type of variation of the desired position or load disturbances after the methods of Fourier. The use of this method of analysis permits the servomechanism designer to utilize the entire field of functions of a complex variable which enables the stability and accuracy of the system to be predicted and the characteristics of any necessary corrective networks to be set forth.

In addition to the problem of performance of servomechanisms, there are other requirements which are of concern to the user. These include the expected life of the apparatus, its serviceability, its reliability under all operating conditions, the workmanship of

its manufacture, restrictions on materials that can be used and economic considerations. There might be other requirements as well concerning the weight, size and resistance to shock and vibration of the servomechanism. Other constraints on the designer are the type of electrical power available, its voltage and frequency variations and the temperature and atmospheric conditions surrounding the servomechanism. All these factors, together with the mathematical considerations given above, give some hint of the complexity of servomechanisms and their design.

See also AUTOMATION; INFORMATION THEORY. (W. R. AH.)

SESAME (BENNE) was one of the first oilseeds grown by man. It probably originated in eastern Africa and later spread to most of the tropical, subtropical, and southern temperate zone areas of the world. Technically *Sesamum indicum*, sesame is an

BY COURTESY OF U.S. DEPARTMENT OF AGRICULTURE

SESAME (SESAMUM INDICUM): (LEFT) GROWING PLANTS AND (RIGHT) PODS AND SEEDS

erect annual plant of many types and varieties. Depending to some degree on the conditions of growth, varieties grow from two to ten feet tall with growth cycles of two to six months; some have branches, others none. One to three attractive flowers (and seed capsules) are borne in the leaf axils. Flowering begins as early as six weeks after planting, continuing until maturity. The white or black seed contains from 45 to 63% edible oil and from 16 to 32% protein. The oil is noted for its stability (resistance to oxidative rancidity). The meal, or oil cake, remaining after oil extraction is a rich source of protein (especially the amino acid methionine), calcium, phosphorus, and the vitamin niacin; it is used chiefly as a cattle feed.

Sesame is used as a salad or cooking oil, in shortening and margarine, and as a carrier for fat-soluble pharmaceuticals. The whole seed is used in a wide variety of confections and as a garnish for bakery goods.

In normal sesame, capsules open when dry, allowing the seed to scatter, making necessary considerable hand labour in harvest to prevent seed loss. Discovery of nonscattering mutant in 1943 led to the possibility of complete mechanized production of this crop. (M. L. K.)

SESOSTRIS, the Greek name of a legendary king of Egypt. According to Greek traditions preserved by Herodotus, Diodorus Siculus (who calls him Sesoosis), and Strabo, he conquered the whole of Asia, divided Egypt into administrative districts or nomes, was a great lawgiver, and introduced a system of caste and the worship of Sarapis. Manetho names him among the kings of the 12th dynasty. During this prosperous period of Egypt's Middle Kingdom, three kings in fact bore the name, the Egyptian form of which was Senusret or Senwosre (formerly read Usertesen).

SESOSTRIS (Senusret) I (reigned 1971–1928 B.C.) inaugurated a policy of conquest in Nubia and extended the frontier to the Second Cataract of the Nile; he engaged in considerable commercial and diplomatic activity in western Asia, opened new mines and quarries, and began to exploit the resources of the Fayyum. He erected monuments to the gods in sites all over Egypt.

SESOSTRIS (Senusret) II (1897–1878 B.C.) is Manetho's Sesostris; little is known of him. The ruins of the town built to house the men working upon his pyramid at Illahun in the Fayyum have produced unique information about the domestic architecture of ancient Egypt and some important legal documents.

SESOSTRIS (Senusret) III (1878–1843 B.C.) was the successor of Sesostris II; under him the dynasty attained the height of its prestige and prosperity. Sesostris III completed the subjugation of Nubia and built border forts at Samnah (Semna), at the southern end of the Second Cataract; he led an expedition into Palestine and subjugated the "sand-dwellers"; it is probable that he reorganized the administration of Egypt, abolishing the excessive power of the regional governors (nomarchs) and centralizing authority under his vizier. In Nubia he was later worshiped as a god.

The memory of this splendid age, in which Egypt's prestige had been higher than ever before, remained associated with the name of its greatest monarchs long after the true scale of their achievement had been forgotten, and the legend of the heroic "Sesostris," embodying also memories of that later and greater age of empire builders, the 18th dynasty, was fostered by a people who, at the hands of the Assyrians and the Persians, had suffered humiliation and defeat. *See* also EGYPT: *History*.

BIBLIOGRAPHY.—W. C. Hayes, "The Middle Kingdom in Egypt," *Cambridge Ancient History*, vol. i, fasc. 3, rev. ed. (1961); Herodotus ii, 102–111; Diodorus Siculus i, 53–59; Strabo xvii, i, 35 ff.
(M. S. DR.)

SESSA AURUNCA, a town of Campania, central Italy, in the province of Caserta, lies on a lava deposit of the extinct Roccamonfina volcano, 33 mi. (53 km.) NNW of Naples. Pop. (1961) 27,941 (commune). The town, built on the site of the ancient Suessa Aurunca, has Roman ruins which include a large cryptoporticus (a private gallery) and a bridge with 21 arches. The 13th-century Romanesque Duomo (cathedral), with the plan of a basilica, has three aisles, a mosaic pavement with geometrical designs, and a paschal candelabrum. The singers' gallery is supported by nine lion-based columns, whose capitals are carved with prophets and sibyls by the sculptor Taddeo da Sessa. Baroque churches are SS. Annunziata, with a cupola, and S. Agostino, with an adjoining monastery. The earliest Latin satirist Gaius Lucilius (*q.v.*) was born there about 180 B.C. One of the town's two schools is a state boarding school. Agriculture is the chief occupation of the inhabitants, and there is trade in cereals, fruit, and wine. The town is near the main Naples-Rome railway and on the Sparanise-Formia branch line.

The ancient Suessa Aurunca was the chief city of the Aurunci and was punished by the Romans for refusing to pay tribute. Later it was made a municipium. During the civil war (88 B.C.) between Marius and Sulla it sided with Sulla. Augustus set up a colony there and changed its name to Colonia Julia Felix Classica. Its subsequent importance has been due only to its rich agricultural land. (M. T. A. N.)

SESSIONS, ROGER HUNTINGTON (1896–), U.S. composer of symphonic and instrumental music who played a leading part in the formation of modern musical taste among his contemporaries, was born in Brooklyn, N.Y., on Dec. 28, 1896. He studied at Harvard University and at the Yale School of Music, and later took composition lessons from Ernest Bloch. He spent several years in Europe, returning to the United States in 1933, though between 1928 and 1931 he had presented in New York, in association with Aaron Copland, an important series of concerts of modern music (the Copland-Sessions Concerts). He taught composition at Princeton University from 1935 to 1945, at the University of California, Berkeley, from 1945 to 1952, and again at Princeton from 1953. In 1965 he was appointed to the staff of the Juilliard School of Music, New York City. His first important work, an orchestral suite from the incidental music which he wrote for Leonid Andreev's play *The Black Maskers* (1923), is marked by rhapsodic eloquence and even flamboyance, and his

remained his most popular score. His five symphonies (1927, 1946, 1957, 1958, 1963) are severe in idiom, using a polyphonic technique and asymmetrical rhythms. Similar qualities mark his two string quartets (1936, 1950), string quintet (1957), two piano sonatas (1930, 1946) and his Chorale Preludes for organ (1925). His one-act opera *The Trial of Lucullus* (1947) and his *Idyll of Theocritus*, for soprano and orchestra (1954), have affinities with the central European expressionist style. His opera *Montezuma* was produced in West Berlin on April 19, 1964. His critical works include *The Musical Experience of Composer, Performer, Listener* (1950) and *Reflections on the Music Life in the United States* (1956). (N. Sy.)

SESTINA, a most elaborate form of verse, employed by the medieval poets of Provence and Italy, and occasionally used by modern poets. The scheme was the invention of the Provençal troubadour, Arnaut Daniel, who wrote many sestinas. Dante, a little later, wrote sestinas in Italian. In the *De vulgari Eloquentia*, Dante admits that he imitated Daniel.

The sestina, in its pure medieval form, consists of six stanzas of six lines each of blank verse: hence the name. The final words of the first stanza appear in varied order in all the others, the order laid down by the Provençals being: *abcdef, faebdc, cfdabe, ecbfad, deacfb, bdfeca.*

After these stanzas followed a *tornada*, or *envoi*, of three lines, in which all the six key words were repeated in the middle and at the end of the lines in the following order: *b-e, d-c, f-a.* Petrarch cultivated a slightly modified sestina, but later the form fell into disuse, until it was revived by the poets of La Pléiade, in particular by Pontus de Tyard. In the 19th century, it was assiduously cultivated by the comte de Gramont, who, between 1830 and 1848, wrote a large number of examples, included in his *Chant du passé* (1854).

A sestina in English was published in 1877 by Edmund Gosse; this was in the style of Daniel. It was subsequently frequently employed by English and American writers, particularly by Swinburne, who composed some beautiful sestinas on the French pattern; of these, that beginning "I saw my soul at rest upon a day" is perhaps the finest specimen in English. Swinburne's astonishing tour de force, "The Complaint of Lisa," is a double sestina of 12 stanzas of 12 lines each. The sestina was cultivated in Germany in the 17th century, particularly by Martin Opitz von Boberfeld and Georg Weckherlin. In the 19th century an attempt was made, not without success, to compose German sestinas in dialogue, or even double sestinas. (G. W. A.; X.)

SÈTE (CETTE), a seaport of southern France in the *département* of Hérault, Languedoc region, and the principal commercial port on the south coast after Marseilles, lies 200 mi. (322 km.) S. by W. of Lyons and 76 mi. (123 km.) S.W. of Avignon by road. Pop. (1962) 35,910. The older part of the town occupies the foot and slope of the isolated Mont St. Clair (574 ft. [175 m.]), on a tongue of land between the Mediterranean and the large lagoon of Thau (29 sq.mi. or 75 sq.km.). This quarter is bounded on the east by the Canal de Sète, which leads from the lagoon to the outer harbour. Between this canal and the Canal Maritime are two islands divided by a wet dock; the two canals are joined at the north by the Canal Latéral. A huge breakwater protects the entrance to the harbour, and a south mole and east jetty enclose the Old basin (the fishing harbour) and the outer harbour. At the outer end of the Canal Maritime are the New and Mediterranean basins, and the oil basin from which refined oil products from Frontignan are exported by pipeline. Damage done during World War II has been repaired. Facilities for the bulk discharge and storage of wine are extensive, with mechanical loading and unloading of wine through 50 pipelines, and Sète is probably the world's foremost wine port. Export and import traffic increased from 1,630,000 tons in 1938 to more than 4,000,000 tons in the early 1960s. There is a fishing industry and shellfish are intensively cultivated in the lagoon. On one island is a marine biological station. Local industries include chemical products, phosphates and cement, and the manufacture of spirits and vermouths. Sète is well served by sea, rail, and road. A sandy bathing beach stretches for about 9 mi. S.W. to Agde. Sète has become a considerable tourist resort, with many festivals including the ancient sport of water jousting.

The port was created in 1666 on the instructions of Colbert, minister of Louis XIV, and laid out according to plans made by Paul Riquet; its development was aided by the opening of the Canal du Midi later in the century. The present form of its name was fixed in 1927. Paul Valéry, the poet, was born and is buried in Sète, and a hall of the municipal museum is named after him.

SETEKH (SUTEKH or SET; Greek SETH), an ancient Egyptian god, patron of the 11th nome or district of Upper Egypt. His name was probably pronounced Suta or Seta in the late period, hence the Greek form. His worship centred at Ombos, the ancient Nubt, near Ballas in Upper Egypt, whose early importance gave Setekh his leading role as the partner and rival of Horus (*q.v.*). The archaic kings of Egypt were thought to be incarnations of both gods, and during the 2nd dynasty Setekh may even have replaced Horus for a time as the paramount royal deity. He was represented as a composite creature with a greyhound's body, long curved snout, and slanting eyes, his square-tipped ears and long forked tail held stiffly erect. Various animals (*e.g.*, dog, pig, donkey, giraffe, okapi, and anteater) have been suggested as prototype of this remarkable monster, whose zoological original had been long forgotten.

The vicissitudes of the cult of Setekh reflect the changing political scene in Egypt. During the rule of the Hyksos, he was worshiped at Avaris in the guise of the Canaanite storm-god Baal. As god of the eastern delta he was the patron of the Ramesside kings, and he appeared in mythological scenes standing at the prow of the barque of the sun-god and stabbing the wicked monster Apophis with his lance. He was the red god, the patron of the deserts and oases and counterpart of his brother Osiris, the god of vegetation. In the late pharaonic period, when the cult of Osiris gained great popularity, Setekh was execrated as his murderer; Horus, son of Osiris, fought with Setekh and avenged his father's death. The worship of Setekh was proscribed and his image effaced from monuments. As the personification of evil, Setekh was depicted in Ptolemaic temples as a hippopotamus or crocodile; the Greeks identified him with the demon Typhon. (M. S. DR.)

SETH (SHETH), one of the sons of Adam and Eve. According to the Priestly source he was the eldest son (Gen. 5:3; *cf.* also I Chron. 1:1 and Luke 3:38). The main tradition of the Yahwist source, on the other hand, makes Cain the firstborn son and Seth the third (Gen. 4:25). Num. 24:17, the Moabites are called "sons of Sheth"; this may be due to confusions with the Sutu of the cuneiform records, a nomad people of the north Syrian desert. Seth is much more prominent in Jewish tradition than in the Old Testament and many fancies gathered around the name. In the Old Testament it is said of him only that he was the father of Enosh and other children.

SETI I (GR. SETHOS) (d. 1304 B.C.) was a warrior pharaoh of the 19th dynasty who reigned over Egypt from 1318 to 1304 B.C. His father, Ramses I, reigned only two years, and it was Seti who was the real founder of the greatness of the Ramessids. In the early years of his reign he led his army northward to restore Egyptian prestige, partly lost during the troubled years of the late 18th dynasty. A punitive campaign in northern Palestine is recorded in the temple at Beth-shan (Beisan; *q.v.*), near the Sea of Galilee, and his campaigns in Syria are depicted in the vigorous reliefs in the temple of Karnak. He fought at least one battle with the Hittite king Muwatallis, and subsequently concluded a peace treaty which may have established the frontier at Kadesh on the Orontes River between Lebanon and Anti-Lebanon.

Seti did much to promote the prosperity of Egypt. He fortified the frontier; opened mines and quarries; dug wells and rebuilt temples and shrines which had fallen into decay or been damaged in the years of the monotheistic heresy under Ikhnaton (*q.v.*); and he continued the work begun by his father on the construction of the great hypostyle hall at Karnak, which is one of the most impressive monuments of Egyptian architecture (*see* THEBES). His greatest memorial is perhaps his mortuary temple at Abydos (*q.v.*) which he dedicated to Osiris and six other deities and deco-

rated with reliefs of great delicacy on which some of the original colour remains. Behind this temple is a curious building thought to be his cenotaph. His tomb is the finest in the Valley of the Tombs of the Kings in western Thebes (*see* THEBES); it was excavated by Giovanni Belzoni in 1817, and the fine alabaster sarcophagus is now in the Soane Museum in London. The king's mummy had been moved in ancient times and was found in the cache of royal mummies at Dayr al Bahri; it is now in Cairo. *See* also EGYPT: *History*. (M. S. DR.)

SÉTIF, chief town of the *département* of the same name in the province of Constantine, Algeria, 69 mi. (111 km.) SE of Bougie (Bejaïa) and about 1 mi. (2 km.) from the left bank of the Bou Sellam Oued, a tributary of the Sahel-Soummam Oued. Pop. (1960) 82,340. The town, laid out in a grid pattern of wide streets, stands at an altitude of 3,596 ft. (1,096 m.) on a fertile plateau largely given over to cereal cultivation. Sétif is connected by road and rail with Constantine, Bougie, and Algiers.

The ancient Sitifis became an important town when the emperor Nerva established a colony of veterans there *c*. A.D. 97. Sitifis became the chief town of the province of Mauretania Sitifensis (created at the end of the 3rd century) and remained so under Byzantine rule. In the military quarter, in the north of the town, are parts of the walls of the great Byzantine fortress. An open-air museum in the Garden of Orléans has statues, bas-reliefs, and inscriptions from Roman Sitifis. In 1959 a rich Roman necropolis was discovered near the centre of the city, between the synagogue and the wall of the citadel. (A. AM.)

SETON, a celebrated Scottish family deriving its name from the Seatown (Seyton) of Tranent in East Lothian. The house of Seton has in Scottish history been remarkable as being perhaps the most outstanding exponent of splendour in art, architecture, and heraldry, and its members have been consistently noted for the charm and "nobleness" of their lives and characters. Family honours have included the earldoms of Winton and Dunfermline; Seton scions became earls of Eglinton through marriage with the Montgomeries; and, through alliance with a Gordon heiress, ancestors of the earls and marquesses of Huntly.

The first recorded ancestor was ALEXANDER DE SEATON, who witnessed a royal charter of about 1150. His son, PHILIP DE SETON, held Setune, Wintune, and Winchelburgh by a charter granted between 1177 and 1185, and Philip's son, ALEXANDER DE SETON (d. *c*. 1245) had confirmation of these lands by a charter dated about 1195. Some three generations later, SIR ALEXANDER DE SETON (d. *c*. 1349), lord of Seton (who, with the chiefs Sir Gilbert Hay and Sir Neil Campbell, pledged support to Robert I, "the Bruce," in the "band" of 1308), contributed to the victory at Bannockburn in 1314, subscribed to the barons' declaration of 1320 to Pope John XXII, which asserted the independence of Scotland, and acquired special fame as the defender of Berwick against the English in 1333. All his four sons fell in the service of Scotland, THOMAS (d. 1333), the eldest, being hanged by Edward III before the walls of Berwick, since his father would not yield the town. SIR JOHN (d. *c*. 1328), the youngest, left a son SIR ALEXANDER (d. *c*. 1346) with whom the direct male line failed, whereupon the succession passed to MARGARET, lady of Seton, apparently the only child of the third son SIR ALEXANDER (d. 1332), who was killed while trying to repulse the landing of Edward Balliol. Margaret was romantically abducted in 1347 by Alan de Wyntoun, traditionally a cadet of the Seton line, who had Winton as appanage. This exploit, and their marriage, created a family feud in Lothian, and Alan went on crusade to the Holy Land where he died.

Their son, SIR WILLIAM DE SETON (d. *c*. 1409), lord of Seton and Tranent, carried on his mother's line as chief of the name of Seton, and with Winton recovered, restored the power of the family. From his second son, ALEXANDER, who married Elizabeth, heiress of the Gordon chiefs, descended the earls and marquesses of Huntly (chiefs of clan Gordon, George, the 2nd earl, having taken that surname), and also the Setons of Meldrum, Pitmedden, Mounie, and others. Sir William's eldest son, SIR JOHN SETON (d. *c*. 1433) of that ilk, the next chief, and his son SIR WILLIAM (who was killed at Verneuil in 1424), were famous in chivalry. SIR GEORGE DE SETON (d. *c*. 1479), the only son of Sir William, who succeeded before 1434, was created a lord of Parliament, as Lord Seton, before 1445 (a dignity sometimes stated to have begun in his grandfather). "Ane grit hous haldar, and all gevin to nobilnes," he established the family as Scotland's patrons of splendour, made their palace of Seton the most magnificent seat in Scotland, and built the choir of the parish church of Seton, which became a collegiate church in 1493.

Sir George's grandson and heir, GEORGE (d. *c*. 1508), 2nd Lord Seton, also a patron of arts and letters, bought a warship, the "Eagle," from James IV, and kept her manned at sea to harry the Flemish traders. He was succeeded by his son GEORGE (d. 1513), 3rd Lord Seton, who fell at Flodden and whose widow Janet, eldest daughter of the 1st earl of Bothwell, built Niddrie Castle, West Lothian, and the convent of St. Catherine of Siena, near Edinburgh, during the minority of her son GEORGE (d. 1549), 4th Lord Seton, in whose time the palace of Seton was burned by the English (1544). JOHN, his second son, founded the house of Seton of Carriston, in Fife, while MARIE, the youngest daughter, is the most celebrated of the four "Maries" attendant on Mary Stuart, queen of Scots. George's eldest son, GEORGE (1531–85), 5th Lord Seton, was the queen's most trusted supporter and under him the family reached the height of its fame. In June 1567 Mary and the earl of Bothwell visited Seton, which had been restored and enlarged, and after her escape in May 1568 from Lochleven Castle (to which Marie Seton contributed), Mary went to Niddrie Castle. George's fourth son ALEXANDER (1555–1622), created Lord Fyvie (1598) and 1st earl of Dunfermline (1605), was lord president of Scotland from 1593 to 1605 and chancellor from 1605 to his death. He was one of the most upright and distinguished judges of his time and the patron-architect of perhaps the greatest period in Scottish domestic architecture: Seton Palace, Winton House in Midlothian, Fyvie Castle in Aberdeenshire, Pinkie House in Midlothian, and Muchalls Castle in Kincardineshire were developed or completed by him or under his influence. His line expired with his junior grandson, JAMES (d. 1694), 4th earl of Dunfermline, a Jacobite attainted in 1690.

The 5th Lord Seton's second son, ROBERT (*c*. 1552–1603), 6th Lord Seton, was created earl of Winton (Nov. 1600). By his wife Lady Margaret Montgomerie, daughter of Hugh, 3rd earl of Eglinton, he had a large family, of whom his third son SIR ALEXANDER (1588–1661) of Foulstruther took the name Montgomerie and under a family resettlement became 6th earl of Eglinton. GEORGE (1584–1650), 8th Lord Seton and 3rd earl of Winton, the second son of the 6th lord, completed Seton Palace and in 1620 built Winton House. Of his large family his fourth son ALEXANDER (1620–91) became Viscount Kingston in 1651 and his tenth son JOHN (1639–1686) was created a baronet in 1664, being styled "of Garleston." GEORGE (1642–1704), 9th Lord Seton and 4th earl of Winton, the eldest grandson of the 3rd earl, led a regiment of his vassals in 1679 to Bothwell Bridge, Lanarkshire, where he assisted the royalist troops under the duke of Monmouth to defeat the Covenanters. His son GEORGE (*c*. 1678–1749), 10th Lord Seton and 5th earl of Winton, was attainted for joining the Jacobite rising in 1715 and forfeited his estates in 1716. Seton Palace was subsequently demolished. SIR GEORGE SETON (*c*. 1685–1769), of Garleston, 3rd bart., the grandson of the 1st bart., renounced his own title on the death of the 5th earl in 1749 and assumed the forfeited dignity of the earldom of Winton, but the title was never confirmed. He died unmarried in 1769.

Since 1749 the chiefship of the name of Seton has been dormant. It has been suggested that the lordship of Seton passes in the female line but no claim to this or reversal of the attainder has been formally made. A creation for Archibald William, 13th earl of Eglinton, chief of the Montgomeries, of an earldom of Winton on June 17, 1859, did nothing for the "name of Seton." Two Seton baronetcies still survive: Seton of Pitmedden, conferred in 1684 on SIR ALEXANDER SETON (d. 1719), a distinguished judge and cadet of Seton of Meldrum, of which the present holder is Sir ROBERT JAMES SETON (1926–), 11th bart; and Seton of Abercorn, conferred in 1663 on SIR WALTER SETON (d. 1692), descending from Seton of Touch, hereditary squire of the body of the kings of Scots and under an entail, heritable armourbearers

The present holder is SIR BRUCE LOVAT SETON (1909–), 11th bart.

See Sir Richard Maitland, *The Historie and Chronicle of the Hous of Seytoun,* continued by Alexander, Viscount Kingston (1829); G. Seton, *A History of the Family of Seton,* 2 vol. (1896). (T. I.)

SETON, ELIZABETH ANN (née BAYLEY) (1774–1821), foundress of the American Sisters of Charity, was born in New York City on Aug. 28, 1774. Her concern for the sick and poor early merited her the title of "a Protestant Sister of Charity." In 1794 she married William Seton who died in 1803 in Italy, where the family had gone for reasons of health. Left with five children, destitute in a foreign land, Mrs. Seton turned to the Filicchi family, old acquaintances, whose devotion was to eventually lead her into the Roman Catholic Church. She returned to the United States and after a great spiritual struggle was received into the Catholic Church in 1805. At the invitation of Louis G. B. Dubourg, she opened a grade school in Baltimore in 1809. Soon several young women were placed under her care; this led to the taking of simple vows as the "Sisters of St. Joseph." The community decided to take as its model the Sisters of Charity of St. Vincent de Paul, but it was not until 1812, after the community moved to Emmitsburg, Md., that the constitution and rules were formally adopted. Elizabeth, now known as Mother Seton, was named first superior but was allowed to remain legal guardian of her children. Mother Seton has been called the first of the American Catholic sister-school nuns, and the mother of the parochial school system of the United States, although hers was not the first Catholic grade school. She died Jan. 4, 1821. On March 17, 1963, she became the first native-born American to be declared blessed; her cause for canonization was introduced in 1965.

BIBLIOGRAPHY.—R. Seton, *Memoirs, Letters and Journals of Elizabeth Seton,* 2 vol. (1869); H. Bailly de Barberey, *Elizabeth Seton* (1927); K. Burton, *His Dear Persuasion* (1940). (E. R. V.)

SET THEORY (THEORY OF AGGREGATES). Set theory was created, almost without forerunners and against the opposition of most leading mathematicians of the time, between 1874 and 1897 by Georg Cantor (1845–1918). Since the beginning of the 20th century it has developed enormously and gained ever-increasing recognition and importance in both mathematics and logic, being applied in almost all fields of mathematics in view of the set-theoretical analysis of general mathematical concepts such as number, correspondence, function, order, etc. This article deals with the theory of abstract sets; *i.e.,* sets without discrimination as to the nature of their members.

Two principal attempts were made to define the concept of set: (1) as an arbitrary collection of "definite and distinct" objects (Cantor); and (2) by "comprehension" of the objects satisfying a given property. Both proved to be too general and led to contradictions, the antinomies of set theory. Although Bertrand Russell showed that the antinomies were of a logical or semantical rather than mathematical nature, a new foundation became imperative. The main direction taken involved restricting the extent of sets, either axiomatically or by silent limitation. The term "class" is used for extremely comprehensive sets, but, to avoid confusion with the logical concept of class, it will not be used here.

Fundamental Concepts.—The fundamental relation of set theory is membership: a is a member of the set s; in symbols, $a\epsilon s$. For instance, 2 is the only member of the set of even prime numbers; 1,2,3, . . . are among the (infinitely many) members of the set of positive integers. In principle it is sufficient for the needs of mathematics to restrict the members of sets to sets only, because numbers, functions, points, etc., may also be conceived as sets (see *Cartesian Product; Functions as Sets,* below). The principle of extensionality states that sets which have the same members are considered equal; hence the set s with the members $a,b,c,$ . . . may be denoted by $s = \{a,b,c, \ldots\}$. It is then presupposed that with respect to every object x, it shall be definite whether $x\epsilon s$ or not—which, however, need not be decided generally.

Other important concepts are subset, equivalence and order. In principle they can be reduced to the membership relation, but usually this is done for the subset relation only: s is a subset of t ($s \subset t$) if every member of s is also a member of t. Hence $t \subset t$ for every set t. If t has at least one member which does not belong to s, s is called a proper subset. To avoid the confusion prevailing up to the end of the 19th century, mathematicians say that a set contains its members, but comprises (includes) its subsets. Usually a subset of s is defined by a property P which is meaningful for the members of s, namely, as the set of those $x\epsilon s$ which satisfy P. For instance, from the set I of all positive integers and the property "prime" there is obtained the set of all prime numbers, which is a subset of I. If no member of s has the property, there must be admitted a set which contains no member, the empty set or null set ϕ.

If there exists a one-to-one correspondence between the members of s and those of t, s is called equivalent to t, in symbols $s \sim t$; then also $t \sim s$, and $s \sim t$, $t \sim u$ together imply $s \sim u$ (transitivity). A set D which is equivalent to the set $I = \{1,2, \ldots ,k, \ldots\}$ of all positive integers is called denumerable (or countable); denoting by d_k the member of D which corresponds to the integer k, D can be written in the form

$$D = \{d_1,d_2, \cdots d_k, \cdots\}$$

The set of all positive integers I as such is a plain set, but it becomes an ordered set when a succession (order) of its members is introduced, subject to the conditions that for any two different members one and only one precedes the other and that this order relation is also transitive. Possible orders of I are, for instance, (1,2,3, . . .), (. . . ,3,2,1), (1,4,7, . . . 2,5,8, . . . 3,6,9, . . .). The parentheses indicate that the order has to be observed. An ordered set is said to be well-ordered if every (ordered and nonempty) subset has a first member. Accordingly, among the ordered sets above, the first and the third are well-ordered but not the second, because the subset of all odd integers, as well as the set itself, has no first member.

If between the members of the ordered sets m and n there exists a one-to-one correspondence such that if a precedes b in m, also the image of a precedes the image of b in n, m and n are called similar ($m \simeq n$). None of the above ordered sets is similar to another, although, conceived as plain sets, they are equivalent and even equal. Yet, for instance, $(1,2,3, \ldots) \simeq (2,4,6, \ldots)$.

Comparability: Transfinite Induction.—The enormous importance of the well-ordered sets, both for set theory itself and for its applications, lies in the comparability of their ordinal and cardinal numbers and in the availability of transfinite induction.

Equivalent sets are said to have the same cardinal (number), and similar ordered sets are said to have the same order type. The order type of a well-ordered set is called an ordinal (number). For finite sets, the three notions essentially coincide. Comparing (ordering) cardinals according to their magnitude causes great difficulties, and comparing order types is in general altogether impossible. In contrast, ordinals and even the cardinals of well-ordered sets can be compared, and one shown to be less than the other. The well-ordering theorem, one of the most important and most discussed theorems of modern mathematics (E. Zermelo, 1904), states that for any set there exists a well-ordered set with the same members; accordingly, the cardinals of plain sets are also comparable.

For well-ordered sets a far-reaching generalization of ordinary mathematical induction becomes possible. If a statement involving the members of a well-ordered set W is true for the first member of W, and also true for any member w if true for all members that precede w, then the statement is true for all members of W. This theorem of proving by transfinite induction is almost trivial; far more profound is the fact that by this induction there can also be defined, for instance, the arithmetical operations between ordinals.

Finite and Infinite Sets.—The distinction between finite and infinite sets is fundamental. Dozens of definitions are in existence, among them many which differ in principle. To show the difference it is sufficient to mention the following two definitions:

1. A set is called inductive if there exists a positive integer n such that the set contains just n members (the null set is also inductive); otherwise the set is noninductive.

2. A set is called reflexive if it is equivalent to a proper subset; otherwise it is nonreflexive. Hence the set I of all positive integers is both noninductive and reflexive; the latter fact follows, for instance, from the correspondence

$$\begin{array}{ccccccc} 1 & 2 & 3 & \cdots & n-1 & n & \cdots \\ \updownarrow & \updownarrow & \updownarrow & & \updownarrow & \updownarrow & \\ 2 & 3 & 4 & \cdots & n & n+1 & \cdots \end{array}$$

which shows that I is equivalent to the subset obtained by dropping the member 1. It is also equivalent to the subset of all odd, even or prime numbers.

It seems as if "noninductive" and "reflexive" were the same property, namely, infinite; and that inductive and nonreflexive were another property, finite. Only in the first decade of the 20th century was it discovered that the equivalence between the definitions mentioned, as well as others, depends on a new principle of purely existential character, the principle (axiom) of choice or "multiplicative principle," which had not before been formulated and even in the 1960s is not yet generally accepted in mathematics and logic. Russell's formulation (1906) states: If S is a set of nonempty sets such that any two of these sets have no common member, then there exists a set which contains a single member out of each member of S. This principle is indispensable for many other purposes both inside and outside set theory: in the former, for the arithmetic of cardinals and order types, and for the well-ordering theorem, for whose proof the principle was explicitly introduced. By means of the multiplicative principle it can be proved that every noninductive set comprises a denumerable subset, and hence that every such set is reflexive. The equivalence between other definitions of finiteness and infinity can also be proved.

The virtual, though not conceptual, coincidence between cardinal and ordinal numbers of finite sets derives from the following theorem, which is proved for inductive sets by mathematical induction: By arranging the members of a finite set in any two different ways two ordered sets are obtained which, although different, are similar; *i.e.*, follow identical schemes of order: first, second, . . . , nth, with the same n terminating. The above examples of (essentially) different orders of the set I show that this theorem does not hold for infinite sets; in fact, to an infinite cardinal belong infinitely many different order types and ordinal numbers.

Boolean Algebra.—Prior to the invention of set theory, Boolean operations had been introduced and applied by G. Boole (1815–64) and others; set theory increased the applications of these operations. Two major and one minor operations between sets are fundamental. The union of two or more (or infinitely many) sets $s_1, s_2, \ldots$ is the set of those members which are contained in at least one set s_n, and is denoted by $s_1 \cup s_2 \ldots$. Corresponding to this operation is the operation of intersection, which yields the set $s_1 \cap s_2 \ldots$ of the members contained in each of the sets s_n. Of minor importance is the difference $t - s$, which is essential only in the case $s \subset t$; then $t - s$ is the set of the members of t which are not contained in s. Here again the significance of the null set becomes manifest; without it the intersection of two sets having no common members, as well as the difference $t - t$, would not exist. In Boolean algebra, but not in general set theory, there is introduced, corresponding to the null set, the "universal set" U, which contains all members considered. The difference $U - s$ is called the complement of s.

Union and intersection are in many respects analogous to sum and product in arithmetic; the operations are commutative (by definition) and associative. The first distributive law,

$$s \cap (t_1 \cup t_2 \cup \cdots) = s \cap t_1 \cup s \cap t_2 \cup \cdots$$

corresponds to the arithmetical law $a(b + c) = ab + ac$; the second distributive law of Boolean algebra, however, viz.,

$$s \cup t_1 \cap t_2 \cap \cdots = (s \cup t_1) \cap (s \cup t_2) \cap \cdots$$

has no parallel in arithmetic, where $a + bc$ in general differs from $(a + b)(a + c)$. The validity of both laws is not accidental but is connected with a law of duality; this latter law may be based on the laws of A. de Morgan, which state that the complement of a union of sets is the intersection of the corresponding complements, and that the complement of an intersection is the union of the complements. From this it is concluded that, A being a set formed from sets $s_1, s_2, \ldots$ by repeated operations of union and intersection, the complement $A' = U - A$ is obtained by replacing the sets s_n by their complements, union by intersection, and intersection by union. Since $A = B$ implies $A' = B'$, any equality resulting from those operations remains true after the replacements have been effected, and if the equality is an identity (*i.e.*, true for any set) the transition to the complements becomes superfluous. For example, from $s \cup s' = U$ follows $s \cap s' = \phi$ and from $s \cup \phi = s$ follows $s \cap U = s$ for any set s; one distributive law follows from the other. From $s \subset s \cup t$ follows that $s \subset t$, $s \cap t = s$, $s \cup t = t$ are equivalent statements, as are $s = t$ and $s \cap t = s \cup t$

Usually the term Boolean algebra is reserved for the abstract theory in which no reference is made to the meaning of union, intersection, complement and subset; a few of the statements mentioned above, and others, are then considered to be the axioms of Boolean algebra.

The significance of Boolean algebra in logic mainly derives from the fact that union corresponds to logical disjunction (in the sense of *vel*, not exclusive), intersection to logical conjunction, complement to negation and the subset relation to (material) implication. For instance, $s = U$ corresponds to "everything is an s," $s = \phi$ to "there is no s," $s \cap t \neq \phi$ to "some s are t," $s \cap s' = \phi$ to the logical law of contradiction and $s \cup s' = U$ to the law of the excluded middle.

The operation of union (but not of intersection) is basic for the arithmetic of cardinals, namely, for their addition: a sum of finitely or infinitely many cardinals is defined as the cardinal of the union of sets (representatives) with the respective cardinals, if no two representatives have common members. To be sure, the sum's independence of the particular representatives rests upon the principle of choice if the sum has infinitely many terms.

Cartesian Product; Functions as Sets.—In addition to union and intersection there are two other operations with sets which have great importance not for Boolean algebra but for the arithmetic of cardinals, and to a lesser degree of order types, and for the applications of set theory to analysis. One is the Cartesian product. If s and t are sets, say without common members, the Cartesian product $s \times t$ is defined as the set of all pairs

$$\{\sigma, \tau\} \text{ with } \sigma \in s \text{ and } \tau \in t$$

For instance, if $s = \{1,2,3\}$, $t = \{4,5\}$, then $s \times t = \{\{1,4\}$, $\{1,5\}, \ldots\ldots, \{3,5\}\}$; *i.e.*, a set with $3 \times 2 = 6$ members. Therefore the product of two cardinals is defined as the cardinal of the Cartesian product of suitable set representatives. The definitions of both Cartesian products and of products of cardinals may be extended to any finite or infinite number of factors.

While it follows from the definition that a Cartesian product, among whose factors the null set occurs, is empty—hence a product of cardinals with the factor zero equals zero—the converse statement, for infinitely many factors, rests on the principle of choice or, more precisely, constitutes this principle. In fact, only by "choosing" one arbitrary member out of each factor set, none of which is empty by assumption, can it be generally ensured that the Cartesian product is not empty. This is the origin of Russell's term "multiplicative principle." Incidentally, Russell illustrated the difference between set formation by (constructive) "comprehension" and by purely existential "choice" by confronting a set of infinitely many pairs of shoes with another of pairs of stockings. To obtain a set that contains a single member of each pair, it suffices in the first case to take the set of all left shoes which is defined constructively; yet the existence of a set containing a single stocking of each pair is guaranteed only by the axiom of choice.

Finally, for the exponentiation of sets or cardinals a generalization of the ordinary concept of function is most appropriate. For functions $s = f(t)$, assume the domain of variability to be an arbitrary set T over which the argument t runs, whereas the func-

tion values s are members of a set S. The insertion set (S/T) shall be defined as the set of all such functions; *i.e.*, of all "insertions" of members of S into T. For instance, simultaneously casting four dice (whose sides show from one to six pips) produces a function $s = f(t)$ with $T = \{1,2,3,4\}$ and $S = \{1,2,3,4,5,6\}$; if two casts are considered equal only on condition that the same die shows the same number of pips, (S/T) contains $6 \cdot 6 \cdot 6 \cdot 6 = 6^4$ members. If T is the set of all positive integers and $S = \{0,1,2, \ldots, 9\}$, every function $f(t)$ may be considered to be a decimal fraction of the form $0 \cdot s_1 s_2 s_3 \ldots$ with values s_t taken from S; hence the insertion set contains all such decimals and its (infinite) cardinal is of the form $10 \cdot 10 \cdot 10 \ldots$, the factors forming a sequence. Thus general powers, with infinite bases and exponents, can be defined and calculated.

The most important case is the so-called power set of a set S; *i.e.*, the set whose members are all subsets of S. It may be conceived as an insertion of the pair $\{1,0\}$ into S, where 1 is related to every $s \epsilon S$ contained in the respective subset and 0 to the members not contained in it. The so-called "theorem of Cantor" (1892) states that the power set of S has, for finite and infinite S, a greater cardinal than S; it constitutes the first step in ascending to ever-increasing infinite cardinals.

Independently of the concept of insertion set, a function can always be regarded as a set; for instance, a single-valued function $y = f(x)$ can be regarded as a set of ordered pairs (x,y) in which different pairs contain different values of x but not necessarily different values of y. Furthermore, integers and even infinite numbers may be regarded as sets; by means of the null set ϕ the integers 0,1,2,3, . . . may respectively be defined as the sets

$$\phi, \{\phi\}, \{\phi,\{\phi\}\}, \{\phi,\{\phi\},\{\phi,\{\phi\}\}\}$$

and generally every ordinal number may be defined as the set of those preceding it. An elaborate theory of numbers in this sense was given by J. von Neumann in 1923. (*See* POINT SET.)

Applications.—The applications of set theory penetrate most branches of mathematics; as N. Bourbaki puts it: "Today it is possible to derive almost all contemporary mathematics from a single source, the Theory of Sets." (*Actualités Scientifiques et Industrielles*, vol. 1212, Paris, 1954.) The applications are most spectacular, through the theory of point sets, in analysis; perhaps the greatest analytical achievements of set theory are the modern theories of measure and integration. In geometry and topology many concepts and problems such as "curve" and "dimension," which had defied the cruder conservative methods, were treated set-theoretically; Cantor solved the riddle of the linear continuum (unsuccessfully attacked from Greek antiquity to the 19th century) in terms of a perfect set; namely, one which densely comprises a denumerable subset. In abstract algebra, the well-ordering theorem and transfinite induction (or an equivalent maximum principle) solve problems such as algebraically closed extensions of a field. The theory of probability owes important achievements to set theory, and there are also applications to the theory of games, to physics and even to chemistry.

The modern development of logic has been closely connected with the development of set theory, and the antinomies have neither checked the development of these sciences nor seriously impaired the agreement of most mathematicians with D. Hilbert's description (1925) of the "paradise created by Cantor from which nobody will ever expel us."

See also POINT SET; GROUPS.

BIBLIOGRAPHY.—G. Cantor, *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, ed. by E. Zermelo (1932); J. E. Littlewood, *The Elements of the Theory of Real Functions* (1955); A. A. Fraenkel, *Abstract Set Theory*, 2nd ed. (1960); A. A. Fraenkel and Y. Bar-Hillel, *Foundations of Set Theory* (1958); W. Sierpinski, *Cardinal and Ordinal Numbers* (1958); R. Stoll, *Set Theory and Logic* (1963); W. S. Cooper, *Set Theory and Syntactic Description* (1964).

(A. A. F.)

SETTLE, ELKANAH (1648–1724), English playwright and miscellaneous writer, scathingly satirized by Dryden, whose unworthy rival and antagonist he was. Born at Dunstable on Feb. 1, 1648, he went to Westminster school and entered Trinity college, Oxford, in 1666 but left without a degree. His first tragedy, *Cambyses, King of Persia*, was produced in 1666. He was encouraged as Dryden's rival by John Wilmot Rochester, through whose influence Settle's *Empress of Morocco* was twice acted at Whitehall: it also succeeded on the public stage in 1670–71. In the dedication of the printed edition (1673) Settle referred scornfully to other dramatists' effusive dedications, obviously aiming at Dryden who in 1674 cooperated with John Crowne and Thomas Shadwell in an abusive pamphlet condemning the play. Settle replied in kind.

Neglected by the court after 1675, he continued to have his bombastic plays published and performed. However, in 1679 the earl of Shaftesbury engaged him to organize a "pope-burning" pageant and thereafter he wrote copiously for the Whig interest. He published a reply to Dryden's *Absalom and Achitophel* in 1682 and was ridiculed as "Doeg" in the second part of his rival's satire. After Shaftesbury's downfall, he turned Tory, attacked Titus Oates, and even enlisted in James II's army, but soon rejoined the Whigs on the accession of William III.

"Recanting Settle" was appointed the last poet laureate of the city of London in 1691 and devised ten lord mayor's shows during his term of office. Hoping for patronage, he made a practice of composing occasional poems, elaborately bound and now very rare, which he presented to persons of note. He also wrote entertainments for Bartholomew fair, where in his old age, clad in a green leather suit, he is said to have played the dragon in his own droll, *St. George for England*. Pope makes him a type of the dull writer in the *Dunciad*, and puts into his mouth the couplet:

> "Yet lo! in me what authors have to brag on!
> Reduc'd at last to hiss in my own dragon." (iii, 286–87)

He entered the Charterhouse in 1718, and died there on Feb. 12, 1724.

See F. C. Brown, *E. Settle* (1910); *Five Heroic Plays*, ed. by B. Dobrée (1960).

SETTLEMENT, in law, a compromise or agreement between litigants to settle the matters in dispute between them in order to dispose of and conclude their litigation. Generally, when litigants settle their differences, prosecution of the action is simply terminated and the action is withdrawn or dismissed without any judgment being entered. In such a case the settlement itself, as a binding contract between the parties, prevents reinitiation of the litigation. But the parties may, and often do, incorporate the terms of the settlement into a consent judgment, recorded by the court. Such a judgment may afford the same protection against a reopening of the dispute in litigation as is provided by a court judgment at the conclusion of a fully litigated case.

Settlements commonly provide or are construed to allow either party to enforce their terms or at his election ignore them and reopen the underlying dispute should the other party fail to fulfill the terms and conditions agreed upon. Since in modern litigation by far the greater number of suits brought are either withdrawn or settled, the settlement constitutes an important feature of the litigatory process. (*See* JUDGMENTS AND DECREES; PRACTICE AND PROCEDURE.)

The term "settlement" is also applied to a disposition of property to be held in trust (*see* REAL PROPERTY AND CONVEYANCING, LAWS OF; TRUST). (C. E. CL.)

SETTLEMENT, ACT OF, the name given to that act of Parliament passed in June 1701 (12 & 13 Will. III, c. 2) which has since that date regulated the succession to the throne of Great Britain. The settlement made in the Bill of Rights of 1689 extended no further than the children of Queen Mary, the Princess Anne, and William III. Mary died childless in 1694 and William did not remarry. When the last surviving child of Anne, William, duke of Gloucester, died in 1700 the need to provide for the succession after her death became urgent, and it was decided that in default of issue to either William III or Anne, the crown was to pass to the electress Sophia of Hanover, granddaughter of James I, and the heirs of her body, being Protestants.

In addition to settling the succession to the throne, the act further defined the conditions on which it was to be held in clauses designed to come into force on the death of Queen Anne

which were inspired by Tory zeal and characterized as much by suspicion of the foreigner as by suspicion of the executive. Future monarchs were to join in communion with the Church of England and were not to go abroad without the consent of Parliament. Other clauses prohibited the use of English resources in defense of foreign territories without the consent of Parliament and declared that foreigners were not to hold offices or places of trust, to belong to the privy council or to either house of Parliament, or to hold offices of profit under the crown or receive from it pensions or grants of land.

The possibility that members who held places or pensions from the crown might lessen the independence of the House of Commons was to be prevented by excluding them from it. And by stipulating that "all matters . . . properly cognizable in the privy council . . . shall be transacted there" it was hoped that all those who advised the king could be made legally accountable for their advice by being known to belong to an institution known to the law. Another clause protected the independence of the judges by making their appointments *quamdiu se bene gesserint*, dependent on their continued good behaviour, instead of on the sovereign's pleasure (*durante bene placito regis*) as before. They were to be removable only on address of both houses of Parliament. It was also declared that impeachments by Commons in Parliament were not to be interfered with by pardons under the great seal of England.

Just as the Bill of Rights drew attention to actions of James II which were disapproved of, so the Act of Settlement censured by implication many practices of William III. Like it also, it showed how far a reluctance which Tories may once have felt to restrain a divinely appointed monarch was replaced by anxiety to limit the actions of monarchs appointed by Parliament.

So much, indeed, was this the case, that it was thought necessary, for the sake of effective government, to repeal the clause restricting the crown's advisers to privy councilors before it was due to come into effect; and the clause prohibiting placemen and pensioners from the House of Commons was modified by an act of 1707 (6 Anne c. 41 [otherwise cited as c. 7] paragraphs 25–26) which merely obliged members appointed to offices of profit under the crown to seek reelection. Slowly, and somewhat reluctantly, it was being recognized that increasing the importance of the House of Commons in order to limit more stringently the power of the executive might prevent the executive from acting effectively; and also that it might in practice be necessary, when executive action depended upon the support of the legislature, to run the risk that if the executive could not influence the legislature then the legislature might come to control the executive.

See W. C. Costin and J. Steven Watson, *The Law and Working of the Constitution*, vol. i (1952). (Ro. R.)

SETTLEMENT, SOCIAL: *see* Social Settlements.

SETÚBAL, a seaport of Portugal, in the district of the same name, lies 27 mi. (43 km.) SE of Lisbon by road via the Tagus ferry. Pop. (1960) 44,177. The town is situated on the north shore of a deep estuary formed by the Sado, Marateca, and São Martinho rivers. Setúbal exports salt, oranges, and muscatel wine and grapes; it has sardine-curing and boat-building establishments and produces fish manure, cement, fertilizers and canned foods. Under John II (1481–95) Setúbal was a favourite royal residence, but most of the ancient buildings were destroyed by the earthquake of 1755. The most important remaining ones are the Convent of Jesus (Manueline style), the churches of São Julião (Manueline decoration) and Santa Maria da Graça, and the 17th-century castle of São Filipe (now a hotel). In the sand hills on the left bank of the estuary are the ruins of Cetobriga, which was destroyed by a tidal wave in A.D. 412.

Setúbal District, now being industrially developed, occupies the peninsula between the Tagus and Sado estuaries. Pop. (1960) 376,128. Area 1,989 sq.mi. (5,152 sq.km.). It is remarkable for the profuse vegetation on the southern slope of the Serra de Arrábida (1,634 ft. [500 m.]) along the southern coast, including many Mediterranean species, giant oaks, and strawberry trees (*Arbutus unedo*). In the eastern foothills of the Serra is the town of Palmela with a magnificent castle of Moorish origin. The manor of Bacalhoa at Vila Fresca de Azeitão dates from the 15th and 16th centuries and contains beautiful examples of decoration with Portuguese glazed tiles.

SEURAT, GEORGES PIERRE (1859–1891), French painter, the leading figure among a small group of painters known as the Neoimpressionists, was born in Paris on Dec. 2, 1859. From the very start his search for scientific truth and for an applicable formula carried him beyond the limits of academic training; at the same time, he departed from the concept of fleeting pictorial impressions which Claude Monet and his group had developed. The result was a new and valid form of monumentality for easel painting that reinstated a firm structure of composition and re-established the formal value of the subject.

Seurat's artistic career comprised a bare decade and was outwardly uneventful. He began to study drawing when he was 15 and entered the École des Beaux-Arts in 1878 in the class of Henri Lehmann, a pupil of Ingres. Supplementing his formal training with untiring studies in the Louvre and in the libraries, he copied Ingres, Raphael, Holbein and Poussin; he also pored over scientific studies of colour by M. E. Chevreul, O. N. Rood, H. L. F. Helmholtz, D. Sutter and C. Blanc.

After a year of military training at Brest, where he did many studies of people and the sea, Seurat returned to Paris late in 1880. For some time he concentrated mainly on drawing, developing a completely new technique which enabled him to replace line by volume and to create an intricate pattern of light and shade. He also began making small studies in detached strokes of pure colour. In 1883 Seurat began his first major project in painting, "Une Baignade," a large scene with bathers (Tate gallery, London) for which many drawings and studies are extant. When this painting was rejected by the salon of 1884, Seurat, with Paul Signac, H. E. Cross, C. Angrand, A. Dubois-Pillet, M. Luce and, for a short while, Camille Pissarro, helped found the Société des Artistes Indépendants. These painters were also called the Pointillists, Divisionists or Chromatic Luminarists.

In 1884 Seurat began work on his second large composition, "La Grande Jatte" ("A Sunday Afternoon on the Island of la Grande Jatte"). For months the artist was on the island every day sketching. Back in his studio he developed the theme in endless studies, drawings, sketches of sections and complete and carefully

BY COURTESY OF THE BUFFALO FINE ARTS ACADEMY, ALBRIGHT ART GALLERY

Second Study for "Le Chahut," Painted by Seurat in 1889

executed studies of the whole scene. For the final painting (now at the Art Institute of Chicago) Seurat prepared a palette exactly according to Chevreul's colour disk. Four basic colours and their intermediaries were only to be mixed with varying amounts of white before they were applied in a veritable maze of small, carefully separated dots. When shown at the last exhibition of the Impressionists in 1886 "La Grande Jatte" caused a great controversy in which only a few artists, among them Signac and Pissarro, and the poet and art critic Félix Fénéon defended Seurat's work. In 1887 Seurat exhibited the painting again in Brussels with "Les Vingts" ("The XX"), a small group of independent Belgian painters; here it received a more favourable reception.

After finishing "La Grande Jatte," Seurat was so completely in command of his new method that the works that followed required fewer and fewer preparatory sketches. He painted only five more major compositions: "Les Poseuses," 1887 (Barnes foundation, Merion, Pa.); "La Parade," 1887 (Stephen C. Clark, New York city); "Le Chahut," 1889–90 (Rijksmuseum Kröller-Müller, Otterlo); "Jeune Femme se poudrant," 1889–90 (Courtauld institute, London); and "Le Cirque," begun in 1890 (Louvre). In each of these Seurat further explored new facets of the pictorial and technical problems which interested him: the study of space and perspective, of artificial light and varying moods, the sensation of air moving around objects. During the summer months from 1885 to 1890 Seurat painted out-of-doors a number of pure landscapes and marines. Though done in the same strictly scientific method, they are intimate and yet lasting records of nature.

When Seurat died in Paris on March 29, 1891, he had produced 7 monumental compositions, 40 smaller paintings and sketches, about 500 drawings and several sketchbooks. These show him as one of the greatest artists of his time, contributing immeasurably to the development of 20th-century art.

Bibliography.—J. Rewald, *Georges Seurat* (1943), complete earlier bibliography, and *Post-Impressionism* (1956); Seurat, "Paintings and Drawings," The Art Institute of Chicago, *Exhibition Catalogue*, 1958; D. C. Rich, *Seurat and the Evolution of "La Grande Jatte"* (1935); J. Rewald and H. Dorra, *Catalogue raisonné*. (W. M. V. D. R.)

SEVAGRAM, a village in Wardha District, Maharashtra, India, lies 6 mi. SE of Wardha town. Originally called Segaon, the village was renamed Sevagram (Hindi "village of service") by Mohandas Karamchand Gandhi (*q.v.*). In 1936 Gandhi left his *ashram* ("hermitage") on the Sabarmati River, near Ahmedabad, with a vow never to return there until the achievement of Indian independence. He settled at Sevagram, where he continued to direct the independence movement, and founded another *ashram*. Within this he created a model community, which still flourishes. The inhabitants of the *ashram* live a simple existence, growing their own food and weaving their own clothes. The hut in which Gandhi lived still stands in the *ashram*. Another important feature at Sevagram is the Nai Talimi Sangh, the educational centre established by Gandhi, and whose purpose is "learning by doing." Gandhi wanted to evolve a system of education which would fully develop the individual in body, mind, and spirit, and through cooperative work to bring about a new social order from which all would benefit, irrespective of caste. The task of the Nai Talimi Sangh was to build a self-sufficient community providing its own necessities in food, clothing, shelter, and tools and to establish a society which would be able to fulfill its aesthetic, spiritual, and intellectual needs, creating its own art, music, literature, and drama. (D. G. Na.)

SEVAN, LAKE (Ozero Sevan, or Gokcha), is the largest lake in the Armenian Soviet Socialist Republic, U.S.S.R., covering an area of 547 sq.mi. (1,417 sq.km.). It lies in the basin of the Aras (Araks) River, into which it discharges by way of the Razdan, or Zanga, River. The lake lies 6,279 ft. (1,914 m.) above sea level in a basin enclosed by mountain ranges, the Pambakskiy and Shakhdagskiy ranges to the north and the Gegamskiy and Vardenisskiy ranges to the south. The lake falls into two connected parts, Little Sevan in the northwest, smaller but deeper (325 ft.; 99 m.), and the larger, shallower (174 ft.; 53 m.) Great Sevan in the southeast. There are 28 permanent streams flowing into Sevan and many seasonal streams in spring. The largest tributaries are the Masrik, Kyavar-Chay, and Gedak-Bulag. Rainfall is about 15 in. (381 mm.) a year, but evaporation from the lake is exceptionally high, about 33 in. (838 mm.) a year. This high evaporation accounts for 92% of the water loss and the outflow for only 8%. Since World War II a cascade of six hydroelectric plants has been built on the Razdan and the outfall has been artificially lowered to increase the outflow. This has resulted in a slight drop of the lake surface. Average surface water temperatures vary from 1.6° C (35° F) in winter to 17.7° (64°) in summer, and between 1834 and 1954 the whole lake froze on only ten occasions, although certain bays tend to freeze every year. The annual temperature range decreases with depth and below 160 ft. (49 m.) is 3.8° C (39° F). The lake is rich in varieties of fish, including the Sevan trout. The fish caught are canned at Martuni on the southern shore. Chromite is mined in small quantities on the northern shore. The chief town is Sevan at the Razdan outfall, near the highest barrage of the cascade. There are other, small hydroelectric stations on the streams flowing into the lake, notably at Martuni. (R. A. F.)

SEVASTOPOL, a town and seaport in the Krymskaya (Crimean) Oblast' of the Ukrainian Soviet Socialist Republic, U.S.S.R., is located in the extreme southwest of the Crimea on the southern shore of a long narrow inlet, which forms a splendid natural harbour, the best on the Black Sea. Pop. (1959) 148,033. West of this site stood the ancient Greek colony of Chersonesus, founded at the end of the 6th century B.C. According to the historian Strabo the peninsula on which Chersonesus stood was protected by a wall from the inlet to the south coast of the Crimea at Symbolon (now Balaklava). Originally a republic, Chersonesus, threatened by Scythian attacks, became part of the kingdom of Pontus (*c.* 100 B.C.) and then of the Cimmerian Bosporus. It passed to the Romans, being allowed a large measure of autonomy at times, and eventually to the Byzantine Empire. Tradition has it that the apostle Andrew visited Chersonesus, and that St. Clement was banished there by Trajan to work in the quarries.

In 988 or 989 the town fell to Prince Vladimir of Kiev, who was baptized there and took the Orthodox faith back to Kiev. Vladimir restored the city to the Byzantines, but in the 13th century it passed to the empire of Trebizond (or Trabzon). With the growth of the Genoese factories of Caffa and Surozh (Feodosiya and Sudak), Chersonesus lost all importance. After the 13th-century Mongol invasion, the Crimean Tatars established a small settlement, Akhtyar, on the northern side of the inlet. The remnants of the town were razed by the Lithuanian prince Algirdas (Olgierd) in the 14th century. In the 15th century it came under the Ottoman Empire in the reign of Mohammed II, and in 1783 it was annexed by Russia, under Catherine the Great. In the following year, on the orders of G. A. Potemkin, the site was selected for a naval base to protect Russia's newly acquired gains on the Black Sea coast. The new town was given the name Sevastopol ("the august city"). In 1804 Sevastopol became the main Black Sea base and in 1808 a commercial port was opened. The port was constructed under the direction of Admiral M. P. Lazarev, famed for his Antarctic explorations with F. G. von Bellingshausen. During the Crimean War the Anglo-French armies laid siege to Sevastopol, which had been strongly fortified by the military engineer E. I. Todleben. After an eleven-month campaign, from September 1854 to August 1855, the allies captured the key fortress of the Malakhov and the Russians retreated from Sevastopol to the northern side of the inlet. The town suffered considerable damage, but rebuilding was started and its growth was accelerated by the construction of a railway to the port in 1875. In 1894 the commercial harbour was transferred to Feodosiya, and Sevastopol suffered a considerable economic setback. Its principal function thereafter was that of fortress and naval base.

Sevastopol underwent a second siege in World War II, when the town put up a prolonged and gallant defense against the Germans for 250 days, from October 1941 (when the Perekop Isthmus was forced) until July 3, 1942. When resistance was finally overcome the entire town had been reduced to rubble. Liberated in May 1944, Sevastopol began once more to be rebuilt, on the attractive, planned layout of the 19th century. It once more became a major naval base, with commercial functions considerably less important.

There are shipbuilding and repair yards, and bricks, furniture, and foodstuffs are produced. Passenger services link Sevastopol to Odessa and Sukhumi. The town has a museum (1905) devoted to the defense of Sevastopol, and also a vast panorama of the Crimean War siege, 377 ft. long and painted by F. A. Rubo, which is housed in a special building. *See* also CRIMEA; CRIMEAN WAR; BOSPORUS, KINGDOM OF THE; CHERSONESE, TAURIC. (R. A. F.)

SEVEN AGAINST THEBES, in Greek mythology, the seven champions who were killed fighting against Thebes after the fall of Oedipus (*q.v.*). Oedipus' twin sons, Eteocles and Polyneices, whom he had cursed, failed to agree which of them was to succeed to the Theban throne and decided to rule in alternate years. As Eteocles' turn came first, Polyneices withdrew to Argos, where he married Argeia, daughter of the king Adrastus. Her sister Deipyle married Tydeus, son of King Oeneus of Calydon, who was in exile for homicide. At the end of the year Polyneices claimed to rule Thebes in his turn. When Eteocles refused, Adrastus mobilized an army, whose chieftains in Aeschylus' tragedy about the seven were Tydeus, Capaneus, Eteoclus, Hippomedon, Parthenopaeus, Amphiaraus (*q.v.*), and Polyneices. Other authors count Adrastus as one of the seven and omit Hippomedon. In the fighting Tydeus after killing Melampus gnawed his head, and so forfeited the immortality Athena had intended for him. Polyneices and Eteocles killed each other, fulfilling Oedipus' curse. When the sons of the dead seven, the Epigoni or second generation, had grown to manhood, Adrastus again attacked the city and occupied it after the Thebans had evacuated it by night. He died at Megara on the homeward journey.

The story was a great favourite in antiquity. It is the subject not only of Aeschylus' *Septem contra Thebas* but also of Euripides' *Phoenissae*. (D. E. W. W.)

SEVEN CHAMPIONS OF CHRISTENDOM, the name given in medieval tales to the seven national saints of England, Scotland, Ireland, Wales, France, Spain and Italy—*i.e.*, Saints George, Andrew, Patrick, David, Denis, James and Anthony (*qq.v.*) The classical version of their achievements is that of Richard Johnson, *The Most Famous History of the Seaven Champions of Christendome* (3 parts, 1596, 1608, 1610).

SEVENOAKS, a market town and urban district in the Sevenoaks parliamentary division of Kent, Eng., 24 mi. (39 km.) SE of London by road. Pop. (1961) 17,645. Area 5.8 sq.mi. (15 sq.km.). Residential in character, it is the shopping centre for the surrounding rural district. Sevenoaks school and almshouses were founded under the will of William Sevenoke (1432), a foundling who became mayor of London. St. Nicholas Church nearby dates in part from the 13th century. Cricket has been played on the famous Vine ground for over 200 years.

Knole House, since 1456 owned by monarchs, archbishops, and, from *c.* 1603, by the Sackville family, is one of the finest houses in England. Given to the National Trust in 1946 by the 4th Lord Sackville, who lives in a part of it, it contains rare furniture, tapestries, and historic paintings.

See V. M. Sackville-West, *Knole and the Sackvilles*, new ed. (1947).

SEVEN SLEEPERS OF EPHESUS, the heroes of a famous legend which, because it affirmed the resurrection of the dead, had an enormous and lasting success in both Eastern and Western Christendom and even in Islam during the Middle Ages. According to the story, during the persecution of Decius (A.D. 250) seven (in some versions eight) Christian soldiers were concealed near Ephesus (in western Asia Minor) in a cave of which the entry was later blocked up. They fell into a miraculous sleep, from which they awoke in the reign of the emperor Theodosius II (408–450), and after bearing witness to the doctrine of the resurrection of the body they died like saints. The origin of the legend in the 5th century remained unknown until the publication of E. Honigmann's *Patristic Studies* in 1953. This story of Christian apologetics, with its reminiscences of classical mythology, is extant in a number of versions: Greek, Syriac, Latin, Coptic, Armenian, Ethiopic, and Georgian. The sanctuary of the Seven Sleepers near Ephesus long attracted crowds of pilgrims, both Christian and Muslim, many of whom hoped to sleep their last sleep beside these guarantors of survival and resurrection. Austrian excavations have revealed all that remains of the sanctuary and the cemetery. The feast day of the Seven Sleepers is July 27 among Latins, Aug. 2 or 4 and Oct. 22 or 23 in the Greek Church.

BIBLIOGRAPHY.—M. Huber, *Die Wanderlegende von den Siebenschlafern* (1910); I. Guidi, "Testi orientali inediti sopra i sette dormienti di Efeso," *Atti della reale accademia dei Lincei*, series 3, *Memorie della classe di scienze morali, storiche e filologiche*, vol. 12 (1884–85); P. Peeters, "Le Texte original de la passion des sept dormants," *Analecta Bollandiana*, 41:369–385 (1923), *cf* 72:265–266 (1954) on Honigmann's discovery; *Österreichisches archäologisches Institut, Forschungen in Ephesos*, vol. 4, part 2 (1937). (F. HA.)

SEVENTH-DAY ADVENTISTS: *see* ADVENTISM.

SEVEN WEEKS' WAR, the war fought in the summer of 1866 between Austria, Saxony, Hanover, Bavaria, Württemberg, Baden, Hesse-Kassel, Hesse-Darmstadt, and Nassau on the one side, and Prussia (supported by some minor German states) and Italy on the other. Won by Prussia, it settled the long-standing Austro-Prussian conflict over Germany; and it enabled Italy to annex Venetia.

Causes and Outbreak.—The Empire of Austria (*q.v.*), a conservative form of state, comprising several nationalities loosely articulated, survived in fundamental opposition to the forces of nationalism. These forces were represented in Germany by the ambitious Prussia (*q.v.*) and in Italy by the national monarchy which came into being in 1861 (*see* GERMANY: *History;* ITALY: *History;* also ITALIAN INDEPENDENCE, WARS OF). Within the German Confederation, as it had been established in 1815, Austria and Prussia were rivals for hegemony; and, since only part of the Austrian Empire was included in the Confederation, the rivals were more or less equally balanced. Prussia's scheme for a union of German states without Austria had been quashed by Austria in 1850; but the Austrian plan for uniting the whole Empire with the Confederation had been rejected at the Dresden Conference in 1851. In the Italian war of 1859 Austria had lost Lombardy to Sardinia-Piedmont (nucleus of the Kingdom of Italy formed in 1861), but still retained Venetia.

The emperor Francis Joseph (*q.v.*) on the whole was working toward a lax internal policy and an assertion, against Prussia, of Austria's headship of the German Confederation; from 1865 his chief minister was Richard, Graf Belcredi (*q.v.*), with Alexander, Graf Mensdorff-Pouilly, as foreign minister. In Prussia policy grew in unity and clarity after King William I took Otto von Bismarck (*q.v.*) as his prime minister in 1862. Bismarck set out to create a new Germany and was prepared to destroy the old Confederation. Despite differences about military policy, he found able collaborators in H. K. B. von Moltke (*q.v.*), chief of the general staff, in Albrecht von Roon (*q.v.*), minister of war, and in Edwin von Manteuffel (*q.v.*), chief of the Military Cabinet.

The ostensible cause of the Seven Weeks' War was disagreement over Schleswig and Holstein (*see* SCHLESWIG-HOLSTEIN QUESTION); and the chain of events leading to war began when the Italian, the Schleswig-Holstein, and the German questions were associated in autumn 1864. In September, Napoleon III (*q.v.*) undertook to withdraw the French garrison from Rome within two years, thus giving Italy a free hand to plan the "redemption" of Venetia from Austria. In October, Denmark ceded Schleswig-Holstein jointly to Austria and Prussia. Differences between the two occupying powers raised the German question. Austria aimed at a final settlement in consultation with the German Confederation and was willing that Frederick, duke of Augustenburg, should rule the duchies. Prussia denied the right of the Confederation to consultation and countered the Augustenburg claim with an alternative Oldenburg one. Prussia's next proposal —annexation of the duchies to Prussia—failed when Austria required territorial compensation. In February 1865, also unavailingly, Prussia proposed independence for the duchies, provided that their armed forces were incorporated in Prussia's and that Kiel was ceded to Prussia as a naval base. In March war seemed possible, but tension was relaxed by negotiations.

The Convention of Gastein, on Aug. 14, 1865, divided the duchies. Thenceforward Prussia in Schleswig firmly checked agitation on Augustenburg's behalf, whereas Austria treated Holstein as a temporary trust and respected local feeling. Meanwhile

Napoleon III discussed Germany with Bismarck at Biarritz and St. Cloud, though he supported Austria with a loan publicly subscribed on the Paris *Bourse*. Italy, failing to buy Venice from Austria (Alessandro Malaguzzi's mission to Vienna), signed a trade treaty with the Prussian *Zollverein* (Customs Union) on Dec. 31.

Prussia protested in Vienna against Austria's policy in Holstein (Jan. 26, 1866). Austria next tried to save peace by diplomatic pressure from Saxe-Coburg-Gotha and from Great Britain on Berlin and by a slow *rapprochement* with France; but Prussia took more warlike measures, including negotiations with Italy, which led on April 8 to a treaty binding Italy to join Prussia if German considerations necessitated a Prussian attack on Austria within three months. But the Austrians also put themselves in the wrong: by asking Bismarck (March 16) whether he intended war; by making precautionary troop movements (Austria needed seven or eight weeks to mobilize, Prussia only four); by insisting on territorial compensation if Prussia took the duchies; and by requiring guarantees for the pope if Italy took Venice (which Mensdorff on April 30 privately offered, through Paris, for after the war). The Prussian government improved its standing by proposing on April 9 a German parliament elected by universal suffrage. Francis Joseph put himself further in the wrong when, alarmed by Italian preparations, he decided to mobilize the southern army. This was ordered on April 21, and mobilization for Bohemia on April 27. The archduke Albert (*q.v.*) was appointed to command in the south, and Gen. L. A. von Benedek (*q.v.*) in the north. Prussian mobilization was ordered between May 3 and May 8. Meanwhile Saxony had started arming (April 14); and when the Prussians failed to satisfy the Diet of the Confederation that they had good cause to arm, other pro-Austrian states, namely Bavaria, Württemberg, Baden, and Hesse-Darmstadt, also mobilized.

Napoleon III, with British and Russian support, tried to organize a European congress to settle the three questions, but this plan collapsed on June 3. He subsequently announced a convention with Austria binding himself to neutrality (June 12).

Austria, offering to disarm if legality were restored in the duchies, remitted the question to the Diet (June 1). This was a technical infraction of an Austro-Prussian agreement of Jan. 16, 1864, to keep the settlement to themselves. Prussian troops from Schleswig entered Holstein on June 7. Austria proposed "federal execution" against Prussia; *i.e.* action by the Confederation as a whole. Then Prussia, faced by an adverse vote on a subsidiary motion in the Diet, declared the Confederation abolished (June 14).

On June 15, when Prussian ultimatums to Saxony, Hanover, and Hesse-Kassel, demanding benevolent neutrality, were rejected, Prussian troops invaded the three states; and on June 21, there being no formal declaration of war, the Prussian crown prince Frederick William (the future German emperor Frederick III) informed the nearest Austrian commander that a state of war existed.

The Campaigns.—Making full use of railways for military purposes for the first time, the Prussians had concentrated their troops in the south in an arc 270 mi. long, stretching from the Saxon frontier eastward into Silesia, where there was a common frontier between Prussian territory and Austrian Bohemia. The Elbe Army on the Prussian right drove the Saxons before them into Bohemia, and then joined the 1st Army under Prince Frederick Charles, advancing into Bohemia from positions farther along the arc. On the far left was the 2nd Army, under the crown prince. The Prussians numbered some 285,000 men, of whom about 50,000 were detached for work from Hanover to the Rhineland, and 9,000 for the protection of Upper Silesia. They were armed with the breach-loading needle gun. The Austrians, with some 74,000 men in Italy and some 241,000 in Germany, used a muzzle-loading rifle, and depended much on the bayonet charge.

The Austrians retained their supremacy against the Italians: the archduke Albert defeated Gen. A. F. la Marmora at Custoza (June 24); and Adm. Wilhelm von Tegetthoff destroyed the Italian Navy under Adm. Carlo Persano off the island of Lissa (July 20). In Bohemia, Benedek had the strategical advantage, since he might have concentrated his massed forces against one fraction of the Prussian troops before they united. But the Prussians won early engagements on the Iser and at Gitschin (Jičín) and, in the east, at Náchod and at Trautenau (Trutnov); and by July 1, Moltke had realized his plan to unite the three Prussian armies.

The Battle of Königgrätz (Hradec Kralove), or Sadowa (Sadova), on July 3, 1866, engaged the whole forces of both sides on the Bohemian front. But, since Prussian reconnaissance had been faulty and since Benedek had accepted battle against his better judgment, both sides fought from positions which they had not chosen. Austrian losses were heavy: 20,000 killed and wounded and nearly as many prisoners. The Prussian losses were less than 15,000. The Austrian retreat to Olmütz (Olomouc) was confused, but the Prussians, in almost equal confusion, did not follow up. Both armies had assembled outside Vienna by July 22, when an armistice was signed at Nikolsburg (Mikulov), to be followed on July 26 by preliminaries of peace.

Meanwhile Prussian forces had also been launched against the pro-Austrian states in western Germany. One from Lauenburg on the lower Elbe and one from Westphalia converged on Hanover, while more Prussians, from Wetzlar, struck northeastward at Kassel. The Hanoverian Army (19,000 men) retreated southward into Thuringia, but was prevented from effecting a junction with the Bavarians and so took up a defensive position at Langensalza. There, on June 27, it repelled an attack by a Prussian force of 8,200 from Gotha; but on June 29, it was obliged to capitulate.

Farther to the south, on the Main River, the states loyal to the old Confederation had still two separate army corps, together about 96,000 men, representing Bavaria, Württemberg, Baden, Hesse-Darmstadt, and Nassau. A Bavarian advance toward Thuringia was defeated at Dermbach on July 4 by the victorious Prussians from Langensalza, who then went on to defeat the Bavarians again at Kissingen (July 10) before turning westward to capture Frankfurt am Main. Turning next southward and eastward, the Prussians overcame the Federal army's resistance near Würzburg, which fell on July 27. A few days later an armistice was concluded.

The Results of the War.—The Austro-Prussian Peace of Prague (Aug. 23, 1866) assigned Schleswig-Holstein to Prussia. But Bismarck left Austria itself territorially intact: reconciliation and then alliance with Austria were to become the basis of Germany's later European hegemony. Art. IV of the Peace of Prague, however, recorded Austria's consent to "a new organization of Germany" without Austrian participation.

Prussia annexed Hanover, Hesse-Kassel, Nassau, and Frankfurt outright, thus acquiring the territory which had separated the eastern and the western parts of the Prussian state. Peace treaties, including war indemnities to be paid to Prussia, were concluded with Bavaria, which ceded the small districts of Gersfeld and Orb; with Hesse-Darmstadt, which ceded Hesse-Homburg; with Württemberg; with Baden; and with Saxony.

By the Peace of Vienna (Oct. 3, 1866) Austria ceded Venetia to the French emperor, for transfer to Italy, but retained the southern Tirol and Trieste.

The main issue of the war had been German, and in Germany its most significant results appeared. Prussia could organize the North German Federation (*see* GERMANY: *History*). At the same time the liberal majority in the Prussian lower chamber, which had been defying the government since 1862 (*see* PRUSSIA), capitulated to Bismarck and passed an Act of Indemnity, thus weakening liberalism throughout Germany.

The Austrian Empire, withdrawing from Germany and Italy, was enabled to rearticulate itself. The result was the Austro-Hungarian *Ausgleich* or Compromise of 1867 (*see* HUNGARY: *History*). Austrian stability rested thenceforward on the balance between two partners in the dual monarchy.

Napoleon III suffered a worse defeat than Austria. While the issue of the war was undecided he had been imprecise in demands for territorial compensation. His first impulse after Königgrätz was intervention to curb Prussia, but he stopped short of military pressure. He then formulated demands (July 23) too late. France, in the face of Prussia's new Germany, lost the continental security enjoyed since 1815.

See also references under "Seven Weeks' War" in the Index.

BIBLIOGRAPHY.—O. von Lettow-Vorbeck, *Geschichte des Krieges von 1866 in Deutschland*, 3 vol. (1896–1902); C. W. Clark, *Franz Joseph and Bismarck: the Diplomacy of Austria Before the War of 1866* (1934); G. A. Craig, *The Politics of the Prussian Army, 1640–1945* (1955). (A. R.)

SEVEN WISE MASTERS, THE, a widely spread cycle of stories of oriental origin. Other names sometimes used are *The Seven Viziers* and *The Story of the Seven Sages*. The framework of the stories is that an oriental king entrusted the education of his son to a wise tutor named Sindbad (hence the further alternative titles for the cycle, *Sindbad-name, The Book of Sindbad*, etc.). An occasion arose when the prince was ordered to keep silence for a week. During this time his stepmother, who had tried to seduce him, accused him before the king and sought to bring about his death by seven stories which she related to the king; but her narrative was each time confuted by the tales of the craft of women related by seven sages. Finally the prince's lips were unsealed and the truth was exposed.

It is disputable how far the Middle Persian text (no longer extant) went back to Indian originals. No old Arabic text has survived, but there are modern versions (one included in the *Arabian Nights: see* Nights 578–606 in Sir R. Burton's translation, vol. vi, 1886). The Arabic text gave rise to the Hebrew translation (*Mishle Sindbar*), a Spanish translation (13th century), and the Syrian translation from which derives the Greek version called *Syntipas* (11th century). Of the Persian versions the most important is that of al-Samarqandi (12th century). The Greek version was translated into Latin in the 12th century by Jean de Hauteseille under the title of *Dolopathos* (ed. by H. Oesterley, 1873); this was translated into French by a *trouvère* named Herbers. The German, English, French, and Spanish chapbooks of the cycle are generally based on a Latin original. The Middle English metrical romances are based on the French; the most important of these is *The sevin Seages Translatit . . . in Scottis meter* . . . (1578; ed. by G. F. Black, 1932, Scottish Text Society), by John Rolland of Dalkeith (fl. 1555–60).

BIBLIOGRAPHY.—D. Comparetti, *Researches Respecting the Book of Sindibad*, Eng. trans. by H. C. Coote, with Spanish text (1882); *The Seven Sages of Rome* (southern version) ed. by K. Brunner (1933); B. E. Perry, *The Origin of the Book of Sindbad*, offprint from *Fabula*, vol. iii (1960). (S. M. SN.)

SEVEN WISE MEN OF GREECE (SEVEN SAGES), certain historical (late 7th–early 6th century B.C.) Greek figures subsequently grouped together for their practical wisdom. They first appear in Plato (*Protagoras*, 343 a) as Thales, Pittacus, Bias, Solon, Cleobulus, Chilon (*qq.v.*), and Myson of Chen. Elsewhere Periander (*q.v.*) often replaces the obscure Myson, but there are many other variations of the list. The numerous references to them in ancient writers are mostly on three main topics: a contest for a tripod to be given to the wisest, finally dedicated by the winner to Apollo; a banquet held by Croesus or Periander (as in Plutarch's *Symposium of the Seven Sages*, where his place as sage is taken by Anacharsis [*q.v.*]); and their proverbial sayings, including the famous "Know thyself" and "Nothing in excess."

BIBLIOGRAPHY.—*Fragments:* F. W. A. Mullach, *Fragmenta Philosophorum Graecorum*, vol. i, pp. 203 ff. (1860). *See* also Diogenes Laertius, book i; K. Freeman, *The Pre-Socratic Philosophers*, pp. 44–45 (1946).

SEVEN WONDERS OF THE WORLD, the seven preeminent sights of the ancient world. The seven wonders usually so called (except for the Pharos of Alexandria) are those listed in an epigram of Antipater of Sidon (mid-2nd century B.C.) and in an incomplete prose list that goes under the name of his older contemporary, the engineer Philo of Byzantium, but that may well be much later. Some later writers give variant lists.

Pyramids of Egypt.—These are the oldest wonder and the only one substantially in existence today. There are pyramids elsewhere in Egypt, but the best-known lie along a 50-mi. stretch of the Nile Valley west of the river near ancient Memphis (*q.v.*) and date from the Old Kingdom (*c.* 2700–2300 B.C.). The three great pyramids of Khufu, Khafre, and Menkaure at Giza are the most famous, but there are many other pyramids in this region, notably at Abu Roash, Abu Sir, Saqqarah, and Dahshur. (*See* PYRAMID.)

Hanging Gardens of Babylon.—These were not literally "hanging" but "up in the air"—that is, they were roof gardens laid out on a series of terraces irrigated by pumps from the Euphrates. Tradition sometimes made them the work of Nebuchadrezzar, built to console his Median wife who missed the mountains of her homeland, or of the semilegendary queen Semiramis. No certain traces survive.

Statue of Zeus at Olympia.—This was one of the two masterpieces of Phidias of Athens (the other being the statue of Athena in the Parthenon). The statue, over 30 ft. (over 9 m.) high and plated with gold and ivory, represented the god sitting on an elaborate throne with a Victory standing on one hand and a sceptre topped by an eagle in the other. It was famous for the divine majesty and goodness it expressed. The discovery in the 1950s of Phidias' workshop (*see* OLYMPIA) confirmed the statue's date as *c.* 430 B.C. The temple was destroyed in A.D. 426; the statue may have perished then or in a fire at Constantinople 50 years later. No copies survive.

Temple of Artemis at Ephesus.—Little remains above ground of this temple, built by Croesus, king of Lydia, *c.* 550 B.C. and rebuilt after being burned down in 356 B.C., but there are fragments in the British Museum, and excavation revealed traces of both Croesus' and the 4th-century temple as well as of three earlier smaller temples. The temple was famous not only for its great size (over 300 by 150 ft. [about 110 by 55 m.]) but also for the works of art that adorned it. It was destroyed by invading Goths in A.D. 262 and was never rebuilt. Copies survive of the statue of Artemis, an un-Greek representation of a mummylike goddess, festooned with breasts or eggs and ornamented with bees and animals, and wearing a high pillared headdress. (*See* EPHESUS.)

Mausoleum of Halicarnassus.—The monumental tomb of the Carian dynast Mausolus was built, beginning in 353 B.C., by his sister and widow, Artemisia. The architect was Pythius or Pytheos, and the sculptures were the work of four leading Greek artists, Scopas, Bryaxis, Leochares, and Timotheus (*see* GREEK ART: *The Classical Period*). Fragments preserved in the British Museum include the frieze of battling Greeks and Amazons and a ten-foot statue, probably of Mausolus. The Mausoleum was probably destroyed by earthquake between the 11th and the 15th century A.D., and the stones reused. (*See* HALICARNASSUS.)

Colossus of Rhodes.—The Rhodians commemorated the raising of Demetrius Poliorcetes' long siege (305–304) of Rhodes by erecting a bronze statue (reinforced with iron) of the sun god, Helios, over 100 ft. (30 m.) high. It was the work of Chares of Lindus and stood by the harbour, perhaps shielding its eyes with one hand, as a relief suggests; the idea that it straddled the harbour entrance is medieval and impossible. It took 12 years to build (*c.* 292–280), but *c.* 225 it broke off at the knees in an earthquake. The fallen Colossus was left in place until A.D. 653 when the Arabs raided Rhodes and had it broken up and the bronze sold for scrap; it was said there were 900 or more camelloads of metal from it. (*See* COLOSSUS.)

Pharos of Alexandria.—Though famous earlier, this does not appear in any list of wonders until the 6th century A.D. (the earliest list gives the walls of Babylon instead). It was the most famous ancient lighthouse, built by Sostratus of Cnidus for Ptolemy II of Egypt *c.* 280 B.C. on the island of Pharos off Alexandria. It was said to be over 440 ft. (135 m.) high and built in three stages, all sloping slightly inward; the lowest was square, the next octagonal, and the top cylindrical. A broad spiral ramp led to the top, where a fire burned at night. In the Middle Ages the Arabs replaced the beacon with a small mosque. The Pharos was still standing in the 12th century, but in 1477 the sultan Qait Bey built a fort from its ruins. *See* also ALEXANDRIA.

BIBLIOGRAPHY.—*Pyramids:* I. E. S. Edwards, *The Pyramids of Egypt*, rev. ed. (1961). *Hanging Gardens:* Diodorus Siculus, ii, 10. *Zeus:* Pausanias, v, 11. *Artemis Temple: see* EPHESUS: *Bibliography*. *Mausoleum:* Pliny, *Natural History*, xxxvi, 5, 30. *Colossus:* H. Maryon in *Journal of Hellenic Studies*, 76:68 ff. (1956). *Pharos:* G. Sarton, *A History of Science*, vol. ii, pp. 21–26 (1959). *Lists:* J. Lanowski in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft*, supplementband x, 1020–30 (1965).

SEVEN YEARS' WAR (1756–1763), the name given to the last great conflict involving all the major powers of Europe, except Turkey, before the French Revolutionary Wars. One aspect of the conflict, that is, the specifically Franco-British aspect, can be understood as an extension of the so-called French and Indian War (*q.v.*), a colonial struggle which broke out in North America and in India before 1756 but ended simultaneously with the Seven Years' War. It is possible to regard the French campaigns in western Germany, against electoral Hanover (dynastically united with Great Britain) and its allies, Brunswick and Hesse-Kassel, as well as the Spanish entry and the consequent Portuguese involvement in the last phase of the war, as by-products of the colonial struggle. From such a point of view, the Seven Years' War properly so named is the Austro-Prussian War of 1756–63, in which Russia was also deeply committed against Prussia till the beginning of 1762 and in which Sweden likewise played a part against Prussia. On the other hand a British alliance with Prussia and a French one with Austria, together with the geographical contiguity of Hanover and Prussia, interwove the fortunes and problems of the two sets of belligerents so closely that, for a practical appreciation, the European campaigns at least should be surveyed as factors of a single war, complex although not coherent. British and European historians, in fact, are often inclined, regardless of chronology, to use the expression Seven Years' War to include also the French and Indian War.

THE DIPLOMATIC REVOLUTION

The Peace of Aix-la-Chapelle (1748), at the end of the War of the Austrian Succession (*q.v.*), left wide grounds for discontent among the powers. In the first place, it did nothing to allay the colonial rivalry between Great Britain and France in North America or in India. Secondly, it confirmed the conquest of the rich province of Silesia (*q.v.*) from Austria by Frederick II (*q.v.*) the Great of Prussia, which the Habsburg heiress Maria Theresa (*q.v.*) resented as an iniquitous price to pay for the recognition of her consort, Francis I, as Holy Roman emperor. Thirdly, Russia was irked by the aggrandizement of Prussia, which constituted a challenge to Russian designs on Poland and on Poland's dependencies along the southeastern coast of the Baltic. Russia, however, under the Treaty of St. Petersburg of Dec. 9, 1747, had accepted a British subsidy and had supplied mercenary troops to the British for use against the French in the last stage of the war; and the French, in resentment, had vetoed any representation of Russia at the peace congress.

The War of the Austrian Succession had seen the belligerents aligned on a time-honoured basis. France's traditional enemies, Great Britain and Austria, had coalesced just as they had done against Louis XIV; and Prussia, the leading anti-Austrian state in Germany, had been supported by France. Neither group, however, found much reason to be satisfied with its partnership: British subsidies to Austria had produced nothing of much help to the British, while the British military effort had not saved Silesia for Austria; and Prussia, having secured Silesia, had come to terms with Austria in disregard of French interests. Even so, France had concluded a defensive alliance with Prussia in 1747; and the maintenance of the Anglo-Austrian alignment after 1748 seemed essential to the duke of Newcastle, British secretary of state in the ministry of his brother Henry Pelham. The collapse of this system and the aligning of France with Austria and of Great Britain with Prussia constitute what is known as the "Diplomatic Revolution," or the "Reversal of Alliances."

The Interests of the Powers.—King George II (*q.v.*) of Great Britain was passionately devoted to his electorate of Hanover whose interests to him counterbalanced those of the British colonies overseas: if war against France for colonial expansion was to be resumed, then Hanover had to be secured against Franco-Prussian attack. France was very much interested in colonial expansion and might exploit the vulnerability of Hanover in war against Great Britain, but wanted not to have to divert forces to central Europe for Prussia's sake. French policy was, moreover, complicated by the existence of the *secret du roi*, that is, the private diplomacy conducted by King Louis XV (*q.v.*), unbeknown to his foreign minister, mainly with the purpose of winning the Polish crown for his kinsman, the prince de Conti (Louis François de Bourbon), and of keeping Poland, Sweden, and Turkey as client states of France, in opposition alike to Russian and to Austrian interests.

Austria and Russia had already concluded a defensive alliance (June 2, 1746) covering their own territory and Poland against attack by Prussia or by Turkey, with a secret clause promising the restoration of Silesia and the countship of Glatz (Klodzko) to Austria in the event of hostilities with Prussia; but their real desire was to destroy Frederick's power altogether, reducing him to his electorate of Brandenburg and giving East Prussia to Poland in exchange for the cession to Russia of Polish rights over the Duchy of Courland. A. P. Bestuzhev-Ryumin (*q.v.*), grand chancellor of Russia under the empress Elizabeth, was no less hostile to France than to Prussia, in view of French activity in Poland, Sweden, and Turkey; but the Austrian statesman W. A. von Kaunitz (*q.v.*), though he was firmly determined to recover Silesia, would not commit himself to Bestuzhev's offensive designs against Prussia so long as Prussia could rely on French support.

Frederick the Great saw Saxony and Polish West Prussia as potential fields for expansion, but could not expect French support if he started an aggressive war for them; and if he joined the French against the British in the hope of annexing Hanover he might fall a victim to an Austro-Russian attack. The hereditary elector of Saxony, Frederick Augustus II, was also elective king of Poland as Augustus III; but Saxony and Poland were physically separated by Brandenburg and Silesia, and neither state could pose as a great power, since Saxony was merely a buffer between Prussia and Austrian Bohemia, while Poland, despite its union with the ancient lands of Lithuania, was a prey to pro-French and to pro-Russian factions. The Prussian scheme for compensating Frederick Augustus with Bohemia in exchange for Saxony obviously presupposed further spoliation of Austria.

Preliminary Negotiations and the Outbreak of the Colonial War.—To gratify Austria, the British government proposed Hanoverian support for the election of Maria Theresa's son Joseph as king of the Romans, that is, agreed successor to his father as emperor; but this proposal foundered on the opposition of Frederick the Great (elector of Brandenburg as well as king of Prussia), whom the other German electors did not dare to antagonize. In 1750, Great Britain acceded to the Austro-Russian defensive alliance of 1746, but without subscribing to the secret clause on Silesia and without obtaining a guarantee of Hanover by the two empires.

Kaunitz in 1750 went as Austrian ambassador to France to urge French participation in Austro-Russian plans against Prussia. France, however, was neither ready to resume diplomatic relations with Russia (severed in 1748) nor willing to connive at the destruction of Prussia, which would have restored Austria to incontestable hegemony in Germany. By 1753, when Maria Theresa recalled him to Vienna to be chancellor, Kaunitz had achieved only a vague atmosphere of Franco-Austrian goodwill.

Meanwhile the French and the British East India companies had been engaged in continuing hostilities, despite the peace of 1748 (*see* INDIA-PAKISTAN, SUBCONTINENT OF: *History: European Settlements*); and in North America, likewise, relations between the colonists had deteriorated steadily from 1752. In 1754 French aggression in North America reached a point which the government in London could no longer pretend to ignore (*see* FRENCH AND INDIAN WAR). The British admiral Edward Boscawen attacked French ships in the Strait of Belle-Ile, off the Breton coast, in June 1755; but before declaring open war on France the British government had still to ensure Hanover, so that British naval superiority could be used to advantage in a largely maritime struggle while France's superior land forces in Europe were held in check by some continental ally of the British.

The Defensive Alliances.—Obsessed with Silesia, Austria was most reluctant to become implicated in the Anglo-French quarrel. Kaunitz took the view that Great Britain should hire German and Russian mercenaries to defend not only Hanover but also the southern Netherlands, whence in previous wars Austro-British and

Dutch operations against France had been launched. The decline of the Dutch as a military force in any case compromised the defense of the Austrian Netherlands, whose value to Austria was furthermore greatly reduced by Anglo-Dutch commercial restrictions. France on the other hand always coveted the Netherlands, and Kaunitz was in fact willing to consider ceding them in return for French help over Silesia. The strength of Austrian troops which Kaunitz was prepared to supply for the defense of Hanover or of the Netherlands against France was far less than what the British required from him.

Rebuffed by Austria, the British sought a new treaty with Russia, like that of 1747; and on Sept. 30, 1755, a preliminary agreement was signed in St. Petersburg by Bestuzhev and the British ambassador, Sir Charles Hanbury Williams. This stipulated that Russia should maintain 55,000 men on the Livonian-Lithuanian frontier, so that they could be promptly moved to defend King George's territory or his allies in Europe in the event of an attack; that in such an event Great Britain would grant to Russia a yearly subsidy of £500,000; and, by a secret article, that Russia in the meantime was to receive £100,000 a year. Bestuzhev, interpreting the treaty as aimed at Prussia, was delighted to have British money to spend on his own projects.

At the same time, without Russia's knowledge, the British were making overtures to Frederick the Great. Afraid of Austro-Russian intentions and alarmed at the Anglo-Russian negotiations, Frederick welcomed these overtures, though the result was unlikely to please his French ally. On Jan. 16, 1756, the Convention of Westminster was signed, whereby Great Britain-Hanover and Prussia agreed to respect one another's territory in Europe and undertook to resist jointly any invasion of "Germany" by a foreign power; but the Austrian Netherlands were expressly excluded from this guarantee.

The Convention of Westminster dismayed Bestuzhev and his empress, who had not yet ratified the British treaty. The empress peremptorily informed the British that the common enemy envisaged in the treaty could only be Prussia; and when the British rejected this interpretation the whole Russo-British arrangement came to nothing. The French government was no less angry at the duplicity of its one ally, Prussia.

The French, in order to get more information about the Anglo-Russian talks and to test the ground for reopening official intercourse with Russia, had sent a Scottish Jacobite refugee, Alexander Mackenzie, known as the chevalier Douglas, on a clandestine mission to St. Petersburg in autumn 1755. Douglas had been briefed by the prince de Conti, for the *secret du roi*, as well as by the French foreign ministry; but the chief agents of the *secret* in Poland had been kept unaware of his mission, lest they should regard the approach to Russia as a betrayal of the anti-Russian line to which their activity had been dedicated. Conti, indeed, seems to have been ready to accept the Duchy of Courland, under Russian protection, or perhaps even marriage to the empress Elizabeth, instead of the precarious Polish crown. The Russian vice-chancellor, M. I. Vorontsov, a persistent enemy of Bestuzhev, received Douglas very favourably; and Elizabeth's indignation at the Convention of Westminster served to accelerate a Franco-Russian *rapprochement*. In April 1756 the Russians offered 80,000 men to Austria for an attack on Prussia.

To Kaunitz the Convention of Westminster gave obvious reasons for self-congratulation. It justified his view that the British alliance was no longer worthwhile; and at the same time it obliged France to accept Austria's overtures, for fear of isolation now that Prussia was defecting. Franco-Austrian negotiations, which had been resumed in summer 1755 between Austria's ambassador Georg Adam, Graf von Starhemberg, and the French abbé de Bernis (*q.v.*), a protégé of Louis XV's mistress, Madame de Pompadour, had reached a stalemate in December; but the announcement of the Convention of Westminster gave them new impetus, and on May 1, 1756, the First Treaty of Versailles was signed (not at Versailles but at Jouy) by the French foreign minister A. L. Rouillé, Bernis, and Starhemberg. This was a defensive alliance between France and Austria, either party undertaking to send 24,000 men to support the other in the event of attack; but it exempted Austria from any obligation to join in war against Great Britain.

The Convention of Westminster and the First Treaty of Versailles are generally taken as the constituent factors of the Diplomatic Revolution, but they did not make war in Europe inevitable. Both being expressly defensive, they might well have had the contrary effect, though Kaunitz at least could see the Austro-French agreement as a step toward enticing France into an Austro-Russian offensive alignment against Prussia. The French seizure of Minorca from the British, achieved in the month April 19–May 20, 1756, did not oblige Prussia to war on the British side, and was of course no concern of Austria's.

Frederick the Great had tried, unavailingly, to present the Convention of Westminster as not inconsistent with his French alliance. He had, accordingly, to profess to regard the First Treaty of Versailles as harmless to Prussia. But that treaty was clearly advantageous to Austria, and so, indirectly, to Russia; and both Austria and Russia were now massing troops on their frontiers nearest to Prussia. In July and as late as Aug. 20, 1756, Frederick appealed to Maria Theresa for assurances of her good intent toward him; but he received no satisfactory reply. On Aug. 29, 1756, Frederick led his army into Saxony, on the way to Austria's Bohemian frontier.

The motive of Frederick's decisive action, which started the European war, has been much debated. Was he frightened into a preventive war, intended only to seize what military advantage he could in the face of imminent aggression by Austria and Russia? Or did he think that the moment had come for another war of annexation?

However much the British were annoyed at the prospect of having to support Frederick if his war went ill, the French were aghast at his action. Whereas they had signed their Austrian treaty in the belief that their hands would be free for the vital war against the British and that they could later choose whether or not to abet an Austrian offensive against Prussia, they now found themselves committed to defend Austria against the unforeseen aggression.

THE COURSE OF THE WAR

The Anglo-French campaigns in North America are described in the article FRENCH AND INDIAN WAR. For those in India see INDIA-PAKISTAN, SUBCONTINENT OF: *History;* also CLIVE, ROBERT. The present article is concerned mainly with the campaigns involving Prussia in central and eastern Europe; but the European campaigns of the war between Great Britain-Hanover and France, with the naval war on the French coast, are also noticed here.

1756.—Crossing the Saxon frontier with 70,000 Prussians on Aug. 29, 1756, Frederick entered Dresden, the Saxon capital, on Sept. 10, while the Saxon Army, not more than 20,000, fell back on Pirna to the southeast. The elector and his minister, Heinrich von Brühl, were offered assurances of Prussia's good intentions, which they naturally mistrusted. An Austrian force of 32,000 under M. U. von Browne, was moving from Bohemia to unite with the Saxons. To prevent this junction, Frederick advanced southward into Bohemia; and Browne was heavily defeated at Lobositz (Lovosice, midway between Dresden and Prague) on Oct. 1. Returning to Saxony, Frederick received the capitulation of the Saxons at Pirna (Oct. 16), whereupon he took nearly all of them into his own service. The elector and Brühl retired to the former's kingdom of Poland.

Russia might have sent forces to Austria's help at once but for the problem of the route which they should take—namely the route across Poland. Poland was a sphere of French influence, which had maintained itself by opposition to Russian designs. For the perfect achievement of an anti-Prussian coalition, it was most desirable, as Kaunitz saw, for Russia and France to come to terms. The Russians, however, saw the new contingency as an occasion for extracting concessions from France with regard not only to Poland, but also to Sweden and to Turkey. The views of the French foreign ministry, ready to admit a swift passage of Russian troops across Polish territory in order to smash Prussia and so to relieve France of the obligation to help Austria, came into

conflict with the *secret du roi*, the primary purpose of which had been to exclude the Russians from Poland at any cost.

In Great Britain, the accession of the elder William Pitt (*see* CHATHAM, WILLIAM PITT, 1st Earl of) to office in November 1756 was to have a decisive effect on the development of the war.

1757.—After weeks of negotiation at cross purposes, Douglas, having returned to St. Petersburg as France's official agent, obtained Russia's accession to the First Treaty of Versailles by signing a secret promise of French help to Russia in the event of an attack by Turkey (Jan. 11, 1757, new style; Dec. 31, 1756, old style). This contradiction of the long standing Franco-Turkish entente was immediately disavowed by the French government; and a personal letter from Louis XV to Elizabeth, the first of an important series, finally secured Russia's accession to the treaty without the objectionable appendix (April 19). An Austro-Russian offensive alliance against Prussia was concluded on Feb. 2, 1757, the parties each undertaking to put 80,000 men into the field and forswearing any separate peace, while secret articles provided for a partition of Prussia.

On May 1, 1757, Austria and France signed the Second Treaty of Versailles, an offensive alliance against Prussia, with further provision for territorial adjustments elsewhere. Austria was to recover Silesia, of course, but would cede the Netherlands for partition between Louis XV and his Spanish Bourbon cousin, Philip, duke of Parma, Piacenza, and Guastalla, whose Italian possessions should then revert to Austria. Militarily, France was to maintain 105,000 men in Germany, in addition to the contingent to be supplied to Austria (which was raised from 24,000 to 30,000), and was to grant an annual subsidy of 12,000,000 livres to Austria. Bernis became French foreign minister in succession to Rouillé.

By a large majority of votes in the Council of Princes of the *Reich*, Austria had secured the declaration of a "war of the Empire" against Prussia. Though Hesse-Kassel, Brunswick, and, naturally, Hanover opposed it, some Protestant states supported Austria, despite Frederick's attempt to pose as the champion of Protestantism against an Austro-French Catholic coalition.

In central Europe the Prussians in April 1757 advanced into Bohemia again. In the Battle of Prague, on May 6, the 66,000 Austrians under Browne and Prince Charles of Lorraine were routed by Frederick's force of 64,000 before they could be reinforced by another Austrian force under L. J. von Daun (*q.v.*). The Austrians lost more than 14,000 men; 16,000 escaped to join Daun, the rest took refuge in Prague itself, which the Prussians, whose losses were about the same, proceeded to besiege. A month later Daun, with more than 50,000 men, moved forward to relieve Prague; and Frederick went to meet him with 34,000. The Battle of Kolin, on June 18, was a great victory for Daun (8,000 Austrian losses, 13,000 Prussian). Raising the siege of Prague, the Prussians evacuated Bohemia.

Prussia meanwhile was exposed to attack from several directions. The French had begun their campaign in the spring by sending L. C. d'Estrées with 100,000 men against the Hanoverians and their allies, who formed the so-called Army of Observation under William Augustus, duke of Cumberland, a younger son of George II. Defeated at Hastenbeck, on the Weser southwest of Hanover, on July 26, 1757, Cumberland withdrew to Stade, near the Elbe estuary, abandoning the defense of the electorate and of Brunswick. A few weeks later, on Sept. 8, the duc de Richelieu (L. F. A. du Plessis), to whom the French command had been transferred, forced Cumberland to sign the Capitulation of Klosterzeven, which stipulated the disbanding of the Army of Observation. Richelieu then advanced on Prussia's western frontier; and at the same time another French army of 24,000, under the prince de Soubise (Charles de Rohan) was crossing Franconia, to join the Army of the *Reich*, namely Austria's German allies, under Joseph of Saxe-Hildburghausen. Furthermore, Sweden, having signed an alliance with France and Austria on March 21, invaded Prussian Pomerania in September, with the intention of annexing it, whether Russia approved or not.

A Russian army of 90,000 men, which had begun to cross Polish territory in May, at last entered East Prussia in August 1757; and on Aug. 30 its general, S. F. Apraksin, inflicted a crushing defeat on the Prussians under Hans von Lehwaldt at Gross-Jägerndorf west of Gumbinnen. To the stupefaction of the world, Apraksin then began a retreat, pleading difficulties of supply. It seems that his conduct was caused, partly at least, by a consideration which was long to bedevil Russian affairs—the fact that the health of the empress Elizabeth, who hated Prussia, was notoriously uncertain, while her heir, the future emperor Peter III, was a self-confessed adorer of Frederick and an opponent of the anti-Prussian war. Any general or statesman who did too much harm to Prussia was therefore risking the displeasure of his future master.

Frederick, with Saxony as his main base, decided first to confront the danger from the west, leaving Frederick Francis of Brunswick-Bevern to face the Austrians in Silesia. To prevent Richelieu's joining Soubise and Saxe-Hildburghausen, he marched first toward Halberstadt; but Austrian successes in Silesia, where Brunswick-Bevern was defeated at Moys (Zgorzelec) on Sept. 7, made him turn eastward again, while Ferdinand (*q.v.*) of Brunswick remained to observe Saxe-Hildburghausen. A daring Austrian raid on Berlin caused further diversion of Frederick's forces. Finally, hearing that Soubise and Saxe-Hildburghausen were together on the move in Thuringia, Frederick went to meet them.

The Battle of Rossbach followed on Nov. 5, 1757. The combined strength of the French and the Army of the *Reich* was at least 41,000, against the 21,000 Prussians; but the aggressive Saxe-Hildburghausen and the more cautious Soubise were at variance; and when at last they began the engagement for which Frederick had hoped, the greatly superior mobility of the Prussians, with the brilliant cavalry leadership of F. W. von Seydlitz (*q.v.*), won the day. In less than two hours' fighting, the Allies lost 7,000 men, the Prussians only 550.

Encouraged by the news of Rossbach, the British government repudiated Cumberland's Capitulation of Klosterzeven. It decided to reinforce the Hanoverians and to transfer the command in western Germany to Ferdinand of Brunswick. In September, a British naval expedition against the French base of Rochefort had been a failure.

In Silesia, the Austrians took Schweidnitz (Świdnica) on Nov. 11 and Breslau (Wroclaw) on Nov. 22. Then Frederick arrived by forced marches from Thuringia to support Brunswick-Bevern. In the Battle of Leuthen (Lutynia), on Dec. 5, 1757, he won the greatest of his victories. With 43,000 men, he attacked the 72,000 under Charles of Lorraine and utterly routed them by an unexpected cavalry charge followed by a heavy bombardment from his artillery. As against Frederick's losses of rather more than 6,000, Charles lost 22,000 men, including 12,000 taken prisoner. Breslau fell to the Prussians again.

In the course of the winter, Lehwaldt drove the Swedes back into their own part of Pomerania, where they were able to hold the Prussians outside Stralsund. Frederick's miraculous energy and masterly handling of well-disciplined troops, combined with Apraksin's retreat, saved Prussia from a situation which, after Kolin, had appeared desperate.

1758.—William Fermor, a Scottish *émigré* in the Russian service, had taken Apraksin's place in autumn 1757. On Jan. 22, 1758, the East Prussian capital, Königsberg, surrendered to him. With the onset of spring, however, the thawing of the snows made the northern roads impassable, so that Fermor was temporarily immobilized. In Russia itself, the still anti-French Bestuzhev was arrested, and power came into the hands of his rival Vorontsov.

Ferdinand of Brunswick, with his Anglo-Hanoverians, launched a successful offensive against the French in Westphalia; and on March 27 he crossed the Rhine at Emmerich, near the Dutch frontier. On June 23, with 40,000 men, he defeated the comte de Clermont (Louis de Bourbon-Condé), with 70,000, at Krefeld; and the effect of this victory, which enabled him to hold all northwestern Germany, was scarcely offset by subsequent French successes farther to the south, in Hesse and in Thuringia. Apart from reinforcing Hanover, the British on April 11 signed a new treaty with Prussia, promising an annual subsidy of 4,000,000 talers (£670,000), while both parties undertook not to make a separate peace with any of the belligerents.

Frederick began the year's campaign with an offensive in Silesia,

where Schweidnitz fell on April 16. He then advanced into Moravia, to lay siege to Olmütz (Olomouc). In July, however, by threatening Frederick's communication with his supply bases, the Austrians forced him to abandon the siege. In the north, meanwhile, a new Swedish attack on Prussian Pomerania was being fended off by Lehwaldt; but the Russians were on the march again, going southwestward from East Prussia toward the Oder River and Brandenburg.

To evade interception by the Austrians, Frederick had to march first northwestward into Bohemia, then northward across Silesia. Fermor's 52,000 Russians, having reached the Oder, began a siege of Küstrin (Kostrzyn) on Aug. 15, but Frederick was at Frankfurt an der Oder by Aug. 20. He then moved round Fermor's east flank and, with a total of 36,000 men, attacked the Russians at Zorndorf (Sarbinowo) on Aug. 25. In the bloodiest battle of the war, the Russians lost 42,000 (21,000 killed), the Prussians 13,500.

Leaving Christoph von Dohna to pursue the defeated Russians, Frederick hastened back to Saxony, to save his brother Prince Henry from attack by superior Austrian forces under Daun, outside Dresden. Daun fell back till he found a strong position at Kittlitz, where he decided to stand with his 90,000 men. Frederick, with 37,000 advanced as far as Hochkirch, never thinking that Daun would venture an offensive. Daun's attack, in the early morning of Oct. 14, took the Prussians by surprise; but Hochkirch was an expensive victory for Daun, as he lost 7,500 men (the Prussians 9,500), and he was unable to interfere with Frederick's retreat into Silesia. Daun advanced on Dresden again, but the news of Frederick's approach through Lusatia caused him to withdraw to Pirna in November.

Hochkirch in any case put new spirit into the French, who after Krefeld and Zorndorf had been inclined to despair of their European war. The duc de Choiseul (*q.v.*) became foreign minister in December, in the place of the exhausted Bernis, whose overtures for peace had been scorned by the British government.

1759.—The Third Treaty of Versailles, already drafted by the end of December 1758, was signed in March 1759 and ratified in May. By this, the French obligations of direct help to Austria in men and money were considerably reduced, and the plan of 1757 concerning the Netherlands and Parma was discarded. The French, however, were still to maintain 100,000 men in Germany.

On April 13 Ferdinand of Brunswick, who had advanced against the French in southwestern Germany, was defeated by the duc de Broglie (Victor François) at Bergen, near Frankfurt am Main; and on July 9 Broglie took Minden on the Weser, which opened the way into Hanover again. When Marshal L. G. E. de Contades joined Broglie, the French had 60,000 men against Ferdinand's 43,000 Anglo-Hanoverians; but on Aug. 1, in the Battle of Minden, Ferdinand contrived to lure Contades into an engagement which, thanks partly to accident, partly to extraordinarily stubborn fighting by British regiments, resulted in the complete rout of the French.

Choiseul had relieved France of the heavier commitments toward Austria in order to prosecute the war against Great Britain with greater vigour. He planned an invasion, with landings around London and in Scotland. To transport and escort the expeditions, the Mediterranean fleet from Toulon was summoned to join the Atlantic fleet at Brest; but on its way northward the former fleet was attacked and scattered by Boscawen in the Battle of Lagos (Aug. 19), off the Portuguese coast. Meanwhile, Edward Hawke (*q.v.*) was blockading Brest. Later, when Hawke withdrew to English waters (Nov. 9), the Brest fleet took to sea. Then Hawke reappeared, and in the Battle of Quiberon Bay (Nov. 20–21) the French suffered a decisive defeat, with most damaging losses. Not only had the project of invasion to be dropped, but also British naval superiority was established for the rest of the war.

Minden, Lagos, and Quiberon Bay, together with outstanding successes in North America—the capture of Ft. Niagara on July 24 and, of far more significance, that of Quebec on Sept. 13 (*see* French and Indian War; also Wolfe, James)—made 1759 the *annus mirabilis* or "wonderful year" for the British. From its strong position, the government began negotiations for peace with France. But on the one hand its terms were too stiff, and on the other its proposals were made in association with its Prussian ally, who at the end of the year was in no situation to expect favourable treatment from his enemies. Furthermore, Austria and Russia strongly objected to France's treating separately with the British.

In May 1759 the Russian command had been transferred from Fermor to P. S. Saltykov. Advancing across Poland through Poznan into Brandenburg with 70,000 men, Saltykov defeated 26,000 Prussians, under C. H. von Wedel, at Züllichau (Sulechów) east of Krossen (Krosno), on July 23. He then moved down the Oder toward Frankfurt, while Daun, from Saxony, sent 35,000 Austrians, under E. G. von Laudon (*q.v.*), northward to join forces with him. Frederick, who had been facing Daun, promptly marched northward likewise, hoping to prevent the Austro-Russian junction; but he failed to do so. Having joined forces with Wedel and with another Prussian force under F. A. von Finck, so that he finally disposed of about 50,000 men, Frederick boldly assailed the enemy in their strong position at Kunersdorf (Kunowice), east of Frankfurt, on Aug. 12. The result was an appalling disaster for Frederick, who in six hours lost more than 18,000 men. Saltykov, however, made no immediate use of his victory; but Daun, advancing against the diminished Prussian forces in Saxony, took Dresden on Sept. 14.

Attempts at junction between Daun and Saltykov were frustrated by Frederick's skilful movements after Kunersdorf; and when Saltykov was forced by lack of supplies to retire from the scene, Frederick turned on Daun again. Finck, however, sent with more than 12,000 men to attack Daun's rear, was surprised by 42,000 of Daun's men at Maxen, south of Dresden, and had to surrender (Nov. 21). The year had been a bad one for Frederick, and the necessity of reinforcing him after Kunersdorf had precluded the full exploitation of Ferdinand's victory at Minden.

1760.—For the campaign of 1760, Russia and Austria chose Silesia as the main field of operations: Saltykov, from Poznan, was supposed to march southward (instead of westward into Brandenburg) and to join forces with Laudon. The latter, on June 23, destroyed a Prussian force at Landeshut (Kamienna Góra) and, on July 26, captured the stronghold of Glatz. Frederick meanwhile, having been watching Daun in Saxony, had first moved eastward against Laudon, then had turned back to besiege Dresden (July 12) on hearing that Daun was moving to Laudon's support. When Daun likewise turned back, Frederick raised his siege and marched in haste through Meissen and Lusatia into Silesia. While 20,000 Russians, under Z. G. Chernyshev, occupied Prince Henry of Prussia in the vicinity of Breslau, the Austrians were converging on Frederick; but on Aug. 15, at Pfaffendorf near Liegnitz (Legnica), a sudden attack launched by Laudon on Frederick's columns, in the hope of preventing their escape from encirclement, was beaten off with heavy Austrian losses. A ruse of Frederick's tricked Chernyshev into retreating, and the Austro-Russian plan for a decisive victory in Silesia came to nothing.

Most of Saxony remained defenseless against Daun, and Brandenburg was open to the Russians. A detachment under G. G. Totleben took Berlin in the night of Oct. 8–9 and was able to retire unmolested on Oct. 13, when Frederick was approaching from Silesia. Daun was able to concentrate 64,000 men around Torgau, against whom Frederick marched with about 45,000. The Battle of Torgau, Frederick's last major victory, began on Nov. 3. The Austrian artillery devastated his attacking troops, but he sent wave after wave of them forward till at last the Austrian line was broken; and on the arrival of his general, H. J. von Zieten (who should have appeared earlier), the Austrians gave up the struggle and retreated. Frederick had lost 13,000 men, Daun 11,000 (including 7,000 prisoners).

In western Germany, Broglie won a victory at Korbach on July 10, 1760, but this was offset by Ferdinand's victory at Warburg on July 31. Hanover was again saved from French invasion; but the subsequent advance of Ferdinand's troops across the Rhine was reversed by C. E. G. de Castries at Klosterkamp on Oct. 16.

George II of Great Britain died on Oct. 25. His grandson and successor, George III (*q.v.*) was far more attached than he to British, as distinct from Hanoverian, interests and had a strong

dislike for Pitt, who was the foremost exponent of the Anglo-Prussian alliance. Without the British subsidies, Prussia could not have fought on.

1761.—By March 1761, when George III's favourite, the earl of Bute (*q.v.*), became British secretary of state for the northern department, the members of the anti-Prussian coalition were at variance in their attitudes toward the war: France wanted a negotiated peace with the British; Austria desired a general congress of the powers, at which the retrocession of Silesia might have been obtained from Prussia; and the Russian empress was still bent on war till Prussia could be carved up. Formal discussions between the French and the British broke down in July because Pitt finally insisted that, whereas the British would go on supporting Prussia, the French should reduce their support of Austria to a minimum and virtually abandon Germany; and that the British should keep all their colonial conquests. Choiseul would not submit to this dictation. To counter it, he fell back on a plan which he had long had in mind—the introduction of Spain into France's war against Great Britain. In August a "Family Compact" between the two Bourbon kings, Louis XV of France and Charles III of Spain, was concluded: Spain would declare war on Great Britain if France had not obtained peace by May 1, 1762; and France would see that Spanish claims against Great Britain were met at the final peace-making. When the British government refused to declare immediate war on Spain, Pitt resigned office (Oct. 5, 1761).

Meanwhile the fighting in western Germany went on as usual. Ferdinand advanced southward from Westphalia but was repulsed by Broglie at Grünberg, between Fulda and Marburg, on March 21, 1761; and a French counterthrust into Westphalia was checked by Ferdinand at Vellinghausen, in the middle valley of the Lippe River, on July 15. By October, however, the French had made very considerable progress eastward. The capture of Belle-Île-en-Mer, off the Breton coast of France, was achieved by a British expedition in April–June.

For Prussia, Frederick's first concern was to prevent the junction, in Silesia, of Laudon's 72,000 Austrians, based on Glatz, with a Russian army of 50,000 under A. B. Buturlin. He concentrated his available forces around Schweidnitz, but after two months of skirmishing and marching the Allies effected their junction between Liegnitz and Jauer (Jawor) on Aug. 23. Cut off from the north and outnumbered by three to one, Frederick entrenched himself at Bunzelwitz, where his enemy did not dare to attack him. When Buturlin withdrew to the north in September, leaving only 20,000 Russians under Chernyshev in Silesia, Frederick could move toward Brandenburg; but Laudon took Schweidnitz on Oct. 1, so that the Austrians could count on wintering in Silesia. In Saxony, meanwhile Daun made gradual progress against Prince Henry; and on the Pomeranian coast the fortress and harbour of Kolberg (Kolobrzeg) fell to the Russians under P. A. Rumyantsev on Dec. 16. No longer sure that he could rely on a British subsidy (since Pitt had gone) Frederick saw that only luck could save him from destruction next year.

1762.—Frederick's salvation came from an event which had so often been expected in error that it could scarcely have been predicted with confidence any more—the death of the empress Elizabeth, which took place on Jan. 5, 1762. This brought the Prussophile Peter III to the Russian throne. On May 5 Peter made peace with Frederick; and on May 22 the Treaty of Hamburg was concluded between Prussia and Sweden, through Peter's mediation. Next, in June, Peter not only allied himself with Frederick for action against Denmark over his ancestral Holstein, but even instructed Chernyshev to help Frederick to expel the Austrians from Silesia. In July, when Peter was deposed and murdered, his widow and successor Catherine II countermanded his positive measures against Denmark and Austria; but she did not renew the war against Frederick.

Daun had been given the Austrian command in Silesia in Laudon's place. Before Catherine's recall of Chernyshev had become effective, when Frederick was trying to recapture Schweidnitz, Daun marched to relieve it, but was defeated at Burkersdorf (Burkatów) on July 21. His second attempt at relief was likewise defeated, at Reichenbach (Dzierżoniów) on Aug. 16; and on Oct. 9 Schweidnitz fell again to the Prussians.

In Saxony, Prince Henry and Seydlitz together won a considerable victory at Freiberg, on Oct. 29, over the Austrians and their German allies. Austria and Prussia signed an armistice on Nov. 24.

In the west, the British government declared war on Spain on Jan. 2, 1762, three months after its rejection of Pitt's advocacy of the same measure and four months ahead of the Family Compact's deadline for Spain's intervention. The Spaniards then attacked Portugal, which the British promptly reinforced. The Portuguese fortress of Almeida fell to the Spaniards on Aug. 25; and overseas they took Colonia do Sacramento, on the estuary of the Río de la Plata, opposite Buenos Aires. These Spanish successes were very heavily counterbalanced by the British capture of Havana, in Cuba, which capitulated on Aug. 13, and of Manila, in the Philippines (Oct. 5). In the same year three important West Indian islands had fallen to the British: Martinique and St. Lucia in February, Grenada in March.

In western Germany, Ferdinand of Brunswick won victories over Soubise at Wilhelmsthal (June 24) and over Prince Xavier of Saxony at Lutternberg (July 23) and took Göttingen (Aug. 16). The French had a success at Johannisberg, near Nauheim, Aug. 30, but lost Kassel on Nov. 1.

Russia's defection from the anti-Prussian alliance convinced Austria that nothing was to be gained from prolonging the war. After the removal of Austria's objections, France could soon come to terms with Great Britain, which in turn had no interest in continuing to back Prussia in a quarrel about Silesia with Austria alone. France in October induced the disappointed Spaniards to join in the negotiations with the British; and on Nov. 3, 1762, anticipating the Austro-Prussian armistice by three weeks, Great Britain and France signed preliminaries of peace, at Fontainebleau.

THE TREATIES OF PEACE

Just as there had, in theory, been two wars, the Franco-British and the Austro-Prussian, so there were two final treaties of peace.

The definitive Treaty of Paris, of Feb. 10, 1763, was concluded between Great Britain-Hanover, France, and Spain, with Portugal expressly understood to be included. By this treaty, France renounced to Great Britain all the mainland of North America east of the Mississippi, excluding New Orleans and environs, the West Indian islands of Grenada, St. Vincent, Dominica, and Tobago, and all French conquests made since 1749 in India or in the East Indies, while Great Britain restored to France the West Indian islands of Guadeloupe, Martinique, Marie-Galante, and Désirade (St. Lucia was also ceded), the Atlantic islands of St. Pierre and Miquelon, the West African colony of Gorée (Senegal), and, of course, Belle-Île-en-Mer. Spain at the same time recovered Havana and Manila, ceded Florida to the British, and received Louisiana, including New Orleans, in compensation from the French. The French moreover evacuated Hanover, Hesse, and Brunswick. The British concessions to France in the West Indies were made, partly, in order to secure the evacuation by the French of Prussian exclaves in western Germany which France claimed to be obliged to occupy pending Austria's settlement with Prussia; but a vociferous section of the British public would have preferred to retain the West Indian islands, or to retrocede Canada instead.

The Treaty of Hubertusburg, between Austria, Prussia, and Saxony, was signed on Feb. 15, 1763, at a hunting lodge between Dresden and Leipzig. Negotiations had started there on Dec. 31, 1762. Frederick, who during his alliance with Peter III had considered ceding East Prussia to Russia if Peter would help him to secure Saxony, finally insisted on excluding Russia (in fact no longer a belligerent) from the negotiations; and at the same time he refused to evacuate Saxony until the elector had renounced any claim to reparation. The Austrians wanted at least to retain Glatz, which they had in fact reconquered; but Frederick would not allow it. The treaty simply restored the *status quo* of 1748, with Silesia and Glatz reverting to Frederick, Saxony to its own elector. The only concession that Prussia made to Austria was to consent to the archduke Joseph's being elected king of the Romans.

Bute's settlement with France was mild by comparison with what Pitt's would have been, both because he hoped for a lasting

peace with France and because he was afraid that, if he took too much, the whole of Europe would unite in envious hostility against Great Britain. But Choiseul had no intention of making a permanent peace; and when France went to war with Great Britain during the American War of Independence (*see* AMERICAN REVOLUTION), the British found no support among the European powers.

Prussia emerged from the war as a great power whose importance could no longer be challenged. Frederick the Great's personal reputation was enormously enhanced, since his debt to fortune (the Russian *volte-face*) and to the British subsidy were soon forgotten, while the memory of his energy and of his military genius was kept strenuously alive.

Austria's prestige was diminished by Prussia's success. Russia, on the other hand, made one great invisible gain from the war—the elimination of French influence in Poland. The First Partition of Poland, in 1772, was a Russo-Prussian transaction, with Austria only reluctantly involved and with France simply ignored.

See also references under "Seven Years' War" in the Index.

BIBLIOGRAPHY.—E. I. Masslowski, *Der Siebenjährige Krieg nach russischer Darstellung*, 3 vol., Ger. trans. (1889–93); W. L. Clowes (ed.), *The Royal Navy*, vol. iii (1898); R. Waddington, *La Guerre de sept ans*, 5 vol. (1899–1914); Prussian General Staff, *Die Kriege Friedrichs des Grossen*, 12 vol. (1901–13); Sir R. Lodge, *Great Britain and Prussia in the 18th Century* (1923); P. Muret, *La Prépondérance anglaise, 1715–1763*, vol. xi of the series "Peuples et Civilisations," 3rd rev. ed. (1949); Z. E. Rashed, *The Peace of Paris, 1763* (1952); L. J. Oliva, *Misalliance: a Study of French Policy in Russia during the Seven Years' War* (1964). *See* also works cited in the articles on the sovereigns, statesmen, and generals mentioned in the text above.

SEVERINUS, pope from 638 to 640, was elected to succeed Honorius I (d. Oct. 638) but was not consecrated till May 28, 640 (little more than two months before his death, Aug. 2), as the emperor Heraclius hesitated to ratify his election pending his acceptance of the *Ecthesis* (*see* HERACLIUS). Meanwhile the exarch of Ravenna, supported by the Roman soldiery, occupied the Lateran and seized the church's treasure.

SEVERN (Welsh, HAFREN; Roman SABRINA), the longest river in Great Britain (about 180 mi.), rises near the Wye on the northeastern slopes of Plynlimon (Welsh Pumlumon) in Montgomeryshire and follows a semicircular course to the Bristol Channel. It drains an area of 4,350 sq.mi. (11,270 sq.km.) and its average discharge at Bewdley is 2,255 cu.ft. a second. Its course is at first southeasterly, descending in 15 mi. from 2,000 ft. to 500 ft. at Llanidloes.

At this point its tributary, the Tylwch, following the strike of the rocks, is 300 ft. lower than the Wye, less than 2 mi. away. At Llanidloes the river turns sharply to the northeast and follows the glaciated valley of the Vale of Powis (Powys) past Newtown and Welshpool. At Llanymynech it is joined by the Vyrnwy River which has also followed the strike for 12 mi. parallel to the Severn. The headwaters are dammed to form the reservoir of Lake Vyrnwy to supply Liverpool with drinking water.

The united rivers now turn eastward over a plain; the old town of Shrewsbury was here protected from invasion within one of its loops. There is no doubt that the river formerly continued in this direction, joined the Dee and found its way to the Bay of Liverpool, but during the Pleistocene Epoch (Ice Age) a large lobe of ice entered the Cheshire Lowlands from the north, ponding up a lake in the Shropshire Lowlands. The lowest escape for the lake waters might have been by the col at Newport in which case the Upper Severn would have joined the Trent, giving continuous river from Plynlimon to the east coast, but the Newport col was closed by ice and the escape occurred over the high land between Wenlock Edge and the Wrekin at Ironbridge and joined a tributary of the Stour, rapidly deepening the channel and maintaining that course after deglaciation. Through the gorge at Ironbridge the current of the Severn flows swiftly, a matter of some importance in the early iron industry of Coalbrookdale under the Abraham Darbys, the third of whom built the first cast-iron bridge, now disused by traffic (*see* BRIDGES). The Severn continues across the soft red marls of the Triassic, receiving the Stour at Stourport and passing through Worcester, whose cathedral stands on a cliff rising from its steep left bank. The Teme enters from the west just below Worcester, and the Avon from the northeast at Tewkesbury, a yachting and motorboat centre. At Gloucester the river becomes tidal and meanders in increasing amplitude to the sea.

Navigation on this part of the Severn is difficult and is bypassed by a ship canal, more than 10 ft. deep, which leaves the estuary at Sharpness. Opened in 1827, it admits vessels carrying cargo to Gloucester. At Worcester the river is joined by the Worcester and Birmingham Canal and at Stourport by the Staffordshire and Worcestershire Canal connecting with the Midland Canal System, now little used. The connections with the Thames navigation and with the Kennet by the Kennet and Avon Canal are disused. The estuary widens gradually between South Wales and Devon, and the tide, approaching from the west, gradually increases in height and finally the spring tides and the tidal bore (a wave of about 5 ft. in height which reaches up to Newnham) check the outward flow of the river as far as Gloucester. The estuary broadens to more than 2 mi. just above the confluence of the Wye and then narrows to less than 1 mi. where there is a ferry from Aust cliff to Beachley. At spring tides the ferry is suspended for 3 or 4 hr. because of the low water exposing mud. The last railway bridge is the Berkeley viaduct (1,194 yd.) between Sharpness and Lydney, but the Severn Tunnel, 15 mi. further downstream, is 4½ mi. long and carries the railway traffic to South Wales under the English stones which narrow the estuary there. A 3,240-ft.-span suspension bridge over the estuary was constructed in the 1960s, forming part of a road link connecting London with the South Wales industrial area.

The total annual discharge of the Severn is the largest of British rivers and the level of the river rises and falls rapidly in accordance with the variation of rainfall, causing sudden floods, especially in winter. In addition, 78 cu.ft. per sec. are sent to Liverpool from the Vyrnwy. The atomic power station (opened 1962), built on the flats at Berkeley, uses Severn water for cooling purposes. The Severn is a good salmon river and is noted for lampreys. Many of its tributaries, such as the Teme, Vyrnwy, and Avon abound in trout and coarse fish. (AU. M.)

SEVERNAYA ZEMLYA ("Northern Land"), an archipelago belonging to the U.S.S.R., and lying in the Arctic Ocean north of Cape Chelyuskin, the most northerly point of the Eurasian mainland, from which it is separated by Vil'kitskogo Strait. It consists of four large islands—Oktyabr'skoy Revolyutsii, Bol'shevik, Komsomolets, and Pioner—and a number of smaller ones, with a total area of 14,286 sq.mi. (37,000 sq.km.), of which 6,680 sq.mi. (17,300 sq.km.) is ice-covered. The highest elevations, reaching about 3,100 ft. (960 m.), are found at the summits of the ice rises (domes of ice covering land). The ice is mainly in the form of gently sloping ice caps, some of which reach the sea, especially on the east coast, and give birth to icebergs. Komsomolets has an ice cap occupying 65.5% of the surface. The ice cover decreases farther south: on Oktyabr'skoy Revolyutsii it is 55.3% and on Bol'shevik only 29.3%. The islands are chiefly of Paleozoic sedimentary rocks (sandstones, limestones, dolomites), with small intrusions in the south. Vegetation is mainly lichens and mosses, and the fauna includes lemming, arctic fox, polar bear, and wild reindeer. Many sea birds come to nest.

The islands were first reported by the Russian Arctic Ocean Hydrographic Expedition under B. A. Vilkitskiy in 1913, and were explored by a Soviet party under G. A. Ushakov in 1930–32. Soviet weather stations have been permanently manned since the 1930s; there are no other inhabitants.

See G. A. Ushakov, *Po nekhoshdenoy semle* ("In Unexplored Land") (1951). (T. E. A.)

SEVERODVINSK (formerly MOLOTOVSK), a town and seaport of Archangel Oblast' of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the shore of the Dvinskaya Guba (Dvina Gulf) of the White Sea, the so-called Summer Coast, at the western edge of the delta of the Northern (Severnaya) Dvina. Pop. (1959) 78,657. The town was founded after the Revolution as an outport of Archangel, with which it is linked by rail. Timber and timber products from the Northern Dvina basin form a large part of its exports. There are small-scale food processing industries and a fishing fleet is based there. (R. A. F.)

SEVERO-OSSETIAN AUTONOMOUS SOVIET SOCIALIST REPUBLIC: *see* OSSETIA.

SEVERUS, FLAVIUS VALERIUS, Roman emperor 306–307, a Pannonian officer, was appointed *Caesar* to Constantius I on May 1, 305, to rule Pannonia, Italy, and Africa. On Constantius' death on July 25, 306, he was made *Augustus* by Galerius, but he made himself so unpopular in Rome by holding a census of the hitherto tax-free population that on Oct. 28 a revolt broke out headed by Maxentius. Severus invaded Italy next spring but his troops deserted him and he took refuge in Ravenna. He surrendered to Maximian on condition that his life would be spared, but was shortly executed. (A. H. M. J.)

SEVERUS, LUCIUS SEPTIMIUS (146–211), Roman emperor 193–211, was born in Leptis Magna in Tripolitania, North Africa. His grandfather (or perhaps great-grandfather), whose names he bore, is among the earliest fully romanized Africans known to us, described by the poet Statius, in the reign of Domitian (81–96), as an outstanding young man of the equestrian order, nothing foreign in his outlook, through and through an Italian, not African. He owned considerable estates near Rome. When Trajan made Leptis a Roman colony, the elder Severus was chief magistrate in the year of the change, the last *sufes* of the old native town, the first duumvir of the new Roman colony. His two elder sons became senators, one of them receiving a second consulate, an honour reserved to the most distinguished senatorial political leaders. The third son, the future emperor's father or grandfather, remained an equestrian in Leptis.

Young Severus already belonged to the upper ranks of the senatorial aristocracy when he entered the Senate, about 173. He served the normal junior offices of a young senator in Rome, Spain, Africa, and probably Syria. After the early death of his first wife, he married Julia Domna, daughter of a priestly family of Syria. During the political predominance of Commodus' minister Perennis, both Severus and the future emperor Pertinax were out of office, Severus spending his retirement studying the antiquities of Athens. Severus and Pertinax both returned to office on Perennis' fall in 185. Severus governed Gallia Lugdunensis during the critical years of the rising of deserters and poor men, and after a year's quiet proconsulate in Sicily became consul in 190.

In the following year, the irresponsible excesses of the emperor Commodus (*q.v.*) began to verge on madness, to the dismay of the effective heads of the government, the prefect Laetus, the chamberlain Eclectus, and the emperor's mistress Marcia. In 191 Pertinax became prefect of the city; Clodius Albinus and Pescennius Niger were appointed to the most powerful armies of the West and East, Britain and Syria; while Severus took over the largest Danubian army, in Upper Pannonia, with his brother Geta in command of another Danubian army. These appointments were scarcely accidental, since those who made them knew that Albinus was "most loved" of the senatorial Latin aristocracy of the West, that Niger was the "darling of the East," and that the brothers Severus and Geta were politically associated with Pertinax. Commodus' growing insanity made a vacant throne a dangerous probability; West and East both received the senior commander each respected, but between the potential rivals stood Pertinax and the superior forces of Severus and his brother.

The crisis came on the last day of 192, when Commodus' lunacy constrained his advisers to murder him. It passed without disturbance; Pertinax was acclaimed before dawn on Jan. 1, 193, and accepted by the legions and the guards. But when Pertinax was murdered three months later by mutinous guards, who then sold the throne to Didius Julianus, Severus marched swiftly on Rome as the avenger of Pertinax. Niger was proclaimed emperor in the East, but Albinus was bought off with the title of Caesar, heir apparent. Severus disbanded the guards, formerly recruited from the cities of Italy and Noricum, and replaced them with new guard regiments, whose members were promoted from the legions. He marched east to defeat Niger (194), and then fused the victorious and defeated legions in a campaign in northern Mesopotamia. But he could not prevent a significant number of Niger's troops, including numerous technicians, from taking refuge in Parthia, where their knowledge contributed to the increased military efficiency of the enemy in the next generation.

In 197 came the rupture with Albinus, who marched on Rome from Britain. After a rapid march through German territory and the Alps, Severus caught him from the rear and defeated him near Lugdunum (Lyons). Severus now applied the lessons of the civil war: 30 or 40 senatorial and equestrian partisans of the rebels were executed, their confiscated property forming a new treasury department; the three large armies of Britain, Syria, and Upper Pannonia were divided, so that no commander should again control more than two legions; the army's pay was increased, for the second time in 250 years, and its outstanding grievance, that a soldier's marriage was not legally valid, rectified—for a disaffected army was an unreliable defense and a political danger.

The immediate conflicts resolved, Severus heavily emphasized that his rule was to be a return to the golden days of the Antonines. His son Bassianus, better known as Caracalla, had become joint emperor under the name of M. Aurelius Antoninus, and Severus himself took the name of Pertinax, became the putative son of Marcus Aurelius, and asserted a long pedigree back to Nerva on innumerable inscriptions. In practice, the Antonine policies continued. After 197 the Senate and senators were respected, while the work of the great Severan jurists, especially Papinian and Ulpian, continued and consolidated the administrative and legal practice of the Antonines. The lively intellectual activity of the empress Julia Domna strove to reconcile the still discordant concepts of Greek and Latin thought. The political quiet was disturbed only by the dismissal and execution, in 205, of the overweening prefect Plautianus, a near relative of the emperor, and by a war against brigands in Italy. Abroad, a Parthian campaign in 198–202 was conducted as a methodical consolidation and extension of the frontier defenses. After six years in Italy, where the murderous mutual enmity of his two sons, Caracalla and Geta, became openly apparent, Severus took personal command in Britain, reviving Agricola's abandoned dream of subduing the whole island. His armies penetrated at least as far north as Aberdeenshire but were defeated by bog and forest, harassed by guerrilla enemies who would never stand to battle, and killed by disease rather than by weapons, until Severus died at York in February 211.

Severus' reputation suffers because it is depicted by two senatorial contemporaries, Herodian and Dio, who had known the happy times before Commodus and blamed Severus for the disasters of his successors' reigns. To modern writers also his rule appears as the turning point between stability of the 2nd century and the chaos of the 3rd. The essence of that judgment is that he tried without success to put the clock back.

See also references under "Severus, Lucius Septimius" in the Index. (JN. R. M.)

SEVERUS, MARCUS AURELIUS ALEXANDER: *see* ALEXANDER SEVERUS.

SEVERUS ALEXANDER, MARCUS AURELIUS: *see* ALEXANDER SEVERUS.

SEVIER, JOHN (1745–1815), American frontiersman, Revolutionary War soldier, and first governor of the state of Tennessee, was born in New Market, Va., Sept. 23, 1745. In 1773 he moved westward across the Allegheny Mountains with his wife and children to settle in the region that is now eastern Tennessee. The following year he served as a captain in Lord Dunmore's War against the Indians. During the American Revolution he took part in the Battle of King's Mountain in 1780 and in the next year served under Gen. Francis Marion against the British in the Carolinas and Georgia. In 1784 Sevier took part in the settlers' revolt against North Carolina that led to the formation of the separate state of Franklin and was elected its first governor. Many of the settlers, however, were hostile to Sevier and by 1788 the state of Franklin had collapsed and Sevier had fled into the mountains. The following year, with the help of friends in the East, he regained favour, was elected to the North Carolina senate, and in 1789–91 served in the U.S. House of Representatives.

After North Carolina finally ceded its western territory to the new federal government in 1790, Sevier was a leader among the settlers of the region and when it was admitted to the Union in

1796 as the state of Tennessee he became its first governor. He served for three terms (1796–1801), the maximum allowed by the constitution, and then, after an interval, for three more terms (1803–09). After retiring from the governorship he was elected to the state senate and then to the U.S. House of Representatives, where he served until his death on Sept. 24, 1815. He died near Fort Decatur, Ga., while serving as commissioner to determine the boundary of Creek lands in Georgia.

See C. S. Driver, *John Sevier: Pioneer of the Old Southwest* (1932).

SÉVIGNÉ, MARIE DE RABUTIN-CHANTAL, MARQUISE DE (1626–1696), the writer whose letters not only constitute one of the finest monuments of French literature, but also stand as an epoch-making model for the epistolary genre in any language. She was born on Feb. 5, 1626, in the Place Royale (now Place des Vosges) in Paris. Her father, Celse Bénigne de Rabutin-Chantal, belonged to the old nobility of Burgundy and was a son of the future saint, foundress of the Order of the Visitation (*see* CHANTAL, SAINT JANE FRANCES DE); but his family had been shocked at his marriage, in 1623, to Marie de Coulanges, whose lowborn father, Philippe I de Coulanges (d. 1636), had made a fortune out of tax collecting.

Though her father was killed in battle in 1627 and her mother died in 1633, the future letter-writer's childhood was happy. Brought up by her uncle Philippe II de Coulanges (1595–1659) and his wife Marie Lefèvre d'Ormesson, in the great house in the Place Royale or on the magnificent Coulanges domain at Sucy-en-Brie, she spent her youth surrounded by uncles almost young enough to be her brothers and by cousins (including the songwriter Philippe Emmanuel de Coulanges) only slightly junior to her. She was educated at home and learned Italian perfectly, with some Latin and Spanish. The famous Jean Chapelain and Gilles Ménage (*qq.v.*) were among her teachers; but her uncle Christophe de Coulanges (1607–87), the "most kind abbé" of the letters, had nothing to do with her upbringing, as he already had the abbacy of Livry at the time of her birth.

Marie's background was thus entirely bourgeois. She was introduced into court society and the *précieux* world of the Hôtel de Rambouillet (*see* RAMBOUILLET, CATHÉRINE DE VIVONNE, Marquise de) only after her marriage, in 1644, to Henri de Sévigné, a Breton gentleman of old nobility who bore the courtesy title of marquis and owned the château of Les Rochers, near Vitré. A lamentable creature, he squandered her money before being killed in a duel for a strumpet on Feb. 5, 1651. He left his widow with two children of great beauty: Françoise Marguerite (1646–1705), the future comtesse de Grignan; and Charles (1648–1713), baron de Sévigné.

At this point the abbé Christophe de Coulanges assumed control of his niece's affairs and remained her trusted adviser and friend for the rest of his life. For some years she enjoyed life in the fashionable world, exposed to its dangers but never succumbing to them. Then she devoted herself to bringing her children up. If she was unduly tolerant of their shortcomings, at least she gave them an admirable grounding of culture.

In 1658 Mme de Sévigné broke a promise to lend money, for military equipment, to her cousin Roger de Rabutin, comte de Bussy (*q.v.*). He avenged himself by composing a satire on her, scathing in its cumulative effect though each individual stroke may have been justifiable. Written for his friends' eyes only, this satire was included in the *Histoire amoureuse des Gaules*, published without his consent in 1665, when he and she had made up their quarrel. Their reconciliation had come about after the arrest of the superintendent of the French finances, Nicolas Fouquet (*q.v.*), in 1661. Some quite innocent letters from Mme de Sévigné (on family business) were found among Fouquet's papers, and scandalmongers alleged that they were part of Fouquet's hoard of love letters. Though Chapelain, Ménage, Madeleine de Scudéry, and others rallied to her defense, Bussy alone succeeded in vindicating her reputation, though she had not appealed for his help.

Mme de Sévigné's beloved daughter, whose beauty had caused a sensation at Louis XIV's court, was married in January 1669 to François Adhémar de Monteil, comte de Grignan (d. 1714). Ugly, already twice a widower, and heavily in debt, Grignan was of the most ancient nobility, a former army officer with a splendid bearing and polished manners. Though he owned a palatial château in Provence, he resided in Paris, and Mme de Sévigné hoped that her daughter would remain near her. Soon after marriage, however, Grignan was appointed king's lieutenant general of Provence; and on Feb. 4, 1671, his wife, having stayed behind to bear her first child, left Paris to join him.

Separation from her daughter meant the beginning of acute loneliness for Mme de Sévigné—in the midst, still, of kinsmen and friends. From this loneliness grew the greatest and most important part of her literary achievement, her letters to Mme de Grignan, which, however, she wrote without any literary intention or ambition. Though she and her daughter were parted for an aggregate of less than ten years out of the total from 1671 to 1696, the volume of the letters and the intensity of the writer's grief give the impression of a much longer time. Without her daughter she felt that she had nothing.

Of course there is more than maternal solicitude in the letters, numbers of which, moreover, are addressed to other correspondents than her daughter. They recount current news, events in worldly society, and details of every sort of the writer's life from day to day—her household, her acquaintances, her visits, and her taste in reading (with an inclination to Jansenism paradoxical in one so ready for amusement). For the historian she in fact provides little that he could not discover elsewhere, but her manner of telling her stories makes her version of them unforgettable. Three examples, which she relates not as an eyewitness but from hearsay, may be cited: the suicide of the master cook François Vatel, during a visit of Louis XIV to Chantilly (letter of April 26, 1671); the death of Turenne (letters of July 31 and Aug. 2, 1675); and the great promotion of the Knights of the Holy Spirit at Versailles (letter of Jan. 3, 1689). Once her imagination had been caught by an incident, her sensibility and her powers as an artist were released in narrative.

Mme de Sévigné took no literary model for her artistry. She is known to have been a brilliant talker; and she wrote just as she talked. While she can gallop into a period no less confidently than the duc de Saint-Simon, she can also produce startling effects by suggestive ellipsis. Her style is one of swift movement, even in her descriptions of natural scenery. Before her, critics had held that epistolary literature should conform to certain rules of composition and should observe a unity of tone ("serious" or "playful," "eloquent" or "didactic"). Her letters demonstrate that the only rule is that of spontaneity, a completely natural disorder, the loose manner of conversation, the mixture of the tones; and occasional lapses from linguistic or syntactical correctness can be passed without comment.

When she was not in Paris, Mme de Sévigné spent her last years on visits to Les Rochers, to Livry, to watering places (Vichy and Bourbon), or to Provence. At Grignan from the summer of 1694, she died there of a "continuous fever" (influenza, not smallpox) on April 17, 1696.

BIBLIOGRAPHY.—*Editions:* Six letters from Mme de Sévigné were printed in her cousin Bussy's *Mémoires* (1696), and more than 100 in his *Lettres* (1697). Meagre selections of her letters to Mme de Grignan appeared in two separate editions in 1726, one by Thiériot, the other by Bussy's son, Celse Roger de Bussy-Rabutin. Denis Perrin produced a more ample edition in 1734–37 (614 letters, in 6 volumes) and another, still more ample, in 1754 (772 letters, in 8 volumes). Perrin however "improved" the text according to his own notions, dressing it up to conform to the later standards of Louis XIV's reign, whereas the author's style was really that of an earlier generation. Apart from the publication of three previously unknown groups of letters to particular correspondents in 1756, 1773, and 1814, the next important edition was that by L. J. N. Monmerqué in 1818–19 in 10 volumes (1,40 letters, but not all from Mme de Sévigné); with textual criticisms and annotations which made a great advance on Perrin.

The so-called Grosbois manuscript, an 18th-century transcription came to light in 1820. Full of miscopyings and far from complete, yet contained new letters and new readings of others which seem authentic. Monmerqué went to work again, to produce, for the series Grands Écrivains de France, the monumental edition which Adolphe Regnier completed for him, in 14 volumes (1862–65). In 1872, however, Charles Capmas discovered another manuscript transcription, in six volumes, from which the entire contents of the Grosbois had demonstrably been selected; but extracts from this new material were

published in a strange and misleading presentation (1876), partly out of respect for the "definitive" text of the *Grands Écrivains.* At last, in 1953–57, a three-volume edition by E. Gérard-Gailly appeared in the *Pléiade* series, containing 1,155 letters, with innumerable new passages and thousands of textual restorations.

Biographies and Studies: Of fundamental importance to the study of Mme de Sévigné are C. A. Walckenaër, *Mémoires touchant la vie et les écrits de Marie de Rabutin-Chantal* . . . , 5 vol. (1842–45), with a 6th vol. by A. Aubenas (1865); P. Mesnard, "Notice" prefixed to the *Grands Écrivains* edition (1862); and J. Lemoine, *Madame de Sévigné, sa famille et ses amis* (1926–). There are general appreciations by C. A. Sainte-Beuve (in his *Portraits de femmes,* in his *Causeries du lundi* and in his *Nouveaux lundis*), by A. de Lamartine (1864), by G. Boissier (Eng. trans. 1887), by R. Vallery-Radot (1888), by E. Faguet (1910), by C. Lecigne (1912), by A. Hallays (1921), by Mme M. Saint-René Taillandier (1938), and by A. Bailly (1955). There are also specialist researches in learned periodicals, one group covering Mme de Sévigné's childhood, the other her connections with Brittany. Useful works in English are E. Fitzgerald, *Dictionary of Madame de Sévigné* (1914); P. A. Grouvelle, biographical sketch in the "Carnavalet" translation of the *Letters,* 7 vol. (1928); R. Aldington, preface to a reprint from the translation of 1811 (1937); and W. S. Maugham, preface to the select translation by V. Hammersley (1955).

(E. G.-GA.)

SEVILLE (SEVILLA), the capital of the Spanish province of that name, is the chief city of Andalusia and of all southern Spain. The town is situated on the left bank of the Guadalquivir River about 54 mi. (87 km.) from the Atlantic, about the same distance north of Jerez de la Frontera, the centre of the sherry trade, and 355 mi. (571 km.) SW of Madrid by rail. Seville is an archiepiscopal see and ranks as the fourth city in Spain. It lies low in the river valley, few parts of the city being more than 30 ft. (9 m.) above sea level, and was long subject to floods; these have been mitigated by modern dikes and embankments and by the diversion of the Tagarete, a tributary of the Guadalquivir. The climate is relaxing and in summer it is apt to be excessively hot. Pop. (1960) 442,300 (mun.).

Seville is famous for the magnificently impressive ceremonies and processions of Holy Week, and for the gaieties of the *feria* ("fair") which immediately follows Easter. During the *feria,* Seville is alive with bright costumes, gypsy music, and the flamenco dancing of Andalusia. Important bullfights are held, and throughout these two contrasting weeks the city is full of visitors and many normal activities come to a standstill. Yet at all times the city has great character and a strong feeling of the south, with its many *bodegas* ("bars") open to the pavements, and restaurant meals served out of doors. Typical of many charming touches are the ceramic plaques bearing street names and other plaques indicating houses and streets which figure in the works of Cervantes. Velázquez and Murillo were both natives of Seville, and a house in the Barrio de Santa Cruz was the birthplace, in 1802, of Nicholas Cardinal Wiseman, the first archbishop of Westminster.

The City.—Old Seville is mostly an irregularly planned city, a maze of narrow, twisting streets and small enclosed squares; its houses, built and adorned in the Moorish manner, offer glimpses through doorways to beautiful little arcaded patios. There is a more spacious layout in the district near the cathedral and the Alcázar and in a few of the streets and squares. The broad Alameda de Hércules, laid out under Philip II, with its two Roman pillars bearing statues of Julius Caesar and Hercules, its gardens and rows of trees, is an exception to the normal Seville pattern. More spacious and regular planning is found outside the limits of the old Moorish and medieval city, and the Maria Luisa Park is a particularly beautiful district; its avenues lead to the elaborate buildings of the Hispanic-American Exhibition of 1929. But east of the Alcázar the Barrio (or quarter) de Santa Cruz, which was once the ghetto, is typically picturesque in the Sevillian manner; it is popular with tourists for its associations with Doña Elvira, Don Juan, and Figaro. Over the river is the centre of Seville's ceramic industry, the Barrio de Triana, with its attractive narrow streets; in the modern suburb of Los Remedios next to it is the tobacco factory built to replace the one associated with the opera *Carmen.*

In the Marcarena quarter there is a much altered section of the Roman walls, and some of the Roman aqueduct is also preserved on the eastern side of the city. The chief Roman remains, however, are those of the large town of Itálica, the birthplace of the

JOSEF MUENCH

THE TORRE DEL ORO (TOWER OF GOLD), PART OF THE ALCÁZAR PALACE FORTIFICATIONS, ON THE GUADALQUIVIR RIVER, SEVILLE. THE GIRALDA, THE BELL TOWER OF THE CATHEDRAL, CAN BE SEEN IN THE BACKGROUND AT LEFT

emperors Trajan and Hadrian, at Santiponce 5 mi. NW. The ruins of the amphitheatre are especially imposing, and the outlines of streets and houses are also visible. By mid-20th century much of Itálica had yet to be excavated from beneath fields and olive groves.

From the Arab or Moorish period the most splendid survival is the Alcázar Palace, which was begun in 1181. A decagonal, brick tower, the Torre del Oro (1220), part of the outer fortifications, is a striking feature on the riverbank; it houses a maritime museum. The splendid courts and state apartments in the Alcázar, from the Muslim or the Mudejar period when the Moorish style continued under Christian rule, may be compared in beauty with those at Granada. Other examples of Moorish building are the tower of the Church of San Marcos (once the minaret of a mosque), the Puerta del Perdón, two sides of the cathedral's Patio de Naranjos (court of orange trees), and the lower part of the Giralda, or belfry, of the cathedral. This was the minaret of Muslim Seville's chief mosque. Built about 1180–1200 by Ahmad ibn Baso for Yusuf I, it has surfaces almost entirely covered with beautiful yellow brick and stone paneling of Moorish design; its upper stages were added during the reign of Philip II. Seville also has several fine Mudejar buildings, the Casa de Pilatos and the Casa de las Dueñas being among the best.

The building of Seville's churches started soon after the conquest of 1248. Among them, San Marcos and Santa Ana are good specimens of comparatively early Spanish Gothic. The exquisite chapel of the Old Seminary is of the 15th century. The brickwork of the Convent of Santa Paula is late Gothic and is brightly decorated with Italian majolica.

For a century and a half the principal mosque, with adaptations, served Seville as its cathedral. The new Cathedral of Santa Maria de la Sede, the second largest in area of all Gothic churches, was started in 1402; most of the building was finished by 1506. It has stylistic affinities both with the late Gothic of France and with English Perpendicular. The nave and choir are flanked by vast double aisles, the external breadth being 295 ft. and the nave vault 100 ft. above the pavement. The royal chapel at the east end, with the tombs of Ferdinand III and Alfonso the Wise, was completed by 1575 in the Plateresque style, and at the northwest corner is the Baroque chapel, added in the 17th century, which serves as the parish church. The numerous windows of the cathedral range in date from the 15th to the early 19th century. Pictures include works by Alejo Fernández, Luis de Vargas, and

Murillo. The numerous chapels are enclosed by iron screens (*rejas*) of varying dates, many of them of sumptuous beauty, and many of the altars are backed by fine altarpieces of the Baroque and earlier periods. In the choir the richly carved 15th-century stalls lead on, past two huge iron screens, to the high altar and its great Gothic reredos, a towering masterpiece of carved woodwork started by the Fleming Hyamson Dancart in 1482 and finished half a century later. Before the high altar, at the festivals of Corpus Christi and the Immaculate Conception, the altar boys (*seises*) stage their unique ceremony of a dance with castanets, an old custom the origin of which is obscure.

Renaissance and Baroque churches in Seville have domes or belfries picked out with brightly coloured *azulejos*, or decorative tiles. La Magdalena, San Luis, and San Salvador are among Seville's most imposing Baroque churches. The Palace of San Telmo, once a naval academy and now the diocesan seminary, has a splendid Churrigueresque (*see* CHURRIGUERA, JOSÉ) portal, and Rococo art is brilliantly represented by the ornate work in Santa Maria la Blanca and in the chapel of San José in the very middle of the city. Severer Renaissance architecture is seen in the Casa Lonja, built from designs by Juan de Herrera, the architect of the Escorial, and finished in 1598. First used by the merchant community, the building has housed the Archivo General de Indias since 1785; the impressive staircase of red marble was built for this new purpose by Charles III. The Archivo has a superb collection of books, plans, manuscripts, and several million documents, all bearing on the history and administration of Spain's empire in the Americas.

The university, originally a school but given university status in 1502, now uses the large, imposing fortified buildings of the old tobacco factory, completed in 1757 under Ferdinand VI. It has a charming chapel and a richly sculpted Baroque portal. The museum, formerly the Mercedarian convent, has a fine collection of paintings of the Seville school, which flourished during the 16th and 17th centuries. Among the artists represented are Velázquez, Murillo, Goya, and Martin de Vos.

History.—Seville appears originally to have been an Iberian town. In the Roman period it flourished from the 2nd century B.C. onward, and was captured by Julius Caesar in 45 B.C., after the defeat of the Pompeians at Munda. Under the Empire it became the headquarters of one of the four judicial subdivisions (*conventus juridici*) set up in the province of Hispania Baetica, the capital of which was at Córdoba. Early in the 5th century the Silingian Vandals made it the seat of their empire; later the town passed under Visigothic rule. The Moors besieged and took the city in 712, and under their rule Seville flourished greatly. Idrisi speaks in particular of its great export trade in the olive oil of Aljarafe. About 741 much of this part of Spain was occupied by Syrian Arabs from Emesa, and a member of one of these Emesan families, Abu'l-Qasim Mohammed, cadi of Seville, headed a revolt of the townsmen against the Berbers in 1023 and became the founder of the Abbadid dynasty whose capital was Seville. This regime lasted under his son al-Mu'tadid (1042–69) and his grandson al-Mu'tamid (1069–91), after which the city was taken by the Almoravides. The Muslims of Spain found the later years of Almoravide rule highly oppressive, and were even ready to welcome the victorious arms of Alfonso VII who was crowned emperor in 1135. Eleven years later all Andalusia rebelled, and Almohade troops took Seville in 1147. Under the Almohades, their capital, Seville, was the scene of great prosperity and artistic endeavour, but after the decline of that dynasty Ferdinand III captured it for Christendom in 1248; temporary ruin followed as many thousands of its people are said to have gone into voluntary exile.

Seville's geographic position was too favourable for it to decline altogether as a trading port, however, and by the 15th century it was again able to derive full benefit from Columbus' discoveries and Spanish colonization in the Americas. The Casa de Contratación, or House of Trade, was established there in 1503, and Seville thus became the temporary residence of the casa's *pilotos mayores*, or chief navigational advisers, who included Amerigo Vespucci and Sebastian Cabot. For two centuries Seville held a dominant position in Spain's new world commerce, and many emigrants to the Americas sailed from its quays. Hydrographic conditions in time favoured the Atlantic port of Cádiz, and Seville, despite its great silk factories, shared in Spain's 17th-century economic decline. In 1800 an outbreak of yellow fever killed about 30,000 people, and in 1810 the city was severely plundered by the French under Marshal Soult.

Politically, Seville long had a reputation of peculiar loyalty to the throne and was therefore much favoured by the Spanish monarchs. For its loyalty in the revolt of the Comuneros, Charles V gave it the motto *Ab Hercule et Caesare nobilitas; a se ipsa fidelitas* ("Nobility came from Hercules and Caesar; loyalty from within itself"). The Central Junta against Napoleon (the committee to organize resistance to invasion by the French) was formed in Seville in 1808, and in 1823 the *Cortes* brought the king with them from Madrid to Seville. In the civil war of 1936–39 it was early occupied by the nationalists and remained in their hands; as a result its churches and other fine buildings escaped virtually intact. Since the civil war Seville has retained its position as the leading commercial centre of Andalusia and has fully regained its attraction for tourists and lovers of architecture.

(BN. L.)

Commerce and Industries.—Although an inland port, Seville was always one of the chief outlets for the wealth of Spanish trade. The windings of the Guadalquivir, from its mouth to Seville, where it is still tidal, render it dangerous for shipping. The construction of a 4-mi.-long ship canal from the Punta de los Remedios to the Punta de los Verde—two points between which navigation was especially dangerous—was therefore undertaken in 1907. On its completion, vessels drawing 25 ft. were enabled to reach Seville. Especially after World War II shipbuilding was greatly developed downstream from the town. One of the chief shipyards is owned by the Empresa Nacional Elcano and is supported by the government. The left bank quay, 4,500 ft. long, is equipped with powerful cranes and has large storage sheds.

The principal exports are sherry and other wines; Seville oranges and lemons, olives, oil, and cork; pyrites, mercury, and iron and lead ores. Hemp, jute, and farm implements are also manufactured, and there are iron foundries and a royal artillery works. Seville has a famous tobacco and cigar factory associated with Carmen in Bizet's opera of that name. Pottery has been the characteristic industry of the Triana from time immemorial, and there is a porcelain and earthenware factory in the Carthusian convent. Justa and Rufina, the city's patron saints, are said to have been potters. Nearby butane gas deposits have also increased Seville's commercial potential.

The main-line railway from Seville to Madrid goes via Córdoba and there are branch lines to Badajoz and Lisbon; other lines connect Seville with Huelva and Cádiz. It is linked by road to other main cities. From the airport at San Pablo, 7½ mi. from the city, there are services to Madrid, Valencia, and Morocco.

(SH. H.)

SEVILLA PROVINCE, one of the eight provinces forming Andalusia, centres along the lower Guadalquivir Valley. Area 5,406 sq.mi. (14,002 sq.km.). Pop. (1960) 1,234,435. The province has varied relief, being bordered northwest by the Sierra Morena and south and southeast by the Sub-Baetic ranges of the Sierras de Algodonales (3,704 ft. [1,129 m.]) and de Yeguas. Between these ranges the Guadalquivir trough forms rolling tablelands (*campiñas*) at about 150–600 ft. (50–200 m.). At Palma del Río, at the confluence of the Genil River, the Guadalquivir Valley broadens and the river more than doubles its volume. The Genil, fed by the generous rainfall and snows of the Sierra Nevada, and the Guadaira, fed from the Sub-Baetic ranges, provide a striking contrast to the meagre seasonal flow of the minor tributaries from the Sierra Morena. The rugged relief and thin soils of the Sierra Morena support few people, unlike the richer tablelands of the southern *campiñas* with their olive groves, cereals, and cotton. Cultivation in large estates (*latifundia*) preponderates around the large centres of Écija (49,762 [mun.]), Carmona (28,216 [mun.]), Marchena (20,600 [mun.]), Utrera (41,126 [mun.]), Morón de la Frontera (35,248 [mun.]), and Osuna (20,775 [mun.]). Very different are the numerous small towns along the north bank of the

Guadalquivir, many on Roman sites, practising horticulture in smaller holdings, but isolated from the main lines of communication. Below Seville, capital of the province, the navigable Guadalquivir falls only 30 ft. (9 m.) to the sea 88 mi. (142 km.) downstream. Forming a lake in classical times (Lacus Ligustinus), the marshy plains of Las Marismas are shut off seaward by littoral sand dunes, Arenas Gordas. These plains are the haunt of wild fowl, bulls are pastured there, and land reclamation projects have long been in progress. Apart from its agricultural wealth, the province yields copper, iron ore, and coal. (J. M. Ho.)

SÈVRES, a town of northern France, in the Région Parisienne, lies on the Seine River midway on the main road from Paris to Versailles. Pop. (1962) 20,119. It is famous for the porcelain factory, Manufacture Nationale de Porcelaine, established there in 1756 and moved in 1876 to buildings between Sèvres Bridge and Saint-Cloud Park. The old building contains a mixed *lycée* and also the international centre for pedagogical studies. (*See* also POTTERY AND PORCELAIN.) There is also a ceramics museum and a technical school of ceramics. There are two ammunition (percussion caps) factories. In 1920 the Treaty of Sèvres (later superseded by that of Lausanne) between the Allies and Turkey was signed at Sèvres. (RE. H. B.)

SEWAGE DISPOSAL constitutes one of the main barriers between disease organisms and man. Every city, village, and town is faced with the problem of disposing of water-carried wastes. The degree of purification that must be attained is dictated by the amount of treatment that can be accomplished by unaided natural processes—physical, chemical, and biological. The relative volumes of treated sewage and the quantity of water in a stream, lake, sea, or ocean into which the treated sewage is discharged and the use of the mixture for industrial water, recreation, or as a water supply govern the degree of treatment.

Methods of sewage treatment may be grouped broadly into (1) primary, such as sedimentation or passage through fine screens, to remove the grosser solids, and (2) secondary, such as trickling filters, sand filters, or the activated-sludge method, by which the dissolved and colloidal organic matter in the sewage is collected by microorganisms and oxidized to stable forms that will not cause odour or nuisance.

Sewage is generally divided into two classes: domestic or sanitary sewage and industrial wastes, carried in what are called "separate sewers"; and storm water or storm sewage, carried in "storm-water drains." In some cities both classes are collected in one system, in "combined sewers." In the United States no attempt is made to treat storm sewage unless it is mixed with domestic sewage; in such cases it is feasible to attempt treatment of only a small portion of the combined sewages.

The underground conduits, or sewers, constituting the system through which sewage is conveyed to the point of disposal are usually laid under public streets and form an intricate network of mains and branches. House sewers connect the plumbing systems of buildings to lateral sewers in the streets. The sewage from laterals discharges into submain or main sewers, from which it discharges into trunk and outfall sewers. Interceptors are generally laid transversely to the submain and trunk sewers to intercept the dry-weather flow of sewage and such additional surface and storm water as may be necessary. Relief sewers are provided to carry a portion of the flow from a district already served by sewers of insufficient capacity, and thus preventing overtaxing the latter. Storm-water overflow sewers are provided to carry flows in excess of the capacity of the main or intercepting sewers.

The average volume of domestic sewage is about 100 gal. per person per day. The measure of strength of the sewage is commonly based on the suspended solids and on the biological oxygen demand (BOD), which indicates the amount of oxygen required for biological decomposition of the organic matter. Both strength and volume may be markedly influenced by industrial wastes.

Cities situated on large rivers, which thus have a large quantity of water to dilute their waste waters, do not need to achieve the high degree of sewage treatment that is required, for example, of a city that must discharge treated wastes into a lake from which its drinking water is obtained.

The use of specially constructed sewers dates to the time of Babylon and ancient Greece, but only during the 19th and 20th centuries has the water-carriage system for removal of household and factory wastes been gradually adopted in the Western world. Cities such as Paris, London, Boston, and New York had sewers early in the 19th century. About the earliest application of engineering principles occurred at Hamburg, Ger., in 1842, at London in 1852, and at Brooklyn, N.Y., in 1857.

With the adoption of the water-carriage system, the transfer of wastes from home and factory to the nearest watercourse brought serious nuisance and health problems. Often the stream served the dual purpose of sewage disposal and water supply, and hence there were frequent, disastrous epidemics of cholera, typhoid fever, and other waterborne diseases. Gradually, corrective measures were devised, but at a slower rate than sewers were being built. Effective methods of sewage treatment had their beginnings only in the last quarter of the 19th century, and the greatest development did not occur until the second quarter of the 20th.

SEWAGE TREATMENT

Primary Treatment.—*Screening.*—Pieces of wood, bundles of rags, pieces of wire, and other sizable material are often found in sewage. Such material must be removed because it can damage the pumps. In most cases screens with clear openings of about 1¼ in. are used. The screens are cleaned by a mechanism actuated by an electronic timing device. The screenings are only a small part of the total solids found in sewage, but in a large city they may amount to several truckloads per day. Disposal may be accomplished by burial, incineration, or by grinding and returning the screenings to the sewage.

Grit Removal.—Gritty materials such as sand and ashes must be removed to reduce wear of pump surfaces and accumulation of grit in parts of the treatment plant where it will be a nuisance. Grit is separated from lighter sewage solids by regulating velocities of flow through channels so that the fast-settling grit will be deposited while the lighter solids are carried on. Grit is removed from the channels and buried; if it has been freed of organic material by washing, it can be used as fill dirt. The equipment used in washing grit is the same type as that used in the mining industry to separate ore from debris.

Sedimentation.—Sewage from which screenings and grit have been removed still contains suspended solids, about one ton per 1,000,000 gal. A city of about 500,000 persons has a sewage flow of nearly 100,000,000 gal. per day and consequently has the problem of disposing of approximately 100 tons per day of solids. Some of the solids will settle if the liquid is allowed to remain quiescent for a short period.

The standard method is sedimentation in a tank through which the sewage flows continuously for two to four hours. The tank units are eight to ten feet deep, round or rectangular. Round tanks employ a central feed, and the settled sewage is discharged over a peripheral weir. Rectangular tanks, generally four or five times longer than they are wide, have an inlet at one end and discharge at the other. Various methods are used to minimize currents that may interfere with the sedimentation. The solids (or sludge) that settle are concentrated in a central hopper in circular tanks, or in a hopper at the inlet end of rectangular tanks, by means of scrapers or collectors that may be operated either continuously or several times a day.

Approximately 40 to 55% of suspended solids can be removed by sedimentation, reducing the biological oxygen demand by nearly the same amount. Practically none of the dissolved solids can be removed by the process.

The German engineer Karl Imhoff developed a tank with an upper compartment for settling and a lower compartment where solids are decomposed by bacteria. The Imhoff tank is a simple device containing no moving mechanical parts. It has been used successfully in many parts of the world, but in new plants it has been gradually replaced by settling tanks with mechanisms that permit more rapid handling of sewage. Imhoff tanks are often used for settling prior to filtration through sand beds.

Chemical Precipitation.—Some cities on large rivers have

been able to treat sewage sufficiently by primary sedimentation, disposal of the settled solids, and chlorination of the effluent, the liquid remaining after sedimentation. Other cities have found it necessary to provide for chemical precipitation during periods when the rivers are low. The chemicals coagulate some of the colloidal solids into particles of a size and weight that will settle. The chemicals—commonly aluminum sulfate, ferric chloride, and ferric and ferrous sulfates—are added to the sewage and quickly mixed with it in a small tank. The mixture then passes to a flocculating tank where it is gently agitated with stirring paddles. The floc is formed in 20–40 minutes and the flocculated sewage flows to a settling tank. The amount of chemicals required depends upon the characteristics of the sewage, varying from 400 to 700 lb. per 1,000,000 gal. Addition of chemicals makes it possible to remove 80–90% of the suspended solids and reduce the biological oxygen demand by 65–75%. The amount of dissolved organic solids removed by chemical treatment is very small.

Secondary Treatment.—The effluent from primary sewage settling processes contains some suspended and much dissolved and colloidal matter that will decompose, often creating nuisance and odour. Various methods of secondary treatment are used, among which trickling filters, the activated-sludge process, and sand filters are commonest.

Trickling Filters.—A trickling filter is a six-to-eight foot deep bed of stones two to four inches in diameter. Purification is accomplished by a film of bacterial slime and other organisms on the surfaces of the stones, rather than by mechanical filtering action as is suggested by the name given to the equipment. The sewage is applied to the stones in a thin sheet or spray, with short intervals between applications, and allowed to trickle down through the bed. Application is principally by means of rotating arms onto a circular bed, although some plants used fixed-spray nozzles and square or rectangular beds. The standard rate of application is 300,000 gal. per acre per day per foot of depth, though some installations operate at rates twice as great. Trickling filters are usually enclosed in concrete walls. Openings in the walls or vents to the underdrains assure an adequate supply of oxygen for the microorganisms. After passing through the trickling filter the sewage usually goes to a secondary settling tank for removal of organic matter that has sloughed off from the stones of the filter. Trickling filters, together with primary treatment and final sedimentation, will remove 85–95% of suspended solids and reduce biological oxygen demand by the same amount.

High-rate trickling filters are so named because 10,000,000 to 30,000,000 gal. per acre per day of liquid load may be applied to their surfaces. This rate is attained by recirculating effluent over the filter along with the settled sewage. Application of sewage and effluent to the stones is continuous and the rate is usually high enough to eliminate accumulations of large quantities of slime.

The sludge that accumulates in the settling tanks of both types of trickling filters must be removed and disposed of.

Activated Sludge.—The activated-sludge method utilizes biologically active sludge mixed with treated sewage and agitated in the presence of an ample supply of air in an aeration tank for four to ten hours. Suspended solids and many organic solids are quickly absorbed or adsorbed by the activated sludge, while organic matter is oxidized by the microorganisms in the sludge. The sludge, representing about 25% of the total volume, is separated from the liquid in a secondary settling tank and then is passed through the aeration tanks again.

Aeration may be accomplished by means of compressed air blown through porous plates or tubes or admitted through specially designed jets, by means of mechanical agitators, or by a combination of these methods. Porous plates, tubes, or jets are generally set in rows either across the aeration tank or along one side. In mechanically aerated tanks, the commonest arrangement is a revolving mechanism that agitates the sewage and sludge mixture and aerates it by throwing a portion of the mixed liquor through the air or by drawing air into a downdraft tube and discharging it under the surface in bubbles.

Advantages of the activated-sludge method include flexibility, permitting almost any desired degree of treatment by varying the period of aeration, and efficiency: it removes about 95% of bacteria and more than 90% of suspended solids and of organic matter. Accumulated solids must be removed from the system and disposed of in one of several approved ways.

Sand Filters.—These are beds of sand, about 36 in. deep and underdrained, on which raw or settled sewage or the effluent from trickling filters is applied for final treatment. Sand beds produce a clear, sparkling, and highly stable effluent. The beds are periodically cleaned by backflushing.

Sludge Disposal.—Sewage sludge is the accumulated settled solids deposited in tanks or basins and containing more or less water to form a semiliquid mass. The various sludges are: (1) primary sludge; (2) chemically precipitated sludge; (3) trickling filter secondary sludge; (4) excess activated sludge; and (5) digested sludge.

Sludge may be hauled or pumped to sea, dewatered on vacuum filters, and incinerated or sold for fertilizer immediately after removal from the sewage-treatment process, or it may be digested by microorganisms under controlled conditions to reduce the organic content prior to ultimate disposal. Digestion tanks are usually circular, with diameter as required and 15 to 25 ft. deep. They are usually heated to about 95° F (35° C), optimum temperature for the microorganisms that decompose the organic matter. A standard design provides for two or three cubic feet of heated digestion tank per person contributing to the sewers; twice that volume is required for unheated tanks. About 30 days ordinarily is needed for complete digestion in a heated tank.

In the process of digestion the microorganisms produce carbon dioxide and methane—from 0.5 to 1.3 cu.ft. per day for each person contributing to the sewers—which may be collected and used to heat the digestion tanks and buildings and in gas engines that drive generators, sewage pumps, and air compressors.

At the end of the digestion period the sludge may be used in liquid form as fertilizer, but usually it is dried first, on sludge-drying beds. These consist of about 12 in. of sand supported on 6 to 8 in. of gravel, the whole being underdrained by open-joint tile spaced 8 to 10 ft. apart. The sludge can be removed from the beds after two to three weeks. If properly digested, it is not objectionable and can be used for fertilizer, or dumped. It is excellent fertilizer for grass and shrubs but should not be used to fertilize vegetables that are to be eaten raw.

Before sludge can be dried and sold as fertilizer, or incinerated part of the water must be removed by vacuum filtration. A vacuum filter consists of a drum covered with a filtering medium, a vacuum being maintained inside one quadrant. As the drum is rotated through a tank containing sludge (to which coagulating chemicals have been added), the liquid is drawn into the drum while the solids are collected as a mat on the filtering medium. The process is continuous, the sludge mat being peeled off as the drum revolves. A vacuum filter will dewater about five pounds of dry solids per hour per square foot of filter surface, and the moisture content of the sludge cake will range from 70 to 85%. Vacuum filters will dewater digested, raw, activated, or partly digested sludge.

Digested sludge or activated sludge filter cake can be dehydrated in heat driers and then bagged or sold in bulk as organic fertilizer. All sludge cake can be incinerated, and there is usually sufficient heat value in the sludge to carry on the incineration process without further addition of fuel, once the incinerator has been brought to operating temperature. Sludge may be hauled to dumping grounds at sea or pumped to suitable areas at sea.

SEWER DESIGN AND CONSTRUCTION

Separate versus Combined Sewers.—Combined sewers have been used where disposal of sewage by dilution is permissible. However, as population increases the receiving waters may become so seriously polluted as to require the installation of sewage-treatment plants; then it becomes necessary to ensure that the plants have adequate capacity to treat substantial quantities of both storm water and sewage. Moreover, combined sewers may have to be made much too large for the ordinary dry-weather flow of sew-

age alone; as a result, flow velocities may be too low to prevent formation of objectionable deposits of solids in the sewer.

Against these disadvantages, combined sewerage systems have the advantage of requiring only a single conduit in each street. In addition, the quantity of domestic sewage usually is no greater than the margin of error in estimates of the quantity of storm water; therefore, the combined sewer need not be appreciably larger than a storm-water drain for the same district. Consequently, there is a substantial saving in building the single conduit instead of two.

Design.—The arrangement of a sewer system is governed by the topography of the area to be served and the relation of the point of disposal to various points in the system. Unlike water-distribution systems, which are under pressure, sewer systems must be laid out, within the limits imposed by the terrain, so that sewage can flow by gravity from the point of origin to the point of discharge. However, since economic and engineering factors control the depth to which sewers can be laid, it becomes necessary in flat terrain to provide pumping stations in the system.

Estimates must be made of the volumes of sewage and storm water, in the case of combined systems, and the pipe sizes and slopes must be such as to provide sufficient capacity to handle maximum flows and at the same time give adequate scouring velocities at minimum flows to prevent deposits. Five factors determine the amount of liquid a sewer can convey: slope or fall; cross-sectional area and shape; character of interior walls; depth of sewage; and presence or absence of obstructions such as curves, air resistance, and the like. Formulas for the mean velocity of flow are based upon the fundamental laws of fluid mechanics. Minimum velocities of design flows should be at least two feet per second and maximum velocities not over eight feet per second. The amount of sewage depends upon the water consumption of the contributing population and the amount of groundwater infiltration and storm runoff.

Sewer Pipe.—The pipe used in sewer systems is made of vitrified clay or concrete; it is usually circular in cross section, although for large combined sewers a variety of shapes may be used to meet special conditions. Clay pipe is generally used for sizes up to 42 in. in diameter and concrete pipe for larger sizes up to 108 in. Large pipe may be of reinforced concrete cast on the site.

Miscellaneous Appurtenances.—Access manholes are provided at frequent intervals to permit inspection and cleaning. In the case of drains and combined systems, catch basins at street curbs admit surface runoff and, in instances where high flow may at times exceed sewer capacity, regulating devices are provided to direct overflows to the nearest stream. Sometimes, to pass such obstacles as building foundations and rivers, sewers must be depressed for short distances; this is done in such a manner that the sewer flows under pressure, and the depressed sections are called inverted siphons. Under certain conditions special provision must be made for ventilating equipment to remove corrosive gases which may attack concrete pipe.

See also PLUMBING.

BIBLIOGRAPHY.—H. E. Babbitt and E. R. Baumann, *Sewerage and Sewage Treatment*, 8th ed. (1958); W. A. Hardenbergh, *Sewerage and Sewage Treatment*, 3rd ed. (1950); K. Imhoff and G. M. Fair, *Sewage Treatment*, 2nd ed. (1956); K. Imhoff *et al.*, *Disposal of Sewage and Other Water-Borne Wastes* (1956); E. V. Bevan and B. T. Rees, *Sewers: Theory, Design, Specification and Construction*, 2nd ed. (1953).

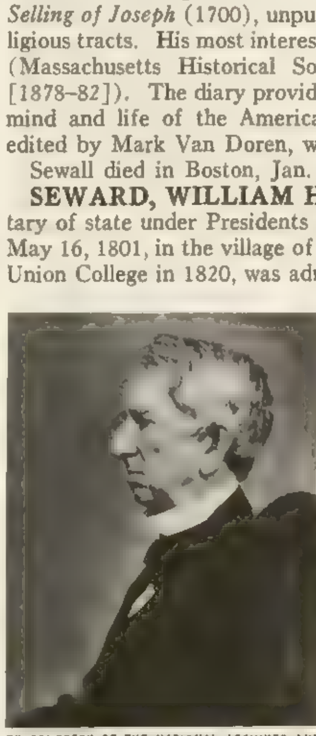
(D. E. BL.; E. S. CH.; X.)

SEWALL, SAMUEL (1652–1730), American judge of the Salem witchcraft trials, a colonial merchant and noted diarist, was born at Bishopstoke, Eng., March 28, 1652. His parents, who had previously lived in New England, returned there in 1661 and settled in Boston. After graduating from Harvard College in 1671 Sewall became a student of divinity and tutor at Harvard until he received his master's degree in 1674. A long and distinguished public career began for Sewall when he was made a "freeman" in 1679. He was manager of the colonial printing press (1681–84), member of the Council (1684–1725), and chief justice of the Superior Court (1718–28). In addition, he was a commissioner of the Society for the Propagation of the Gospel in New England, an overseer of Harvard College, captain of the Ancient and Honorable Artillery Company, and one of the special commissioners appointed to try the Salem witchcraft cases in 1692. He acknowledged, in 1697, that he had done wrong in these trials, in which 19 persons were sentenced to death.

Sewall's writings include his famous antislavery appeal *The Selling of Joseph* (1700), unpublished verses, and political and religious tracts. His most interesting and appealing work is his diary (Massachusetts Historical Society, *Collections*, 5 ser., V–VII [1878–82]). The diary provides an incomparable insight into the mind and life of the American Puritan. An abridged edition, edited by Mark Van Doren, was published in 1927.

Sewall died in Boston, Jan. 1, 1730. (RA. MU.)

SEWARD, WILLIAM HENRY (1801–1872), U.S. secretary of state under Presidents Lincoln and Johnson, was born on May 16, 1801, in the village of Florida, N.Y. He graduated from Union College in 1820, was admitted to the bar at Utica in 1822, and in 1823 began the practice of law at Auburn, which was his home for the rest of his life. Attaining distinction in his profession, he moved into politics, for which he had a greater liking, and early became associated with Thurlow Weed, then a member of the state legislature.



BY COURTESY OF THE NATIONAL ARCHIVES AND U.S. SIGNAL CORPS

W. H. SEWARD, PHOTOGRAPHED DURING THE AMERICAN CIVIL WAR BY MATHEW B. BRADY

After 1828 Seward became allied with the Antimasonic Party (*q.v.*), attending its national conventions in 1830 and 1831; as a member of this party he served in the New York senate from 1830 to 1834. By 1833, when the Antimasonic movement had run its course, he allied himself with other opponents of the Jackson Democrats and became a Whig. In 1834 he received the Whig nomination for governor of New York but was defeated by William L. Marcy. Four years later he was elected and served until 1843.

As governor, Seward favoured a policy of internal improvements, *i.e.*, building roads and canals at public expense. His administration was disturbed by antirent agitation among farmers (caused by dissatisfaction with the terms of leases) and by incidents growing out of the Canadian rebellion of 1837. During this period he attracted much attention by his liberal and humane policy; he did much to promote prison reform and proposed to admit Roman Catholic and foreign teachers into the public schools of the state. Laws were passed during his term putting obstacles in the way of recovering fugitive slaves, and Seward soon became recognized as the leader of the antislavery Whigs. He was one of the earliest political opponents of slavery, as distinguished from the radical Abolitionists who devoted themselves to moral agitation rather than to political action.

When the Whigs secured control of the state legislature in 1849 they elected Seward to the U.S. Senate where the antagonism between free labour and slave labour became the theme of many of his speeches. In his first set speech in the Senate, on March 11, 1850, he opposed the pending compromise measures (*see* COMPROMISE OF 1850) and attracted the attention of the whole country by his assertion that "there is a higher law than the constitution" regulating "our authority over the domain" (*i.e.*, the Territories). When the Democrats declared such language incendiary Seward tried to explain it away, and in so doing offended his friends without appeasing his opponents.

In the presidential election of 1852 Seward supported Gen. Winfield Scott but rejected the platform of the Whig Party because it declared the Compromise of 1850 a finality. He opposed the Kansas-Nebraska bill of 1854, which repealed the Missouri Compromise and established the principle of popular sovereignty in the Territories. As a senator he actively supported the free-state cause in Kansas. In 1854–55, when it became evident that the Whig

Party in the North was moribund, Seward helped to lead its scattered remnants into the Republican fold. As the recognized leader of this new party, his nomination by the Republicans for the presidency in 1856 and 1860 was regarded as certain; but each time he was put aside for another. The heterogeneous elements of the new organization could not be made to unite on a man who for so many years had devoted his energies to purely Whig measures, and he was considered less "available" than Frémont in 1856 and Lincoln in 1860. It was during this period, in 1858, that he referred in a speech at Rochester, N.Y., to the "irrepressible conflict" between the free and slave sections of the nation.

After Lincoln was elected in 1860 he chose Seward to be his secretary of state. The new president was a man comparatively unknown outside the state of Illinois, and many of his supporters, doubtful of his ability to deal with national problems, looked to Seward as the most experienced man of the administration and the one who should direct its policy. Seward himself apparently shared these views (not entirely out of vanity) and at first possessed an unbounded confidence in his ability to influence the president and his cabinet. He believed that the Union could be saved without a war, and that a policy of delay would prevent the secession of the border states, which in turn would gradually coax their southern neighbours back into proper relations with the federal government. He even suggested embroiling the United States in a war with England or France as a means of restoring the Union. In informal conferences with commissioners from the seceded states he assured them that Fort Sumter would be speedily evacuated. Lincoln overruled him and sent a relief expedition to Sumter without Seward's knowledge. On most issues, however, Seward remained Lincoln's closest and most influential adviser.

As secretary of state Seward rendered services of inestimable value to the nation. To prevent foreign governments from giving official recognition to the Confederacy was the task of the hour, and in this he was successful. While he did not succeed in preventing the French occupation of Mexico or the escape of the Confederate cruiser "Alabama" from England (*see* "ALABAMA" ARBITRATION), his diplomacy prepared the way for a satisfactory adjustment of the difficulties with these powers. While his treaty with Lord Lyons in 1862 for the suppression of the slave trade conceded to England the right of search to a limited extent in African and Cuban waters, he secured a similar concession for U.S. war vessels from the British government, and by his course in the "Trent" affair (*see* AMERICAN CIVIL WAR: *Politics, Economics and Foreign Affairs*) he virtually committed Great Britain to the U.S. attitude with regard to the right of search.

On April 5, 1865, Seward was accidentally thrown from his carriage and severely injured. Nine days later, while lying ill at his home in Washington, he was attacked by Lewis Powell, alias Payne, a fellow-conspirator of John Wilkes Booth, at the same time that Lincoln was assassinated. Seward's wife, an invalid, received such a shock that she died within two months, and his only daughter, who witnessed the assault, never recovered from the effects of the scene and died within the year. Seward gradually regained his health and remained in the cabinet until the expiration of President Johnson's term in 1869. In the struggle between the executive and Congress over the method of reconstructing the southern states, Seward sided with Johnson and thus shared some of the obloquy bestowed upon that unfortunate president. His greatest work during this period was the purchase of Alaska from Russia in 1867 for $7,200,000, a purchase which at that time was characterized by some as "Seward's Folly." After returning to private life, Seward spent two years in travel and died at his home in Auburn on Oct. 10, 1872.

BIBLIOGRAPHY.—The best biography of Seward is that by Frederic Bancroft, *The Life of William H. Seward*, 2 vol. (1900). *See* also G. E. Baker (ed.), *The Life and Works of William H. Seward* (1883); F. W. Seward (ed.), *William H. Seward: an Autobiography from 1801 to 1834, with a Memoir of His Life and Selections from His Letters*, 3 vol. (1877–91). (W. B. HE.)

SEWARD, a town of Alaska, U.S., at the head of Resurrection Bay, about 1,800 mi. (2,897 km.) NW of Seattle, Wash. Transportation is the chief industry. Founded in 1903 as the terminus of a railway to the Yukon Valley and named for William H. Seward, who as secretary of state negotiated for the purchase of Alaska by the U.S., Seward was chartered in 1907. In 1913 the federally owned Alaska Railroad was created and took over the railway development. The town was severely damaged by the March 1964 Alaskan earthquake and the resultant seismic waves.

The small boat harbour and airfield are used extensively by chartered hunting and fishing parties. The August silver-salmon derby is a special attraction for fishermen. Hunting is varied; moose and bear, mountain goats, small game, and birds abound. The population is approximately 2,000. (J. E. CL.)

SEWING MACHINE. The sewing machine was the first mechanical home appliance to have wide distribution and has served as a forerunner of industrial civilization in many remote areas. It remains one of the most important labour-saving devices for both industry and the home. (*See* CLOTHING MANUFACTURE; HOME EQUIPMENT: *Sewing Machines*.)

History.—The earliest recorded aid to hand sewing was the double-pointed needle with the eye at one end, patented in England by Charles F. Weisenthal in 1755. The needle was held in the middle, eliminating the need to turn it for each reentry into the fabric.

Many features of the modern sewing machine were specified in a patent secured in England by Thomas Saint in 1790, in which he described a horizontal cloth plate or table, an overhanging arm carrying a straight needle, and a spool for a continuous supply of thread. A forked needle pushed the thread through a hole made by a preceding awl, the thread being caught below by a looper and held until the needle descended again. Had Saint hit on the idea of the eye-pointed needle, his machine would have been a complete anticipation of the modern chain-stitch machine.

Early 19th-century efforts were directed toward embroidery machines. The first machine for utilitarian stitching was a chain-stitch machine, using a barbed needle and simulating a manual operation, devised and patented in 1830 by a French tailor Barthélemy Thimonnier. In 1841 Thimonnier and his associates provided 80 machines to make clothing for the French Army, but a mob of angry tailors, fearing competition, destroyed them. Thimonnier later obtained patents in England and the U.S. but gained nothing financially from either venture.

Walter Hunt of New York, about 1832–34, constructed a machine having a vibrating arm, at the extremity of which was fixed a curved needle with an eye near its point. In stitching, this needle formed a loop of thread under the cloth; another thread was passed through the loop by an oscillating shuttle, making a lockstitch. This new concept of machine sewing enabled inventors to depart from the idea that a sewing machine must imitate a hand operation.

Apparently unaware of Hunt's invention, which was never patented, Elias Howe (*q.v.*) of Spencer, Mass., invented an amazingly similar machine, which was patented in the U.S. in 1846. In both machines the fabric was held vertically by pins on a baster plate. The curved eye-pointed needle moved in an arc to a horizontal position as it entered the fabric, interlocking the needle thread with a second thread carried by a shuttle running to and fro on a track. The Howe machine could stitch straight seams the length of the baster plate, after which the fabric had to be reset.

Howe's invention was sold in England to William F. Thomas, a corset manufacturer, who in 1846 secured the English patent in his own name and engaged Howe on weekly wages to adapt the machine for his manufacturing purposes. The career of the inventor in London was unsuccessful, and, having pawned his first machine and his American patent rights, he returned in 1849 in poverty to the U.S. There, in the meantime, the sewing machine was beginning to excite public curiosity, and various inventors were making machines which Howe believed infringed on his patent rights. Among them were Charles Morey and Joseph Johnson, who used the eye-pointed needle and a hook to form the chain stitch; John Bachelder, who added an endless leather belt as a continuous feed device; and Sherburne C. Blodgett and John A. Lerow, who used a circular feeding bar that allowed continuous sewing.

The most prominent manufacturer, if not inventor, ultimately

appeared in Isaac Merrit Singer (*q.v.*), who in 1851 secured a patent for his machine, which made a lockstitch by means of a straight eye-pointed needle moving vertically and a reciprocating shuttle moving horizontally. His machine also featured a table to support the cloth, with the toothed wheel-feed protruding through the table surface. A vertical presser foot held the cloth down as the needle retracted. Alarmed by Singer's invention, Howe, having recovered his American patent rights, moved to vindicate his rights by instituting suits against the infringers of his patent. An enormous amount of litigation ensued, but in 1856 four manufacturers (including Howe) joined the "Sewing Machine Combination," whereby they agreed to pool their patent interests and pay Howe royalties, thus providing the first large-scale example of a patent pool.

Allen B. Wilson, one of the most ingenious inventors, also worked without knowledge of previous efforts. In 1849 he devised the rotary hook and bobbin combination and with the financial support of Nathaniel Wheeler obtained patents for them in 1851 and 1852, and then patented in 1854 the important and effective four-motion feed for moving the fabric after every stitch. In 1851 William O. Grover and William E. Baker, two Boston tailors, patented a double chain-stitch machine that used two threads directly from the spools.

Very few sewing machines were manufactured before 1850, yet in 1860 more than 111,000 machines were produced by 74 U.S. companies. Although thousands of patents have been issued in the U.S. and Europe for sewing machine improvements and attachments, the main principles have not been affected.

Modern Sewing Machines.—In the manufacturing industries there are more than 2,000 varieties of modern sewing machines, primarily designed for specific operations. Straight-stitch sewing machines commonly used in the home have undergone little change in the basic engineering since the early 20th century, but the use of electricity and such refinements as the hinged presser foot and the sewing light have greatly increased their popularity.

Although sewing machines are manufactured in several European countries and in the U.S., the world's largest producer is Japan. Japanese manufacturers also pioneered development of "zigzag" machines, which permit machine embroidery. Zigzag stitching is basically a lockstitch of bobbin and spool threads that zigzag as the threaded needle shifts from side to side. Manufacturers in other countries produced models of zigzag machines and straight-stitch machines with zigzag attachments. (G. R. Co.)

SEX, the sum of the features by which organisms exhibit a complementary difference, recognized in higher plants and animals as maleness or femaleness.

Beginning with mechanisms for ensuring occasional genetic recombination in primitive microorganisms, sex has undergone an evolution of its own throughout the whole history of life on earth. Elaborate patterns of behaviour, strange and complex structures, and delicate physiological equilibriums have all been developed under the influence of natural selection, ensuring the continued existence of species with an abundant supply of new combinations of genes in each generation.

This article is concerned strictly with the biological aspects of sex. Helpful background information can be found in REPRODUCTION, which is concerned with, among other things, the evolution of sex. For the anatomy and physiology of the human sex organs, *see* REPRODUCTIVE SYSTEM. For details of the fusion of gametes, *see* FERTILIZATION. The following sections constitute this article:

I. TYPES OF SEXUALITY

Although structural differences between male and female are not always apparent, as in many lower organisms, functional differences can always be distinguished. In the majority of organisms above the level of complexity of viruses and bacteria, a new individual is brought into being by the cooperation of a male and a female organism, each of whom contributes a special cell, a gamete or sex cell, toward that creation: the male donates a spermatozoon, or sperm; the female, an ovum, or egg. The affinity that sperm and egg have for each other may result in their fusion (fertilization). The single-celled product, a zygote, carries its own blueprint for development derived from each of its parents.

Sexuality is not always an either-or determination. Organisms exist that produce both sperms and eggs; these, called hermaphrodites, occur naturally in many groups of lower animals and in most groups of higher plants. Among fungi and protozoa, sex is often found in more than two genetic forms, which are referred to as mating types. In higher animals and plants, intersexes, gynandromorphs, and sexual mosaics are not uncommon; the occurrence of such intergraded forms reflects the range of possibilities that may proceed from variable action of genetic endowment and hormonal influence.

A. ANIMAL

1. Sex Cells.—The condition in which the gametes are equal in size and shape, as in some protozoa, algae, and fungi, is termed isogamy. It is a phenomenon that occurs only at the morphological level, for regardless of appearance the gametes exist in different functional forms of a complementary nature. In the higher animals anisogamy is the rule; the gametes are regularly of two kinds. The egg is in most cases a very large nonmotile cell containing considerable food reserves in the form of yolk or other substances that contribute to the development of the embryo. The sperm is a relatively small cell that swims by means of a tail or flagellum. Some exceptions exist, such as motile amoebalike ova in sponges and tailless sperms in various groups of invertebrate animals (*e.g.*, nematode worms, crabs, and lobsters).

Usually, the number of sperms produced is vastly in excess of the number of eggs, the ratio often being millions to one. Such an arrangement maximizes the probability that the egg will be fertilized and, because the sperms are so small and lack specialized food reserves, this is done with biochemical economy.

2. Bisexuality.—In most higher animals the sperms and eggs are produced by males and females, respectively. The sexes are distinguished by primary sex characters, the ovaries or testes and their ducts, and may also show secondary sex characters such as the comb and wattles in cocks, beard in man, etc. The extent of obvious sex differences (sexual dimorphism) depends on the degree of development of secondary sex characters. At one end of the scale are species in which the sexes are not distinguishable externally (certain beetles, some birds), while at the other end are forms in which there is so little resemblance between the sexes that, in some instances, they were originally described as distinct species. In the extreme case of certain marine worms and deep-sea angler fishes, the male is a relatively minute individual that lives a parasitic existence on or in the much larger female.

3. Hermaphroditism.—In an earthworm or a land snail, ova and sperms are produced by the same individual, which is re-

ferred to as monoecious or hermaphroditic. Copulation usually occurs between two different individuals, and fusion of their ova and spermatozoa gives rise to zygotes that develop into the individuals of the next generation. Obviously, however, there exists in such cases a possibility that self-fertilization may occur; *i.e.*, that the ova and spermatozoa produced by a single individual may fuse. Self-fertilization does in fact happen in a number of species of plants and animals. There are, however, various anatomical, physiological, or genetic mechanisms that may limit self-fertilization or prevent it altogether in particular species. One of these is consecutive hermaphroditism, in which the individual produces sperms and eggs at different times in its life cycle. Periodic sex reversal of this kind is widespread in marine mollusks such as oysters and slipper limpets.

Hermaphroditism is found in almost all flatworms, including the free-living planarians and the parasitic flukes and tapeworms (but with a few exceptions such as the blood-inhabiting *Schistosoma* or *Bilharzia* flukes, which are bisexual); in earthworms and leeches; and in land and freshwater mollusks. In the arthropods, including insects, hermaphroditism is only very rarely encountered. It does occur, however, in the scale insect *Icerya purchasi*, a pest of citrus trees in many countries. In the vertebrates bisexuality is almost universal, but a few species of fishes may show consecutive hermaphroditism. (*See* also HERMAPHRODITE.)

4. Variations in Sexual Behaviour.—In all higher vertebrates psychological stimuli seem to play a part in the control of anterior pituitary activity, which in turn determines the activity of the sex organs, or gonads. There is thus a very complex chain of relationships with "feedback" mechanisms of several kinds (*e.g.*, the gonadal hormones tend to depress pituitary activity). This elaborate series of highly labile physiological mechanisms is developed in each individual as a result of genetic equilibria involving several complexes of genes on the sex chromosomes and the autosomes; it is therefore not surprising that sexual behaviour should be highly variable. In the human species, where it is undoubtedly modified and influenced by psychological factors to an even greater degree than in other higher vertebrates, sexual behaviour is especially variable.

To distinguish between "normal" and "abnormal" behaviour is almost meaningless, except insofar as the term "normal" may be used to designate a statistical average. The strength of the sex drive in both sexes of the human species exhibits an extreme range, from very low to very powerful activity, and this primary sex drive tends to be associated with other creative activities and outlets. The human sex drive may be concentrated entirely on individuals of the opposite sex, but not infrequently it may be directed toward individuals of the same sex (homosexuality; *q.v.*).

Many mammals exhibit occasional homosexual behaviour and in some species it may be manifested quite regularly; *e.g.*, in the giraffe. In the human species, the frequency of homosexual activity and its relation to heterosexual behaviour vary greatly from one culture to another: thus the Siriono of eastern Bolivia are said to hardly ever engage in homosexual practices, while among the Arunta (Aranda) of central Australia almost all the men are reported to practise homosexuality to some extent. According to A. C. Kinsey and his collaborators, about 10% of U.S. males had shown exclusively or mainly homosexual behaviour for three years or more prior to interview. The incidence of a comparable degree of homosexuality among women (lesbianism; *q.v.*) is believed to be only about 1%. F. J. Kallmann's data on homosexuality in monozygotic (identical) and dizygotic (nonidentical) twin pairs strongly suggest that genetic factors play a considerable role in the determination of homosexuality. If homosexuals, who have few or no children through heterosexual alliances, do not transmit any genes to future generations, why has homosexuality not disappeared? A possible explanation (which, if at all valid, is only part of the answer) may be that a portion of the heterosexual population bears the genes for homosexuality in a masked, hybrid state and thus continues to replenish the homosexual population. There is little evidence of gross hormonal imbalance in homosexuals, as was earlier believed.

See SEXUAL BEHAVIOUR; SEXUAL DEVIATIONS.

B. PLANT

Sexual differentiation in the higher plants is essentially different from that in animals, although the final result is similar. The individual plant is a sporophyte or spore producer.

1. Spores.—There are two kinds of spores, large megaspores and small microspores or pollen grains. Unlike gametes, these spores do not undergo direct fusion; instead, they develop, within the pistil of the flower, into microscopic individuals known as gametophytes, the pollen grain forming the male gametophyte and the megaspore the female gametophyte. The gametophytes each produce a gamete and the two gametes fuse and give rise to a sporophyte, thus completing the life cycle. The seed in a flowering plant is simply an embryo sporophyte, its development temporarily arrested, together with nutritive materials and a protective coat.

2. Sexual Variation.—In most species of higher plants the sporophyte is monoecious or hermaphroditic; *i.e.*, it produces both ovules and pollen grains. But in such plants as hop, hemp, date palm, papaya, and sorrel dock (to name only a few), two kinds of individual sporophytes exist, which may be called males (producing only pollen) and females (ovule producing). Plants like this are said to be dioecious. In some other cases the sporophyte is hermaphroditic, but the pollen and ovules are borne on different kinds of flowers (*e.g.*, in maize, melon, and cucumber).

In certain plants, such as primrose, different kinds of individuals may exist ("pin" with long style and short stamens and "thrum" with long stamens and short style), and the system may be such that under natural conditions pin cannot pollinate pin or thrum fertilize thrum. But this is not really sexual dimorphism since the two kinds of sporophytes each produce both ovules and pollen grains; it is, however, a phenomenon analogous to sexual differentiation. Many species of plants possess mechanisms of self-sterility without any visible difference between kinds of individuals, such as exist in the primroses. And, on the other hand, there are some plant species, such as cultivated wheat, in which self-fertilization almost always occurs and cross-fertilization is a rare accident. *See* PLANTS AND PLANT SCIENCE: *Morphology of Plants*.

II. VALUE OF SEX

Typically, the zygote or fertilized egg contains a double or diploid set of chromosomes in its nucleus, each chromosome being represented twice. The individual that develops from the zygote likewise has two sets of chromosomes in every cell. During the formation of the gametes (in animals) or of the spores (in plants) a reductional process called meiosis occurs, whereby the chromosome number is halved. Thus gametes and spores contain only a single or haploid set of chromosomes. The diploid number is restored at fertilization by the fusion of two haploid nuclei.

In the higher plants, with their characteristic alternation of sporophyte and gametophyte generations, the sporophyte is typically diploid, the gametophyte haploid. In ferns the conspicuous leafy sporophyte is likewise diploid, the inconspicuous but free-living hermaphroditic gametophyte, haploid. In the mosses, the familiar plant is the haploid gametophyte, which is monoecious in some species and dioecious in others, the diploid sporophyte being small and parasitic on the gametophyte. In both mosses and ferns the sperms are flagellated and swim to the egg.

In many algae and fungi meiosis occurs immediately after fertilization, so that the zygote is the only diploid cell in the life cycle.

1. Genetic Significance.—In sexual reproduction every individual receives a set of chromosomes from each of two parents and passes on mixtures of these chromosomes to its offspring. Thus, in each generation the genetic material is recombined in new ways. Actually, the "shuffling" of the hereditary determinants, or genes, is even more thorough than the above oversimplified statement would indicate, since during meiosis a process of crossing over between the maternal and paternal chromosomes takes place, which leads to the formation of chromosomes containing some genes of maternal origin and others of

paternal origin. The number of genes is so great that in a sexually reproducing population of higher organisms no two individuals are precisely alike, all having unique combinations of genes that will not recur in populations of even many trillions of individuals.

The uniqueness of the individual is thus a consequence of sexual reproduction. The significance of sex can be fully appreciated only if the instances in which it has been lost in the course of evolution are considered. There exist numerous species of animals and plants in which the egg regularly develops into an embryo without having been fertilized—a phenomenon known as parthenogenesis (*q.v.*). Obviously in such cases sperms are superfluous and, in fact, males may be entirely lacking from the population, so that the species consists exclusively of females. In species of this kind the chromosome number is not halved by meiosis during the formation of the egg, or, if it is halved, a compensatory doubling takes place to maintain the chromosome number constant from one generation to the next. In such parthenogenetic species all the progeny of a female, in general, carry exactly the same genes as their mother and in consequence resemble one another and their parent almost exactly, like the so-called identical twins in the human species, which owe their origin to a nonsexual form of reproduction in which the zygote divides vegetatively to produce two embryos (*see* MULTIPLE BIRTHS).

2. Biological Diversity.—The great biological diversity characteristic of sexually reproducing species is thus not found in parthenogenetic forms, and the adaptive genetic qualities of the two different parents can never be combined in their offspring in a nonsexual genetic system. In parthenogenetic species, natural selection, the force that has molded the whole evolution of life, lacks the constructive and creative properties it possesses in sexual species. It is not surprising, therefore, that asexual forms of reproduction, including both parthenogenesis and asexual reproduction by simple fission, budding, runners, bulbs, etc., seem to have been relatively unsuccessful in evolution, except as accessory mechanisms. They may permit the rapid multiplication of individuals on a large scale, but at the expense of evolutionary adaptability, so that every asexual species is in a "blind alley." Sooner or later its uniformity and lack of plasticity and adaptability will lead to extinction.

Sex is certainly not necessary for reproduction, and many species of animals and plants seem to get on quite well without it, but it is surely essential for long-term evolutionary success. From this standpoint there is not much difference between unisexual (hermaphroditic) and bisexual reproduction. What is important is the fusion of genetically unlike gametes; whether they are produced by separate male and female individuals is of secondary importance. Even in bacteria and viruses, which were formerly believed to be strictly asexual, it has been known since 1946 that various kinds of phenomena analogous to sex occur, which result in bringing together in one individual genetic materials derived from two different ones. The occurrence of this system in such primitive organisms indicates that sexual processes may have existed from the earliest period of life on earth.

In various groups of organisms, some kind of compromise between sexual and nonsexual forms of reproduction is found. Thus the typical life cycle of aphids consists of a number of parthenogenetic summer generations, followed by a single bisexual winter generation. The parthenogenetic generations permit very rapid multiplication of individuals under favourable conditions, while the sexual generation produces new combinations of genes once a year, upon which natural selection can operate. A different kind of compromise between sex and parthenogenesis exists in certain insects such as wasps, ants, and bees. In these, the sex of the individual depends upon whether it has developed from a fertilized or from an unfertilized egg. In the former case a female results, while in the latter a male is produced. Females thus have two parents (but only three grandparents), while males have a mother, but no father. In these insects the members of the "worker" caste come from fertilized eggs, show some female secondary sex characters, and are essentially females whose ovaries have remained in a rudimentary condition. In termites, on the other hand, the "worker" caste seems to consist of both sterile males and sterile females.

In some forms of parthenogenesis, sperms as well as eggs are produced, and copulation occurs, the sperms being even necessary for triggering off the development of the egg although they do not contribute any chromosomes to it. This type of reproduction, known as pseudogamy, occurs in some earthworms, where it has clearly been derived from the usual hermaphroditism of the group in the course of evolution. An analogous process is also known in some plants. Pseudogamy also exists in a small Mexican freshwater fish, the Amazon molly (*Mollienisia formosa*), a species that consists exclusively of females whose eggs have to be stimulated to develop by the sperms of two closely related species; the developing embryos of *M. formosa* do not inherit any of the characteristics of the pseudo fathers.

III. SEX DETERMINATION

1. Chromosomal Mechanism.—In the great majority of bisexual organisms some kind of genetic mechanism of sex determination is operative, causing certain zygotes to develop into males, others into females. The "switch" mechanism that controls the direction of development consists in most cases of a special pair of sex chromosomes, present in all the cells of the body. In one sex the two sex chromosomes are alike in size and shape (*i.e.*, are homogametic) and are known as X chromosomes. In the other, or heterogametic sex, there is a pair consisting of an X chromosome and a chromosome that differs from it genetically (and sometimes also in size and shape), known as a Y chromosome. Chromosomes other than the sex chromosomes are known collectively as autosomes. The gametes produced by the homogametic sex all carry an X, while those produced by the heterogametic sex are of two kinds, bearing X and Y chromosomes respectively. Thus at fertilization (if the X- and Y-bearing gametes are equally likely to take part in fertilization, which is probably only approximately true) equal numbers of XX and XY zygotes will be produced.

In most groups of animals and in the majority of those higher plants that are dioecious, it is the male sex that is heterogametic, so that there are two kinds of sperms, or pollen grains, and only one kind of egg, or megaspore. But in certain fishes and amphibians and in all birds, all lepidopterans, and all caddis flies, the situation is reversed: the female is heterogametic and the male homogametic, so that there are two kinds of eggs and only one kind of sperm. Some authors speak of Z and W chromosomes instead of the usual X and Y terminology, to distinguish the species with female heterogamety. Usually, the X and Y chromosomes are visibly different in size or shape, when examined under the microscope, but in certain fishes and amphibians there is convincing evidence, from breeding experiments, that X and Y chromosomes exist in spite of the fact that they cannot be distinguished from one another or from the remaining chromosomes under the microscope. In some other groups the genetic evidence is still lacking, so that in the absence of cytologically distinguishable sex chromosomes it is not certain which sex is heterogametic (this is the situation with regard to some reptiles).

Before proceeding to discuss how the switch mechanism of sex determination actually works, two further types of systems may be mentioned that have arisen as evolutionary derivatives of the XY-XX type. (1) In certain animal species the Y chromosome is altogether absent, so that the diploid number in the heterogametic sex is an uneven one, the X being without a mate. This is the case in most grasshoppers, crickets, roaches, dragonflies, and in many species of beetles and nematode worms (all groups with male heterogamety). In such forms, the heterogametic sex may be called XO in constitution, where O represents simply the absence of a chromosome. (2) In certain other species of animals and in some dioecious plants, sex chromosome mechanisms are found in which the X or the Y is represented by several chromosomes (X_1, X_2, X_3 . . . , or Y_1, Y_2, Y_3 . . .). Most of the larger green species of praying mantises are X_1X_2Y in the

male and $X_1X_1X_2X_2$ in the female (while the smaller species and some larger ones have XO males and XX females). Most species of spiders have X_1X_2O males and $X_1X_1X_2X_2$ females, while four species of mammals, including the common European shrew (*Sorex araneus*) and the Tasmanian rat kangaroo, or potoroo, are known to have XY_1Y_2 males and XX females. In all these instances, the several X or Y chromosomes are inherited as a group, so that, for example, the X_1X_2Y praying mantises produce sperms with both an X_1 and an X_2 chromosome and sperms with a Y but no X. The eggs of the same species carry both an X_1 and an X_2.

There appear to be at least two basically different genetic mechanisms of sex determination. In species with a Y chromosome, this element may determine the characters of the heterogametic sex. This seems to be the case in the human species, where the diploid number is known to be 46 in both sexes (22 pairs of autosomes and a pair of Xs in the female, the same autosomes plus an X and a Y in the male). Persons who possess two Xs and a Y exhibit a type of intersexuality known as Klinefelter's syndrome. They may be regarded as individuals that would have been females if they had not possessed the extra Y chromosome that masculinized them to some extent—or as individuals that would have been normal males had they not been feminized by an additional X. Persons with 45 chromosomes who are XO in constitution are sterile females exhibiting ovarian dysgenesis (Turner's syndrome). The conclusion to be drawn from this work on abnormal sex chromosome situations in the human species is that the Y is actively male determining and the X female determining. Obviously one X cannot do the work of two—otherwise the 45-chromosome individuals would be normal females—and it seems fairly certain that the remaining chromosomes (the autosomes) are, on balance, male determining. Normal sexual development in the human species thus depends on an equilibrium between X chromosomes and autosomes in the female and between the X, the Y, and the autosomes in the male. With such a complex genetic basis it is perhaps not surprising that there should be a wide range in the type and degree of expression of sex behaviour in the human species.

Mammalian species in general have XY males, so that the same kind of genetic sex-determining mechanism as in man may exist throughout the whole class.

In the fruit fly *Drosophila*, a different kind of system operates. The Y chromosome, although present in the normal male, is not actively sex determining, since experimentally obtained XO individuals are entirely male in appearance. However, the Y does carry some genetic factors needed to ensure male fertility, since such XO flies do not form motile sperms and are hence sterile. *Drosophila* females that carry one or two Ys in addition to the normal chromosome set are not masculinized in any way. Since individuals with three sets of autosomes (instead of the usual two sets) and two Xs are intersexual, it may be concluded that sex in *Drosophila* depends on a balance between male-determining genes in the autosomes and female-determining genes in the X chromosome. Two Xs are sufficient to override the male-determining effect of two sets of autosomes, so that a normal female results, but they are incapable of completely overriding the effect of three sets of autosomes. YY individuals of *Drosophila* or mammals, which entirely lack X chromosomes, have never been obtained and are presumably nonviable.

In the Mexican axolotl, sex determination takes place in a manner closer to the mammalian type; but in this amphibian, the female is the heterogametic sex (XY), so that the role of the sex chromosomes is reversed. Some females were experimentally transformed by R. R. Humphrey into males by implanting testes of another species at an early developmental stage. Such implanted testes do not become genetically functional, but they produce male sex hormone, which masculinizes the XY (*i.e.*, genetically female) host tissues. The offspring of the mating between normal XY females and the sex-reversed salamanders consist of three kinds of individuals in a Mendelian ratio 1 XX : 2 XY : 1 YY. The XX individuals are males, the XY and YY ones are females (in this case the YY constitution is not lethal). XY and YY females cannot be distinguished externally, but when a YY female is mated with a normal XX male all the progeny are XY females. It may be concluded from this experiment that axolotls carrying one or more Y chromosomes are female and those lacking a Y are in all cases male. It should be pointed out that the X and Y chromosomes of the axolotl are not cytologically distinguishable; the interpretation is based on genetic evidence alone. In the human species, mouse, and *Drosophila*, direct cytological observation of X and Y chromosomes can be used to check the genetical interpretation.

In all the above instances the sex of the zygote is determined at fertilization by a genetic mechanism involving special sex chromosomes. There exist, however, a few species of animals with mechanisms of sex determination that seem to depend on factors in the egg cytoplasm rather than on nuclear genes. It was shown by E. Korschelt that the little marine worm *Dinophilus apatris* lays large and small eggs in the same cocoon and that the large eggs develop into females, the small into males. There is no evidence of sex chromosomes in this species, and the determination of the amount of egg cytoplasm takes place long before the meiotic divisions occur. Another animal that lays two kinds of eggs is the scale insect *Pseudaulacaspis pentagona*, in which the fertilized females first lay eggs containing coral-coloured female embryos and later eggs containing whitish male embryos; the most probable interpretation seems to be a nonchromosomal mechanism of sex determination. A similar state of affairs may exist in other scale insects: in *Stictococcus*, for example, eggs whose cytoplasm is invaded by a symbiotic microorganism give rise to females while eggs not invaded develop into males.

Sex chromosomes are, of course, absent in species of animals and plants that are normally hermaphroditic, early accounts of X and Y chromosomes in such forms as land snails being based on erroneous observations. The ovarian and testicular tissues of such animals arise by processes of histological differentiation analogous to those that produce different somatic tissues such as liver and kidney.

Genes borne on the sex chromosomes are inherited in special sex-linked ways. In the human species a gene on the Y is transmitted by a father to all his sons, but to none of his daughters. There is no certain instance of this so-called holandric ("completely masculine") inheritance in man, various supposed examples having been shown by C. Stern to rest on doubtful or erroneous records. Genes borne on the X chromosome in the human species include the well-known ones for hemophilia and colour blindness. (*See* GENETICS, HUMAN.)

2. Endocrine Mechanism.—In the vertebrates the primary switch mechanism consists of the sex chromosomes, which determine whether the individual will develop with an ovary or with a testis. These organs (or, rather, certain cell types within them) together with some other tissues such as the adrenal cortex secrete into the bloodstream steroid male- and female-determining hormones that influence the development of the secondary sex characters and the whole pattern of sex behaviour of the individual.

In mammals and birds the main androgen (male hormone) is testosterone. The site of production of testosterone in the male is the so-called interstitial tissue of the testis; other androgens are formed in the adrenal cortex. Castration of the male thus reduces the amount of androgen in the bloodstream but there is still a quantity left. If carried out in childhood the operation leads to eunuchism, a condition in which the appearance of the male secondary sex characters at puberty is more or less suppressed. The physiological independence of the interstitial (endocrine) and spermatogenetic tissue in the testis can be proved by the operation of vasectomy (cutting and ligaturing the *vasa deferentia* or sperm ducts): this leads to sterility, since there is regression of the spermatogenetic tissue and in any case the sperms could not pass down the duct, but the production of testosterone by the interstitial tissue is not diminished, so that the secondary sex characters and sex behaviour are unaffected.

There are a number of female hormones or estrogens, which are closely related chemically to testosterone. Although andro-

gens and estrogens act antagonistically in controlling the sexual dimorphism of the vertebrates, it is important to realize that they are not restricted to one sex only; *i.e.*, in mammals the testis probably secretes estrogen as well as androgens and the ovary produces a certain amount of androgen in addition to estrogen. Thus the hormonal difference between the sexes is merely a matter of degree. The not infrequent appearance of male secondary sex characters (*e.g.*, facial hair) in women is an evidence of minor shifting of the equilibrium between androgen and estrogen production, the cause of the abnormality being most often an overactive adrenal cortex.

The development and physiological activity of both the testis and the ovary in the higher vertebrates are under the control of the gonadotrophic hormones, secreted into the bloodstream by the anterior lobe of the pituitary body, an endocrine body on the ventral side of the brain. These hormones, of which there are several, promote the growth of the ovarian follicles and of the germinal epithelium in the testis. They are apparently the same in the two sexes and their action is to bring about the maturation and activity of the ovary or testis. Unlike the sex hormones, which are relatively simple steroid molecules, the gonadotrophic hormones are complex proteins. Two main types have been studied: follicle-stimulating hormone (FSH) and luteinizing hormone or interstitial cell-stimulating hormone (LH or ICSH). The relation between the activity of the gonads and that of the pituitary is to some extent a reciprocal one, since the sex hormones have an inhibiting effect on the formation of gonadotrophins by the pituitary. Castration may thus result in an abnormally large production by the pituitary of LH and FSH, which are excreted in large amounts in the urine. The rate of secretion of the pituitary is under neurohormonal control mediated through the hypothalamus.

Both male and female sex hormones seem to stimulate cell division, not only in the gonads and accessory sexual organs but also in other tissues, including the epidermis and intestinal lining. However, male organs such as the penis, sperm ducts, seminal vesicles, and various glands producing accessory secretions are especially responsive to androgens, while the corresponding female organs react more readily to estrogens. The sexual development of an individual is hence a complex equilibrium in which not only the relative amounts of male and female hormones but also the presence of tissues and organs differentially reactive to them is involved. *See* ENDOCRINOLOGY.

IV. SEX REVERSAL AND INTERSEXUALITY

1. Hormonal Agency.—By shifting the balance between androgens and estrogens the sexual status of the individual can be more or less profoundly modified in the vertebrates. A partial change may produce an intersexual condition, while a more radical one may lead to complete functional sex reversal in some cases.

One of the classic examples of hormonal intersexuality is the phenomenon of bovine "freemartinism." Twins in cattle may consist of two males, two females, or a male and an intersexual calf known as a freemartin. In the freemartin, the external genitalia are mainly female in type, although the internal reproductive system is generally malelike and testes may be present. It is known that the embryonic membranes (chorion and allantois) of cattle twins become partially fused in development. This permits the sex hormones of one twin to flow into the bloodstream of the other and also allows for an actual exchange of embryonic cells, which become permanently established in the co-twin. Thus all twin calves are chimeras—they contain cells derived from two separate zygotes. It has generally been supposed that freemartin calves are genetic females more or less masculinized as a result of androgens derived from the male co-twin. It is possible, however, that this is not the whole story, and that the presence in them of some genetically male (*i.e.*, XY) cells may have something to do with the development of this very characteristic type of intersexuality.

Intersexuality or sex reversal in birds may arise differently. In the female of most bird species, what is called the ovary is simply the gonad of the left side, the organ of the right side remaining in a rudimentary condition throughout life. However, upon destruction of the ovary by disease or after its removal by surgery, the right gonad may regenerate and develop into an organ that is testicular in character. The individual may, under these circumstances, undergo a more or less complete sex reversal, acquiring male plumage, wattles, and behaviour. There is one case on record of a chicken that had reputedly been the mother of many offspring before she suffered avian tuberculosis; the disease destroyed the ovary, and the chicken underwent sex reversal and eventually became a fertile rooster and fathered two chickens.

2. Genetic Agency.—Although the above instances of intersexuality are hormonal rather than genetic in origin (if the traditional interpretation of freemartinism is accepted), genetic intersexuality is quite a well-known phenomenon. Two types of human intersexuality which are due to unusual sex chromosome combinations have already been referred to (see *Chromosomal Mechanism*, above). In insects, however, many instances are known where single gene mutations lead to the production of intersexuality or even (in the case of the transformer gene in *Drosophila*) to the complete transformation of one sex into the other.

One of the most thoroughly studied cases of genetic intersexuality is that of the offspring of certain interracial crosses in the gypsy moth (*Porthetria dispar*). Here the strength of the genes determining maleness and femaleness varies from one geographic race to another. Within each race the two are in balance, so that only normal males and females are produced, but in laboratory-bred hybrids between different races, various types of unbalance may lead to intersexuality or sex reversal. Genetic intersexuality seems to exist also in certain vertebrates, such as the pigs of the New Hebrides.

No circulating sex hormones seem to exist in insects. Thus the removal of the testis or ovary from a young caterpillar or other insect larva, even if followed by implantation of gonads of the opposite sex, does not lead to any modification of the secondary sex characters and the adult that develops from such a larva, although sterile, still acquires the sex characters and behaviour that correspond to its original chromosomal constitution. The intersexes that have been described in insects are thus, in general, all of genetic rather than hormonal origin. In the wasp *Habrobracon* the male normally arises from an unfertilized egg and has only a single set of chromosomes. Such males carry any one of nine or more genetic sex factors (x_1, x_2, x_3 . . .). Females arise from fertilized eggs (*i.e.*, they have two chromosome sets) and carry two different sex factors (*e.g.*, x_1 x_2, x_5 x_7 . . .). By inbreeding, one can obtain individuals with two chromosome sets that carry two identical sex factors (x_1x_1, x_2x_2 . . .). These are sterile "biparental" males. Thus femaleness depends on some kind of interaction between different x factors in the same cell. It is also possible to obtain, in special genetic experiments, individuals of *Habrobracon* that have arisen from binucleate eggs. Some of these individuals have no more than a single set of chromosomes in any of their cells, but they carry different x factors on the right and left sides of the body. Such individuals would be expected to be males, but in fact they have a streak of feminized tissue down the middle, along the meeting ground of the two kinds of tissue. Thus some local cell-to-cell diffusion of sex-determining substances produced by the x factors seems to take place.

3. Parasitism.—Another phenomenon that modifies secondary sex characters (amounting in some instances to conditions similar to true intersexuality) may result from infection by parasites. This phenomenon, which occurs in crustaceans and insects, has been called parasitic castration, since it is usually accompanied by some degree of degeneration or destruction of the gonad, resulting in sterility.

Four examples may be cited. (1) Certain species of crabs are liable to infestation by parasitic barnacles (*Sacculina, Thompsonia*), degenerate crustaceans whose bodies are reduced to a saclike mass with a branching system of rootlike processes that

penetrate the tissues of the host. Male crabs so infested may become partly feminized in their external features, as their claws become relatively smaller, their abdomens narrower, and hairs develop on the swimmerets (to which, in a true female, eggs would be attached). Infested female crabs, on the other hand, are masculinized only to a slight extent, or not at all. (2) Males of some species of bees, wasps, and leafhoppers are similarly affected when parasitized by insects belonging to the order Strepsiptera, a phenomenon which has been called stylopization, from the name of the principal genus of these parasites, *Stylops*. (3) Males of certain species of grasshoppers may fail to acquire male secondary sex characters and may undergo a certain degree of feminization following parasitization by certain nematode worms (Mermithidae). (4) In certain chironomid midges a similar change, but in this case from femaleness to maleness, may likewise be caused by mermithids. The changes in the secondary sex characters in all these instances are probably due to profound changes in the general metabolic pattern produced by parasitization rather than to any specific action of sex hormones.

4. Gynandromorphism.—In all multicellular organisms it is possible to obtain, from time to time, abnormal individuals whose chromosomal constitution is different in different parts of the body. Such monstrosities are referred to as genetic mosaics. The most clear-cut genetic mosaics are found among the types called gynandromorphs, individuals having part of the body chromosomally male, the remainder chromosomally female. Gynandromorphs have been studied extensively among the insects and spiders, in which there are no circulating sex hormones to mask the expression of the genetic constitution of certain tissues. In many other groups of invertebrates with clear sexual dimorphism, gynandromorphs seem not to be known. Where one side of the body is male, the other female, there is a bilateral gynandromorph; where the dividing line between the two regions is transverse there is an anteroposterior one. In many instances, however, the distribution of the male and female parts is more irregular. In fruit flies of the genus *Drosophila*, gynandromorphs arise mainly by loss of one X chromosome from one of the two daughter nuclei at the first cleavage division in the zygote. In other insects they may arise in different ways; *e.g.*, in moths from binucleate eggs, one nucleus carrying an X, the other a Y. In bees and other Hymenoptera, gynandromorphs may likewise arise from binucleate eggs, one nucleus being fertilized by a sperm while the other remains unfertilized and gives rise to male tissue.

In all these cases the dividing line between the male and female parts is a sharp frontier; there are no tissues or regions that show an intermediate expression of the secondary sex characters. Gynandromorphs may exist but be unrecognizable as such in the vertebrates, where the circulating sex hormones would inevitably blur the sharpness of the frontier between the genetically male and female tissues and would probably lead to a condition indistinguishable from genetic or hormonal intersexuality. *See* also GYNANDROMORPH.

V. SEXUAL CYCLES

1. Periods.—Almost all animals and plants go through cycles of development, somatic growth, sexual activity, and reproduction. In many organisms death may follow closely on the single sexual cycle, while in others there may be a long series of cycles, recurring at more or less regular intervals. Annual plants have a single sexual season, perennial ones usually have annually recurring flowering periods, while certain desert agaves increase in size for many years before they put up a reproductive spike, after which the whole plant dies.

In most animals the reproductive cycle is related to the seasons. This is most obvious in temperate climates, but seasonal cycles occur even on the Equator. The reproductive cycles of many species of birds are controlled by the length of the day, so that in temperate regions sexual behaviour (courtship, nest building, etc.) is initiated when a certain critical day length is reached (usually in spring, but sometimes in the fall). In desert regions, however, rainfall may be the initiating stimulus. Some reproductive cycles of marine organisms are related to the phases of the moon: the Pacific palolo worm spawns only in October and November; at daybreak on the day when the moon enters the last quarter the sea may become milky white with the eggs and sperms. Moonlight is also known to control the time of spawning of many mollusks.

In mammals, the mature females pass through a series of estrous cycles in the course of the year and will only mate and become pregnant during the period of estrus, or "heat." If pregnancy occurs, the sequence of estrous cycles is interrupted and is usually resumed after the termination of the subsequent lactation period. In the higher primates a modified type of estrous cycle, the menstrual cycle, occurs. It differs from the more typical mammalian cycle in two ways: the female will generally mate throughout the cycle instead of only at a sharply defined period of "heat" and at a certain stage in the cycle there is a pronounced vaginal bleeding (menstruation), due to breakdown of the lining of the uterus.

The timing of both the seasonal and estrous cycles is undoubtedly under the control of the anterior pituitary, but the details of the control mechanisms are not fully understood. As far as the seasonal cycles are concerned, it is clear that visual stimuli play a major role in triggering off pituitary activity in many species.

2. Courtship.—More or less elaborate and stereotyped patterns of courtship behaviour occur in many groups of animals. They are usually highly species-specific, thereby preventing or strongly discouraging the mating of different but related species. However, in the fruit flies of the genus *Drosophila*, H. T. Spieth (1952) noted that closely related species seem to display identical mating behaviour (*i.e.*, identical to the human observer; it might not appear identical to another *Drosophila* or to a human observer equipped with special recording instruments).

In moths the males of certain species are attracted to the females over very considerable distances by minute amounts of special chemical substances carried as aerosols. In some other groups of insects, such as grasshoppers, crickets, and cicadas, sound may play a major part in bringing the sexes together; the same is true of course, in certain vertebrate groups such as the frogs. Some of the more specialized types of courtship or precopulation behaviour in insects may involve the presentation of a special food morsel to the female, which is eaten by her as part of the sexual ritual.

Courtship behaviour can be quite elaborate in some fishes (sticklebacks, swordtails) and in many birds (ruffs, grebes, bowerbirds). In the higher vertebrates, not only sexual activity but many other forms of behaviour are under the control of the gonadal sex hormones. Sexual display, nest building, territorial defense, etc., may all be induced in immature or castrated males by injection of testosterone. Testosterone seems to be responsible for aggressive behaviour and social dominance in both sexes in birds. In a flock of chickens the "peck order" is determined by the testosterone levels of the individual birds, and hens injected with testosterone undergo a dramatic increase in their social status in the flock. In general, female sex hormones injected into males produce very little physiological effect, either in birds or in mammals. (*See* COURTSHIP, ANIMAL.)

3. Insemination.—The sperm is almost always a delicate cell which requires a fluid medium in which to swim and which is readily killed by desiccation. In a great many marine and freshwater organisms the seminal fluid is simply shed into the water by the male, in close proximity to the eggs. The success of fertilization is ensured, partly by the enormous number of sperm and partly by the fact that they are chemotactically attracted toward the eggs by special secretions. This is external fertilization, which obviously cannot occur in strictly terrestrial animals. Internal fertilization, generally characteristic of land animals, involves copulation and the deposition of the male gametes within the genital tract of the female; it occurs also in some marine animals such as sharks and rays, where the mature egg has a hard shell that could not be penetrated by the sperm, or where the young are born alive. In all vertebrates which have hard eggshe-

(sharks, rays, reptiles, birds, and monotremes), fertilization occurs before the shell is formed. In insects, on the other hand, fertilization takes place after the formation of the shell, the sperm penetrating through a small pore.

In animals with internal fertilization, the sperms may be conveyed either in a seminal fluid secreted by special glands or in spermatophores with a protective envelope that is dissolved or broken open in the female genital tract. Fluid semen occurs in dipterous flies, spiders, birds, and mammals, while spermatophores are found in grasshoppers, scorpions, octopuses, and salamanders —to pick a few examples at random. In insects, the sperm may be stored in the female for relatively long periods, up to several or many years in the case of the queen bee.

In most animals with internal fertilization the male possesses a penis, which is thrust into the female vagina in copulation, depositing the sperm therein. In many insects, the penis is concealed within the terminal abdominal segments and is only protruded or everted at copulation. There may be a variety of accessory male structures such as hooks and processes of different kinds that serve to clasp the female genitalia and interlock with them in complex ways. In mammals, the penis is flaccid except when the individual is sexually excited, a condition that leads to the erection of the penis, which becomes enlarged and stiffened as the spongy corpora cavernosa are distended by blood under pressure. A penis "bone," or baculum, is also present in many species of mammals. In some animals there are two penises: in sharks they are usually called "claspers"; in lizards and snakes, they are known as hemipenes. (*See* Reproduction: *Physiology of Reproduction.*)

There are several groups of animals in which the sperms are not transferred directly from the genital aperture of the male to the female, but undergo a double transfer. Three instances may be cited. (1) In the dragonflies the male genital aperture is on the ninth abdominal segment, but there is a pouch on the ventral side of the second and third abdominal segments to which the male transfers the sperm for storage. When copulation takes place the sperms are conveyed from this pouch to the female genital aperture. (2) In spiders the terminal segment of the male palps is enlarged and contains a seminal pouch. The male sheds a drop of semen from his genital aperture onto the web and then dips his palps into it, drawing up the sperm into the cavity in the palp; later, in copulation, the palps are used as "sperm guns" to inject the semen into the female aperture. (3) A somewhat similar arrangement exists in the cephalopods (squids and octopods), in which one of the arms of the male (the so-called hectocotylus) is specialized for conveying groups of very elaborate spermatophores to various parts of the female's body or into her mantle cavity. In *Argonauta* the male is minute by comparison with the female and lacks the characteristic shell; its largest organ is the hectocotylus arm which is detached in copulation and penetrates into the mantle cavity of the female, where it moves about autonomously for a while.

Dwarf males, parasitic on the females of their own species, occur in a number of groups of invertebrates. Certain peculiar mollusks (family Entochonchidae), which are endoparasitic in sea cucumbers, have much simplified wormlike bodies with a cavity containing the eggs and a number of minute males which consist of a testis with hardly any other tissues. Minute dwarf males also exist in the worm *Bonellia*, where they live in a special section of the uterus of the female. In this instance it is known that the free-swimming larvae that settle down on the sea bottom remote from a female themselves develop into females. If, on the contrary, they settle down on the proboscis of an adult female they develop into the minute parasitic males and migrate to the uterus later. A secretion from the female proboscis is apparently responsible for masculinizing the larvae.

Some unusual types of copulation occur in certain hermaphroditic invertebrates. In snails each individual discharges a large calcareous "love dart" into the flesh of the other, which stimulates the partners to complete the mating pattern. In most hermaphroditic mollusks each individual of a pair fertilizes the other, but in mollusks of the genus *Aplysia* copulatory chains are formed, each individual acting as a male to the one in front of it and as a female to the one behind. In the parasitic worm *Diplozoon paradoxum*, which occurs on minnows in Europe, two individuals become permanently fused together in sexual union.

(M. J. D. W.)

VI. SEX RATIO

The relative proportion of males and females at birth or at some other specified stage of the life cycle is usually expressed as number of males per hundred females. In man and the higher animals generally, sex is normally determined at fertilization by the chromosome constitution of egg and sperm (see *Sex Determination*, above); this leads to an expected sex ratio of equality at conception—the primary sex ratio. What is observed and recorded is usually the secondary sex ratio—the sex ratio at birth. The primary sex ratio is usually unknown, although in the case of the laboratory mouse and the fowl, when all eggs fertilized can be accounted for, the primary ratio has been found close to equality. The secondary sex ratio, however, frequently departs from equality, and sometimes the disparity is large. Interest in the sex ratio stems mostly from this difference. Search for the causes of disparity may elucidate problems connected with sex determination, differential mortality of the sexes before birth or maturity, and variations in sex ratio in different human populations. Ability to change the sex ratio experimentally may lead toward controlling it in favour of one sex or the other, an age-old ambition of animal breeders.

In lower forms of life, the sex ratio is greatly modified by environmental agents such as weather and population conditions. Several instances of such influence have already been described under *Sex Reversal and Intersexuality*. This section is primarily concerned with the sex ratio in man and certain other mammals.

1. Newborn.—In all races of mankind, more boys than girls are born. The ratio is generally considered to average about 106 males per 100 females (51.5% male), a figure that applies to whites of the United States. Among U.S. Negroes, the sex ratio is 102.6 to 100, with a slighter excess of males. This difference is not peculiar to the United States: W. Heape (1908) reported sex ratios in Cuba to be 108.4 for whites and 101.1 for Negroes, and S. de Jastrzebski (1919) found sex ratios in Cape Province to be 105.4 for whites, 102.6 for Negroes. At the upper end of the scale are Greece and Korea with sex ratios of 113.2 and 113.1 (W. T. Russell, 1936); Spain with a ratio of 111.1 (C. Gini, 1908); and the Jews of Austria with a ratio of 109.1 (De Jastrzebski, 1919). A more complete listing with references is provided by P. S. Lawrence (see *Bibliography*).

No one has given a completely satisfactory biological explanation for the excess of human males at birth. The disparity is particularly puzzling since a number of other species having the same type of sex-determining mechanism as man show an excess of females. The sex ratios reported from different countries are usually based on large numbers; thus differences are statistically significant.

2. Stillbirths.—If the mortality that occurs between conception and birth affects the two sexes differentially, then the secondary sex ratio would differ from the primary sex ratio. Such differential mortality is believed to occur, though the evidence is conflicting.

Figures based primarily on early U.S. census data showed an increasingly greater male mortality the younger the stillborn fetus and point to an even greater disparity in the proportion of the sexes at the time of fertilization than at the time of birth. The reliability of such data from registration reports has been challenged, however. For example, A. T. Hertig's examination of more than 1,000 spontaneous abortions submitted by physicians and hospitals in the Boston area suggests that about half the embryos cannot be classified for sex. Moreover, as many as half of all abortions are induced and rarely submitted for examination. The second difficulty is the one of determining the sex of young embryos, particularly from examination of the external genitalia. Mistakes can easily be made up to the fifth month, and even the opinion of experts is unreliable for embryos

younger than 11 weeks. Since most writers agree that the incidence of abortions is higher in the early months of pregnancy (at least one-half of all abortions occurring in the first three months), the sex ratio of "all abortions" remains unknown and hence the sex ratio at conception remains an open question. If more girls were lost than boys, then the primary sex ratio could be close to equality. With no differential mortality, or greater loss of boys, the primary sex ratio should be as high as or higher than the secondary sex ratio.

Museum specimens are in a special category, and the extensive collection of the Department of Embryology of the Carnegie Institution of Washington has been examined a number of times. C. Tietze (1948) made microscopic sections of the gonads (ovaries or testes) of 5,787 embryos ranging in prenatal age from three to seven months and found no significant variation in the proportion of males at any stage. The sex ratio of the whole sample is 107.9 males to 100 females (51.9% males). The discrepancy between these findings and census data (the latter admittedly less accurate for early months) is not easy to reconcile. It has been pointed out that the Carnegie collection is not a random sample of abortions and hence may not represent the general population. T. McKeown and C. R. Lowe (1951) made a study of the sex ratio of stillbirths related to cause of death and duration of gestation; they classified the 7,066 stillbirths from records in Birmingham, Eng., for the years 1936–49 as follows: (1) disease in, or accident to, the mother; (2) fetal malformation; (3) difficult labour; (4) ill-defined or unknown causes. The sex ratio of stillbirths (all causes) increases with the duration of pregnancy (from 99.2 at seven months to 133.1 at nine months and over), with the sex ratios of the four classes of stillbirth differing from one another.

McKeown and Lowe's data indicate that the relative proportions of the sexes change from month to month, and that after death male fetuses are retained in the uterus longer than females. It must be concluded that the actual primary sex ratio cannot be inferred from a combination of figures for secondary sex ratios and figures for abortions. Although the mechanism of sex determination indicates a 1 to 1 ratio at conception, the real ratio is unknown. In the absence of facts, the hypotheses that have been advanced concerning unequal production of X- and Y-bearing sperm, differences in sperm motility, or ability to penetrate the egg have not been confirmed.

3. Human Populations.—Every continent except Europe seemingly has an excess of males, but in every continent except Australia the numbers of the two sexes approach equality. Even countries like India, which have not been greatly affected by migration, show an excess of males.

The geographical distribution of the sexes is far from uniform and is subject to changes from both long range and short range forces. In the United States the influence of the frontier has been decreasing, whereas cities have proved a stronger attraction for females, who now make up more than half the urban population. The difference between city and country is characteristic of civilizations of Western Europe and its offshoots. It has been largely explained by demand for female labour in cities, especially for clerical positions and domestic service, and for male labour in the country. However, other influences cooperate with migration, since the female excess in urban populations generally applies to all ages. There is also evidence for a higher sex ratio at birth in urban than in rural areas in countries of Western Europe but not in the United States.

There is an excess of male deaths even in infancy, when social causes, like the greater exposure of males to accident, exert little effect. This must be biological in nature just as is the slight but uniform excess of males at birth. Also, except for limited periods related to reproduction, male mortality is higher at all ages.

4. Variations in Human Ratios.—When data do not fit a generally accepted hypothesis, the procedure is to look for special causes to explain the anomalous cases. For human beings, not subject to experimental control, frequently there is cause to question the data themselves. Then, again, there may be disturbing elements peculiar to human populations. One factor that could influence sex ratios in sibships is concerned with the effectiveness of a preference for one sex over the other, or a preference for a family containing both sexes. Presence of only one sex in a given family is not, of itself, an indication of anything unusual, for the presence by chance of a family of all boys, or all girls, has a probability of $(1/2)^n$ where n represents family size. Thus, out of 64 families of six, we expect, on the average, one family to contain only boys and one family to contain only girls. Results of statistical tests for family planning do indicate, however, that there is an effort to balance the sexes, but deviations from what would be expected if there were no planning are low. There seems to be no effective preference for boys.

Family planning, with preference for sibships containing both sexes, will not change the proportions of the sexes, but merely increase the proportion of families with longer than otherwise expected sequences of one sex. There is the added complication that large families that deviate from expectation because of real genetic influences on sex ratio cannot readily be separated from families that are large because of a continuing desire to have a child of the sex not yet represented. There is evidence that a sex preponderance in one generation may be followed by the same tendency in the next, and in some cases explanations invoking special chromosomal or cytoplasmic conditions have been put forward.

Numerous attempts have been made to correlate the secondary sex ratio with environmental influences such as season (low proportion of males in autumn and winter), birth order (usually more boys in first births compared to later ones), socioeconomic class (usually a higher sex ratio in upper levels), illegitimacy (fewer males in illegitimate births), effect of war (more boys born during or after wartime). The usual explanation for most of the above cases has been differential prenatal mortality, but the scientific validity of this is by no means established. Attempts to explain the significant increase in sex ratio toward the end of, or immediately after, World Wars I and II, in England, France, and Germany, on a basis of more first births, younger mothers, greater intervals between births, and other factors believed to reduce incidence of stillbirths, have not withstood close scrutiny by statisticians. The claim that the United States sex ratio was unaffected by World War II has been refuted by B. MacMahon and T. Pugh (1954), who found a significant rise in the secondary sex ratio for the years 1945 to 1947 inclusive, with a peak occurring in 1946. M. E. Bernstein (1958) advanced the hypothesis that in wartime, when opportunity for fertilization is decreased for a large segment of the population, proportionately more children will be born to the "more quickly fertile" parents, a group that has a relatively high incidence of male births at any time.

Interest in relations between disease incidence and the distribution of the ABO blood groups has been extended to the sex ratio with the first report, by L. D. Sanghvi, appearing in 1951. Real differences in sex ratio seem to occur in independent samples; these include high sex ratio (greater excess of males) among B offspring of B mothers and low sex ratio in A offspring of A mothers. Two explanations for the results have been advanced: one involves departures from randomness in egg and sperm unions; the other relies on known immunological phenomena concerning passage of antibodies through the placenta and postulates selective elimination of fetuses. However, studies on the relation between sex ratio of offspring and physique and temperament of parents have strengthened the conclusions from statistics that the father has an influence on sex ratio. This influence may be brought about through metabolic processes of some sort acting on the two kinds of sperms.

5. Experimental Alteration.—If natural selection adjusts the sex ratio to a certain optimum level for the species, then artificial selection should be effective in changing the ratio. A successful experiment using a mammal (rat), and applying selection directly to the sex ratio, is that of H. D. King (1918). She obtained, after six generations of selection, a line of rats with a high sex ratio (55% males) and a line with a low sex ratio (45% males). An attempt to duplicate the experiment using mice (D. S. Falconer, 1954) was not successful in changing sex ratio. J. A. Weir (1953, 1955, 1958) obtained high and low sex ratio lines

of mice by selection not for sex ratio but for high pH (low hydrogen ion concentration) of venous blood. Unlike King's rats, the strains are still existent. In reciprocal crosses between the high line (55% males) and the low line (45% males), the sex ratio of the progeny tended to follow the sex ratio of the male parent's line. The inbred lines have significantly different blood lactic acid levels and different activity levels. Artificial inseminations and double matings seem to rule out effects of seminal fluid, and no evidence could be found for differential mortality of embryos. The theory has been advanced that there is differential viability of sperm. H. G. Wolfe (1960) obtained an opposite response (high pH, excess females; low pH, excess males) when selection was for pH of arterial blood. Evidence for effects of differential mortality on sex ratio of mice has been found in certain strains.

For farm livestock, significant deviations may occur within the same species when different breeds, strains, or families are considered, and so pooled figures are not homogeneous. Jersey cattle averaged 52% males at birth, Guernseys 44%; most breeds of sheep show an excess of females. The dog resembles man in having a consistent excess of males, while the thoroughbred horse has exact equality of sexes. Sex ratio figures for wild species are frequently unreliable due to difficulties in sampling. Frequently hybrids show anomalous sex ratios; an excess of females is commonly encountered when the parents belong to different species.

There have been attempts to determine sex by methods designated to favour one kind of sperm over the other. Acid and alkaline douches have been tested on several different mammalian species. The results have been conflicting and negative for the most part. M. J. Gordon (1957) had some success in separating rabbit sperm into fractions by use of an electrophoretic technique. If a difference in the concentration of electrical charge between X-bearing and Y-bearing sperms exists, it should be possible to collect one kind of sperm at the anode, the other at the cathode. Gordon obtained 62 females and 25 males by artificial insemination from sperm collected at the anode and 29 females and 51 males from sperm that migrated to the cathode. This technique applied to cattle was not successful in deflecting the sex ratio from the theoretical value.

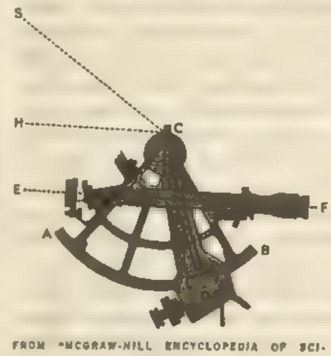
See also references under "Sex" in the Index. (J. A. We.)

Bibliography.—R. Goldschmidt, *The Mechanism and Physiology of Sex Determination* (1923), and *Die Sexuellen Zwischenstufen* (1931); M. Hartmann, *Die Sexualität* (1956); K. Ponse, *La Différenciation du sexe et l'intersexualité chez les vertébrés* (1949); F. H. A. Marshall, *Physiology of Reproduction* (1952); E. Allen *et al.* (eds.), *Sex and Internal Secretions* (1939); C. P. Swanson, *Cytology and Cytogenetics* (1957); M. J. D. White, *Animal Cytology and Evolution* (1954); M. E. Bernstein, "Studies in the Human Sex Ratio 5," *Amer. J. Hum. Genet.*, vol. 10 (1958); A. Ciocco, "Sex Differences in Morbidity and Mortality," *Quart. Rev. Biol.*, vol. 15 (1940); W. A. Craft, "The Sex Ratio in Mules and Other Hybrid Mammals," *Quart. Rev. Biol.*, vol. 13 (1938); D. S. Falconer, "Selection for Sex Ratio in Mice and Drosophila," *Amer. Nat.*, vol. 88 (1954); M. J. Gordon, "The Control of Sex," *Sci. Amer.*, vol. 199 (1958); H. D. King, "Studies on Inbreeding III," *J. Exp. Zool.*, vol. 27 (1918); P. S. Lawrence, "The Sex Ratio, Fertility, and Ancestral Longevity," *Quart. Rev. Biol.*, vol. 16 (1941); T. McKeown and C. R. Lowe, "The Sex Ratio of Stillbirths Related to Cause and Duration of Gestation," *Hum. Biol.*, vol. 23 (1951); J. W. MacPherson and S. D. Vesselinovitch, "Electrophoresis of Bovine Semen IV," *Canad. J. Comp. Med.*, vol. 23 (1959); C. Tietze, "A Note on the Sex Ratio of Abortions," *Hum. Biol.*, vol. 20 (1948); J. A. Weir, "Association of Blood-pH with Sex Ratio in Mice," "Male Influence on Sex Ratio of Offspring," "Sex Ratio Related to Sperm Source in Mice," *J. Hered.*, vol. 44 (1953), vol. 46 (1955), vol. 49 (1958); W. C. Young, *Sex and Internal Secretions*, 3rd. ed., 2 vol. (1961).

SEXTANT, one-sixth of a circle, 60°. The name is applied especially to an optical instrument for measuring angular distances, invented by John Hadley in 1731. Hadley's original instrument was, strictly, an octant, employing a graduated arc of one-eighth of a circle. The arc was later enlarged to one-sixth, to meet the needs of navigation. In current usage the term sextant often is applied to instruments with arcs of other than 60°.

The sextant is mainly used at sea, and the angle that is measured is the altitude of the sun (or a star) above the horizon (*see* Navigation). A familiar sight on an ocean steamer is the officer on the bridge "shooting the sun" at noon, in order to determine his latitude. The officer is looking through a small telescope straight at the sea horizon; but he sees also an image of the sun (dimmed by an interposed dark glass) which has been reflected into his field of view by an arrangement of mirrors described below. He moves an arm which turns one of the mirrors until the solar image appears just to touch the sea horizon.

The figure shows the construction of the sextant. ABC is a light framework of brass in the shape of a sector of 60°, the limb AB having a graduated arc of silver inlaid. It is held in the hand by a small handle at the back, either vertically in a position in front of the eye to measure the altitude of an object, or in the plane passing through two objects the angular distance of which is to be found. It may also be mounted on a stand. CD is a radius movable round C, where a small plane mirror of silvered plate glass (called the "index glass") is fixed perpendicular to the plane of the sextant and in the line CD. At D is a vernier read through a microscope, also a clamp and a tangent screw for giving the arm CD a slow motion. At E is another mirror (called "the horizon glass"), also perpendicular to the plane of the sextant and parallel to CB. F is a small telescope fixed across CB, and pointed to the mirror E. As only the lower half of E is silvered, the observer can see the horizon in the telescope through the unsilvered half, while the light from the sun or a star S may be reflected from the index glass C to the silvered half of E and thence through F to the observer's eye. If CD has been moved to make the image of a star or of the limb of the sun coincide with that of the horizon, it is seen that the angle SCH (the altitude of the star or solar limb) equals twice the angle BCD. The limb AB is graduated to avoid the necessity of doubling the measured angle, a space marked as a degree on AB being in reality only 30′.



FROM "MCGRAW-HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY"; REPRODUCED BY PERMISSION OF MCGRAW-HILL BOOK COMPANY

SEXTANT

If the sextant is used on land, an "artificial horizon" is required instead of the sea horizon. This consists of a trough containing a shallow layer of mercury, which gives a truly horizontal reflecting surface. The telescope F is now pointed downward to view the sun's image reflected in the mercury trough; an image of the sun reflected by the sextant mirrors appears as before, and the two images are made to touch. The reading now gives the angle between the sun and its image in the mercury trough, which is double the angle between the sun and the horizon. In the air, however, the visible horizon is of no use, since its "dip" (below the truly horizontal direction) is large and unknown. The mercury trough is obviously unsuitable for use in an airplane. Hence some form of "bubble sextant" is used, in which a spirit level is reflected into the field of view in such a way that the centre of the bubble indicates the true horizon. Air sextants are often equipped with periscopes to eliminate the need for astrodomes.

SEXUAL BEHAVIOUR refers to all behaviour through which sperm is brought to the egg (*q.v.*). The mere existence of eggs, sperm and accessory reproductive organs does not ensure fertilization (*see* Reproduction). There must also be behavioural tendencies and reactions through which the male approaches and inseminates the female. This is accomplished in so many ways that it precludes a statement of the nature of sexual behaviour that will apply to all animals. However, the sexual behaviour of the different species has a common function: to secure the fertilization of the egg. The evolutionary influence of natural selection ensures that behavioural mechanisms are adapted to fulfill a function, but it does not limit the manner in which it is carried out. Also, the different parts of the overall sexual pattern in any one species are interrelated to make a coherent, integrated whole. Thus sexual behaviour, along with other aspects of animal biology, can be regarded as developing in the course of evolution and as being highly adapted to the environment, including other members of the species.

Although a good deal is known about the sexual behaviour of invertebrates, this discussion is limited to vertebrates.

VERTEBRATES (EXCLUDING MAN)

Establishment of Sexual Relations

Monogamy and Polygamy.—Among different species of animals reproductive relationships vary from casual and promiscuous mating to lifelong monogamous pairing. In general, the different classes of vertebrates (fishes, amphibians, reptiles, birds, mammals) are not characterized by particular types of sexual relationships. Widely differing patterns often can be found within the same class. In some fishes (*e.g.*, guppies and swordtails) mating is entirely promiscuous, and pairings endure briefly. Parental behaviour is absent and the young are in danger of being eaten by the parents. In other families of fishes, such as the cichlids, regular pairs are formed, nests may be established, and mates may co-operate in guarding the eggs or even in caring for the young. Some cichlids can distinguish their own mates from others of the same species and sex. In other species, such as the sticklebacks, a male may mate with several females, each of which lays her eggs in a nest he provides; he then immediately chases her away, to guard the eggs himself. Relatively little information is available about mating systems in amphibians and reptiles, but it appears that they are usually promiscuous, stable pairs occurring only rarely. Among many lizards one male may have several mates that stay near him for some time during the breeding season; but it is not clear if this relationship rests on mutual attraction or merely upon common reactions to the habitat.

In such birds as the European ruff and some species of grouse, the sexes may meet only for a brief period of courtship and promiscuous copulation, after which the females build nests, lay eggs and rear young alone. Many species of geese are strictly monogamous and the pair may endure over a number of migrations. Sexual relationships in these birds involve very strong bonds based upon highly developed individual recognition. Commonest in birds, however, are monogamous relationships for a single breeding season. This is observed in many species of songbirds, such as the American robin. There is little evidence that such pairs continue between breeding seasons; in the relatively few cases in which a pair becomes reestablished in successive seasons, it is not known if this is a function of individual recognition or a tendency of both birds to return to the same place.

Among mammals, too, various patterns may be found; most rodents appear to be casual and promiscuous; in wolves, monogamous families seem to prevail. In many mammals that live in herds (*e.g.*, antelope, deer, seals) males may have "harems"; it is not uncommon for one bull seal to have 40 or 50 cows.

Territory.—Sexual behaviour does not occur randomly in any location but is usually restricted to a specific site. Such an area is often defended by the male against competing males during the breeding season; the breeding area is usually weakly defended, if at all, at other times. Such a defended area is called a territory.

Birds.—The concept of territory was first elaborated for birds, and the classic and best-defined cases are described for avian species. Such songbirds as the snow bunting migrate in flocks, the males arriving at the breeding grounds days or weeks before the females. Shortly after the males arrive the flock gradually disintegrates as they take up residence in individual territories. A bunting on his territory perches on a prominent point such as a rock or bush, advertising his sovereignty with a characteristic song that warns other males away and attracts females. He vigorously guards against intrusion by other males and establishes boundaries by fighting with resident neighbours. While physical encounters may occur, stereotyped postures and vocalizations, which appear to have intimidating effects, play a dominant role. On his territory the male is almost always capable of dominating intruders; in boundary disputes each fighter becomes more effective as the battle moves closer to the centre of his territory.

Snow buntings and many other species of birds live during the breeding season entirely within their own boundaries. Many species of sea birds, however, breed in very crowded colonies on shore, leaving daily to feed from the ocean. The territory in such colonies may be as small as the area that can be defended by a bird sitting on its nest and stretching its neck; in other cases, the territory may be yards wide. Such territories do not protect the holders' food source, but they are defended as vigorously and elaborately as are the larger territories of land birds. In the sexually promiscuous European ruff, many males may gather in an area called a lek, and each defends a territory only a few feet wide. These territories serve solely for mating; nests are built elsewhere. (*See* also Bird.)

Fishes.—Although reproductive territories are not so widespread among lower vertebrates as among birds, many fishes and reptiles, and probably also amphibians, defend breeding areas. Many male salmon and trout defend well-defined areas of stream bottom, in which one or more females may spawn. The three-spined stickleback builds a nest, vigorously driving other males from the vicinity, and eventually leading a female to spawn there. Many other species of fishes defend territories during the breeding season, differing with respect to such factors as the sharpness of boundaries and the size of the area. The territory is not necessarily fixed, it may be the area around a movable object. Bitterling eggs live as parasites in the shell of a mussel, the male defending the surrounding area; his "territory" can be moved experimentally by moving the mussel. (*See* also Fish.)

Reptiles.—Such reptiles as small lizards of the genus *Anolis* show distinct territorial behaviour. Males of these species defend a definite area against other males during the breeding season; at other times of year they may attack other males that approach closely, but the aggressiveness is not limited to the territory, and the animals do not patrol boundaries as they do during the breeding season. (*See* also Reptiles.)

Mammals.—Information about the behaviour of mammals under natural conditions is meagre, but it is certain that many exhibit territorial behaviour. However, it is not always clear that defense is limited to the breeding season, as is characteristic of lower vertebrates. Some male mammals, including bears, hyenas and many wild relatives of the dog and cat, mark boundaries by depositing scent signals on rocks, tree trunks, hillocks and so on. These signals (or flags) may be either urine or the odorous secretions of special glands (located, in different mammals, on such parts as the head, lower abdomen and anus) to which other members of the species are highly sensitive. Among such polygamous herd-living animals as deer and antelope the territory may move with the harem, which may range considerably. (*See* also Mammal.)

Function of Territory.—The area defended during the breeding season does not serve exactly the same function in all species. It may be the complete living area; a mating area only; only the site of the nest; or it may be the vicinity of the female or an object required for egg laying. The common feature seems to be that the territory reserves for the defending animal something required for reproduction.

Mating and Copulation

The term mating is restricted here to behaviour immediately associated with the transfer of sperm to egg. Among vertebrates fertilization may take place inside or outside the body of the female. The term copulation is limited to animals in which fertilization is internal and in which mating requires insertion of the male copulatory organ into the body of the female. The term mating may be applied both to copulatory behaviour and to its functional equivalents in external fertilization.

Mammals.—Fertilization is internal in all mammals; copulatory movements are relatively stereotyped and fairly similar in most mammalian species. In four-footed mammals the female characteristically stands still, depressing the centre of her back, simultaneously elevating and exposing her genital region; the tail (if present) is usually raised or deflected to the side. This position, in which the depression of the back contributes to the rotation of the hindquarters, is called lordosis. The male mounts from behind, with forelimbs encircling or pressing the sides of the female's body. Camels, llamas and similar animals often copulate with the female sitting upright on the ground, legs folded beneath

while the male mounts from the rear. Among monkeys, the female bends sharply forward at the hips and the male mounts from the rear, holding himself erect by grasping her hips or waist.

Females of most mammalian species are relatively passive during copulation, although in a few cases the hindquarters may be moved from side to side or pushed against the male. The male makes pelvic thrusts that assist insertion and movement of the penis in the vagina, contributing to the stimulation of ejaculation. The behaviour of many male mammals during and after copulation suggests that something similar to an orgasm occurs. The male rabbit utters a sharp cry and falls off the female after delivering sperm. In male rats and guinea pigs there is a mild convulsive reaction during ejaculation, followed by a period of substantially lowered sexual responsiveness.

Birds.—In birds fertilization is always internal, but the males of most species lack an organ of copulation. Instead, both sexes have an external opening (the cloaca) under the base of the tail. The cloaca serves as the common outlet for the digestive and urogenital systems. During copulation the female stands still, typically tilted forward from the usual fairly upright position, with the back more or less level. The male hops or flutters to her back, maintaining balance by fluttering and flapping his wings. His tail is depressed while hers is raised, both simultaneously being rotated in opposite directions to bring the cloacas into contact. Transfer of sperm is very rapid and the cloacal contact usually lasts only a few seconds. In many species such postcopulatory displays as preening, ruffling of the feathers and pecking at the ground, suggest something analogous to orgasm. In such birds as ducks copulation usually is in the water, the female being somewhat submerged as the male swims from behind and clambers to her back.

Some groups of birds, such as ostriches and ducks, have a male organ of copulation partly everted from the cloaca, with similar functions to those of a mammalian penis. Copulation posture and behaviour in these species, however, is similar to that in birds that lack the organ.

Reptiles.—Most male reptiles have an organ of copulation (the hemipenis) normally located inside the body cavity and everted only during copulation. Snakes closely intertwine with sexual organs in contact; the male lizard usually straddles the lower body and grips the neck of the female in his jaws. Bending the hind part of his body to one side and downward, he gradually inverts his tail until their undersides meet, and the hemipenis is everted to enter the female cloaca. Copulation may last from several minutes to an hour or more.

Amphibians.—Among amphibians two basic types of mating behaviour are found: one in tailless amphibians (frogs and toads), the other in tailed amphibians (newts and salamanders). In frogs and toads fertilization is external. The male swims or climbs upon the back of the female and clasps her with his forelimbs at the level of her chest. She releases eggs in spurts, a few at a time, and the male delivers sperm in ejaculations synchronized with the release of eggs. The duration of clasp (amplexus) by the male may be greatly prolonged: in some cases, he may remain clasped for a full day before she begins to release eggs; in other cases she may begin to release eggs in a few minutes. He usually unclasps within a few minutes after the eggs are laid.

Tailed amphibians typically exhibit a curious form of internal fertilization. Following a more or less elaborate precopulatory courtship display, the male deposits on the ground a spermatophore, a small gelatinous stalk surmounted by a mass of sperm. The female, walking behind him during this ceremony, enfolds the cap of the spermatophore with the lips of her cloaca, thus taking it into her body where fertilization is accomplished. This type of fertilization takes place on land or in the water, depending on the species.

Fishes.—Among fishes fertilization may be internal or external, depending on the species. For example, the modified anal fin (gonopodium) of male live-bearing fishes serves as a guide for the sperm when it is inserted into the female genital opening. In most species, however, fertilization is external, the eggs and sperm both being released into the water. In the spawning of salmon, for instance, the female scoops out or cuts a shallow depression at the bottom of the home stream by quick, darting movements over the gravel, accompanied by sweeping movements of the anal fin. The male accompanies her and both may sweep over the nest, then circle back to repeat the process until there is a definite hollow in the gravel. Then both lie side by side over the nest and simultaneously extrude eggs and sperm. Although the two salmon may not come into actual contact, their spawning movements are synchronized, simultaneously becoming more vigorous and reaching a climax followed by reduction of activity and of apparent tension, reminiscent of orgasm in higher animals. Other fishes show a variety of methods for external fertilization. The male Siamese fighting fish blows bubbles to the surface where they adhere to form a so-called bubble nest. Following a series of preliminary activities, he swims over the female and bends himself into a rough U-shape, with head and tail down, enclosing her body. With the mates in this position, eggs and sperm are simultaneously released. The male swims down and retrieves a fertilized egg in his mouth, then swims upward and blows a bubble with the egg inside; the bubble rises to become part of the nest. The male three-spined stickleback leads the female into his tubular nest, swims through, then circles behind her while she is inside. He then moves forward and backward in quick jerks to tap her back with his snout. This stimulates her to deposit her eggs in the nest as she moves through; he follows and fertilizes them.

Nature and Function of Sexual Stimulation

The Nature of Sexual Stimulation.—Stimuli involved in courtship and sexual behaviour may be in any sensory modality; different species of animals differ widely in this respect. Mammals in zoos and in the wild clearly show that visual, auditory, olfactory and tactual stimuli may all arouse sexual excitement, and aid in the recognition of the mate. Olfactory stimuli are very important in most mammals; rats and dogs may respond to the odour of sexually receptive females while not being aroused by that of non-estrous females. In most mammals copulation is preceded by such activity as licking, sniffing and nibbling directed at the body of the mate, and particularly at the genital area. This behaviour undoubtedly represents for both sexes a combination of olfactory, tactual and other stimuli that is sexually arousing. Visual stimuli are also effective in mammals; the characteristic mode of locomotion of a female rat in heat may play a role in inducing the male to follow her, and female monkeys and chimpanzees often incite males to sexual activity by assuming postures characteristic of copulation.

Birds appear relatively insensitive to odours, and no reliable experimental evidence indicates that any of their sexual reactions are stimulated or guided by olfactory stimuli. However, both visual and auditory stimuli are important for birds. Females of many species appear attracted by male courtship songs and, in the case of the common parakeet, the female may be induced solely by the sounds made by other parakeets to lay eggs. Birds are highly visual in their behaviour generally; in a number of species the male can react differently and appropriately to stuffed and mounted specimens of either sex from its own species. The striking development of such features as special structures and patches of colour, which in many birds are displayed by courtship movements, is prima facie evidence of the importance of visual stimulation in their sexual life. Tactual stimuli also may be significant since some birds bill and preen each other before copulation.

Visual, auditory, tactual and chemical stimuli all operate among the lower vertebrates (fishes, amphibians, reptiles), several stimuli sometimes being effective at different stages in the sexual behaviour of the same species. For example, the males of some species of frogs approach and clasp on the basis of other frogs' visual characteristics, and then either maintain the clasp during emission of eggs and sperm or discontinue it immediately, depending upon whether the clasped animal emits the sound and has the body shape characteristic of another male or of an egg-laden female. Female salamanders are sometimes stimulated to follow the male during courtship by the secretion of glands on the surface of his body.

Many aspects of the courtship of fishes vary with visual stimulation. Experiments with artificial models, for example, show that a male three-spined stickleback on its territory acts differently depending upon whether the colour and shape of an approaching model are characteristic of a male or of a female.

Functions and Effects of Sexual Stimulation.—Animals of different species are subjected to a wide variety of stimuli from the mate and from the environment generally. This stimulation can induce a broad range of sexual effects, including not only immediate behavioural responses but also, more indirectly, pervasive physiological changes:

1. Most direct and obvious is the behavioural response immediately aroused or elicited by a specific stimulus. When another stickleback approaches a male on his territory he responds immediately. The response is likely to be an attack or an attempt to lead the caller to his nest, depending upon the visual characteristics of the visitor. The female stickleback, in turn, follows the male into the tubular nest. Male fence lizards adopt a characteristic posture when another lizard approaches; the approaching animal responds with a similar posture or by sexually receptive behaviour, depending on its sex. A territory-holding male bird similarly attacks or courts approaching birds, depending upon their visual characteristics and upon their immediate response to his singing. Sexually receptive female mammals of various species can be made to adopt copulation posture by touching or stroking them to imitate stimuli normally provided by the male just before copulation.

2. In addition, stimulation may sometimes contribute to the development of a mood, or temporary condition of sexual motivation. This is particularly important when mating requires close synchronization in the behaviour of the partners. In such cases, it is essential that both be aroused before mating occurs; many premating activities contribute to this. For example, the elaborate and protracted prespawning behaviour of the male and female salmon seems to culminate in the arousal of intense, synchronized spawning activity. In birds precopulatory displaying and posturing is widespread. In many species of pigeons and doves copulation is almost always preceded by billing, in which the mates peck and preen each other about the bill and head; often one bird will next place its bill in the mouth of the other, the heads of both move up and down in a pumping motion, and there may be an exchange of regurgitated food. Although billing sometimes occurs without being followed by copulation, copulation rarely occurs without this introduction. The nibbling, sniffing, licking and biting before copulation in most mammals appears to be needed to bring the mates simultaneously to the pitch of arousal required for effective copulation.

3. In many cases stimuli from features of the environment other than the partner appear essential for sexual behaviour. For example, experimental evidence indicates the nest of the male Siamese fighting fish to be essential in arousing sexually receptive behaviour in the female. Some species of toads that breed in the spring appear to depend on the appearance of ponds following heavy rainfall, without which they show no sexual behaviour. Species of weaver finches in Australia and in Africa breed only in the rainy season; in central Australia these birds live in large flocks, showing no sexual behaviour in dry weather. The sudden appearance of green vegetation, after rainfall occurring irregularly at any time of year, is followed by an outburst of courtship and breeding activity among the weaver finches in that neighbourhood; in dry areas some distance away birds of the same species may remain sexually inactive. It is commonly observed in zoos that the arousal of breeding behaviour very often requires not only the presence of a mate but such aspects of the normal breeding situation as cover, nesting material and appropriate places in which to nest.

In some animals the holding of territory is essential for sexual activity. When two pairs of the same bird species are kept in an aviary large enough for only one territory, only the pair that becomes dominant over the other is likely to develop breeding activity.

4. Finally, stimuli from the mate may induce changes in the activity of the endocrine glands (*q.v.*), sometimes with pervasive behavioural effects. A female pigeon may be induced to lay an egg merely by seeing a male court her through a glass plate. The production of an egg by the ovary occurs only in response to pituitary hormones, suggesting that the sight of the courting male stimulates the female to appropriate endocrine secretion. In other birds, such as sparrows, the presence of the courting male induces the female to secrete hormones that prompt her to build a nest.

The onset of puberty in female mice and the regular occurrence of estrous cycles thereafter are partially regulated by olfactory stimuli from male mice. The ovaries of minks produce growing egg cells earliest in the breeding season if other minks are visible.

The Role of Sex Hormones.—The gonads (testes and ovaries) secrete hormones that have important effects on mating behaviour. Among animals that breed at only one season of the year the gonads tend to regress to an inactive condition between breeding seasons. Increased activity of these glands during the breeding season is, in many animals, stimulated by environmental changes. For example, the springtime growth of gonads in birds is stimulated by the increased length of the day; when daylight is artificially prolonged males will develop large testes and sing vigorously in the fall, when wild birds of their species become sexually inactive. Such observations suggest that gonadal hormones may be influential in stimulating the sexual behaviour. Similar conclusions may be drawn from such mammals as the rat, in which the female undergoes a regular cycle of ovarian changes, culminating in the release of an egg (ovulation); in such species the female is sexually receptive only for the short time the egg is available to be fertilized.

Experimental removal of the testes in animals of many types is followed by loss or substantial reduction of sexual behaviour. Replacement of the missing hormones by injection acts in most animals to reinstate sexual behaviour. Removal of the ovaries in most mammals has an even more striking effect; sexually, the animal usually becomes completely unreceptive. Artificial replacement of ovarian hormones restores receptivity; in fact, such treatment may maintain continuous sexual receptivity over long periods even in such species as the rat, in which it normally occurs only for a very short period during each estrous cycle.

In such higher mammals as monkeys and apes sexual behaviour is not quite so closely tied to the sex hormones. Although apes have regular estrous (menstrual) cycles similar to those of lower mammals, receptivity is not so sharply restricted to the period near ovulation. In the case of male chimpanzees castration does not effect nearly so great a reduction in sex behaviour as it does in lower mammals.

The Role of Experience.—A persistent question concerns the contribution of individual experience, or learning, to animal behaviour. The answers vary from species to species and from one kind of behaviour to another.

In most lower mammals, as well as in birds and lower vertebrates, both sexes can be expected to show the normal pattern of sexual behaviour without previous sexual experience. However some aspects of experience before sexual maturity may contribute and, even though mating is adequately performed on the first possible occasion, in some species it tends to become more efficient with increasing experience. For example, the sexual behaviour of male guinea pigs reared in isolation from other guinea pigs is different from that of males reared in groups. When such animals are scored for quickness, vigour and efficiency of sexual approach the socially reared animals score persistently and significantly higher than those reared in isolation. In such species as the rat, males reared in isolation do not appear to differ in sexual behaviour from those reared socially. It is apparent that the role of early experience in the formation of such patterns varies from species to species.

In some fishes, males reared in isolation are slower and less efficient in mating approaches than are normally reared males. Pairs of inexperienced ring doves take longer to build a nest and lay eggs than do experienced pairs of the same age. It has been reported that a male ring dove reared in isolation from other birds preferred to mate with the hand of its keeper.

Learning that contributes to these patterns need not be specifically sexual, but may involve the establishment of general relations with other members of the species. For example, among the guinea pigs in which sexual behaviour is less effective in males reared in isolation, the difference from those reared in groups appears only if the animals are taken from their mothers at about 10 days of age. If they are kept with their mothers until the age of 25 days, and then reared as isolated or grouped animals, no difference appears in their later sexual behaviour. It appears that experiences during a very early period of life are crucially involved in this aspect of mating patterns.

Among primates the effect of learning is much more striking. Mature sexually inexperienced male monkeys sometimes have difficulty during initial contacts with receptive females. Adult chimpanzees without copulatory experience rarely succeed in achieving sexual union, no matter how co-operative the female may be. Erotic arousal is easily evoked in the naïve male ape, but he needs a great deal of experience and practice before this generalized excitement leads smoothly into a well-integrated and complete copulatory pattern. Female monkeys and chimpanzees, on the other hand, appear to show much more nearly adequate patterns of sexual behaviour at their initial experiences.

The effects of experience and of hormones may interact in some species. For example, if male guinea pigs (reared both in groups and in isolation) are castrated, the sexual behaviour of all drops to a very low level. If after about 10 weeks they are injected with male hormone, sexual activity rises, but only to the level characteristic of each group before castration. If male cats are castrated, the effect on sexual behaviour is in part a function of the animal's experience. Cats that have had no sexual experience before castration fail to show more than minimal attempts at copulation which, if they appear at all, die away rapidly. In those with sexual experience before castration, the ability to perform sexual activities may persist for some time and then die away very slowly. It appears that sexual activity in these cases is a joint function of hormonal and experiential influences.

Sexual Behaviour and Evolution

Evolution of Sexual Behaviour.—The elaborate organization of sexual behaviour is in part a consequence of evolutionary selection, having evolved in adaptation to the environment (which includes other members of the species). Thus each pattern depends partly upon the kinds of stimulation to which the given species is sensitive. In many animals one or the other sex may be highly sensitive to very specific kinds of stimuli; in such cases the sensory capacities of the mate provide a powerful selective influence upon the evolution of sexual behaviour and structures. For example, such structures as sexually erectile plumes and crests in some birds undoubtedly evolved because they accentuate and enhance sexual behaviour.

Sometimes the adaptive necessity for striking courtship behaviour may conflict with the equally pressing need for avoiding predators. Thus among some birds drab, inconspicuous males may have concealed patches of brilliant colours visible only when they spread their wings, erect crests, and so on.

Evolutionary relationships among species are sometimes shown very well by their sexual behaviour patterns. In some genera of doves, for example, the courtship movements may be very similar indeed, in spite of considerable interspecies variation in such characteristics as plumage colours and song. In this case the sexual pattern may be a more reliable guide to species relationships than other more traditional criteria. Students of evolution have used sexual behaviour patterns as criteria of interspecies relationships among such groups as fishes, lizards and ducks.

Role of Sexual Behaviour in Evolution.—Sexual behaviour also plays a role in the formation of species. For one species to evolve into two descendent species, it is usually held essential that it be divided into two populations isolated by a geographical barrier; it then is possible for them to evolve in different directions. If the populations are later merged by new migrations or by the disappearance of geographical barriers, the two incipient species may interbreed to reduce their genetic differences. In other cases, however, the disparate development of two species can continue, because both have evolved so differently that they will no longer interbreed. There may be differences in habitat preference that prevent members of the different groups from meeting; anatomical differences may make mating impossible; or differences in courtship behaviour may bar interspecies stimulation. There are many cases of closely related species living in the same area with little or no hybridization; differences in the pattern of sexual behaviour often play an essential role. Field observations and laboratory experiments often demonstrate that the failure of various species to interbreed arises from a lack of responsiveness of one species to stimulation offered by the other, or to a lack of correspondence in normally interlocking aspects of sexual behaviour. *See* Evolution, Organic. (D. S. Le.)

HUMAN SEXUAL BEHAVIOUR

Taboos in western culture and the immaturity of the social sciences have impeded research concerning human sexual behaviour. In the late 19th and early 20th centuries such European writers as Sigmund Freud, Havelock Ellis, Richard Krafft-Ebing (*qq.v.*), Albert Moll, Wilhelm Stekel and Magnus Hirschfeld were among the first to present studies of human sexual activity based on extensive case histories. By the 1920s they had laid the groundwork for more extensive statistical studies in the U.S. by R. L. Dickinson, K. B. Davis, L. M. Terman, G. B. Hamilton, C. Landis and A. C. Kinsey. Of the two major organizations for sexual study the Institut für Sexualwissenschaft in Berlin (established in 1897 by Hirschfeld) was destroyed by the Nazis in 1933. The Institute for Sex Research (begun in 1938 by Kinsey at Indiana University) continued in the 1960s with the study of many aspects of human sexual behaviour. Much of this discussion rests on the findings of the Institute for Sex Research (hereafter abbreviated I.S.R.). These data, largely based on interviews, were the most comprehensive available at that time for the U.S.; comparable bodies of data did not exist for other countries.

The I.S.R. data have been the subject of criticism that can be directed at all such interview research. It has been stated that the people interviewed may not have constituted a fully representative sample since some persons refused to be interviewed. It has also been observed that people do not always tell the truth, either with conscious intent or as the result of inaccurate remembering and sincerely held fantasies.

Being acutely aware of the problems of sampling, reliability and validity, the I.S.R. took some precautionary measures. Reliability was estimated by interviewing some respondents again after a lapse of at least two years. Statements were validated in many instances by separately interviewing sexual partners, chiefly husbands and wives. In a small number of cases reported data were corroborated by observation. In an attempt to overcome volunteer bias in sampling, defined groups were selected and worked with until 90% or more of their members had been interviewed. Nevertheless the sample does have deficiencies; the data, at least in some cases, should be regarded simply as approximations. Basically the conclusions deal with the approximate prevalence of specific forms of sexual behaviour; since moral or ethical evaluations are beyond the scope of the scientific method, none have been attempted.

Mammalian sexual patterns clearly demonstrate evolutionary origins for most of man's sexual behaviour. For example, masturbation, foreplay, homosexual behaviour and interspecies copulation are found, sometimes abundantly, in other mammals. However, the most prevalent sexual activity in all mammals is heterosexual.

All human societies face the problem of regulating sexual behaviour, since unrestrained sexuality would disrupt established social patterns. Yet severe restraints also can be disruptive since sexual expression contributes to the psychophysiological well-being of most people. Societies have sought solutions through widely varying systems of reward and punishment, encouraging behaviour held to be beneficial and discouraging activity considered detrimental. There are great differences among societies as to what is judged to be beneficial or detrimental. Among pre-

literate societies marriage is the primary social device for regulating sexual behaviour; many societies also encourage or tolerate premarital intercourse as a prelude to marriage. Adultery is more severely regulated, particularly for females; incest and forced sexual relations are condemned with few exceptions. (*See* MARRIAGE, PRIMITIVE.)

Sexual behaviour in the European-American culture (especially in the U.S.) has been much better documented statistically than in other cultures. Subcultures within a given country may show wide sexual variation; in the U.S. some of these differences as reported by I.S.R. can be summarized in the following way. The upper social level, chiefly people with higher education and occupations of marked social prestige, tended to show acceptance of and seemed most likely to engage in foreplay, heterosexual petting, masturbation, nudity, mouth-genital contact, and a variety of coital postures. The lower social level, typically those with a minimum of education and in occupations of lesser prestige, exhibited more acceptance of premarital coitus, prostitution (*q.v.*) and a wide variety of sexual partners.

While it has been widely held that the considerable differences in sexuality between males and females depend upon variation in gross anatomy and physiology, some authorities believe that other factors account for most of the divergence. One difference, the relative slowness of females to become aroused and to reach orgasm, has been discussed since the days of Ovid. Kinsey has suggested that coitus is a poor criterion by which to measure this difference; when speed of arousal and sexual climax in self-stimulation were considered, this difference was considerably narrowed. One of the most striking contrasts is the reported interest of males in a variety of sexual partners and of females in long-time association with one partner. I.S.R. data indicated that such differences derive from a male tendency to be more easily conditioned to respond sexually (especially to visual stimuli) and were interpreted as being both cultural and inherent.

One important measure of sexual activity is total sexual outlet, defined as the sum of orgasms derived from all sources including self-stimulation, dreaming or interaction with any other living being. While some individuals reported they derive their total outlet from only one source, most named two or three sources. A few interviewees of both sexes claimed the use of as many as six different sources in a relatively short period of time. Total reported sexual outlet for the sample of males studied by I.S.R. reached its maximum between ages 16 and 20, decreasing gradually but steadily thereafter. In this respect, total outlet followed the same course as most other physiological functions. The total outlet reported by unmarried males averaged about three orgasms per week between ages 16 and 20 and declined to slightly less than once per week by age 60; married men in the sample indicated a decrease from nearly five per week to about once per week. The opinion that males are sexually most active in their late 20s and early 30s was not verified. While the data indicated a peak of total outlet in the late teens, the peak of activity, in terms of rapid response and repeated climax, may be found even earlier in life.

Unmarried females showed a different aging picture; unlike the males, they did not seem to begin frequent sexual activity shortly after puberty. The total outlet of unmarried females studied rose progressively with age until the late 20s and then remained at essentially the same level (four to six orgasms per month) until the 50s. There appeared in most cases no cessation, or even diminution, of sexual activity or response as a direct result of menopause (*q.v.*). The gradual increase in the total outlet reported by these unmarried females is partly due to the fact that females apparently require some time in which to learn to achieve orgasm. Virtually all of the males said they had experienced orgasm within two years after puberty while only about a quarter of the females reported such experience. Only about three-quarters of the females who had never married by age 40 had ever had an orgasm from any source.

The total reported outlet of the married female showed the same aging picture found in males; this effect seems largely dependent on the aging processes of her spouse.

Nearly all preadolescents apparently have the ability to respond sexually, and many have the ability to achieve orgasm. Orgasm, apparently identical with that of adults, has been observed in infants of both sexes. I.S.R. data on preadolescent sexual activity and response supported the Freudian view of the early development and prevalence of preadolescent sexuality (*see* PSYCHOANALYSIS).

Premarital intercourse is acceptable behaviour in most societies of the world; in the U.S. it is legally punishable. About two-thirds of college-educated males interviewed by I.S.R. reported premarital coitus, in contrast to almost all of the lower-social-level males. Reported frequencies were much greater for the lesser-educated male, averaging once to twice a week in younger years, whereas for college-educated males they were only about once a month. Data on the number of men who go to prostitutes remained fairly constant throughout the 20th century, but the frequency of visits apparently decreased by half. About half of the females who married by 20 years of age reported premarital intercourse as opposed to two-thirds of the females who delayed marriage until after age 30.

Petting to orgasm as a premarital sexual activity appears to have been elaborated and extended in the upper social levels of the U.S. Roughly one-sixth of the lower-level males indicated this experience, in contrast to about half of the upper-level males, according to I.S.R. findings. The reported frequency of this behaviour reached its peak by age 25, when it averaged about once in three weeks. It never accounted for more than 3% of the total outlet of any segment of the male sample. The females studied reported petting to orgasm somewhat less commonly. However, petting represented a higher percentage of the total claimed premarital sexual outlet among the females: 5%–15% in various subsamples.

Marital intercourse accounted for most of the sexual behaviour reported by both sex groups. In the U.S. more than 90% of both sexes eventually marry; I.S.R. interviews indicated that marital coitus accounted for about 85% of the outlet in this group. In upper social levels among older people marital intercourse accounted for less than two-thirds of the outlet for both sexes; in lower levels about 90%. Frequencies stated for the early years of marriage, if the marriage occurred by age 25, averaged about four times per week, but by the 50s they averaged once per week or less. About half of the married males and one-quarter of the married females reported extramarital intercourse. For the lesser educated respondents this usually was placed in the early years of the marriage and for the better-educated interviewees in the later marital years.

Most preliterate societies seem unconcerned about masturbation (deliberate self-stimulation) but consider it a poor substitute for heterosexual activity. Studies in the U.S. find more than 90% of adult males say they have masturbated to ejaculation; this report was even more common among males of upper social levels. Among single males Kinsey found that masturbation apparently accounted for 50%–80% (varying with social level) of the total orgasms experienced during adolescence to age 15, but that this proportion dropped to 20%–45% for single males by age 30. Married males said that masturbation accounted for 2%–10% of their sexual outlet.

In contrast, reports by Davis, Dickinson and Beam, Landis and Kinsey showed that only 50%–60% of U.S. females interviewed say they have ever masturbated to orgasm, and the frequency is also considerably less than that volunteered by males. Differences in incidences of masturbation among females by social level were found by Kinsey, but these differences were less pronounced than among males. Masturbation apparently accounted for about 10% of the total outlet of married females at all ages. There is no evidence that masturbation is physiologically harmful; it may be helpful to females in avoiding frigidity in coitus.

Anthropology offers almost no information on orgasm during sleep for preliterate societies. In the U.S. studies almost all the upper-social-level males say they have experienced such nocturnal emissions as contrasted with only about three-fourths of the lower

level males. In I.S.R. interviews it accounted for about 15% of the outlet for upper-level unmarried males but only for about 5% for lower-level males. In marriage, however, such dreams constituted 2%–6% of the total outlet. Only about one-third of the females in the U.S. sample said they had experienced orgasm while asleep, with little difference among social levels. Such dreams accounted for only 2%–3% of the outlet of both the single and married females.

Orgasms during sleep seem to involve other factors besides the release of tensions resulting from lack of waking sexual outlet. Older males particularly have reported that a reduction or cessation of sexual outlet is not followed by a compensatory increase in nocturnal emissions. In some cases coital orgasm seems to induce or increase nocturnal emission. Discharges during sleep were rarely judged by males to be sufficiently frequent to meet their needs.

Surveys have shown that homosexual activity is more often allowed in preliterate societies than forbidden. G. P. Murdock reported that of 193 different societies, 28% accepted the practice among males and 11% accepted it in females. Only 15% rejected male homosexuality and 10% rejected female homosexuality. The remaining groups tolerated such behaviour under special circumstances or were doubtful in their attitudes. In the U.S., where such activity even between consenting adults with few exceptions is illegal, Kinsey found that nearly one-third of the males and one-fifth of the females claimed some postpubescent overt homosexual experience.

The vast majority of those reporting homosexual experience say they have also had heterosexual experience, usually within the same period of their lives. Only a small percentage of either sex appear exclusively homosexual. Reported frequencies of homosexual activity to orgasm in the I.S.R. sample averaged about once in three weeks for the single males of grade-school and high-school education in their late teens and early 20s and about once in ten weeks for the single, college-educated males at the same ages. The frequency decreased with age and was, of course, considerably less for the married males. Females said their frequencies of homosexual activity to orgasm were nearly as high as did the males with the same social-level differences. About 1%–2% of the married females admitted overt homosexual experience. I.S.R. findings failed to verify the concept that there is inevitably an overt, conscious homosexual phase or stage in early life. Nor was there evidence that such behaviour necessarily accompanied general symptoms of neuroticism, immaturity or emotional maladjustment. (*See* also HOMOSEXUALITY; LESBIANISM.)

Sexual contacts between humans and other animals seem relatively uncommon and ordinarily appear to occur in the first two decades of life. As many as 50% of young unmarried rural males, however, reported such contact (17% to orgasm) according to Kinsey (*see* also SEXUAL DEVIATIONS).

In summary, most postpubescent humans seem to require some sexual activity (including orgasm) for their psychophysiological well-being; but with great interindividual variation. Some apparently need only minimal activity, while others seem to require more than daily orgasm.

Physiological factors (*e.g.*, ill-health, age and hormonal changes) clearly play a role in human sexual behaviour. However, what are loosely called psychological and social elements seem to be of paramount importance. For example, degree of religious devoutness has definite correlates among sexual attitudes and behaviour, as do social customs. Within the past century Terman, Kinsey and others have reported pronounced changes in U.S. social customs; for example, such investigators judge that premarital intercourse has become more common, especially among females, and that the double standard which imposes more restrictions on women than men is being replaced not too slowly by a single standard.

Human sexuality is extremely complex and many of its aspects are poorly known. Much more scientific study will be needed before it can be said that adequate knowledge of the subject is available. (W. B. Po.; P. H. G.)

LEGAL ASPECTS

Human societies attempt to regulate individual sexual behaviour through education, religion, informal social pressure, medicine and, ultimately, law. Education, religion and informal social pressures formulate standards of action primarily by precept and persuasion, although elements of coercion may be present as in the threat of exclusion from religious communion, punishment in afterlife, disgrace and ostracism. The medical approach to sexual deviation may be psychotherapeutic or physical; *e.g.*, administration of hormones to homosexuals, emasculation of violently aggressive males. Some of these techniques, including possible confinement for indefinite periods in mental hospitals, undoubtedly exert an *in terrorem* effect which, however, is purely incidental to other objectives. Legal techniques for control of individual sexual behaviour are, by contrast, predominantly coercive. It is true that a convicted rapist or sodomist may receive psychiatric treatment in prison. A long sentence or delayed parole may be justified as necessary to prevent the dangerous convict from attacking others; *i.e.*, the purpose here is to incapacitate rather than to punish. But, by and large, penal provisions relating to sexual activity seem designed to deter people from violating prescribed standards, by exemplary punishment of those who break the rules. The drastic consequences of penal violation and the availability of less punitive methods of regulating sexual behaviour suggest that the criminal law should be reserved for the most serious departures from standards supported by virtually unanimous opinion in the community.

The principal laws designed to control sexual behaviour are those penalizing rape, sodomy, adultery and fornication, incest, prostitution, indecent assault, indecent exposure and obscenity (*see* RAPE; ADULTERY; INCEST). These offenses may be divided into two groups. One involves physical aggression against unwilling victims, as in forcible rape, violent homosexual attack or indecent assault. The other involves such acts as illicit intercourse or voluntary homosexual relations, acceptable to the immediate participants but offensive to relatives, neighbours or a substantial group in the community. It is with the latter group of offenses that this discussion is primarily concerned, since it presents distinctive problems in law enforcement. Voluntary participants in forbidden sexual behaviour do not usually complain to the authorities and, since the offenses are committed in private, only a small proportion of the activity comes to the attention of law-enforcement officials. Complaints by willing participants may be suspect as blackmail efforts. Attempts to enforce laws against homosexuality, fornication or prostitution may lead to such undesirable police practices as entrapment (luring the suspect into proposing or committing a violation) or unconstitutional intrusion upon premises. The most striking consequence of employing the penal law to suppress behaviour that is condemned primarily because it affronts nonparticipants is the great variety of laws and enforcement policies that necessarily ensue. Nations and states, and social and religious groups within each community, differ in the extent to which they are outraged by violations of sexual mores, in their estimates of the seriousness of the police problem and in their concern for individual freedom to deviate from ordinary norms. Thus, while laws against theft, assault and murder are universal and basically uncontroversial, sex laws vary enormously and are constantly in debate.

For example, fornication (heterosexual intercourse of unmarried persons) is not criminal in England and other countries outside the U.S. About a dozen states in the U.S. take the same position, but the majority penalize a single act of illicit intercourse or at least an "open and notorious" illicit relationship. If a married person is party to an illicit relationship, the offense becomes adultery which is quite generally punishable in the U.S., although not in England, Scandinavia, the Soviet Union or Japan, to name a few examples. In France and some other countries a distinction is drawn between adultery by a wife and adultery by a husband. The adulterous wife and her paramour are punishable generally, but adultery of the husband is punishable only in specified circumstances; *e.g.*, where the offense occurs in the marital abode, or in the case of a prolonged illicit relationship. The variety of atti-

tudes toward penal repression of illicit intercourse is further shown by the range of punishments that may be imposed in those states that do punish the activity: from petty fines up to five years imprisonment. The Louisiana Criminal Code of 1942 and the American Law institute's Model Penal Code exclude illicit intercourse from the category of criminal offenses.

Anglo-American codes provide penalties as severe as life imprisonment for homosexual intercourse, even if the relations are voluntary and between legally competent adults. So-called sodomy laws, actually proscribing a variety of sexual contacts, appear to apply even to married couples. No comparable regulations are found in the codes of Denmark, France, Italy, Mexico, Sweden or Switzerland, among others. The Wolfenden committee in England and the American Law institute in the U.S. have recommended abolition of criminal provisions in this area, except in cases involving violence, children or public solicitation to commercial vice.

Having sexual relations with a child generally is punishable as rape even if the child participates willingly. The critical age ranges from 7 to 18, with 16 the most common. Often the statutes recognize several age brackets and provide higher penalties (approaching or equaling those for violent rape) in the lower age brackets. Obviously a mature male who takes advantage of a child lacking comprehension of the physical and other consequences of intercourse is objectively dangerous as well as flouting strongly held views of propriety. However, the statutory rape laws appear to have gone well beyond the objective of dealing with such dangerous characters. Generally, it is no defense that the male himself was a child, perhaps younger than his "victim," or that he mistakenly but reasonably supposed that his partner was over the critical age, or even that the young female was a prostitute (a minority of states do make previous unchastity a defense). Such legislation penalizes (with severity appropriate only to rape) behaviour that is criminologically indistinguishable from illicit intercourse.

Prostitution is penalized throughout the U.S., although informal police toleration in certain districts is not uncommon. In England it is not criminal for a woman to sell her sexual favour, but she may not solicit business in public places to the annoyance of passersby.

Under the expanding influence of psychiatry, a number of U.S. state legislatures have enacted so-called sex-psychopath laws. The common feature of these laws is provision for commitment of sex offenders for therapy until they may be safely released. The concept of sex offense is variously defined, but generally extends to rape (including statutory rape), consensual homosexuality, indecent assault and indecent exposure. The commitment may be authorized before or after conviction, upon request of the prosecutor and certification by psychiatrists. These laws have been criticized on a number of grounds. The definition of sex offense is in most cases so broad as to include some activity not regarded as manifesting significant psychopathology. The indeterminate sentence is held warranted neither by the dangerousness of the offender (for example, exhibitionists or homosexuals typically do not progress to violent assault) nor by substantial prospect of cure. Not only are psychiatrists divided on diagnosis and treatment of these cases but often a legislature that is quite ready to authorize indefinite commitment fails to appropriate funds sufficient to provide adequate therapeutic effort. Thus in actual administration the sex-psychopath law may result in what amounts to a life sentence for what many consider a petty infraction of sexual conventions.

Obscenity laws may be considered as relating to sexual behaviour since obscenity is generally defined as material that appeals to or arouses prurient or morbid sexual interest. However, the chief impact of the laws is upon those who commercially disseminate alleged pornography; among these entrepreneurs it is money making, rather than sexual behaviour, that is being regulated. Nevertheless it is clear that the legislative purpose is to curtail the opportunities of ordinary people to look at erotic material. The idea is that such stimulation leads to overt illegal sexual behaviour. However, evidence linking exposure to obscenity with overt misbehaviour is far from conclusive, and some scholars believe that the voyeur (like the exhibitionist) is substituting a comparatively harmless outlet for sexual drives that might otherwise be expressed more aggressively. In any event, repression of obscenity raises formidable legal problems in view of constitutional guarantees of freedom of speech and other forms of intellectual and artistic expression. When legislatures use the concept of obscenity to prohibit dissemination of information and devices relating to contraception, as is often the case, the impact upon private sexual behaviour is obvious.

Legal aspects of sexual behaviour include not only the penal measures reviewed here but also much nonpenal legislation; *e.g.*, laws authorizing divorce for adultery or sodomy, or restricting the rights of illegitimate offspring.

See also references under "Sexual Behaviour" in the Index.

(L. B. S.)

BIBLIOGRAPHY.—F. A. Beach, *Hormones and Behavior* (1948), "Instinctive Behavior: Reproductive Activities," in *Handbook of Experimental Psychology*, ed. by S. S. Stevens (1951); C. S. Ford and F. A. Beach, *Patterns of Sexual Behavior* (1951); N. Tinbergen, *Social Behaviour in Animals* (1953); W. C. Young (ed.), *Sex and Internal Secretions*, 3rd ed., 2 vol. (1961); C. N. Armstrong and A. J. Marshall (eds.), *Intersexuality in Vertebrates Including Man* (1964).

Human Studies: H. Ellis, *Studies in the Psychology of Sex*, 6 vol. (1901–10); A. Moll, *The Sexual Life of the Child* (1912); K. B. Davis, *Factors in the Sex Life of Twenty-two Hundred Women* (1929); G. V. Hamilton, *A Research in Marriage* (1929); R. L. Dickinson and L. Beam, *A Thousand Marriages* (1931), *The Single Woman* (1933), S. Freud, *The Basic Writings of Sigmund Freud* (1938); L. M. Terman *et al.*, *Psychological Factors in Marital Happiness* (1938); C. Landis *et al.*, *Sex in Development* (1940); M. Hirschfeld, *Sexual Anomalies* (1948); A. C. Kinsey, W. B. Pomeroy and C. E. Martin, *Sexual Behavior in the Human Male* (1948); G. P. Murdock, *Social Structure* (1949); A. C. Kinsey, W. B. Pomeroy, C. E. Martin and P. H. Gebhard, *Sexual Behavior in the Human Female* (1953); P. H. Gebhard, W. B. Pomeroy, C. E. Martin and C. V. Christenson, *Pregnancy, Birth, and Abortion* (1958); A. Ellis and A. Abarbanel (eds.), *The Encyclopedia of Sexual Behavior*, 2 vol. (1961); G. Winokur, *Determinants of Human Sexual Behavior* (1963).

Legal Aspects: American Law Institute, *Model Penal Code* (Proposed Final Draft 1962); M. Ploscowe, *Sex and the Law* (1951); M. S. Guttmacher and H. Weifhofen, "Sex Offenses," *J. Crim. Law*, vol. 43 (1952); *Report of the Wolfenden Committee on Homosexual Offenses and Prostitution*, H. M. Stat. Off. (1957); H. L. A. Hart, *Law, Liberty and Morality* (1963); R. Slovenko (ed.), *Sexual Behavior and the Law* (1965).

SEXUAL DEVIATIONS are ways of sexual behaviour that replace normal copulation. This article will be confined to certain practices that either do not terminate in normal intercourse or involve some degree of physical injury. Normal sexual desire has (1) as its object a person of the opposite sex of more or less the same age as the person feeling the desire; and (2) as its mode of expression the love play and intercourse necessary for the fertilization of the female. In the sexual deviations, the object may be unbiological (for instance a person of the same sex, a child or an animal) or the mode of expression only may be at fault.

Deviant conduct occurs in animals as well as in human beings; it is apparent sometimes in the immature, but more often in animals deprived of a normal outlet for their instincts. Conditioning may have a great deal to do with causing it, although much arises spontaneously. In human beings it may be regarded as caused either by some biological residue (apes appear to pass through a stage of undefined sexual activity) or, according to Sigmund Freud, by a fixation of the individual libido resulting from a traumatic experience during an early stage of development. Environment, however, appears to be important as a possible cause.

Sexual deviations may lead to unhappiness in marriage, either because they often result in impotence or because of the very nature of the deviation. They are also a major factor in the demand for prostitution and may further lead to crime. Medically, however, they are regarded as psychosexual diseases, and the law is becoming more and more willing to allow deviant offenders to be treated rather than punished. (*See* also SEXUAL BEHAVIOUR: *Legal Aspects*.) Some deviations are definitely curable; some can be alleviated. Since glandular and other physical conditions play only a slight part in causation, treatment must be mainly by psychotherapy. This treatment may take either a short or a long

time, according to the nature of the cause. Patients should not be encouraged to marry until there is strong evidence of cure. Those who cannot be cured should avoid drinking alcohol, as it may release deviant behaviour and lead to criminal charges. Treatment by conditioning has been tried but has not proved satisfactory. In this therapy the patient is shown objects associated with his perversion and is given an emetic such as apomorphine at the same time. More useful is the administration of estrogens, which suppress the libido in males, simultaneously causing atrophy of the genitalia. This treatment should be used in patients unsuitable for psychotherapy or in whom it has failed.

Homosexuality.—Homosexuality (*q.v.*), the female form of which is known as lesbianism (*q.v.*), is the most important sexual deviation. In it the object of desire is unbiological, so that the mode of expression is perforce at fault.

Sodomy.—This expression may be understood in a number of senses: (1) as denoting any homosexual practices between men, in allusion to the story of Sodom, in Gen. xviii–xix; (2) as denoting anal intercourse; (3) as synonymous with bestiality (*see* below); and (4) as comprehending a number of activities ranging from hand-genital contacts with minors to mouth-genital contacts between persons of the opposite sex. In this article it is understood in the second sense.

Anal intercourse has been observed in apes and is often accepted as normal between men and boys in preliterate societies, notably during initiation rites. In civilized communities it occurs between male homosexuals (though far less frequently than is commonly supposed) and also between men and women. Between men the passive partner as well as the active one may experience pleasure, but if the act is committed by force the passive partner may suffer painful injuries. The man who plays the active role with one partner may assume the passive with another.

The desire for anal intercourse has been explained (1) as arising from lack of opportunity for vaginal intercourse, an explanation that cannot be maintained in view of men's committing it on women; (2) as caused by Pavlovian conditioning, which may be of some importance in the case of boys habituated to the passive role; (3) as caused by fixation of the libido at the stage where sexual interest, after being centred on the mouth, is centred on the anus, from which it would normally, but for the fixation, proceed to the genitalia; and (4) as a dominance phenomenon, a view that receives some support from the behaviour of apes (*e.g.*, when a weaker one turns to be mounted sexually if a stronger one threatens to steal its food) but does not account for human behaviour in all its aspects.

Bestiality.—Intercourse between human beings and animals (Lat. *bestiae*) is tolerated and even approved by some primitive peoples, but many societies condemn it and some even punish it with death. In some parts of the United States it was estimated by A. C. Kinsey that between 40% and 50% of boys brought up on farms have some sexual relation with an animal, vaginal intercourse being commoner than anal. The cause of the desire for it is not known for certain; sometimes, however, there may be some inhibition against relations with another human being, or the animal may symbolize a loved human in the deviant's subconscious mind.

Sadism.—Named after Count D. A. F. Sade (*q.v.;* called the marquis de Sade), sadism is the obtaining of sexual pleasure from acts of cruelty. Sometimes sadism and masochism (*see* below) are regarded as correlative aspects of a single deviant tendency, sadomasochism. Sadism appears to shade off into nonsexual cruelty and may be classified in degrees as (1) cruel or destructive acts performed with pleasure but not appreciated as sexual in nature, including certain sports, arson and lynching; (2) cruel acts without erection or ejaculation but with some sexual satisfaction; and (3) acts of outrage accompanied by full sexual satisfaction and associated with erection and ejaculation, including necrophilia (intercourse with the dead) and rape.

Sadism is the association of aggressive behaviour with sexual arousal. Sigmund Freud suggested that this was caused by fixation of the libido at an early stage when aggression was manifested by biting, scratching, soiling or some such act. L. Bender and F. J. Curran suggested that rivalry was important, deprivation of love and parental aggression being other factors. The sadist uses hate as the emotional currency where the normal man uses love. He injures the parts of the woman a normal man wishes to caress. There is often an obsessional element present, and B. Karpman (*The Sexual Offender and His Offenses*, 1954) states, "The abnormal expression of sadism represents inner tension and anxiety for which the act is attempted relief. He relieves fear by doing to another what he fears might be done to him."

Sadism is most dangerous when the sadist is also a psychopath and therefore has no inhibitions against antisocial conduct. Murders and violent assaults are then done repeatedly until the sadist is caught. Characteristic of the sadistic murder are (1) periodicity; (2) cutting, stabbing and sometimes biting, sucking the blood and even eating the flesh; (3) concomitant sexual excitement; (4) the occasional revisiting of the scene of the crime; and (5) normal behaviour until the next occasion. The mental abnormality of most sadists does not amount to insanity and is no excuse for their crimes, but when sadism shades off into schizophrenia the individual may be regarded as irresponsible.

Masochism.—Masochism, the converse of sadism in the sadomasochistic deviation, is named after an Austrian novelist, Leopold von Sacher-Masoch (1835–95), whose writings are centred on it. It is the seeking of what would normally be painful to oneself to induce sexual pleasure. Often associated with fetishisms, it may take the form of a desire to be dominated, humiliated, degraded, enslaved, bound and even castrated and may also appear unrelated to sexual pleasure, as in the practice of religious flagellation.

Masochism appears to be the impulse to hurt others reversed onto oneself. Freud suggested that it was a manifestation of being beaten in infancy. He also believed that the girl's discovery that she had no penis was a strong factor in feminine masochism. K. Horney suggests conditioning as the cause.

Masochism is less important socially than is sadism. It does not inspire murders or assaults, but sometimes the masochist is injured so severely as to need medical care or even to die. Thus anyone co-operating with a masochist may lay himself open to charges of assault, manslaughter or murder.

Voyeurism.—Voyeurism, or scopophilia, is the reverse of exhibitionism (*see* below), just as masochism is the reverse of sadism. Thus it is often called scopophilio-exhibitionism. It is the obtaining of pleasure by watching other people undress or by watching sexual intercourse. Voyeurs who are not satisfied by vaudeville acts such as strip tease may go to immense trouble to spy on other people and have been known to fall from buildings and trains in doing so.

Voyeurism appears to be derived from the infantile desire to look. Young monkeys, for instance, are absorbed by the sight of their mother's genitals, and children are intensely curious about sexual matters. Psychoanalysts believe that the voyeur has repressed experiences of the primal scene (*i.e.*, observation of parental intercourse) and is trying to overcome a fear of castration. In normal men the wish to look is a sign of sexual attraction and preliminary to further advances, but in voyeurism all the sexual pleasure comes from looking at nakedness or intercourse. Probably those who watch mating couples show considerable empathy, or imaginative involvement in the act. A. C. Kinsey and his associates pointed out that voyeurism and exhibitionism form an important part of preadolescent sex play. This suggests that, like most deviations, it stems from immaturity.

Voyeurs are called "Peeping Toms," from the legend of Godiva (*q.v.*) and are prosecuted for behaviour contrary to common law and, in England, to the Justices of the Peace act of 1361.

Exhibitionism.—Exhibitionism, or sexual exposure, is exposing the genitalia to obtain sexual pleasure. It usually occurs in males. In exhibitionism the mere showing of the genitals is not sufficient: some emotional reaction (disgust, horror) is sought from the person to whom exposure is made. Ejaculation may occur at or after exposure, or the patient may masturbate.

Some psychoanalysts think that exhibitionism is the denial of castration and that the shocked reaction of the onlooker proves to the exhibitionist that his own unconscious fears are unfounded.

This explanation is not altogether satisfactory. Psychoanalysts also believe that the castration shock is greater in women than in men, and if this is so one would expect women to be more sexually exhibitionistic than men. Men with castration fears are not necessarily exhibitionistic.

Biologically, exhibitionism is the seeking of love. It can be seen in the love dances of many animals, including apes, and in the sexual dances of primitive tribes. Clinically it is often seen, as N. K. Rickles suggested, in men who live in a female-dominated environment. Where a man is subservient to an assertive wife or mother he sometimes goes to some place where he exhibits himself as a means of proving his virility. This desire to assert his masculinity may be uncontrollable.

BIBLIOGRAPHY.—C. S. Ford and F. A. Beach, *Patterns of Sexual Behavior* (1951); A. C. Kinsey, *Sexual Behavior in the Human Male* (1948); L. Bender and F. J. Curran, "Children and Adolescents Who Kill," *J. Crim. Psychopath.*, 1:4 (April 1940); B. Karpman, *The Sexual Offender and His Offenses* (1954); K. Horney, "The Problem of Feminine Masochism," *Psychoanal. Rev.*, 22:3 (1935); N. K. Rickles, "Exhibitionism," *J. Nerv. Ment. Dis.*, vol. xcv (1942); S. Freud, *Psychoanalyse* (1917). (C. A.)

SEYCHELLES, a group of islands in the Indian Ocean about 600 mi. NE of Madagascar, which with the Amirante Isles, Cosmoledo Group, and other islands far to the southwest form the British colony of Seychelles. The nearly 90 islands in the colony have a total area of about 145 sq.mi. (375 sq.km.). The capital is Victoria on Mahé Island.

Physical Geography.—The Seychelles Group rise from a submarine ridge where depths are less than 50 fathoms. The largest are high granitic islands, others being low coral islets little above sea level. Innumerable coral heads rise almost to the surface. The high islands are of hornblende granite, gray and red, with dikes of younger igneous rocks, but there are no signs of recent volcanic activity; large boulders litter the lower slopes. All are much weathered and covered with laterite, in places yards thick. The islands and the Seychelles ridge are thought to be remnants of ancient Gondwanaland (*q.v.*).

The largest of the Seychelles is Mahé Island, 16 mi. (26 km.) long and 4 mi. (6 km.) wide, with an area of 56 sq.mi. (145 sq.km.). Its backbone is formed by a continuous rugged ridge with many spurs, in parts bare and precipitous with striking vertical grooves, and there is a bold westward projection, also mountainous. Many heights exceed 2,000 ft., the greatest being Morne Seychellois (2,971 ft. [906 m.]). The land rises steeply from the coast, but there are small areas of lowland at the mouths of many turbulent rivers, which have cut deep ravines. The east coast has an almost continuous coral reef, adjoining it in the south and a mile out in the northeast with an intervening lagoon.

A few miles east of Mahé there are small, hilly islands and rocky islets, the largest being Saint Anne Island, reaching 830 ft. (253 m.), and Cerf Island. Praslin Island, 26 mi. NE of Mahé, is 8 mi. long and rises to 1,260 ft., near it to the east being Curieuse, La Digue, and Félicité islands, all high. Silhouette Island, rising to 2,467 ft. and very prominent, though only 8 sq.mi. in area, lies 10 mi. NW of Mahé. The low Bird and Denis islands, on the northern edge of the coral bank, overlook the ocean, which deepens abruptly to more than 2,000 fathoms. The bank is similarly steep elsewhere, and anchorage is possible at only a few places on its outer edge.

Climate.—The Seychelles have a hot, humid, equatorial climate, with a small temperature range. The period November–April, with two-thirds of the annual rainfall, is very sultry, almost saturated air being brought by light winds between west and north. May–October is drier with constant, stronger winds from between east and south. At Port Victoria mean monthly maximum and minimum temperatures are 27° and 25.5° C (81° and 78° F). Rainfall, very variable, averages 92 in. (2,337 mm.) annually and ranges from 107 in. (2,718 mm.) in the west of Mahé to 55 in. (1,397 mm.) on Denis Island. There are no tropical cyclones, and gales are few.

Vegetation.—Most of the original rain forest of the high islands has been destroyed for timber or to provide land for crops. On Mahé little remains except in the secluded highlands, where locally useful trees include capucin (*Northea sechellarum*), ironwood (*Stadtmannia sideroxylon*), bois rouge (*Neowormia ferruginea*), and bois montagne (*Campnosperma seychellarum*). Many palms have survived, and on the low islands they are prominent from the sea. Many species of screw pine (*Pandanus*) are common, and mangroves are abundant along the coasts. Among the ferns is the beautiful tree fern. Introductions include the coconut palm (ubiquitous up to 1,000 ft. and economically important), breadfruit and jack, litchi, mango, mangosteen, eucalyptus, casuarina, banyan, dragon's blood (a large *Dracaena*), cinnamon, vanilla, pineapple, and bamboo. A notable feature is the number of endemic plants, for old forms survive from the former continent and new ones have had time to evolve. They account for two-fifths of the species and include one family (Medusagynaceae, very local) and about a dozen genera. Among the six native palms is the coco de mer or double coconut.

Animal Life.—Partly as a result of man's interference there is a poverty of animal life. Of mammals only rats and some African and Indian species of bats remain; the dugong no longer frequents the island waters. Reptiles comprise two snakes and seven lizards (three endemic), of which one is a gecko varying in colour between the islands; turtles come ashore to lay their eggs, but the crocodile, once common, has been exterminated. Giant tortoises are found on Mahé. Of amphibians there are two genera of tree frogs. Birds of 15 species (13 endemic) include 3 pigeons, a black parrot, and a weaverbird, all closely allied to species in Madagascar, and an Indian myna. (W. G. KE.)

Population.—Uninhabited when first visited by Europeans, the Seychelles were colonized by Frenchmen from Mauritius, who brought their African slaves. In the 19th century they were joined by deportees from France and a few British and by African slaves freed from Arab dhows by the British Navy. Asians from China, India, and Malaya came in smaller numbers later, mostly as traders and shopkeepers, and miscegenation left few families of pure European descent. The census of 1960 revealed a total population of 41,041 (1964 est. 46,472). About one-third of the islands are inhabited, but four-fifths of the population lives on Mahé. Victoria, the only sizable town, has more than 10,000 inhabitants. In the mid-1960s the crude birth rate averaged 40 per thousand and the death rate 10.

English is the official language and the medium of instruction in schools, although French is allowed in the courts and legislative council. The commonly spoken language is a creole patois, a mixture of clipped French, much of it archaic, with words from African dialects and a few mutilated English words. Most of the population is Roman Catholic. Less than one-tenth is Anglican, and there are a few Seventh-Day Adventists, Hindus, Muslims, Buddhists, Parsees, and pagans. Witchcraft and voodoo also have their devotees. Street processions on saints' days with pilgrimages to a mountain church and picnics are very popular. There is little native culture, but tortoiseshell ornaments, lace work, raffia, and coco-de-mer mats and fans are sold to tourists.

History.—The Seychelles are believed to have been visited in the 12th century by ships from the Arabian Sea and Persian Gulf. They were known to Portuguese navigators early in the 16th century (Vasco da Gama, 1502; Pedro Mascarenhas, 1505) and were marked on Portuguese charts of 1502. The first recorded visit by a British ship was in 1609. The islands were explored by the Frenchman Lazare Picault in 1742 and 1744 and were formally annexed to France. Mahé, the largest island, was named after the then administrator of Île-de-France (Mauritius), Count Mahé de la Bourdonnais, and the group as a whole were called Îles de la Bourdonnais. In 1756 they were renamed Îles des Séchelles after the Vicomte Moreau de Séchelles, Louis XV's controller general of finance, and a *pierre de possession* ("stone of possession") was planted on Mahé.

The permanent settlement of the islands, which dates from 1768, was largely the work of Pierre Poivre, intendant at Mauritius from 1763 to 1772, who had the idea of establishing spice plantations in the Seychelles. However, when war broke out between Great Britain and France in 1778, orders were issued to the administrator at Mahé to destroy the plantations rather than allow them to fall

into British hands. As a result most of the spice trees were burned when in 1780 a French ship which put into Mahé for water and food, fearing to find the British in possession, hoisted British colours. The islands were surrendered to a British naval force in 1794, but they were to seesaw between French and British possession another six times before the final surrender to Britain in 1810. The Seychelles were formally ceded to Great Britain by the Treaty of Paris in 1814.

The Seychelles were at first administered from Mauritius through civil agents and, later, civil commissioners. The first of these was the Frenchman J. B. Quéau de Quincy, who when he died in 1827 had governed the islands continuously for 34 years, serving both the French and the British. In 1888 an order in council nominated an administrator and executive and legislative councils for the Seychelles, and in 1897 the administrator was given all the powers of a governor. On Aug. 31, 1903, by letters patent, the Seychelles were finally separated from Mauritius and became a crown colony. The elective principle was introduced into the legislative council in 1948. Since then two local government districts, Port Victoria on Mahé and Praslin Island, have been established based on adult suffrage and not on the educational, income, and property qualifications required in the case of the electorate for the legislative council. The legislative council was reconstituted in 1960 (see *Administration and Social Conditions*, below).

From time to time the islands have been used by the British government as a place of detention, the most notable instance in later years being the deportation there of Archbishop Makarios of Cyprus from September 1956 to March 1957.

Gen. C. G. Gordon, who was sent out to make a military survey of Mahé in 1881, propounded the theory that the Vallée des Cocos de Mer on Praslin Island was the site of the Garden of Eden. The coco de mer (*q.v.*), which Gordon considered the forbidden fruit, was known from seaborne specimens before the Seychelles were discovered and has (largely because of the impudicity of the shape of the nut) a remarkable historical reputation as an aphrodisiac.

Administration and Social Conditions.—The colony is administered by a governor with an executive council of four *ex officio* members and others appointed by the governor of whom one must be unofficial. The legislative council, for which elections were first held in 1948, has four *ex officio*, five elected, and three nominated (one unofficial) members. There are two elected district councils, on Mahé and Praslin islands, and nominated local boards for north and south Mahé and La Digue.

Justice.—This is administered through a chief justice presiding over the supreme court with civil and criminal jurisdiction and through two magistrates with restricted criminal jurisdiction. Peace officers on outlying islands can deal with minor offences. Civil appeals go to the Supreme Court of Mauritius and criminal appeals to the Seychelles Court of Appeal.

Living Conditions and Welfare.—Many houses are built of coral hewn into square building blocks. A 50% subsidy is provided for houses built by property owners for their employees. In the mid-1960s the minimum wage for agricultural workers was: (1) 43 rupees a month together with housing for a 30-hour week; or (2) 67 rupees a month with housing for a 45-hour week. There are a few registered trade unions. Personal income tax ranged from 5% on the first 1,000 rupees to 65% on any excess over 100,000 rupees. Those in distress can obtain public assistance. Nearly one-third of the children are illegitimate, and many are cared for by the government-aided L'Oeuvre de Sainte-Elisabeth under the Roman Catholic mission. There are five hospitals, a tuberculosis sanatorium, and several clinics (on Mahé, Praslin, Curieuse, and La Digue). The most prevalent diseases are amoebic dysentery, hookworm, tuberculosis, and gonorrhea.

Education.—In the early 1960s there were more than 30 primary schools, mostly government-run, 8 secondary schools, a technical and a vocational school, and a teacher-training college. Some Seychellois receive further education or vocational training in the U.K. In 1965 a new educational system was established giving greater opportunities for secondary education.

The Economy.—The economy of the islands is mainly agricultural and depends largely on the price of copra. The colony is not self-supporting. A development plan for the period 1964–66 envisaged expenditure of more than £765,000 on natural resources, public works, and communications, and on social services.

Production.—Besides coconut plantations of about 23,000 ac. (9,300 ha.), which are very important, cinnamon trees, patchouli, and vanilla are also cultivated for the export trade. Crops for local consumption include breadfruit, bananas, cassava, and sweet potatoes. Small numbers of pigs, cattle, goats, and poultry are raised for local needs. The government reafforestation plan had achieved 4,000 ac. of managed forests by the early 1960s. Fossil guano is worked on the outlying islands. Fish caught include bonito, mackerel, tunny, and shark; and large numbers of terns' eggs are collected, mostly from Desnoeufs Island in the Amirantes. There is little local manufacturing apart from a sisal rope factory, a coir factory, and some essential oil distilling plants.

Trade and Finance.—In the mid-1960s exports to Britain and other Commonwealth countries, chiefly copra (mostly to India), cinnamon bark and leaf oil, patchouli, vanilla, and salted fish, were valued at about 8,500,000 rupees annually. Imports, which exceeded 14,500,000 rupees in value, included flour, maize (corn), rice, sugar, kerosene and petrol (gasoline), and cotton piece goods. The currency is the Seychelles rupee, valued at 1*s.* 6*d.* and divided into 100 cents. Barclay's Bank Limited has a branch at Victoria, and there is a government savings bank. Revenue is derived largely from customs duties, income tax, and licences.

Communications.—There are about 90 mi. (145 km.) of motorable roads, one going almost all around Mahé. There are regular shipping services to Bombay (India), Durban (South Africa), Dar es Salaam (Tanzania), and Mombasa (Kenya); and a twice-weekly government launch service runs between Mahé, Praslin, and La Digue. Occasional passenger-cargo boats run directly from London and Singapore. There are no railways or air services.

The Seychelles Broadcasting Station transmits nightly from Mahé for several hours. (S. S.-Cl.)

Bibliography.—R. Dupont, *St. Pierre, Astove, Cosmoledo, Assumption, and the Aldabra Group of the Seychelles Islands* (1907), *L'Archipel des Seychelles* (1938); J. T. Bradley, *The History of the Seychelles*, 2 vol. (1940); *British Islands in the Southern Hemisphere, 1945–51* (H.M.S.O., 1951); F. D. Ommanney, *The Shoals of Capricorn* (1952); J. W. F. Rowe, *Report on the Economy of the Seychelles and its Future Development* (1959); E M. Compton Mackenzie, *Mezzotint* (1961); biennial *Colonial Report on the Seychelles* (H.M.S.O.).

SEYDLITZ, FRIEDRICH WILHELM, Freiherr von (1721–1773), Prussian Army officer who is unrivaled in history as a trainer and leader of cavalry, was born at Kalkar, near Cleves, on Feb. 3, 1721. His father, a cavalry major, died in 1728, and he was brought up by his mother in impoverished circumstances. At the age of 13 he became a page at the court of the Hohenzollern margrave Frederick William of Schwedt. There he was introduced to horsemanship and soon attracted attention by his daring equestrian feats. In 1740 he was commissioned in the margrave's regiment of Prussian cuirassiers. His first experience of fighting, in the War of the Austrian Succession, was unfortunate, as he was taken prisoner by the Austrians in Silesia (May 1742). In accordance with the custom of the times, however, he was quickly exchanged. By 1753 his enterprise and efficiency had come to the notice of the Prussian king Frederick II (the Great), who placed him in command of the 8th Cuirassiers, which soon became a model for the Prussian Army.

The Seven Years' War (*q.v.*) gave Seydlitz his real chance. After distinguishing himself at Prague (May 1757) and at Kolin (June 18), he was made a major general. At the Battle of Rossbach (Nov. 5, 1757) he commanded the whole of the Prussian Cavalry (38 squadrons) and won the battle almost unaided by the infantry, but a wound received in the battle kept him out of action for some time. His cavalry saved the day at Zorndorf (Aug. 25, 1758); and later, with a cavalry force of 108 squadrons, he covered the Prussian retreat at Hochkirch. He was wounded again on Aug. 12, 1759, in the disaster at Kunersdorf (Kunowice). At Freiberg (Oct. 29, 1762) he commanded a mixed force (infantry and cavalry) for the first time, and his fine leadership decided the battle.

After the Peace of Hubertusburg (1763), Seydlitz was made in-

spector general of the Silesian Cavalry and in 1767 was promoted to the rank of general of cavalry. Toward the end of his life he fell out of favour with Frederick the Great, but the two men appear to have been reconciled just before Seydlitz' death, which took place at Ohlau (Olawa) on Nov. 8, 1773. He was a man of fiery temper, quite incapable of remaining out of the thick of any fight within reach, but he had no opportunities as a strategist or in the higher military art. (C. N. B.)

SEYHAN, the name of an *il* (province) and a river in southern Anatolia, Turkey. The *il* (area 6,660 sq.mi. [17,249 sq.km.]; pop. [1960] 760,803) extends from the Mediterranean coast in the south to the inner slopes of the Taurus Mountains in the north. The northern part, which occupies two-thirds of the total area, is rugged and sparsely populated; agriculture and animal husbandry are the main occupations. The large plains of Adana and Ceyhan in the south are well irrigated and densely populated; the chief crops are cotton, cereal, sesame, flax, rice, and citrus fruits. Adana (*q.v.*) is the capital. The province is well served by rail and road.

Seyhan River (length 348 mi. [560 km.]) rises in the east of the middle Taurus Mountains and flows south-southwest through Adana (forming a delta above it) into the Mediterranean Sea. A dam with a power station, north of Adana, provides irrigation and electricity. (N. Tu.; S. Er.; E. Tu.)

SEYMOUR, HORATIO (1810–1886), governor of the state of New York and Democratic candidate for president in 1868. Born in Onondaga County, N.Y., on May 31, 1810, he attended Geneva Academy (the present Hobart College) and a military school in Middletown, Conn. He was admitted to the bar in 1832; served as military secretary to Gov. W. L. Marcy, 1833–39; was a member of the New York Assembly, 1842–46; and was elected mayor of Utica in 1842. The Democrats of New York nominated Seymour for governor of the state in 1850 but he failed to win election; two years later he succeeded in winning election but lost his bid for a second term in 1854, mainly because of his opposition to a bill prohibiting the sale of intoxicating liquor.

In national politics Seymour, a conservative, supported the policies of Pierce and Buchanan. While strongly supporting the preservation of the Union, he advocated compromise to avoid secession and war. When war came he gave his full support to the Union cause but opposed Lincoln's policies on emancipation, conscription, and military arrests. As governor of New York from 1863 to 1865 he felt that the draft quotas discriminated against New York City and appealed to President Lincoln to correct the situation. When the draft riots occurred in July 1863 he proclaimed the city and county of New York to be in a state of insurrection but in one of his speeches he adopted a conciliatory tone that weakened him politically. After the war, he was nominated as the Democratic candidate for president to run against Gen. Ulysses S. Grant. Though the popular vote was fairly close, in the electoral college Grant won 214 votes to Seymour's 80. In the years that followed Seymour took an active part in politics as a private citizen. He died in Utica on Feb. 12, 1886.

SEYMOUR, JANE: *see* Jane (Jane Seymour).

SEYMOUR OF SUDELEY, THOMAS SEYMOUR, Baron (*c.* 1508–1549), lord high admiral of England from 1547 to 1549, whose political intrigues led to his execution for treason, was the fourth son of Sir John Seymour of Wolf Hall, Wiltshire. His father held office in the royal household; his sister Jane became the third wife of Henry VIII; and on the accession of their son, the young Edward VI (1547), Seymour's elder brother Edward, duke of Somerset, became regent as lord protector. These connections, together with his own reckless and flamboyant character, determined Seymour's fate. From 1536, when he became Henry VIII's brother-in-law, he obtained minor employment at court and on diplomatic missions abroad, but he found the times more propitious after the renewal of war with Scotland in 1542. In 1544 he was appointed master of the ordnance and commanded the Channel fleet in the war with France; when his brother succeeded to the control of the state on the accession of Edward VI (January 1547), Thomas, created Baron Seymour of Sudeley (February), was also made lord high admiral, and at once began to intrigue for power.

His appearance and charm in 1547 won him the hand of Henry VIII's widow, Catherine Parr, and nearly proved dangerous to the adolescent Princess Elizabeth (later Elizabeth I), a guest in the Seymour household, to whom he openly made advances. It also impressed the boy-king Edward over whom the admiral exercised considerable influence. On the other hand, he used his office to enter into profitable, if criminal, relations with the Channel pirates, and by his arrogance and irresponsibility lost his brother's support. When, after his wife's death in September 1548, Seymour made plain his intention to marry Elizabeth, the protector had to act. In January 1549 Seymour was arrested for treason and felony; his guilt was manifest, and he was executed on Tower Hill, London, on March 20, 1549. The protector's Roman sternness toward his own brother was one of the charges raised against Somerset by the politicians who overthrew him later that year. (G. R. E.)

SFAX (Safaqis), second largest town and port of Tunisia, headquarters of Sfax governorate, lies on the Khalif Qabis (Gulf of Gabès) opposite the Juzur Qarqannah (Kerkena Islands) 82 mi. (132 km.) S of Susah (Sousse) by road. Pop. (1956 census) 65,635, including 3,168 Jews and 9,760 Europeans, many of whom left in the late 1950s. The old Arab town (medina), with its 10th-century great mosque and kasbah or citadel, is surrounded by the 9th-century tower-flanked walls pierced by three gates. The modern (former European) quarter, on a geometrical plan, stretches to the port and is flanked by suburbs on the north and southwest. Sfax is the market for the phosphates of the Qafsah (Gafsa; *q.v.*) region with which it is connected by rail. The production of olive oil is the chief industry; the fisheries are also important. There is some light industry, including metal factories and mechanical engineering. The town is the centre of a network of roads linking it with the rest of Tunisia. It is connected by rail north with Susah and Tunis, south with Gabès (Qabis), and southwest with Qafsah and Tawzar (Tozeur). The spring tidal range in Sfax port is 5½ ft. (1.6 m.), which is rare in the Mediterranean. The harbour, to which a channel 1¼ mi. (2 km.) long and of varying width gives access, was built in 1895–97 and has since been deepened. The phosphates enterprise of Gafsa has built huge stores there, and the greater part of the port's exports is phosphates. Other exports include oils, esparto grass, sponges, and dates; the imports are textiles, cereals, sugar, petroleum products, timber and other building materials, and machinery parts. The harbour was an important Axis base from November 1942 and was almost entirely destroyed by Allied air raids. It was captured in 1943 by the British army.

Sfax, built on the site of two small towns of antiquity, Taparura and Thaenae, was of little importance until the 11th century. It remained isolated and was inhabited by nomads. The Sicilians under Roger the Norman took it in the 12th century. In the 17th century it began to be surrounded by gardens and in the early 19th century by olive groves. The town was bombarded by the French in 1881 prior to their occupation of Tunisia.

Sfax Governorate (Safaqis *wilayat*) had a population (1956) of 338,268, including about 100,000 Muslims in the neighbourhood of the town. Area 3,425 sq.mi. (8,871 sq.km.). The governorate includes the wide coastal belt of sparse gardens and olive groves, extends westward over the steppes, and stretches southwestward to the phosphate-bearing Gafsa region. Its administrative area includes the Juzur Qarqannah. (J.-J. Ds.)

SFORZA, the name of a famous Italian dynasty, important for its role in Milan (*q.v.*), especially when that duchy was a field for the international contest between France and the Holy Roman emperors of the House of Habsburg. It was descended from the Attendoli of Cotignola (Romagna), a wealthy and warlike family.

Muzio Attendolo (1369–1424), nicknamed "Sforza" for his strength, became a distinguished mercenary captain in the wars of Italy. Like other *condottieri*, he served many masters and aspired to the possession of a patrimony if not a principality. He was most prominent in the wars of the Neapolitan succession at the time of Joan II (*q.v.*), in which finally he lost his life.

Muzio was certainly successful in acquiring lands and titles in Romagna and in Naples, but it was reserved to two of his sons

FRANCESCO (1401–66) and ALESSANDRO (1409–73), likewise *condottieri*, to achieve the status of princes. In 1445 Alessandro acquired possession of Pesaro, in the Papal States, where he was subsequently made vicar by the pope; his descendants retained the city, except during its conquest by Cesare Borgia (*q.v.*), until 1512, when Pope Julius II expelled them. Francesco had to wait longer and strive harder before winning himself a state. His first attempt, directed likewise against the Papal States, where for 14 years (1433–47) he held the March of Ancona, ended in failure and dispossession; and success, when it came, lay in the north of Italy in the duchy of Milan, which had long been one of the greatest Italian states. After his father's death, Francesco had entered the service of the last Visconti duke of Milan, Filippo Maria, a morose and unpredictable prince. For more than 20 years Francesco fought sometimes for Filippo Maria, sometimes against him. During the periods of uneasy harmony Francesco had succeeded in getting himself betrothed (1433) and then married (1441) to Filippo Maria's illegitimate daughter, Bianca Maria, receiving as dowry Pontremoli and Cremona; and he had attained also some promise of succession in the duchy itself. But Filippo Maria did not nominate Francesco as his successor when he died (1447), and it was primarily by right of conquest that Francesco triumphantly entered Milan in March 1450, overthrowing the unstable republic that had been proclaimed at Filippo Maria's death. For 16 years Francesco ruled as duke in Milan, though without formal recognition from his suzerain, the Holy Roman emperor Frederick III, whose price (in money) for a "legitimate" title was too high. His government, which in the March of Ancona had been burdensome because of the expense of perpetual war, seems in Milan to have been enlightened, though despotic. He himself was frugal, retaining the manners of a soldier, but his court conformed to the lavish fashion of the time and his children were raised under leading men of letters in humanist education.

Francesco left several sons, among them ASCANIO (1455–1505), who became a cardinal in 1484, worldly, magnificent, and lettered, and Galeazzo Maria and Ludovico il Moro, who in turn both ruled the duchy.

GALEAZZO MARIA (1444–76), whose wife, Bona of Savoy, was a younger sister of the French king Louis XI's consort Charlotte, and who desired the emperor to raise his duchy to a kingdom, was in many ways a capable ruler (for example, he introduced the cultivation of rice into the territory); he was also an active patron of letters and the arts. But the character traditionally given him is very different from his father's, showing him vain, extravagant, dissolute, and often cruel. He was murdered on Dec. 26, 1476, by three citizens of Milan, who hoped thereby to free their city from despotism; but no popular rising followed to endorse their deed, and the duke was rather mourned in his dominions, which in the past had several times stirred against his harsh taxation. Even so the murder did prepare the way for the end of Sforza rule by raising the problems and divisions of a regency.

Galeazzo Maria's son GIAN GALEAZZO (1469–94) never proved himself strong in mind or body, and after a period of conspiracy and violence, effective power passed (1479–80) to his uncle, LUDOVICO IL MORO (1451–1508), whose main concern thereafter was to retain control of the duchy. Ludovico, who in 1491 married Beatrice d'Este, a brilliant personality and no less ambitious than himself, is both celebrated and ill-famed in Italian history: celebrated for his court and for his patronage of such men as Leonardo da Vinci; ill-famed, though perhaps unjustly, for having encouraged Charles VIII (*q.v.*) of France to invade Italy and so begin a series of devastating wars. This was the effect of his design to rule perpetually in Milan and to displace Gian Galeazzo for good. In 1489 Gian Galeazzo had married Isabella of Aragon, granddaughter of Ferdinand I of Naples, and it was to conquer Naples from Ferdinand's heirs that Charles VIII entered Italy in 1494, with Ludovico's reluctant connivance. In September 1494 Ludovico was secretly invested with the duchy of Milan by the emperor Maximilian I, to whom he married his niece Bianca; and in October, after the timely death of Gian Galeazzo, he was chosen duke by the Milanese. His triumph, however, lasted only until 1499, when he was driven from power by Louis XII (*q.v.*) of France. Reinstated for a short time by the Swiss, he was eventually delivered by them to the French (April 1500) and died a prisoner in the castle of Loches.

The two sons of Ludovico, MASSIMILIANO (1493–1530) and FRANCESCO MARIA (1495–1535), took refuge in Germany. The former was restored to Milan by the Swiss in 1512 but, after the defeat of his allies at Marignano (1515), surrendered his rights to the victor, Francis I of France. He died later in Paris. Francesco Maria was put in possession of Milan after the defeat of the French at Bicocca in 1522. His death (Oct. 24, 1535) marked the extinction of the ducal male line of the Sforza.

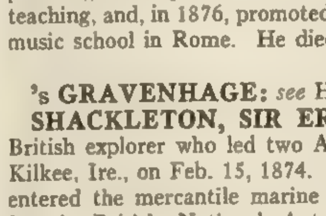
The dukes Sforza-Cesarini (Rome) are descended from Bosio (1411–76), a bastard son of Muzio Attendolo. The counts Sforza, one of whom was the anti-Fascist statesman Carlo Sforza (1873–1952), descend from Sforza Secondo, one of Francesco's younger sons. Gian Galeazzo's daughter Bona married Sigismund I (*q.v.*) of Poland.

BIBLIOGRAPHY.—C. M. Ady, *History of Milan Under the Sforza* (1907); F. Malaguzzi Valeri, *La Corte di Ludovico il Moro* (1913–23); G. Franciosi, *Gli Sforza* (1931); A. Visconti, *Storia di Milano*, 2nd ed. (1952). (P. J. J.)

SGAMBATI, GIOVANNI (1841–1914), Italian pianist, conductor, and composer who promoted a revival in Italy of instrumental and symphonic music during the second half of the 19th century, was born in Rome on May 28, 1841. He studied the piano and harmony at Trevi. Later he returned to Rome where he took piano lessons from Liszt and gave recitals of works by German composers who had hitherto been neglected in Italy. In 1866 he formed an orchestra in Rome and conducted the first performances in Italy of Beethoven's *Eroica Symphony* and Liszt's *Dante Symphony*. He also introduced Beethoven's *Emperor Concerto*, playing the solo part himself. In 1867 he helped to establish the Società Romana del Quartetto. In 1869 he visited Germany with Liszt and became acquainted with the music of Wagner. His own compositions had attracted attention, and Wagner, while in Rome in 1876, secured the publication of two piano quintets by Sgambati and some of his songs. Besides chamber music, piano pieces, and songs, Sgambati composed a Requiem Mass, two symphonies, and a piano concerto. He devoted his later years to teaching, and, in 1876, promoted the foundation of the first public music school in Rome. He died in Rome on Dec. 14, 1914. (DY. H.)

's GRAVENHAGE: *see* HAGUE, THE.

SHACKLETON, SIR ERNEST HENRY (1874–1922), British explorer who led two Antarctic expeditions, was born in Kilkee, Ire., on Feb. 15, 1874. Educated at Dulwich College, he entered the mercantile marine service. He joined Capt. R. F. Scott's British National Antarctic ("Discovery") Expedition (1901–04) as third lieutenant and took part, with Scott and Edward Wilson, in the sledge journey over the Ross Ice Shelf when latitude 82° 16′ 33″ S was reached. His health suffered and he was invalided out on the supply ship "Morning" in March 1903. In January 1908 he returned to Antarctica as leader of the British Antarctic ("Nimrod") Expedition (1907–09). The expedition, prevented by ice from reaching the intended base site in Edward VII Peninsula, wintered on Ross Island, McMurdo Sound. A sledging party, led by Shackleton, reached within 97 mi. of the South Pole, and another, under T. W. Edgeworth David, reached the area of the South Magnetic Pole. Victoria Land Plateau was claimed for the British crown. On his return Shackleton was knighted and was made a companion of the Royal Victorian Order. In March 1914 the British Imperial Trans-Antarctic Expedition (1914–16) left England under his leadership. He planned to



THE BETTMANN ARCHIVE

SIR ERNEST HENRY SHACKLETON, PHOTOGRAPHED ABOUT 1905

cross Antarctica from a base on the Weddell Sea to McMurdo Sound, via the South Pole, but the expedition ship "Endurance" was beset off Caird coast and drifted for 10 months before being crushed in the pack ice. The expedition then drifted on ice floes for another five months and finally escaped in boats to Elephant Island in the South Shetland Islands. Shackleton and five others sailed 800 mi. to South Georgia in a whale boat, and then made the first crossing of the island, to seek aid. He led four relief expeditions before succeeding in rescuing his men from Elephant Island. A supporting party, the Ross Sea party led by A. E. Mackintosh, sailed in "Aurora" and laid depots as far as latitude 83° 30′ S for the use of the Trans-Antarctic party; three of this party died on the return journey.

Shackleton died on Jan. 5, 1922, at Grytviken, South Georgia, at the outset of the Shackleton-Rowett Antarctic Expedition in "Quest"; his exertions in raising funds to finance his expeditions, and the immense strain of the expeditions themselves wore out his strength. He was ambitious both for himself and for the honour of his country, but his outstanding characteristic was leadership. Courage, optimism, and endurance, coloured by a streak of romanticism, earned him the trust and devotion of all his men under the most difficult circumstances.

Shackleton's publications are *The Heart of the Antarctic* (1909) and *South* (1919).

See H. R. Mill, *The Life of Sir Ernest Shackleton* (1923); Margery and James Fisher, *Shackleton* (1957). (L. M. Fs.)

SHAD, fishes of the herring family (Clupeidae) that seasonally leave the sea and swim upriver to spawn. They are characterized by a notch in front of the upper jaw into which the tip of the lower jaw fits. They constitute the genus *Alosa*. Young shad have small teeth, but the adults are toothless. The American shad (*Alosa sapidissima*) ranges from Florida to Newfoundland, and, as a result of an introduction in 1871, from San Diego to British Columbia. It is marine, migrating along the coast to richer feeding grounds during the summer, generally northward and sometimes seaward 100 miles or more. In late autumn they evidently retire into deep water. The sexually mature shad reappear to enter rivers as early as January in the south or as late as June in the northern part of their range. They journey upstream as far as 300 miles or more, produce 30,000 to 156,000 eggs, and then return to the sea. The eggs hatch in 6 to 15 days, depending on temperature; the young remain in the stream until fall and then move downstream to the sea. Shad mature in their third or fourth year, and continue to spawn annually thereafter for several years. They feed on plankton, and attain a maximum length of about 2½ ft. They are excellent game fish, taking an artificial fly or live minnow when running upstream to spawn. They are a favourite food fish, over 10,000,000 lb. a year being caught by United States fishermen.

The Allis shad (*A. alosa*), of the coasts of Europe, has very numerous, long, and slender gill rakers; it attains a length of 30 in. and a weight of 8 lb. The twaite shad (*A. finta*) is smaller and has fewer and shorter gill rakers; the Mediterranean form (*A. f. nilotica*) is distinct from that of the Atlantic, and the species also includes some well-marked forms permanently resident in freshwater, one from Killarney and two from lakes in northern Italy.

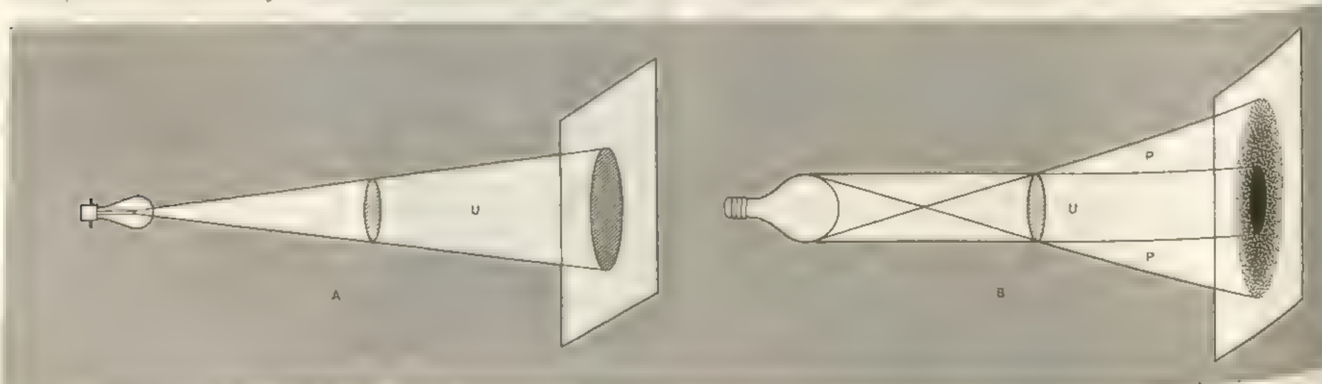
The shad of Black and Caspian seas (*Caspialosa*) have teeth on the roof of the mouth much like the herring. Another important related genus of fishes having a similar life cycle is *Hilsa*, with six species, ranging from east Africa to China. Gizzard shad, which resemble deep-bodied herring, form another family, the Dorosomidae. (L. A. Wd.)

SHADDOCK (Pummelo), a citrus tree, 20 to 40 ft. in height allied to the orange and the lemon, presumably native to the Malay and Polynesian islands. It is *Citrus grandis*, considered the parent type from which the grapefruit (*q.v.*) probably originated. The name "shaddock" is asserted to be that of a captain who introduced the tree to the West Indies. The leaves are like those of the orange, but have broadly winged petioles and are downy on the undersurface, as are also the young shoots. The flowers are large and white, and are succeeded by very large spheroid or almost pear-shaped fruits, resembling grapefruit, lemon yellow in colour, and with a pungent, tart but agreeable flavour. The pulp segments are either pallid or red, and shell out easily. The fruit is highly prized in the Orient. (L. D. B.)

SHADOW usually means that region of an illuminated surface from which the oncoming light rays are shielded by an opaque object. More generally, it can refer to any local changes in the intensity of radiated energy produced by placing some form of matter in its path.

If a source of light is small, as, for example, the filament of an automobile headlight lamp illustrated in part (A) of the figure, the light effectively emanates from a point and the shadow has a relatively sharp boundary between light and darkness. Its shape evidently will be that of the object producing it. If, on the other hand, the source has an appreciable size, as in part (B) of the figure, there is a region U of the total darkness called the umbra and also another P called the penumbra, in which there is a gradual transition from dark to light in going from the inner boundary to the outer. Viewed from the region of the penumbra, only part of the source may be seen, the rest being covered by the obstacle.

Pronounced shadows are not formed if there is more than one light source, except when one of these is much brighter than any other, as in sunlight. Shadows cast by the sun are not sharply bounded because there is always a perceptible penumbra. This is relatively unimportant for shadows cast at small distances; it is possible to read a sundial (*q.v.*) with fair accuracy. The umbra and penumbra assume vast proportions, however, for the shadows of the moon and of the earth. During an eclipse (*q.v.*) of the sun, the moon's shadow falls on the earth, and the penumbra covers the wide area where the eclipse is partial. An eclipse of the moon, on the other hand, occurs when the moon moves into the earth's shadow. The shadow of the moon on its own surface may make a good part of its sphere invisible, so that it appears from earth as a crescent. Another type of shadow occurs when sunlight penetrates to the ground through very small spaces between the leaves of a tree, or to the wall of a room through a hole in a window shade. Small circular or elliptical patches of light are seen which are actual images of the sun's disk, the spaces between the leaves serving to focus the light. The faithfulness of this image becomes apparent during a partial eclipse of the sun



SHADOWS CAST BY AN OPAQUE DISK USING (A) A SMALL LIGHT SOURCE, AND (B) AN EXTENDED SOURCE SUCH AS A FROSTED LIGHT BULB

when the patches of light become crescent shaped (*see* also CAMERA LUCIDA AND CAMERA OBSCURA).

The nature of a shadow may be modified when the light is deflected from its straight-line path by refraction (*q.v.*). Thus when the moon lies in the umbra of the earth's shadow during a total lunar eclipse, it is still visible by a faint copper-coloured light. This is light that has been bent inward by refraction as it traverses the mantle of air surrounding the earth. The light has a hue similar to that of a sunset because in both cases the light scatters substantially as it traverses a great thickness of air (*see* SKY). Refraction by the atmosphere also has the effect of delaying the time of sunset, that is, the time when the shadowing effect of the horizon is complete. As a result of the downward deflection of its rays, the sun appears to be higher than it actually is by an amount almost equal to its own diameter. Thus, if it were not for refraction, the sun would completely disappear at about the time when it is first seen to touch the horizon.

Shadows can form even without a screening object. A transparent substance may, by reflection or refraction, concentrate light in some places and deflect it away from others. The commonest example occurs when sunlight is reflected on a ceiling by a liquid with a disturbed surface. The shifting patterns of light and shade are produced by surface irregularities acting like curved mirrors to focus the light to a greater or lesser degree. Similar effects produced by refraction may be seen on the bottom of a clear pond when the sun shines through the waves. The air itself, when it has turbulent regions and irregularities of temperature, can act in the same way. Here the deflections of the light are so slight that the effects are only apparent when the source is a bright one of very small dimensions. For example, at the instant before a solar eclipse becomes total, darker areas called shadow bands are observed rushing across the landscape, an effect attributed to refractive effects of the upper atmosphere. On a laboratory scale, the same principle has been put to practical use in studying the flow of air around objects in wind tunnels by the Schlieren method (*see* AERODYNAMICS).

X-ray photographs are shadowgrams produced by absorption of the rays coming from the X-ray tube (*see* X RAYS). Sound vibrations, especially those of high pitch, may be screened off almost completely by large obstacles (*see* NOISE AND ITS CONTROL). Even when the source has small dimensions, the resultant acoustical shadows do not have sharp boundaries. The reason lies in an effect called diffraction, exhibited by all types of wave. It becomes more pronounced for longer waves, but occurs even for the waves of light, where the crest-to-crest distance is only about $\frac{1}{20,000}$ cm.

If the edge of an apparently sharp shadow is examined with a magnifier, the transition from dark to light is found not to be abrupt, several faint bands bordering the edge. These diffraction fringes are completely accounted for as a wave phenomenon, as are some other curious effects such as the existence of a minute bright spot at the centre of the shadow of a circular disk.

Spectacular diffraction fringes are observed when the edge of the moon passes in front of a star. The fringes sweep across the observing telescope at a rate corresponding to the frequency of audible sound waves, and may be recorded photoelectrically. From the clearness of these records, it is possible to deduce the diameter of the star's disk, provided it is greater than the apparent size of a one-cent piece at a distance of 750 km.

See also ELECTRON DIFFRACTION; LIGHT: *Diffraction; Refraction and Double Refraction;* SOUND.

BIBLIOGRAPHY.—F. W. Sears, *Optics*, 3rd ed., ch. 9 (1949); H. W. Liepmann and A. E. Puckett, *Introduction to Aerodynamics of a Compressible Fluid* (1947); F. A. Jenkins and H. E. White, *Fundamentals of Optics* (1957); E. H. Hess, "Shadows and Depth Perception," *Scientific American*, vol. 204 (March 1961); Z. Kopal, *The Moon*, 2nd ed. (1964); B. J. Spenceley *et al.*, "Velocity of Sound in Air Using Schlieren Techniques," *Amer. J. Phys.*, vol. 33 (Jan. 1965).

(F. A. Js.; X.)

SHADOW PLAY (SHADOW SHOW) is drama performed by flat, articulated puppets whose shadows are cast on a translucent paper or cloth screen illuminated from behind.

Shadow plays originated in the Orient. In India, subjects were drawn from the Hindu national epics *Mahabharata* and *Ramayana* (*qq.v.*). These passed to Java where they developed into a popular theatre which performed for all ceremonial occasions. Themes from Javanese legendary history and new ones introduced in the 15th century when Java became Muslim were also used. The figures, made from buffalo hide, were highly stylized, perforated, and painted. The narrator (*dalang*) operated the puppets (usually by thin bamboo rods), spoke their lines, and directed the orchestra. Bali received the shadow play from Java, gave the figures a characteristic stylization, and preserved the Indian themes. In Siam similar stories were presented, and the shadow theatre was also found in Malaya, Burma and Cambodia. The shadow play in China, for which a legendary origin in the Han dynasty is claimed, became a favourite entertainment for court ladies. The figures were made from thin donkey skin, dyed to cast brilliant colours on the screen.

Shadow plays spread westward by way of Persia to Turkey, where the witty and vulgar characters Karagöz and his companion Hadjivad became the heroes of a popular theatre. The puppets were made of camel hide, perforated and coloured to present types such as the opium addict and the idiot, as well as regional characters including the Persian, the Albanian, and the Jew. These plays became popular in Greece, where such showmen as Antonios Mollas gave them a Greek interpretation. Shadow plays were also found in the Muslim countries of northern Africa.

In the 17th century the shadow theatre was known in Italy; from there it spread to other western European countries. The French called them *ombres chinoises*, although they were black silhouettes, instead of coloured like those of the Chinese. Dominique Séraphin opened a shadow show in Versailles in 1774, later moving it to the Palais Royal in Paris. Until 1859 his descendants presented specially written plays, such as *The Broken Bridge*. Sophisticated plays were produced by a group of artists, musicians and writers who regularly met at the Chat Noir cabaret in Montmartre, Paris. Other cabarets imitated them, and shows were given until 1923.

Shadows were popular in America during the 18th century, declined toward its end, and were revived around 1880. At the Chicago World's Fair in 1893, Léon-Charles Marot presented the *Théâtre des Ombres Parisiennes*, similar to the performances of the Chat Noir. Greek emigrants brought their versions of the Karagöz plays to the U.S. where they were seen in Chicago and Detroit coffeehouses.

In modern times Pauline Benton produced shows in the U.S. with authentic Chinese shadows. After World War I, Lotte Reiniger developed in Germany a number of films in stop motion, using black silhouettes against backgrounds of varying tones of gray. Although showmen in Europe and the U.S. occasionally used them, shadow plays were less popular in the 1960s than in the past.

See also PUPPETRY.

See Olive Blackham, *Shadow Puppets* (1960). (M. B. McP.)

SHADWELL, THOMAS (?1641–1692), English dramatist and poet laureate, known for his broad comedies of manners, and as the butt of Dryden's satire, was born at Santon, Norfolk, probably in 1641. He went to Bury St. Edmund's Grammar School in 1655, and to Caius College, Cambridge, in 1656; and entered the Middle Temple in July 1658. After the Restoration he became one of the circle of "court wits" with Rochester, Sedley, and Etherege. In 1664 he spent four months in Ireland, but lived mainly in London, where he benefited from the patronage of the duke and duchess of Newcastle; he is thought to have had a hand in two of the duke's plays. He was also an acquaintance of Sir Robert Howard (*q.v.*) and his brother, Edward, both of whom he satirized in *The Sullen Lovers* (1668).

His friendship with John Dryden (*q.v.*) lasted until the political crisis of 1678–79, when Shadwell espoused the Whig cause with typical blunt aggressiveness, causing offense with the anti-papist propaganda of his play *The Lancashire Witches*. This was held up, partly as a result of Dryden's vigilant hostility, until the autumn of 1681, when an expurgated version was performed. Dryden's attack on Shadwell in *The Medall* (March 1682) was

answered in May by *The Medal of John Bayes:* it used to be thought that *Mac Flecknoe* was Dryden's answer; but this is now known to have been written by 1678. Shadwell continued to attack Dryden with *Satyr to his Muse* (July 1682) and *The Tory Poets* (September), and was satirized as Og in the second part of Dryden's *Absalom and Achitophel* (November). The Revolution of 1688 ousted Dryden from the laureateship in favour of Shadwell, who also became historiographer-royal. He took advantage of the change to obstruct Dryden, and in November 1690 succeeded in having his prologue to Thomas Betterton's *The Prophetess* prohibited. He was less successful with *Cleomenes, The Spartan Heroe,* which was held up on the eve of production (April 1692), but later performed. Shadwell died in London on Nov. 19, 1692.

As a comic dramatist, Shadwell claimed direct descent from Ben Jonson (*q.v.*), whose craft of "humours" he continued, often with considerable success. Though not as great as his master, he is always competent and often brilliant in characterization; he directs his plots with skill; and his prose dialogue is often lively. In general he paints a skilful picture of the Restoration scene. His verse is generally dull, its solidity unrelieved by harmony or elegance; even his translation of Juvenal's *Tenth Satyr* (1687) lacks bite. His official verse as laureate falls even below the mediocrity that has often characterized the output of holders of that office.

See *The Complete Works of T. Shadwell,* ed. by M. Summers, 5 vol. (1927); A. S. Borgman, *T. Shadwell: His Life and Comedies* (1929).
(B. L. J.)

SHAFI'I, AL- (Abu 'Abdallah Mohammed ibn Idris al-Shafi'i) (767–820), one of the greatest jurists of Islam and the founder of the Shafi'i school of canon law, was born in A.H. 150 (A.D. 767) at Gaza or Ascalon, and was brought up by his mother in poor circumstances at Mecca. There, and especially in intercourse with the desert tribe of Hudhail, he gained a knowledge of classical Arabic and old Arabian poetry for which he was afterward famous. About 170 he went to Medina and studied canon law under Malik ibn Anas.

After the death of Malik in A.H. 179 legend takes al-Shafi'i to Yemen, where he is involved in a conspiracy on behalf of 'Ali, carried prisoner to Baghdad, but pardoned by Harun al-Rashid. He was certainly pursuing his studies, and he seems to have gone to Baghdad in some such way as this and then to have studied under teachers of the Hanafi school. He had not yet formulated his own system. After a journey to Egypt, however, he is reported in Baghdad again, as a teacher, between A.H. 195 and 198. There he had great success and turned the tide against the Hanafi school. In 198 he went to Egypt in the train of a new governor, and this time was received as the leading orthodox authority in law of his time. In Egypt he developed and somewhat changed the details of his system, and died in A.H. 204 (A.D. 820). He was buried to the southeast of what is now Cairo, and a great dome (erected *c.* A.D. 1211–12) is conspicuous over his tomb.

The "classical" theory of Islamic law, by which is meant the view that the law is based upon four sources (Koran, *sunna, ijma'*, *qiyas*), was, according to Joseph Schacht, the creation of Shafi'i. The classical theory does not fully explain the facts as unearthed by modern research. Whether the opinion of Schacht, in its full import, is correct has yet to be proved; but his researches are worthy of the most serious consideration, and they give an insight into the system developed by a master jurist. *See* also Islamic Law.

See Joseph Schacht, *Origins of Muhammadan Jurisprudence* (1950); *Shorter Encyclopaedia of Islam,* pp. 512–515 (1953). (A. A.-A. F.)

SHAFT, DRIVING. A driving shaft is a cylindrical, rotating member which transmits motion or power, usually both, from one point to another. The shaft may be made of any material strong enough for the purpose, but nearly all shafts are made of metal, especially steel. Motors, engines and machines of all kinds have shafts, often of a specialized kind, such as the camshaft and crankshaft in an automotive engine. In general, the input to or the output from a shaft may be through a shaft coupling, of which there are various kinds, or through belt-driven pulleys, chain drives, gears, etc. The crankshaft of an automotive engine receives power from the work done by the expanding gases in the cylinder; the camshaft is driven from the crankshaft by gears or a chain and delivers power through the cams that operate the valves.

A line shaft drives more than one machine and is itself driven by a prime mover, the power being taken off via belts or chains, usually at several points along the shaft. Shafts intermediate between a line shaft and a driven machine are variously called countershafts, jackshafts or headshafts. In highly industrialized areas where electricity is readily available the use of individual motor drives for each machine is generally favoured over line shafts. It is common practice to speak of short shafts on machines as spindles, especially shafts that carry cutting tools on metal-working machines. The word axle strictly refers to a stationary member carrying rotating wheels, pulleys, etc., as the axle on a wagon or wheelbarrow; but the drive shafts to the rear wheels of an automobile are also called axles, no doubt a carry-over from horse-and-buggy days.

A shaft is designed to carry without failure or without excessive deformation the forces which act upon it. These forces may cause the shaft to bend, inducing what are called bending stresses within the material, or cause the shaft to twist, inducing twisting or torsional stresses; more often than not, shafts are subject to both bending and twisting forces. Sometimes a force acting in the axial direction causes a tensile or compressive stress. In accordance with theory, these stresses combine and if the resultant combination is too large, the strength of the material will be exceeded and the shaft will fail. Most shafts break because the resultant of the combination of stresses at some point repeatedly exceeds the fatigue strength of the material. The stress at a point is increased by discontinuities such as fillets, etc.

It is quite likely that the breaking strength of a shaft is not the significant criterion. In a large percentage of situations, it is a matter of keeping the deflection of the shaft at a specified minimum. For example, in machine tools, rigidity or resistance to deflection is most important if the machine is to cut metal to close tolerances. This is often true even if the amount of deflection involved is much too small for the unaided eye to perceive. The torsional deflection of most machinery shafts will be less than 0.1° per foot of length. The transverse or bending deflection may be from a few ten-thousandths of an inch, depending upon the requirements. If the transverse deflection is large enough, a shaft may tend to bind in its bearings.

Any rotating shaft has a critical speed at which the shaft vibrates noticeably, perhaps violently. It might be said that at the critical speed the shaft is undecided whether to rotate about its geometric axis or about the axis through its centre of mass and it attempts to do both. This speed can be fairly accurately computed for a given situation; in any case, the shaft must be designed to run either well above or well below the critical speed.

Flexible shafts will bend around obstructions, as a cable bends, or will transmit small amounts of power in a curved form. They are used in dentists' grinding tools, portable drilling and polishing machines, etc.

Shafts are manufactured by cold-rolling or cold-drawing, processes which leave a relatively hard, strong, smooth surface on steel. Because of the residual stresses left within the metal after cold-working, cold-finished shafts will warp when a keyway is cut and must later be straightened. Larger shafts are made from hot-rolled steel, turned on a lathe and polished or ground.
(V. M. F.)

SHAFTESBURY, ANTHONY ASHLEY COOPER, 1st Earl of (1621–1683), English politician who was one of the most prominent and skilful opponents of Charles II's pro-Catholic policy. He was born at Wimborne St. Giles, Dorset, on July 22, 1621, and succeeded his father, Sir John Cooper, as 2nd baronet on March 23, 1631. Largely through his mother, Anne (d. 1628), daughter and heiress of Sir Anthony Ashley, he inherited great estates in the southwest of England, and, although much was lost during his long minority, he remained a wealthy man. Educated partly by tutors, partly at Exeter College, Oxford, and partly at

Lincoln's Inn, he early showed evidence of a restless and acute intelligence, which led to his return to the Short Parliament of 1640 as member for Tewkesbury, Gloucestershire, and in December to the succeeding Long Parliament as member for Downton, Wiltshire. As he was still under age, however, and as his election to the Long Parliament was the subject of a dispute that was not determined until 1660, it seems unlikely that he played any effective part in either Parliament.


DETAIL OF A PORTRAIT BY AN UNKNOWN ARTIST, ABOUT 1672–73; PHOTO, BY COURTESY OF THE NATIONAL PORTRAIT GALLERY

THE 1ST EARL OF SHAFTESBURY, ANTHONY ASHLEY COOPER

Civil War and Commonwealth.—His first marriage, on Feb. 25, 1639, to Margaret, daughter of Thomas, 1st baron Coventry, the lord keeper of England, confirmed his natural inclination to take the royalist side in the English Civil War. Early in 1644, however, he went over to Parliament and took an active part in the concluding stages of the war in the southwest. Thereafter he devoted himself mainly to the problems of local government until appointed a member of the Barebones Parliament in 1653 and returned for Wiltshire to the Parliaments in 1654, 1656, and 1659.

Appointed a member of the Council of State on July 14, 1653, and, after Oliver Cromwell became lord protector, of its successor, the Lord Protector's Council, on Dec. 16 of the same year, he supported Cromwell for some time and opposed the more extreme reformers. At the very end of 1654, however, alienated by the protector's increasing neglect of Parliament, he broke away from him, and in consequence was one of the members denied the certificate from the council without which they could not take their seats in the Parliament of 1656. He strongly protested, the certificates were soon afterward abolished, and he took his seat; but he was not elevated, as in other circumstances he probably would have been, to the new "House of Lords" created in 1657. When his criticisms of that House and of the protector's autocratic methods proved unavailing, he continued his opposition, both for the remainder of Oliver Cromwell's life and under Richard Cromwell.

Restoration of Charles II.—After Richard's fall and the restoration of the Rump of the Long Parliament (May 1659) he was reappointed to the Council of State and came under increasing suspicion of favouring the exiled Charles II. Against the power of the army he upheld the authority of Parliament; but it was a genuine Parliament that he had in mind, and he lent all his weight to secure first his own right to sit in virtue of his election in 1640, then the readmission (December 1659) of the members excluded by Pride's Purge in 1648 and finally the dissolution of the Long Parliament (March 1660) and the election of a new one (the Convention Parliament). Returned to that parliament as member for Wiltshire, he exerted himself to secure the recall of Charles II. On May 7 he was appointed one of the 12 commissioners sent by the Commons to Breda in Holland to invite Charles to return, and after the landing of the king at Dover on May 25 he waited on him at Canterbury and was well received.

Offices and titles were immediately conferred upon him. Admitted to the Privy Council on May 27 and to the Committee for Plantations on July 4, 1660, he was created Baron Ashley of Wimborne St. Giles on April 20, 1661, and appointed chancellor of the exchequer on May 13 of the same year. In the general settlement that followed the return of the king he advocated a policy of moderation, urging lenity toward the "regicides," opposing the more extravagant provisions of the so-called Clarendon Code, and supporting Charles II's declaration of 1662 in favour of toleration. Charles was impressed by his industry and ability, and on the death of the earl of Southampton, the lord treasurer, in May 1667 he was appointed to the commission that took Southampton's place, soon becoming its most influential member.

During these years he developed a deep interest in the problems of commerce and colonization. Part owner of a plantation in Barbados, and member of several committees dealing with colonial affairs under the Commonwealth, he had served a long apprenticeship in the colonial field and was now regularly included in the various committees to which such affairs were entrusted. On March 24, 1663, he was given a grant, along with seven others, of the vast province of Carolina in North America, and found in it an ideal field for experimentation. Making the acquaintance, three years later, of the philosopher John Locke, he entrusted him with the task of drawing up a constitution for the province based upon aristocratic predominance, religious toleration, and other principles he would have liked to see recognized in England. In October 1673 he secured for Locke the position of secretary to the recently established Council of Trade and Foreign Plantations, of which he himself had been appointed president in September 1672.

Meanwhile Ashley's position at court had first improved and then suddenly declined. On the dismissal and impeachment of the earl of Clarendon in 1667 he had attached himself to the duke of Buckingham, with whom he had a common interest in the support of religious toleration. As early as 1670 he became so impressed by the danger of the Catholic James, duke of York (later James II), succeeding to the throne as to give his support to schemes for securing a divorce for the king or legitimizing his bastard son, James, duke of Monmouth. He was therefore excluded from the secret of the real Treaty of Dover (1670), which aimed at the reconversion of England to the Roman Catholic faith. He approved of the sham treaty, which provided only for an Anglo-French alliance to reduce the commercial supremacy of the Dutch, and readily signed it, along with other members of the so-called Cabal, on Dec. 21, 1670.

As a reward he was appointed lord lieutenant of Dorset on Jan. 20, 1672, created earl of Shaftesbury on April 23, and appointed lord chancellor on Nov. 17. As chancellor he was, by general admission, a success; but in his capacity of adviser to the king he was now in a false position, for he was advocating a policy half of which was concealed from him. When Parliament assembled in Feb. 1673 he delivered his famous *"Delenda est Carthago"* speech, justifying the attack on the Dutch as the commercial rivals of England; but as he watched the rising tide of opposition to the king's policy, and began to suspect that there was more in that policy than he had been given to understand, his attitude inevitably changed. Later in the same year he supported the first Test Act, designed to exclude Roman Catholics from office under the crown, and did his best in Parliament to oppose the duke of York's marriage with the Catholic Mary of Modena. On Nov. 9 he was summarily dismissed from the chancellorship and went into opposition. On May 19, 1674, he was also dismissed from the Privy Council and from his local offices.

In Opposition.—It is as a leader of opposition to the court that Shaftesbury is chiefly remembered, but real leadership of the malcontent elements in Parliament and the country long evaded his grasp. His frequent changes of policy and his inclination toward extreme courses were regarded with disfavour by responsible men, and even among his own associates some preferred the leadership of Lord Holles. Successful in obstructing the earl of Danby's nonresisting test bill in the spring of 1675, he failed in the autumn of the same year, even with the assistance of the duke of York, to carry in the Lords an address in favour of a dissolution of the Cavalier Parliament. He then sought supporters among discontented elements in the City of London, elaborated an argument that, in accordance with a statute of Edward III's reign requiring annual Parliaments, a prorogation for 15 months pronounced on Nov. 22, 1675, was in effect a dissolution, and advanced this contention in the Lords when the Houses reassembled on Feb. 15, 1677. He was immediately sent to the Tower of London by the Lords themselves and was not released until Feb. 26, 1678.

The event that gave him his real chance was the appearance of

Titus Oates in the autumn of 1678 with his story of a vast "popish plot" against the king and government. Shaftesbury was in no sense the originator of the tale told by Oates, and indeed began by regarding it with suspicion as a fable inspired by government supporters with the object of rallying the nation round the king; but he soon realized what a useful weapon it might prove in his hands, and thereafter he supported it to the utmost. In the nationwide panic that it engendered, moderate leaders and counsels were forgotten and he was able to secure widespread acceptance as the only man with the courage and skill to ride the storm. By developing an elaborate party organization, based on the famous Green Ribbon Club, he exercised great control over elections, built up a large following both in the Lords and in the Commons, and in three successive Parliaments (1679, 1680, and 1681) endeavoured to secure the exclusion of the duke of York from the succession to the throne. To his efforts at this time was largely due the passing of the Habeas Corpus Act (May 1679), perhaps his greatest and most creditable achievement.

Exclusion Crisis.—With the object of silencing him by giving him a high office Charles on April 21, 1679, remodeled the Privy Council, admitted many of the leading members of the opposition to it, and appointed Shaftesbury president; but the king's failure to follow the advice of his new council soon led Shaftesbury to withdraw his support, and on Oct. 15 of the same year he was dismissed. On Nov. 15, 1680, the Exclusion Bill, which had passed without a division in the Commons, was rejected by the Lords, in spite of some brilliant speeches on his part; this only inspired him to more violent courses. To the succeeding Parliament, which met at Oxford on March 21, 1681, he and his associates rode with an armed following, confident that the king, in his need of money, could not withstand the pressure of the House of Commons any longer. Charles, however, who had just strengthened his financial position by an agreement for a subsidy from Louis XIV of France, dissolved Parliament within a week, and Shaftesbury suddenly found himself practically helpless. Unfortunately the policy by which he had secured his acceptance as the oracle of the nation had almost inevitably been marked by extravagance rather than by statesmanship, and as the panic caused by the "popish plot" declined, his credit had correspondingly diminished. Complete exclusion of the duke of York was a hazardous device which could have been accepted, if at all, only during a moment of panic. The substitution of the duke of Monmouth, at which he aimed, probably could not have been accepted at any time, and was only too obviously inspired by his desire to have on the throne a mere puppet of his own. On the unexpected dissolution of the 1681 Parliament therefore, his following simply disintegrated, and he himself came into a position of considerable danger.

Trial and Flight.—For the short remainder of his life his chief efforts were directed toward his own safety. On July 2, 1681, he was seized, committed to the Tower, and kept there for four months while every effort was made to find or suborn evidence on which to base a charge of high treason against him. So gloomy did his prospects appear that he offered, if released, to retire to Carolina. At his trial on Nov. 24, however, the Whig grand jury selected by the Whig sheriffs of London rejected the indictment, and on Dec. 1 he had to be admitted to bail, from which in turn he was released on Feb. 13, 1682. But the court was in the ascendant, and at the midsummer election contrived to secure the return of two Tory sheriffs for London. Thus deprived of all prospect of protection from a favourable jury at a future trial, Shaftesbury made desperate efforts to organize a rising against the government, and when these proved vain left the country on Nov. 28. Reaching Amsterdam early in December he was granted asylum there and admitted a burgher; but imprisonment and privations had aggravated the diseases from which he had long suffered, and after a short final illness he died in Amsterdam on Jan. 21, 1683. His body was brought back to England, and was buried at Wimborne St. Giles.

Shaftesbury was undoubtedly a very clever man, and particularly adept at turning to his own advantage the weaknesses of others; but he failed to realize that in English politics cleverness alone is not enough, and that men are not composed exclusively of weaknesses. He has been credited with few genuine principles beyond a sincere belief in Parliament and in the desirability of toleration in religious matters. His methods are difficult to defend, involving as they usually did not merely an assumption that the end justifies the means but also a reckless disregard of future consequences in the prospect of immediate gain. Especially toward the close of his life he repeatedly embarked on courses which men much less acute than he was could see were bound to lead to disaster. It is typical of his whole career that he ended his life under the protection of the Dutch, the nation which he had done his best, when at the height of his power, to ruin.

Bibliography.—*The Shaftesbury Papers*, consisting of autobiographical fragments, letters, speeches and documents connected with Shaftesbury's colonial ventures, are in the Public Record Office, London. Many of them are printed in W. D. Christie, *A Life of Anthony Ashley Cooper, First Earl of Shaftesbury*, 2 vol. (1871), which is still the standard biography. H. D. Traill, *Shaftesbury* (1886), and O. Airy's article in the *Dictionary of National Biography* are useful sketches, one unfavourable and the other favourable in its interpretation of Shaftesbury's conduct. L. F. Brown, *The First Earl of Shaftesbury* (1933), devotes attention to his connection with commercial and colonial developments. (A. Bc.)

SHAFTESBURY, ANTHONY ASHLEY COOPER, 3rd Earl of (1671–1713), English politician and philosophical writer, was born at Exeter House, London, on Feb. 26, 1671, the son of Anthony (afterward 2nd earl of Shaftesbury) and grandson of the famous 1st earl (*q.v.*). His early education was directed by the philosopher John Locke. In 1683 he went to Winchester College and in 1686, with a tutor, he toured France, Italy, and Germany, returning to England shortly after the Revolution of 1688. He entered Parliament for Poole, in Dorset, at a by-election in 1695, but his health, never robust, broke down and after the dissolution of 1698 he did not return to the House of Commons.

During a debate in the Commons in 1695 he spoke in favour of a bill to allow counsel to assist those accused of high treason. Too overcome by the occasion to recollect the points he had prepared, he was yet able to make out of his own confusion a powerful argument in support of the bill. "If I," he declared, ". . . am so confounded that I am unable to express the least of what I proposed to say, what must the condition of the man be who is pleading for his life without any assistance?" The bill was passed and became the celebrated Trial of Treasons Act of 1696.

Succeeding as 3rd earl in 1699, in the quieter atmosphere of the House of Lords he was able to attend Parliament regularly for the remainder of William III's reign. He pursued an independent policy, as he had done in the Commons, though his general outlook was that of an advanced Whig. Offered high office, he accepted (June 1701) only the vice-admiralty of Dorset, but after the Tories came to power on the accession of Queen Anne (March 1702) he was deprived of the office in July 1702 and retired from public life.

Having spent most of 1699 in Holland, where the climate suited him better, he now returned there on a further visit (August 1703–August 1704). He was, however, to be an invalid for the rest of his life, and his health again forced him to go abroad in July 1711. He died in Naples on Feb. 15, 1713. He had married Jane Ewer in 1709 and his son Anthony, the 4th earl, was born in 1711.

Shaftesbury's philosophy owed something to the Cambridge Platonists (*q.v.*), who had stressed the existence in man of a natural moral sense. Shaftesbury was at pains to put forward this concept, both against the orthodox Christian doctrine of the Fall and against Hobbes's premise that the state of nature was a state of unavoidable warfare. Against Hobbes's conception of "solitary" man, Shaftesbury maintained that man is essentially social; that his mind (in some degree part of the governing mind of the Creator) is so imbued with the principles of unity and proportion that he can adequately balance the claims of self affection (tending to a private good) with those of natural affection (tending to the public good), while rejecting those of unnatural affection (which furthers neither private nor public happiness). Thus, instead of Hobbes's solution of the setting up of the absolute authority of a Leviathan to maintain order, Shaftesbury contended that this could be achieved by a "right Division and Ballance of Power and

by the Restraint of Good Laws and Limitations."

In his stress upon the natural, Shaftesbury was not a forerunner of the primitivists who exaggeratedly revered the "noble savage"; for him the term had a teleological sense, for he considered the natural state of man to be that in which his faculties were most fully developed.

Shaftesbury also criticized revealed religion for its dependence on the miraculous, which, essentially contravening natural order, thereby insulted the dignity of creation. He also disliked the concept of reward and punishment epitomized in the doctrine of heaven and hell; he maintained that virtuous actions, being more truly pleasant than selfish ones, are therefore literally their own reward. He regarded evil as more apparent than real, explaining those events or tendencies which appear evil to man, as in reality facets of the divine scheme of things which man's limited vision prevents his seeing in their proper perspective.

Shaftesbury's Neoplatonism, his contention that what man sees of beauty or truth is only a shadow of absolute beauty or truth, also dominated his attitude to the arts. Akin to man's innate moral sense is his natural aesthetic sense; but Shaftesbury, in accordance with his view that natural means most fully developed, conceded that taste needs to be trained.

During his lifetime Shaftesbury's fame as a writer was comparatively slight, for he published little before the appearance in 1711 of his *Characteristicks of Men, Manners, Opinions, Times*, in which his chief works were gathered together. The effect of this book was immediate and was felt on the Continent as well as in England. Pope, Joseph Butler, Francis Hutcheson, Mark Akenside, Coleridge, and Kant (*qq.v.*) were among the writers to some degrees affected by Shaftesbury; and his literary criticism did much to form Augustan taste. *See* also ETHICS, HISTORY OF: *Modern Ethics*.

BIBLIOGRAPHY.—A large collection of documents connected with Shaftesbury is among the *Shaftesbury Papers* in the Public Record Office. The best edition of his *Characteristicks* is that of J. M. Robertson, 2 vol. (1900). Various of his letters, originally issued separately, were published in one volume, *Letters of the Earl of Shaftesbury, Author of the Characteristicks* (1746); supplementary writings, mainly concerning art, are in *Second Characteristicks; or, the Language of Form*, ed. by B. Rand (1914). See also the account of Shaftesbury's life, written by his son, *The Life, Unpublished Letters and Philosophical Regimen of . . . Shaftesbury*, ed. by B. Rand (1900); T. Fowler, *Shaftesbury and Hutcheson* (1882); R. L. Brett, *The Third Earl of Shaftesbury: a Study in Eighteenth-Century Literary Theory* (1951); B. Willey, *The Eighteenth Century Background* (1940); E. Wolff, *Shaftesbury und seine Bedeutung für die englische Literatur des 18. Jhs.* (1960). (A. Bo.; X.)

SHAFTESBURY, ANTHONY ASHLEY COOPER, 7TH EARL OF (1801–1885), English philanthropist, and the most famous advocate of social reform in 19th-century England, was born in London on April 28, 1801. He was the eldest son of Cropley, a younger brother of the 5th earl of Shaftesbury, and of Anne, daughter of the 4th duke of Marlborough. He became Lord Ashley when his father succeeded to the earldom in 1811, was educated at Harrow and Christ Church, Oxford, and entered Parliament in 1826. He succeeded his father as earl in 1851.

Although giving general support to the Conservatives, Ashley was always prepared to follow an independent line of action. He opposed the Reform Bill of 1832, but had been a supporter of Catholic emancipation, and his objection to the continuance of resistance to the abolition of the Corn Laws led him, in 1846, to resign his seat for Dorset which he had represented since 1831; he was returned to Parliament for the city of Bath in 1847. His parliamentary conduct was strongly influenced by his evangelical Christianity. He told his biographer that he believed that a man's religion, "if it is worth anything, should enter into every sphere of life, and rule his conduct in every relation." Religion counted with him far more than ambition and led him to devote his parliamentary life to the promotion of social reform. His efforts were concentrated on four objectives—improvement of conditions in lunatic asylums, factory reform, regulation of working conditions in coal mines, and public health.

Ashley delivered his first speech in the House of Commons on the subject of lunatic asylums in 1828. "By God's blessing," he wrote, "my first effort has been for the advancement of human happiness." The reward of this effort came in 1845 when the Lunacy Act of that year for the first time considered lunatics not as social outcasts but as "persons of unsound mind." He became associated with the cause of factory reform in 1832 when a deputation of northern factory reformers asked him to act as their spokesman in the Commons. He took the place of Michael Sadler (advocate of factory reform who died in 1835) and, like Sadler, supported the textile operatives' demand for a statutory limitation of the working hours for women and children in the factories.

Ashley was a lord of the admiralty (1834–35) under Sir Robert Peel, but, unable to obtain Peel's support for the Ten Hours' Bill, he refused to join Peel's administration in 1841. Chiefly by his persistent efforts a Ten Hours' Bill was carried in 1847, but its operation was impeded by legal difficulties, which were only removed by successive acts, instigated chiefly by him, until the law was consolidated by the Factory Act of 1874. During this long struggle his main parliamentary opponent was John Bright (*q.v.*), who opposed all factory legislation. Bright accused Ashley on more than one occasion not only of ignorance of the actual working conditions in the factories but of a lack of concern for rural workers, including those on his own father's estates.

The part Ashley took in the legislation bearing on coal mines was equally prominent, and the opposition he faced from the mine owners was as strong as that he had faced from the factory owners. The result of his initiative was the Mines Act of 1842, which excluded from the coal mines all women and girls, all boys under 13, and all parish apprentices. His main speech in sponsoring the bill was so moving that Richard Cobden, one of his opponents on factory reform, crossed the floor of the house and told him that he had never in the course of his life "been put into such a frame of mind . . . as I have been by your speech."

Ashley's interest in public health brought him new allies as well as more implacable opponents. In 1841, when he was crusading for factory and mine reform, he was taken by a friend to see one of the dingiest and most sordid areas of London's East End. The visit left a deep impression on him and he became an active supporter of national legislation to control health conditions. He welcomed the first national Public Health Act of 1848 and became one of the members of the new General Board of Health set up under the act. He deplored the bitter opposition which led to the dissolution of the board in 1854.

Unlike some of his allies in the struggle for public health, Ashley was keenly concerned about the provision of adequate housing as well as the supply of water and the disposal of nuisances. One of his last acts before leaving the Commons for the House of Lords in 1851 was to introduce bills to encourage the establishment of lodging houses for the working classes and to enforce government inspection and regulation of such houses already in existence. Dickens hailed these measures as among the most altruistic of his age. Shaftesbury continued to take an interest in housing and on his 83rd birthday in 1884 gave evidence before the royal commission on the housing of the working classes. He was also keenly interested in education, particularly the religious education of poor children, and by his zeal gave a new impulse to the movement for the establishment of free schools for destitute children, commonly called ragged schools. For almost 40 years he was president of the Ragged School Union. He was also one of the principal founders of reformatory and refuge unions, Young Men's Christian Associations and Workingmen's Institutes. He supported foreign missions and was president of several philanthropic and religious societies. He was acknowledged leader of the evangelical group within the Church of England, and an outspoken critic both of ritualism and broad church liberalism.

Shaftesbury died on Oct. 1, 1885, at Folkestone.

BIBLIOGRAPHY.—E. Hodder, *Life and Work of the Seventh Earl of Shaftesbury* (1886); J. L. and B. Hammond, *Lord Shaftesbury*, rev. ed. (1936); J. W. Bready, *Lord Shaftesbury and Social-Industrial Progress*, 4th ed. (1933). (A. BRI.)

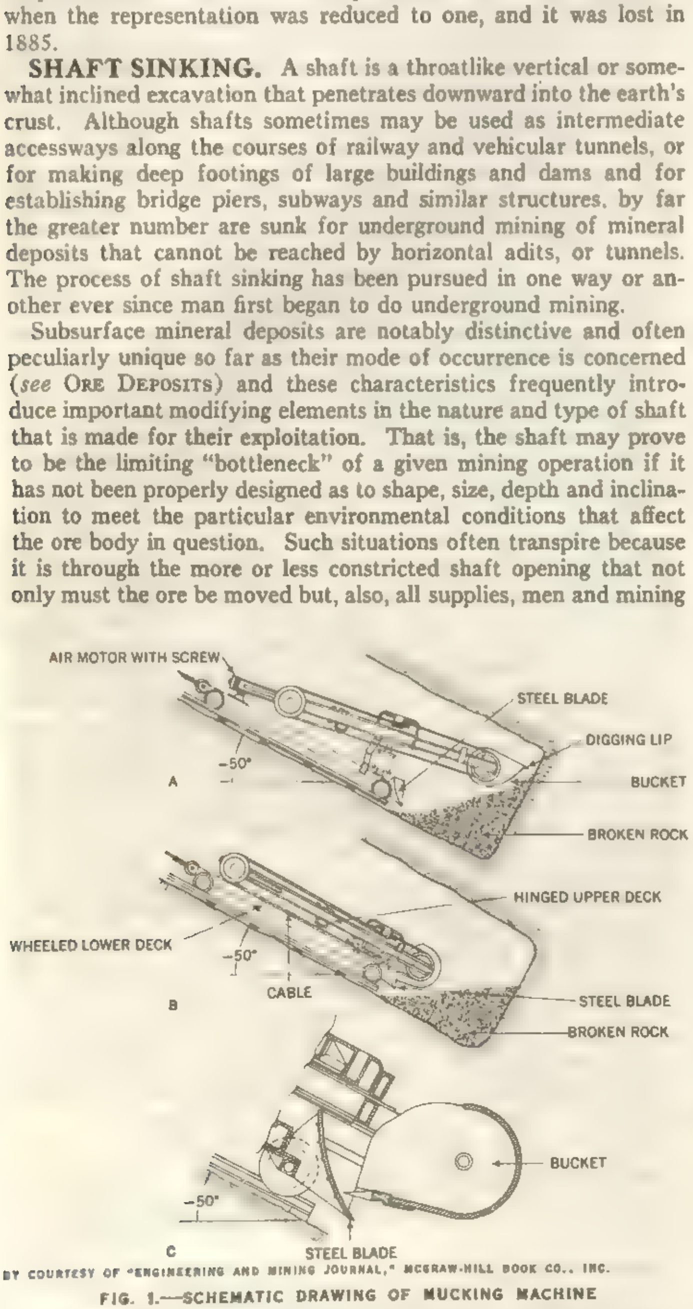
SHAFTESBURY, a market town and municipal borough in the North Dorset parliamentary division of Dorsetshire, Eng., 20 mi. WSW from Salisbury by road. Pop. (1961) 3,372. Area 1.7 sq.mi. (4 sq.km.). It lies high on a hill, on the edge of Cranborne

Chase, above a rich agricultural district. Although traces of British and Roman occupation occur in the neighbourhood, the site of Shaftesbury (Caer Palladur, Caer Septon, Seaftonia, Sceafstesbyrig, Shafton) was probably first occupied in Saxon times. Matthew Paris speaks of its foundation by the mythical king Rudhudibras, while Asser ascribes it to Alfred, who made his daughter Ethelgeofu the first abbess. It is probable that a small religious house had existed there before the time of Alfred, and that it and the town were destroyed by the Danes, both being rebuilt about 888. The foundations of the Saxon abbey and nunnery have been laid bare. In 980 Dunstan took St. Edward's body there from Wareham for burial, and there Canute died in 1035. The relics of St. Edward were discovered in 1931 during excavation.

In 1252 the burgesses received their first charter from Henry III. There is no evidence that Elizabeth I granted Shaftesbury a charter as has been asserted, but she confiscated the common lands in 1585, the town only recovering them by purchase. This probably led to the granting of a new charter by James I in 1604. Yet another charter was granted to the town in 1684. Shaftesbury returned two members to parliament from 1294 to 1832, when the representation was reduced to one, and it was lost in 1885.

SHAFT SINKING. A shaft is a throatlike vertical or somewhat inclined excavation that penetrates downward into the earth's crust. Although shafts sometimes may be used as intermediate accessways along the courses of railway and vehicular tunnels, or for making deep footings of large buildings and dams and for establishing bridge piers, subways and similar structures, by far the greater number are sunk for underground mining of mineral deposits that cannot be reached by horizontal adits, or tunnels. The process of shaft sinking has been pursued in one way or another ever since man first began to do underground mining.

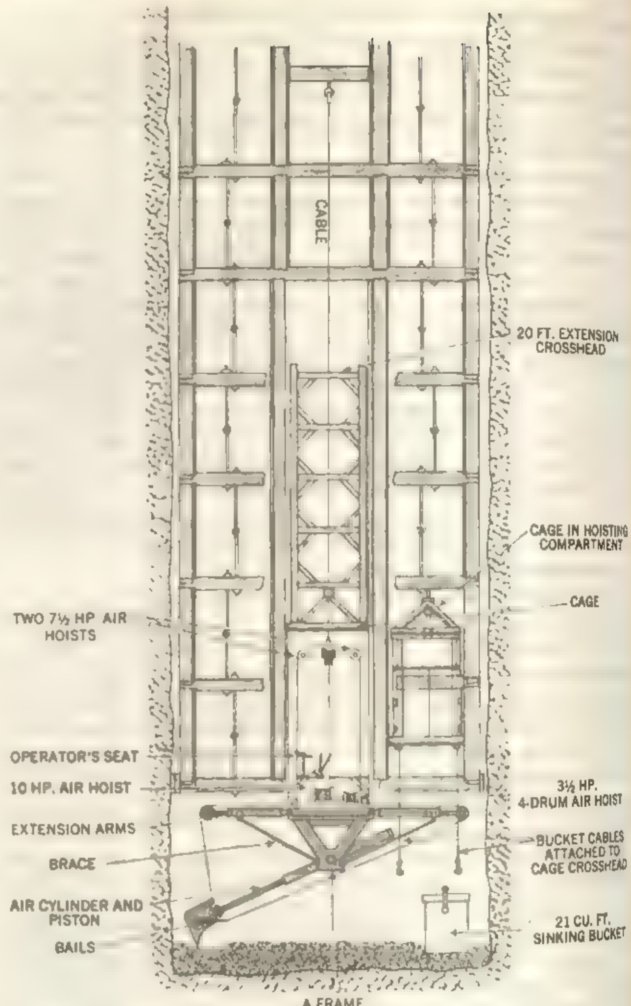
Subsurface mineral deposits are notably distinctive and often peculiarly unique so far as their mode of occurrence is concerned (*see* ORE DEPOSITS) and these characteristics frequently introduce important modifying elements in the nature and type of shaft that is made for their exploitation. That is, the shaft may prove to be the limiting "bottleneck" of a given mining operation if it has not been properly designed as to shape, size, depth and inclination to meet the particular environmental conditions that affect the ore body in question. Such situations often transpire because it is through the more or less constricted shaft opening that not only must the ore be moved but, also, all supplies, men and mining



BY COURTESY OF "ENGINEERING AND MINING JOURNAL," MCGRAW-HILL BOOK CO., INC.

FIG. 1.—SCHEMATIC DRAWING OF MUCKING MACHINE

(A) As extended boom is drawn back, bucket loads itself by digging into broken rock. (B) After digging lip of bucket has closed against steel blade, machine is ready to be hoisted. (C) Detail of digging head shown in (B)



BY COURTESY OF "ENGINEERING AND MINING JOURNAL," MCGRAW-HILL BOOK CO., INC.

FIG. 2.—SHAFT MUCKER USING POWER SHOVEL PRINCIPLE

equipment must be transported. Shafts may range in general type from crude openings to highly specialized and well-constructed excavations, and the several ways of shaft sinking may be of various degrees of complexity and engineering refinement.

The basic cycle of the shaft-sinking operation is similar to that which obtains to most mining ventures, namely: (1) drilling and blasting (breaking); (2) loading (mucking); (3) placing of timbers or other types of wall-support materials; and (4) hoisting.

These activities are interrelated in their application and, to be followed efficiently, they require the energy of co-ordinated mining crews. The several individuals forming a crew must have the faculty of working in close and safety-conscious co-operation, not only with each other but also with supporting associates whose function is to give service to the endeavour as a whole. The physical effort that must be expended ordinarily is great and the space within which the work must be performed is constricted. It often is very wet and of rather high temperature.

At about the beginning of the 20th century, the "breaking" stage of the mining cycle was marked by the progressive perfection and adoption of rock-drilling machines (*see* DRILLING MACHINERY) and this trend, accompanied as it was by the development of better explosives, contributed appreciable progress to the practice of shaft sinking. Most drilling machines are motivated by compressed air and during operation they either are hand-held or more commonly are mounted on a frame known as a "jumbo."

A typical drill hole pattern usually is composed of numerous downward directed holes. The explosives that are employed for blasting generally are water-repellant varieties of high-strength dynamites.

The arduous and expensive procedures of hand-loading and of

"mucking off the rough" were practised for many years after the advent of air-driven drilling machines and therefore the operational cycle of shaft sinking in reality showed but little refinement of consequence until about the time of World War II. That is, from the time when man first began to burrow extensively into the earth until the late 1930s, shafts to be used for mining purposes were sunk almost wholly by hand-labour methods or by a borehole drill rig of some form. In fact, sinking by borehole drilling was not used to noteworthy degree until about 1925.

However, mechanized mucking procedures have been pronouncedly advanced and few shafts are sunk without employing some type of mechanical means of loading. The more widely used and generally accepted types of equipment would fit into the following classification: scraper with 2- or 3-drum hoist (inclined shafts); reverse-action bucket or scoop on sliding boom (inclined shafts) (fig. 1); positive-action power shovel (fig. 2); overshot loader or "rocker shovel"; free-moving clamshell or orange-peel bucket loader; positive-action clamshell unit; and positive-action "backhoe" loader.

Most loading devices discharge into a pan or directly into a car or sinking-bucket sometimes called a "can" and the broken rock thus may be moved at will from the shaft chamber. Except for a few cases, shafts are lined with wood timbering (fig. 2) or with steel or concrete or masonry bricking. The shaft linings are so placed as to serve as protective walls to inhibit the infall of rock from the sides of the excavation and, also, they tend to reduce or to control the inflow of water. In addition, they afford means whereby running guides for the cages can be affixed and aligned.

In giving consideration to the adoption of a shaft-sinking process there are a number of factors that require analysis. First is the question of whether the shaft is to serve chiefly as a means of exploration and prospecting or is to be a large excavation dug so that a major ore body can be mined. Other factors which may bear consequential influence are: (1) availability and degree of skill of man power; (2) location, such as accessibility to sources of power and other supplies; (3) character of surface-plant site, especially with regard to the disposal of muck and the installation of the hoisting engine; and (4) prevailing geologic features and other conditions of the rock environment. This latter factor is particularly pertinent with reference to the shape and general attitude of the body to be exploited. Also, it is related to the existence of possible structures and other ground conditions which, by their influence on rock stability, may tend to induce problems bearing directly on the actual sinking operation.

BY COURTESY OF INTERNATIONAL NICKEL COMPANY

FIG. 3.—DRILLING BY SHAFT-SINKING CREW

There are usually other more incidental elements that must be analyzed. For example, if abnormally heavy water flows are encountered, they often will require the installing of exceptional pumping units or, at least, the building of eavelike ducts called "water rings" at intervals within the shaft. In addition, if the temperature gradient of the earth is steep, it is frequently necessary to furnish cooling and ventilating in a manner and to a degree that otherwise would not be needed.

Where the rock material to be penetrated is loose soil or gravel it may be grouted (cemented) or even frozen before sinking is begun.

In some cases, it is necessary to progress downward by lowering steel caissons or to push or drive wooden planks, called "spiles" or "forepoles," into the rock mass ahead of and beyond the actual excavation. These several practices have the general effect of inhibiting sloughing when the shaft is being made and, also, they usually serve to reduce the lateral percolation and undesirable inflow of subsurface water.

When hard rocks are being excavated, the normal or downward sinking cycle of drilling, blasting, mucking and hoisting ordinarily is followed. However, as before noted, the development of some shafts by various means of borehole drilling or, in some special cases, by "raising," which is a mining process carried upward from below, has proved to be outstandingly economic and efficient.

The operational plant unit needed to accomplish the shaft-sinking work usually is comprised of a hoisting engine and headframe, several buckets or skips, one or more sinking pumps, a satisfactory complement of drilling machines, ventilating fans, a selected loading machine and other accessory items of equipment and structure such as a change house and warehouse. Shaft-sinking costs may run to several hundred dollars a foot.

BIBLIOGRAPHY.—Robert Peele and J. A. Church (eds.), *Mining Engineers' Handbook*, 3rd ed. (1941); *Mining Engineering*, American Institute of Mining, Metallurgical and Petroleum Engineers, Inc.; *Min. Congr. J.*, American Mining Congress. (J. D. FR.)

SHAHABAD, a district in Bihar, India, lies in the angle formed by the Ganges and the Son, and is divided from Uttar Pradesh by the Karamnasa. Area 4,408 sq.mi. (11,416 sq.km.). Pop. (1961) 3,218,017. To the north is an alluvial plain, closely cultivated and thickly populated, constituting about three-quarters of the district. The plain is watered by the Son Canals which irrigate more than 400,000 ac. (161,880 ha.); a further 250,000 ac. is irrigated from other sources. The southern part of Shahabad is occupied by the 1,000-ft.-high (305 m.), undulating, and forested Kaimur Plateau of sandstones underlain by limestone beds along the Son. The chief crops of the district are rice, gram, wheat, oilseeds, barley, and sugarcane. The Grand Trunk Road, the main line of the Eastern Railway, and the Grand Chord line pass through Shahabad, and there are two narrow-gauge railways—one linking Arrah and Sasaram (*qq.v.*) and the other Dehri and Rohtasgarh. Shahabad is also known as Bhojpur, which has given its name to a dialect of the Bihari language. The headquarters of the district is at Arrah. Because of its good communications and its proximity to the limestone beds and to the coal of the Damodar Valley Dehri (pop. [1961] 38,092) has emerged as a modern industrial town, manufacturing cement, sugar, paper, and chemicals. (E. AH.)

SHAH ALAM II (ALI GAUHAR; 1728–1806), nominal Mughal emperor of India from 1759 to 1806, was born on June 15, 1728, the son of the emperor Alamgir II. In 1758 Alamgir's vizier, Imad ul-Mulk (*see* ALAMGIR II), forced him to flee from Delhi to Shuja ud-Daula, nawab of Oudh. In 1759, after his father's assassination, he proclaimed himself emperor. Attempting, with the intention of making a bid for Delhi, to demand tribute from Bihar and Bengal, he came into conflict with the East India Company. After Shuja ud-Daula's defeat at Buxar in 1764, however, he became the company's pensionary, legalizing in return its position in Bengal, Bihar, and Orissa by the grant (1765) of the *diwani* (right of revenue collection). Though comfortably installed at Allahabad, Shah Alam hankered after Delhi, and in 1771 an agreement with the Marathas enabled him to return. During 1772–82, his minister, Najaf Khan, asserted imperial authority over a broad band of Delhi territory from the Sutlej to the Chambal and from Jaipur to the Ganges. In 1788 the Rohilla chief, Ghulam Kadir, seized Delhi and, enraged at his

failure to find treasure, blinded Shah Alam.

Bereft of all save dignity, Shah Alam spent his remaining years under the protection first of the Maratha chief, Sindhia, and then, after the Maratha war of 1803, of the British. He died at Delhi on Nov. 10, 1806.

See T. G. P. Spear, *Twilight of the Mughuls* (1951). (P. H.)

SHAHDOL, a town and district in the Rewa division of Madhya Pradesh, India, and the headquarters of the district of the same name. The town (pop. [1961] 22,196) is located on the Murna River, about 110 mi. (177 km.) NW of Bilaspur on the South Eastern Railway.

Shahdol District (area 5,412 sq.mi. [14,017 sq.km.]; pop. [1961] 829,649) extends across the Upper Son Valley between the Kaimur and Maikala ranges. About two-fifths of the area is forested and nearly one-third arable, rice and sesamum being grown. There are pottery works and cement factories at Kotma and Nowrozabad, and cardboard and lac factories at Umaria. Sohagpur is famous for its Virateshwar temple. Till its merger with Vindhya Pradesh in 1948, Shahdol district was under the authority of the Baghel rajas of Rewa state. (S. M. A.)

SHAH JAHAN (1592–1666), Mughal emperor of India from 1628 to 1658, builder of the Taj Mahal, was born at Lahore on Jan. 5, 1592. He was the third son of Jahangir (*q.v.*), his mother being a Rajput princess. Marrying in 1612 Arjumand Banu Begum, niece of Jahangir's wife Nur Jahan, he became, as Prince Khurram, one of the influential Nur Jahan clique of the middle period of Jahangir's reign. In 1622 Shah Jahan, ambitious to win the succession, rebelled, ineffectually circumambulating the empire until reconciled to Jahangir in 1625. After Jahangir's death in 1627, the support of Asaf Khan, Nur Jahan's brother, enabled Shah Jahan to proclaim himself emperor at Agra (February 1628).

His reign was notable for successes against the Deccan states. By 1636 Ahmednagar had been annexed and Golconda and Bijapur forced to become tributaries. There was also a temporary extension of Mughal power in the northwest. In 1638 the Persian governor of Kandahar, Ali Mardan Khan, surrendered that fortress to the Mughals. In 1646 Mughal forces occupied Badakhshan and Balkh. But Balkh was relinquished in 1647, and in 1649 the Persians reconquered Kandahar. (*See* further Aurangzeb.)

In September 1657 Shah Jahan fell ill, precipitating a struggle for succession among his four sons, Dara Shikuh, Murad Bakhsh, Shah Shujaʿ, and Aurangzeb. The victor, Aurangzeb, declared himself emperor in 1658 and strictly confined Shah Jahan in the Agra fort until his death there on Jan. 22, 1666.

Shah Jahan is famous for his buildings, preeminently for the Taj Mahal (*q.v.*), the tomb at Agra of Arjumand Banu, called Mumtaz Mahal (mother of Aurangzeb; d. 1631), but also for his new city at Delhi, Shahjahanabad (*see* Delhi). In religion, Shah Jahan was a more orthodox Muslim than Jahangir or Akbar but a less orthodox one than Aurangzeb.

See B. P. Saksena, *History of Shahjahan of Dihli* (1932). (P. H.)

SHAHJAHANPUR, a city in the Rohilkhand division of Uttar Pradesh, India, and the administrative headquarters of the district of the same name. Shahjahanpur lies on the left bank of the Deoha River (a tributary of the Ramganga) about 100 mi. (160 km.) NW of Lucknow, and is the site of a cantonment. Pop. (1961) 110,432. It was founded in 1647 during the reign of Shah Jahan by Nawab Bahadur Khan, a Pathan. His mosque is the only building of antiquarian interest. The city is a junction on the Northern Railway and is connected by metaled roads with Lucknow, Nainital, and Pilibhit. The main industries are the manufacture of sugar and oil, and carpet weaving.

Shahjahanpur District forms the southeastern part of Rohilkhand and consists of a long, narrow tract running up from the Ganges toward the Himalayas, and is for the most part level. Area 1,762 sq.mi. (4,564 sq.km.); pop. (1961) 1,130,256. The principal rivers are the Gomati (Gumti), Khanaut, Garai, and Ramganga. The main crops are sugarcane, rice, wheat, and barley. At Rosa, 4 mi. from Shahjahanpur city, there is a large sugar factory and at Lodhipur an important sugarcane research institute. (B. Si.)

SHAHRUKH (Shahrukh Mirza) (1377–1447), the fourth son of the Asian conqueror Timur (*q.v.*) and the first of the Timurid dynasty, was born at Samarkand on Aug. 20, 1377. During Timur's lifetime, Shahrukh played a conspicuous and gallant part in the empire's affairs and wars. As principal Timurid ruler in Iran from 1405 to 1447, he succeeded in restoring some order there, as he did in Transoxiana after the disturbances following his father's death. With the exception of Syria, he ruled at times an empire coterminous with that which Timur had founded. Azerbaijan and Baghdad, however, were subject to inroads from the Black Sheep Turkmens, who did not submit to Shahrukh till after the death of their leader, Kara Yusuf, in 1420. Shahrukh was a reluctant, though successful and brave, warrior; his glory lies in his patronage of the arts and of science. The library and the outburst of miniature painting at Herat, his chief residence, point to this. The tradition was continued, notably by his short-reigned, only surviving son, Ulugh Beg (1447–49). Shahrukh's wars were generally undertaken to settle quarrels between princes of his own house. But settlements thus imposed were temporary and the process which destroyed Timurid power in Iran was not halted. Shahrukh manifested the culture and the quasi-mystical elements latent in his father, who had insatiably devoted his energy to world conquest. Shahrukh maintained relations with China, a shadowy contact which influenced the development of miniature painting. His nominal suzerainty over India heralded his family's eventual revival of power on a stage other than Iran and Transoxiana (*see* Babur). Shahrukh died at Fishawand, Ray province, Iran, on March 12, 1447.

See also Persian History. (P. W. A.)

SHAIVISM: *see* Shivaism.

SHAJAPUR, a town in the Bhopal division of Madhya Pradesh, India, and the administrative headquarters of the district of the same name, lies on the Agra-Bombay road about 50 mi. (80 km.) N of Indore. Pop. (1961) 17,317. The town was founded in the 17th century by the emperor Shah Jahan, when it was called Shahjahanpur, the name being subsequently corrupted to Shajapur. A notable architectural feature is the well-preserved fort; this contains the palace of Tarabai which houses the district offices. The town's educational institutions include a degree college, a high school, and a girls' school.

Shajapur District is situated on the Malwa plateau with its characteristic rolling topography and black cotton soil. Area 2,388 sq.mi. (6,185 sq.km.); pop. (1961) 526,135. The people, who speak Malvi, a dialect of Hindi, are mainly engaged in agriculture, and the chief crops are millet, wheat, cotton, oilseeds, tobacco, and sugar. The district has some cotton ginning and pressing mills, and also oil mills. Other important towns are Agar (11,486) and Shujalpur (13,690), near which is a cenotaph commemorating Ranoji Sindhia, the founder of the state of Gwalior. Fine Jain temples and images are to be found at various places in the district. (D. G. Na.)

SHAKER HEIGHTS, a city of Ohio, U.S., is an attractive residential suburb of Cleveland (*q.v.*). Planned and developed by the Van Sweringen brothers after 1900 in an area once held by the North Union Shaker community, Shaker Heights was incorporated as a village in 1912 and as a city in 1931. After 1920 it was accessible to downtown Cleveland by rapid transit railroad, now city owned. Population increase, after the initial rush of the 1920s, was due to large apartment developments. In the 1960s little vacant land remained and building codes were rigid. There are no industries in the community. Shaker Heights has a top-ranking public-school system, excellent private schools, a public library and a Shaker historical museum. Shaker lakes, once the site of the Shaker community grist and woolen mills, furnish a distinctive setting for the suburb. For comparative population figures see table in Ohio: *Population.* (V. B. S.)

SHAKERS, a celibate communal society that originated in France in the late 17th century, developed in England, and reached its full flower in the United States in the late 18th and early 19th centuries. The Shakers are remembered for their arts and crafts, especially their beautiful furniture, and for the dance that was a part of their worship. *See* Utopia: *Shakers*.

SHAKESPEARE, WILLIAM (1564–1616), English player, playwright, and poet. Ben Jonson, his friend and fellow-dramatist, described Shakespeare as "not of an age, but for all time," and John Dryden, the greatest poet and dramatist of the age after Shakespeare's, spoke for succeeding ages in saying that he "had an Universal mind." Thus, though for a full understanding of Shakespeare, as man, poet, and playwright, it should be remembered that he was a middle-class Elizabethan Englishman, it is his outstanding quality of universality that has made him generally acknowledged as not only England's, but the world's, greatest poet and playwright.

This article is divided into the following main sections:

I. LIFE

1. Birth.—The first of the scanty records of Shakespeare's life is the entry of his baptism in Holy Trinity Church, Stratford-upon-Avon (*q.v.*), in Warwickshire, on April 26, 1564.

2. Ancestry and Origins.—His father, John Shakespeare, was a burgess of the recently constituted corporation of Stratford (it received its charter in 1553), and had already filled certain minor municipal offices. In 1561 and 1562 he was chosen as one of the two chamberlains to whom the finance of the town was entrusted. John Aubrey (*q.v.*), in his life of Shakespeare, written in 1681 and based partly on information given him by the son of the actor Christopher Beeston (*q.v.*), describes him as a butcher, and Nicholas Rowe (*q.v.*), in his *Life of Shakespeare* (1709), as a wool dealer, but it is clear from formal documents that he was a glover, or a "whittawer" (a craftsman who cured the skins used by the glover), although he appears to have dealt from time to time in various kinds of agricultural produce, such as barley, timber, and wool. He is also described as a yeoman, and it is possible that he combined farming with trade. He was living in Stratford as early as April 1552, when he was fined for having a "dunghill" (*i.e.*, rubbish dump) outside his house and shop in Henley Street, but he does not appear to have been a native of the town. He may reasonably be identified with a John Shakespeare of Snitterfield, who administered the estate of his father, Richard Shakespeare, in 1561. Snitterfield is a village near Stratford, and there Richard Shakespeare had been settled as a farmer since 1529.

It is probable that John Shakespeare carried on the farm for some time after his father's death, and possible that by 1570 he had acquired a smallholding called Ingon Meadow in Hampton Lucy. But the farm at Snitterfield seems to have passed subsequently to his brother Henry, who was buried there in 1596. Efforts have been made to trace Shakespeare's genealogy back beyond Richard Shakespeare of Snitterfield, but without success. Certain drafts of heraldic grants of the Shakespeare arms (granted in 1596, reconfirmed and extended in 1599; *see* below) speak in one case of John Shakespeare's grandfather, in another of his great-grandfather, as having been rewarded with lands and tenements in Warwickshire for service to Henry VII. No such grants have been traced; statements about "antiquity and service" in heraldic preambles were, even in the 16th century, looked upon with suspicion.

The name Shakespeare is widespread, and is spelled in an astonishing variety of ways. Paleographers hold that Shakespeare himself, in the six extant examples of his signature, generally wrote "Shakspere" in full or in an abbreviated form, but possibly, in the main signature to his will, "Shakspeare." In the printed signatures to the dedications of his poems, on the title pages of nearly all the contemporary editions of his plays that bear his name, and in many formal documents, it appears as "Shakespeare" or "Shake-speare." This may be in part due to the martial derivation which the poet's literary contemporaries were fond of assigning to his name (as, for example, Ben Jonson's pun on "shake a Lance" in his verses for the First Folio of 1623), and which is acknowledged in the arms he bore (*see* below). Certain forms often used at Stratford, however, such as Shaxpere and Schackspere, suggest a short pronunciation of the first syllable, and thus tend to support Henry Bradley's derivation (in the *Academy*, Feb. 5, 1887) from the Anglo-Saxon personal name, Seaxberht.

On his mother's side Shakespeare was connected with a family of some distinction. At least part of Richard Shakespeare's land at Snitterfield was held from Robert Arden of Wilmcote in the nearby parish of Aston Cantlow, probably a younger son of the Ardens of Park Hall, who were among the leading gentry of Warwickshire. In 1548 Robert Arden married as his second wife Agnes Hill, formerly Webbe; he was already the father of eight daughters by his first wife. To the youngest, Mary, when he died in 1556 he left a freehold farm called Asbies, at Aston Cantlow. He had possibly already settled other property in Wilmcote upon her. Mary, at some date later than November 1556, and probably before the end of 1557, married John Shakespeare, who had, in October 1556, bought two freehold houses in Stratford, one in Greenhill Street, the other in Henley Street. The latter, known as the wool shop, was the more easterly of the two houses which in 1857 were combined to form the property known as Shakespeare's Birthplace. The western house, locally regarded as the Birthplace proper, may already have been in John Shakespeare's hands in 1556, as he seems to have been living in Henley Street in 1552, the date of the prosecution referred to above.

William was not the first child. A Joan was baptized in 1558 and a Margaret in 1562. Margaret was buried in 1563, and Joan must have died young, though her burial is not recorded, as a second Joan was baptized in 1569. A Gilbert was baptized in 1566, an Anne in 1571, a Richard in 1574, and an Edmund in 1580. Anne died in 1579; Edmund (who, like his brother, became an actor) in 1607; Gilbert in 1612; Richard in 1613.

After his marriage John Shakespeare became prominent in Stratford life. In 1565 he was chosen as an alderman, and in 1568 he held the chief municipal office as bailiff. This carried with it the dignity of justice of the peace. He seems to have contemplated the assumption of arms, and usually appears in corporation documents as "Master" Shakespeare.

3. Education.—Shakespeare, therefore, is to be thought of as the son of a leading citizen of an important market town. Stratford stands in agricultural country throughout which enclosed orchards and meadows alternate with open arable fields; not far distant is the wilder wooded district of the Forest of Arden (*q.v.*). Among Stratford's medieval heritages was a free grammar school; and here it is natural to suppose that he obtained a sound education, with a working knowledge of Virgil and Ovid in the original Latin, even though to such a scholar as Ben Jonson this might seem no more than "small Latine, and lesse Greeke" (to quote

his verses for the First Folio: *see* below).

In *c.* 1577, when Shakespeare was about 13, his father's fortunes began to decline. He became irregular in his contribution to town levies, and had to give a mortgage on his wife's property at Wilmcote as security for a loan from her brother-in-law, Edmund Lambert. Money was raised to pay this off, partly by the sale of a small interest in land at Snitterfield which had come to Mary from her sisters, partly perhaps by that of the Greenhill Street house and other property in Stratford. Lambert, however, refused to surrender the mortgage, and an attempt to recover the Wilmcote property by litigation proved ineffectual. John Shakespeare's difficulties increased. He ceased to attend the meetings of the corporation, and in 1586 was removed from the list of aldermen. In this state of affairs it is not likely that Shakespeare's school life was prolonged. He may have been apprenticed to some local trade.

4. Marriage.—Whatever his circumstances, they did not deter him from marriage. Rowe records the name of Shakespeare's wife as Hathaway, and Joseph Greene succeeded in tracing her to a family of that name dwelling in Shottery, one of the hamlets of Stratford. Her monument (next to her husband's in Stratford church) gives her first name as Anne, and her age as 67 when she died (Aug. 6, 1623). Various small trains of evidence identify her with the daughter Agnes mentioned in the will of a Richard Hathaway of Shottery, who died in 1581, and who was owner of a farmhouse called Hewland, later to be known as "Anne Hathaway's Cottage." (Agnes was legally a distinct name from Anne, but custom treated them as identical.)

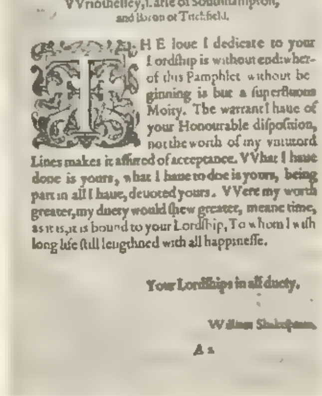
The principal record of the marriage between Anne and Shakespeare is a bond dated Nov. 28, 1582, and executed by Fulk Sandells and John Richardson, two yeomen of Stratford who also figure in Richard Hathaway's will, as a security to the bishop of Worcester for the issue of a licence for the marriage of William Shakespeare and "Anne Hathwey of Stratford," upon the consent of her friends, with only one asking of the banns (*i.e.*, by special licence). The explanation of this form of marriage probably lies in the fact that Anne was already pregnant, and in the nearness of Advent, within which marriages were prohibited, so that the ordinary procedure—by which the banns are called on three successive Sundays—would have entailed a delay until after Christmas. It has been suggested that some form of civil marriage had already taken place, so that a canonical marriage was required only to enable Anne to secure the legacy left her by her father "at the day of her marriage," but there is no evidence of this. On Nov. 27, the day before that on which the bond was executed, an entry was made in the bishop's register of the issue of a licence for a marriage between "Willelmum Shaxpere" and "Annam Whateley de Temple Grafton." However, the bond, as an original document, is the better authority; it is possible that in the register there was a mistake of "Whateley" for "Hathaway." Temple Grafton may have been indicated in the licence as being the place of marriage. There are no contemporary registers for Temple Grafton, and there is no entry of the marriage in those for Stratford. There was apparently a tradition in the 19th century that the marriage had been entered in the registers for Luddington, a chapelry within the parish, but these had already been destroyed when the tradition was recorded.

The Shakespeares' first child, Susanna, was baptized on May 26, 1583; twins, Hamnet and Judith, were baptized on Feb. 2, 1585.

5. The Missing Years.—After this, nothing certain is known of Shakespeare until 1592, when he emerges as an actor and rising playwright in London. There has been much conjecture about what experience helped to fill this gap. One story, first given currency as an addition by Dr. Johnson to his reprint, in 1765, of Rowe's *Life*, says that he "fled to London from the terror of a criminal prosecution," and began his connection with the theatre by holding horses at playhouse doors. A more likely suggestion was put forward by Aubrey, who said that he had heard from an actor who had known Shakespeare, that Shakespeare, in his younger days, had been a schoolmaster in the country. This gains some support from the sort of plays with which Shakespeare began his career as a dramatist: these are strongly academic in style, based closely on Roman comedy, as might be expected of someone who had recently been teaching. The "criminal prosecution" referred to has been connected with an incident mentioned by Rowe, and also by a Richard Davies who became rector of Sapperton, Gloucestershire, in 1695. According to Rowe, Shakespeare was prosecuted for stealing deer from the estates of Sir Thomas Lucy (*q.v.*), and left Stratford to avoid further prosecution; according to Davies, he was whipped and imprisoned by Lucy. Despite discrepancies of detail, likely in a tradition originating in village gossip and handed down orally, there may be a kernel of truth in this story. It seems fairly certain that Shakespeare left Stratford soon after the birth of the twins, and that he went, either immediately or later, to London. It has been suggested—initially by Edmond Malone (*q.v.*), in his valuable life of Shakespeare, that he followed a company of traveling players to London. Later scholars have pointed out that the Earl of Leicester's Men were in Stratford in 1586–87, and that this company may have been taken over on Leicester's death in 1588 by Lord Strange (later earl of Derby), and on Derby's death in 1594 by the lord chamberlain, Henry Carey, 1st Lord Hunsdon, and then by the 2nd Lord Hunsdon, who also became lord chamberlain. In this case, the company, known successively as Strange's Men, Derby's Men, and the Lord Chamberlain's (or Hunsdon's) Men, may have been that with which Shakespeare was connected as player and playwright for the rest of his life. It is not possible, however, to prove a continuity between Leicester's company and Strange's, and, although the names of many of Strange's company in and around 1593 are on record, Shakespeare's is not among them.

6. In London.—What is clear is that by the summer of 1592 he had begun to emerge as a playwright, and had evoked the jealousy of at least one of the group of "school dramatists," or "university wits," who claimed a monopoly of the stage (*see* ENGLISH LITERATURE: *Elizabethan and Jacobean Drama*). This was Robert Greene (*q.v.*), who, in an invective on behalf of the playmakers against the playactors, forming part of *Greenes Groatsworth of Wit* (1592), speaks of "an vpstart Crow, beautified with our feathers, that with his *Tygers hart wrapt in a Players hyde*, supposes he is as well able to bombast out a blanke verse as the best of you: and beeing an absolute *Iohannes fac totum* [*i.e.*, Jack of all trades], is in his owne conceit the onely Shake-scene in a countrey." The play upon Shakespeare's name and the parody of a line from 3 *Henry VI* ("O tiger's heart wrapped in a woman's hide!") make the reference unmistakable.

The London theatres were closed, first through riots and then through plague, from June 1592 to April 1594, except for about a month at each Christmas, and the companies were dissolved or driven to the provinces. Even if Shakespeare had been connected with Strange's Men during their London seasons of 1592 and 1593, it does not seem that he traveled with them. *Venus and Adonis* was published in April 1593, and *Lucrece* in May 1594. Both poems were printed by Shakespeare's schoolfellow, Richard Field, who had gone to London from Stratford as a printer's apprentice in 1579. Each has a dedication to Henry Wriothesley, earl of Southampton (*q.v.*), a brilliant and accomplished court favourite, not yet of age. A possibly supersubtle criticism discerns an increased warmth in the tone of the second dedication, and this is supposed to argue a marked growth of intimacy. The fact of an intimacy between Southampton and Shakespeare is vouched for by the story handed down from Sir William Davenant (*q.v.*) to Rowe that Southampton gave Shakespeare £1,000 "to enable him to go through with a purchase which he heard he had a mind to." The date of this generosity is not specified, and there is no known


TO THE RIGHT
HONOVRABLE, HENRY
VVriothesley, Earle of Southhampton,
and Baron of Titchfield.

THE loue I dedicate to your Lordship is without end: wherof this Pamphlet without beginning is but a superfluous Moity. The warrant I haue of your Honourable disposition, not the worth of my vntutord Lines makes it assured of acceptance. VVhat I haue done is yours, what I haue to doe is yours, being part in all I haue, deuoted yours. VVere my worth greater, my duety would shew greater, meane time, as it is, it is bound to your Lordship, To whom I wish long life still lengthned with all happinesse.

Your Lordships in all duety.

William Shakespeare.

A 2

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

DEDICATION OF THE FIRST EDITION OF "LUCRECE," 1594

purchase by Shakespeare which can have cost anything like £1,000.

There is a surmise, supported by ingenious argument, that Shakespeare's enforced leisure enabled him to make of 1593 a *Wanderjahr*, and in particular that the traces of a visit to northern Italy may be seen in the local colouring of *Lucrece* as compared with that of *Venus and Adonis*, and in that of the group of plays which may be dated in or about 1594 and 1595 as compared with those that preceded it (see *Chronology of the Plays*, below). It must, however, be kept in mind that, while Shakespeare may perfectly well, at this time or earlier, have traveled to Italy, and possibly to Denmark and even Germany as well, there is no direct evidence of this, and that inference from internal evidence is a dangerous guide when a writer of so assimilative a temperament as Shakespeare is concerned.

From the reopening of the theatres in the summer of 1594 Shakespeare's status is in many ways clearer. He had certainly become a leading member of the Lord Chamberlain's company by December, when his name appears for the first and only time in the accounts of the treasurer of the chamber as one of the recipients of payment for their performances at court; there is every reason to suppose that he continued to act with and write for the same associates to the close of his career.

The history of the company may be briefly told. At the death of the lord chamberlain on July 22, 1596, it passed under the protection of George, 2nd Lord Hunsdon, and again became the Lord Chamberlain's Men when he was appointed to that office on March 17, 1597. On his accession James I took this company under his patronage, and during the rest of Shakespeare's connection with the stage they were the King's Men. The records of performances at court show that they were by far the most favoured of the theatrical companies, their nearest rivals being the company known during the reign of Elizabeth I as the Admiral's Men, and then as Prince Henry's. From the summer of 1594 to March 1603 they seem to have played almost continuously in London, although they undertook a provincial tour during autumn 1597, when the London theatres were closed because of the interference of some of the players in politics. They traveled again during 1603 when the plague was in London, and during parts of the summers or autumns of most years thereafter. In 1594 they were playing at a theatre at Newington Butts (then a village not far south of London Bridge), and probably afterward at the Cross Keys Inn in the City. It is natural to suppose that in later years they used the Theatre in Shoreditch (the first public playhouse in London: *see* THEATRES [STRUCTURES]), since this was the property of the father of their principal actor, Richard Burbage (*q.v.*). The Theatre was pulled down in 1598; after a short interval, during which the company may have played at the Curtain (also in Shoreditch), Richard Burbage and his brother Cuthbert rehoused it in the Globe on Bankside, Southwark, built partly out of the materials of the Theatre. There the profits were divided between the members of the company as such and the owners of the building ("housekeepers"); shares in the "house" were held by Shakespeare and some of his leading "fellows." About 1608 another playhouse became available for the company in the "private" (or winter) house of the Blackfriars. This was also the property of the Burbages. A somewhat similar arrangement was made with profits.

Shakespeare is reported by Aubrey to have been a good actor, although, according to one account, he never rose above such a comparatively minor, though impressive, part as that of the Ghost in *Hamlet*, or Adam in *As You Like It*. As a dramatist, however, he was the mainstay of the company for at least 15 years, during which Jonson, Dekker, Beaumont and Fletcher, and Tourneur also wrote for them. On an average he must have written about two plays a year for them, although he seems to have been most productive during the opening years of the period.

7. Prosperity and Fame.—In a worldly sense he flourished, and *c.* 1596, if not earlier, he was able to return to Stratford as a man of substance. There is no evidence to show whether he had visited the town in the interval, or whether he had taken his wife and family to London. His son Hamnet died in 1596. His father's affairs had remained unprosperous. He had incurred debts, partly through becoming surety for his brother Henry's debts, and in 1592 his name was included in a list of recusants at or near Stratford, accused of "not comminge monethlie to the churche," with a note by the commissioners that in his case the cause was believed to be fear of process for debt. William Shakespeare's money must have made a considerable difference. The prosecutions for debt ceased, and in 1597 a new action was brought in chancery for recovery from the Lamberts of the mortgaged Wilmcote property, but apparently without result. Another step was taken to secure the dignity of the family by an application in 1596 to the heralds for the confirmation of a coat of arms said to have been granted to John Shakespeare while he was bailiff. The shield was described in the grant of arms of Oct. 20, 1596, as "Gold upon a bend sables a spear of the first (*i.e.*, gold) steeled argent. And for his crest . . . a falcon, his wings displayed argent, standing on a wreath of his colours; supporting a spar gold, steeled argent, set upon a helmet with mantles and tassels." The motto was *Non sanz droict* ("Not without right"). In 1599 there was a further application, for leave to impale the arms of Arden, in right of Shakespeare's mother; however, no use of the Arden arms by the Shakespeares can be traced. In 1597 Shakespeare made an important purchase for £60 of the house and gardens of New Place in Chapel Street, one of the largest houses in Stratford.

Presumably John Shakespeare ended his days in peace. A visitor to his glover's shop remembered him as "a merry Cheekd old man," always ready to crack a jest with his son. He died in September 1601, and his wife in September 1608, and the Henley Street houses passed to William. Aubrey records that William paid annual visits to Stratford, and there is evidence that he kept in touch with the life of the place. The correspondence of his neighbours, the Quineys (the only contemporary letters to mention Shakespeare), in 1598 contains an application to him for a loan to Richard Quiney on a visit to London, and a discussion of possible investments for him in the neighbourhood of Stratford. In 1602 he rented a copyhold cottage in Chapel Lane, perhaps for his gardener, and in

BY COURTESY OF THE GUILDHALL LIBRARY, LONDON

THE BANKSIDE THEATRES AS THEY APPEARED BETWEEN 1614 AND 1644

This engraving is the most accurate view of the Bankside theatres in Shakespeare's time. However, the engraver accidentally interchanged the names of the theatres, so that the circular building at centre left, labeled "Beere bayting" (bear-baiting); *i.e.*, the Beargarden Theatre, is really the Globe, and vice versa. From *Long View of London* by Wenceslaus Hollar, Amsterdam, 1647

the same year he invested £320 in an estate of 107 ac. in the open fields of Old Stratford, with 20 ac. of pasture and common rights; in 1605 he purchased a lease of certain tithes in Stratford parish, which brought in an income of about £60 a year.

He continued to live in London, near the theatres. There is evidence that in October 1596 he was living in Bishopsgate, and that he later moved to Bankside, perhaps in 1599, when the Globe Theatre on Bankside was opened by the Chamberlain's Men. There is evidence that *c.* 1604 he "lay" (*i.e.*, lodged) in the house of Christopher Mountjoy, a Huguenot tiremaker (maker of head-dresses), at the corner of Silver Street and Monkwell Street in Cripplegate. A note by Aubrey throws light not only on his abode, but on his personality. Like much of Aubrey's information about Shakespeare, it seems to have come from William Beeston, and is as follows: "The more to be admired q[uod] he was not a company keeper, lived in Shoreditch, wouldn't be debauched, & if invited to court; he was in paine." Against this evidence of the correctness of Shakespeare's morals are to be placed an anecdote of a tiring-house amour in 1602, retailed by a Middle Temple student, and a Restoration scandal which made him the father, by the hostess of the Crown Tavern at Oxford, where he stayed en route for Stratford, of Sir William Davenant, who was born in February 1606. His credit at court is implied by Jonson's references to the Sweet Swan of Avon's "flights upon the banks of Thames/That so did take Eliza and our James!"; and by stories of the origin of *The Merry Wives of Windsor* in Elizabeth I's desire to see Falstaff in love, and of a letter (now lost, but thought to have been long preserved by Davenant) written to him by James I. It was noticed by the contemporary playwright, Henry Chettle, however, that his "honied muse" dropped no "sable tear" to celebrate the death of the queen. Southampton's patronage may have introduced him to the brilliant circle around the earl of Essex, but there is no reason to suppose that he or his company was regarded as responsible for the performance of *Richard II* given at the command of some of the followers of Essex as a prelude to the disastrous rising of February 1601. The editors of the first collected edition of his plays (the First Folio, 1623), John Heminge and Henry Condell (*qq.v.*), in their preface to it, speak also of favours received by the author in his lifetime from William Herbert, earl of Pembroke, and his brother Philip Herbert, earl of Montgomery (*see* also *The Poems*, below).

He appears to have been on cordial terms with his fellow-actors and dramatists. An actor in the company, Augustine Phillips, left him a small legacy in 1605, and in his own will he paid a similar compliment to Richard Burbage, and to Heminge and Condell. His relations with Ben Jonson (*q.v.*), whom he is said by Rowe to have introduced to the world as a playwright, have been much discussed. Jests are preserved which, even if apocryphal, indicate a considerable friendship; this is not inconsistent with occasional passages of arms. One of these is referred to in the third of the Parnassus plays (*q.v.*), performed *c.* 1601, in which two actors in the same company as Shakespeare, William Kempe and Richard Burbage, are introduced as characters. Kempe, speaking to Burbage, comments on the style of the university playwrights, who "smell too much of . . . Ovid . . . Why heres our fellow [*i.e.*, fellow-actor] Shakespeare puts them all downe, ay and Ben Jonson too. O that Ben Jonson is a pestilent fellow, he brought up Horace giving the Poets a pill, but our fellow Shakespeare hath given him a purge that made him beray his credit." The reference is to Jonson's *The Poetaster* (performed *c.* 1601), in which Jonson (under the name of Horace) gives John Marston, one of his opponents in the War of the Theatres of 1599–1602, pills to cure him of his windy words. It has been suggested that Shakespeare's "purge" to Jonson may have been the description, in Act I, scene ii, of *Troilus and Cressida* (written 1601–02), of Ajax and his "humours." Jonson, on the other hand, did not spare Shakespeare in either the prologues to his plays or his private conversation. He told William Drummond of Hawthornden, when he visited him in Scotland in 1618, that "Shakesperr wanted Arte"; but the verses he contributed to the First Folio are generous enough to make all amends, and in *Timber: or, Discoveries*, a selection from his notebooks, he praises Shakespeare highly, though, as is to be expected from one who, both as man and dramatist, was as different from Shakespeare as a contemporary could be, deploring in him a facility in writing and a personal levity that he did not share (*see* also *Literary Criticism*, below). The passage is of interest as the considered contemporary judgment of a genius in his own right:

I remember, the Players have often mentioned it as an honour to *Shakespeare* that in his writing . . . hee never blotted out line. My answer hath beene, would he had blotted a thousand . . . I had not told posterity this, but for their ignorance, who choose that circumstance to command their friend by, wherein he most faulted. And to justify mine owne candor, (for I lov'd the man, and doe honour his memory (on this side Idolatry) as much as any.) Hee was (indeed) honest, and of an open, and free nature; had an excellent *Phantsie;* brave notions, and gentle expressions: wherein hee flow'd with that facility, that sometime it was necessary he should be stop'd; . . . His wit was in his owne power; would the rule of it had beene so too . . . But hee redeemed his vices, with his vertues. There was ever more in him to be praysed, than to be pardoned.

In a notebook kept by John Ward (vicar of Stratford 1662–81), there is a reference to a "merry meeting" between Shakespeare, Jonson, and Michael Drayton, at which, presumably from too hard drinking, Shakespeare "contracted" the fever of which he died; and Thomas Fuller (*q.v.*), in his *History of the Worthies of England*, published in 1662, but written considerably earlier, gives a description of the "wit-combates" between Shakespeare and Jonson which, though imaginative rather than actual, is probably based on hearsay about the relationship between them.

Of Shakespeare's literary reputation during his lifetime there is ample evidence. From the time of the publication of *Venus and Adonis* and *Lucrece*, allusions in praise of his work both as poet and dramatist, and often to himself by name, come from writers of every kind and degree. Perhaps the most interesting from the biographical point of view are contained in the *Palladis Tamia* (1598), a kind of literary handbook by Francis Meres (*q.v.*), for Meres not only writes of him as "the most excellent in both kinds [*i.e.*, comedy and tragedy] for the stage," and one of "the most passionate among us to bewaile and bemoane the perplexities of Love," but also lists the 12 plays already written. This list has served as a starting point for all modern attempts at a chronological arrangement of Shakespeare's works. It is, moreover, from Meres that we first hear of "his sugred Sonnets among his private friends." Two sonnets were printed in 1599 in a verse miscellany called *The Passionate Pilgrim*, ascribed on the title page to Shakespeare, but most of Shakespeare's sonnets remained unpublished until 1609 (see *The Poems*, below).

8. Retirement.—In *c.* 1610 Shakespeare seems to have left London, and to have settled at New Place, Stratford. There he lived the life of a retired gentleman, on friendly terms with the richest of his neighbours, the Combes (for whom there is a tradition that he wrote satirical epitaphs), and interested in any local affairs, such as a bill for the improvement of the highways in 1611 and a proposed enclosure of the open fields at Welcombe in 1614, that might affect his income or his comfort. He had a garden with a mulberry tree (if we are to believe Sir Hugh Clopton, the last of the family who owned the house from 1677 to 1751; and according to whom the mulberry tree in its garden had been planted by Shakespeare in *c.* 1609; it was cut down in 1758). Shakespeare's brothers Gilbert and Richard were both still living in Stratford. His sister Joan had married William Hart, a hatter; they had three sons, and in 1616 were living in one of Shakespeare's two houses in Henley Street. In 1607 his elder daughter, Susanna, had married John Hall (1575–1635), a physician of some reputation. They lived in Stratford and had one child, Elizabeth, born in 1608. His younger daughter, Judith, married Thomas Quiney, a vintner, a member of the family who were the Shakespeares' neighbours in Henley Street.

His retirement did not imply an absolute break with London life. In 1613 he devised an *impresa* (or emblem), to be painted by Richard Burbage, and worn by the earl of Rutland in the annual tournament held on Accession Day. In the same year he purchased for £140 a freehold house in Blackfriars—once the gate-house to the lodging of the prior of Blackfriars—and immediately mortgaged it for £60. Both the conveyance (March 10) and the

mortgage (March 11) bear Shakespeare's signature (his only other certain extant signatures being one to a deposition in the Belott-Mountjoy lawsuit, in which he was involved in 1612, and the three on his will). He was probably in London at this time (March 1613) after the festivities for the wedding of Princess Elizabeth to the Elector Palatine, when his company gave about 20 performances, of which 8 were of plays by Shakespeare. He may have stayed on for the production of *Henry VIII* at the Globe in June; there is a tradition that he instructed the actor who played the part of the king. But at the performance on June 29, 1613, the thatched roof of the theatre caught fire from the discharge of one of the guns firing the royal salute in Act I, sc. iv, and the theatre was destroyed. Although it was rebuilt in 1614 (with a tiled roof and in a more elaborate style), the fire is generally taken as being the date of the ending of Shakespeare's theatrical career. His last years were spent in Stratford, perhaps, as Rowe says, "in ease, retirement, and the conversation of friends."

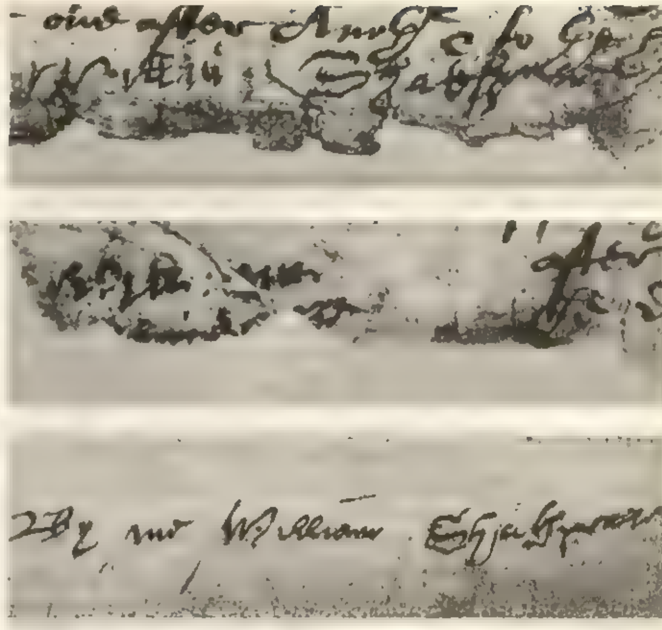
9. Will and Death.—Shakespeare made his will on March 25, 1616, apparently in some haste, as the executed deed is a draft with many erasures and interlineations. It was found in 1747 by Joseph Greene, and is preserved in Somerset House, London. There were legacies to his daughter Judith, his sister Joan, and his granddaughter Elizabeth, and remembrances to friends in Warwickshire and London; but the bulk of the property was left to his daughter Susanna under a strict entail which points to a desire to found a family. Shakespeare's wife, who had, of course, dower in most of the property, is mentioned only in an interlineation, by which the "second best bed with the furniture" was bequeathed to her. Much nonsense has been written about this, but it seems quite natural. The best bed was an important chattel, which would go with the house, in which Anne, as the widow, had the right to live with her daughter. The estate was not a large one: Aubrey's estimate of its value as £200 or £300 a year sounds reasonable. The sum total of Shakespeare's known investments amounts to £960. Sir Sidney Lee's calculation that his theatrical income must have reached £600 a year is a considerable overestimate; it can hardly have been more than about £200. Shakespeare's financial interest in the Globe and Blackfriars probably ended on or before his death.

A month after his will was signed, Shakespeare died, on April 23 (Old Style; May 3, New Style), 1616. As a tithe owner, he was buried in the chancel of the parish church. The doggerel verses on the gravestone in the chancel that local tradition has assigned to his own pen are as follows:

> Good frend for Iesvs sake forbeare,
> To digg the dvst encloased heare!
> Blest be ye man yt spares thes stones,
> And curst be he yt moves my bones.

A more elaborate monument with a bust was set up on the chancel wall (see *Portraits of Shakespeare,* below). Anne Shakespeare followed her husband on Aug. 6, 1623, and was buried next to him. The family was never founded. Shakespeare's granddaughter, Elizabeth, who inherited the entailed property, made two childless marriages, the first in 1626 with Thomas Nash of Stratford, who died in 1647, the second in 1649 with John Bernard of Abington Manor, Northamptonshire, who was knighted in 1661, and died in 1674. Judith Quiney had three sons, all of whom had died unmarried by 1639. There were, therefore, no direct descendants of Shakespeare living after Lady Bernard's death in 1670. Those of his sister, Joan Hart, can, however, still be traced: the 11th descendant of Shakespeare's sister, H. G. Shakespeare Hart (1898–), had a daughter in 1940.

On Lady Bernard's death the Henley Street houses passed to the Harts, in whose family they remained until 1806, when they were sold. In 1847 they were bought by the Birthplace Committees of Stratford and London; in 1857 the Birthplace was restored, and in 1891 it was vested, with New Place (bought by public subscription in 1862), in the Trustees and Guardians of Shakespeare's Birthplace. In 1892 the Trustees bought Anne Hathaway's Cottage; in 1930, Mary Arden's house at Wilmcote; and in 1949, Hall's Croft, the home of Shakespeare's son-in-law. Lady Bernard had disposed of the Blackfriars house. The rest of the property



BY COURTESY OF THE PUBLIC RECORDS OFFICE

ENLARGEMENTS OF THE THREE SIGNATURES FROM SHAKESPEARE'S WILL, DATED MARCH 25, 1616

was sold after her husband's death under the terms of her will, and New Place passed, first to the Cloptons (who had originally built it in *c.* 1490), who rebuilt it, and then to the Rev. Francis Gastrell, who pulled it down in 1759. The site was laid out as a public garden in 1872.

The Shakespeare Memorial Theatre, on the bank of the River Avon, was opened in 1879; a season of performances of Shakespeare's plays is given there every year. The original theatre was destroyed by fire in 1926; the new theatre was opened on April 23, 1932. The money (half of it from the United States) had been raised by public subscription.

10. Appearance, Personality, and Opinions.—Numerous attempts have been made to deduce from the plays, and from the information available, some impression of Shakespeare's personal appearance, character and personality, and opinions. None of the portraits is sufficiently well-authenticated to provide a reliable guide to his appearance (see *Portraits of Shakespeare,* below). Aubrey asserts that he was "a handsome well shap't man: very good companie and of a very readie and pleasant smooth witt," and this is probably based on hearsay, and is sufficiently moderate and particular to be acceptable as truth, not conventional praise. The lameness attributed to him by some writers has its origin in a literal interpretation of references to spiritual disabilities in the *Sonnets* (37 and 89).

There has been much conjecture about Shakespeare's religious views. The Richard Davies, from whom the least likely of the stories of Shakespeare's deer stealing derives (*see* above), is the only early authority for the idea that Shakespeare "dyed a papist," and later biographers have not generally accepted this, although some critics have claimed to find indirect evidence of a change of religious allegiance in the late plays, especially *The Tempest.* The suggestion has been made that Shakespeare's father was, or became, a Catholic. This is based partly on the alleged discovery in the mid-18th century of a "spiritual testament," or will, professing the Catholic faith, and said to have been written by John Shakespeare. Malone published it in his 1790 edition of Shakespeare's works, but later declared that he had evidence (not stated) that it could not have been connected with the Shakespeare family. It has also been suggested that the will was a forgery. The only other evidence of Shakespeare's own religious convictions—apart from that to be derived from the plays—is a record in the accounts of the Stratford corporation that in 1614 "twenty pence" was paid for sack and claret for the entertainment of a visiting preacher at New Place.

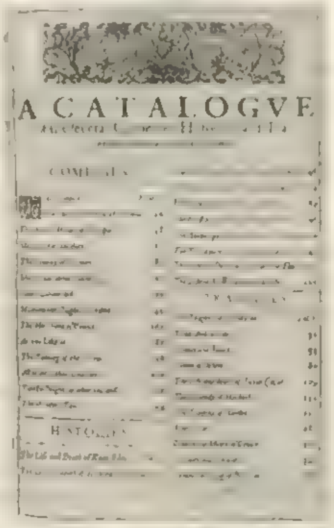
II. THE PLAYS

A. The First Folio and the Quartos

The first collected edition of Shakespeare's plays was printed at the press of William Jaggard (*q.v.*), and issued by a group of booksellers in 1623. This is known as the First Folio (F1): its title page title is *Mr. William Shakespeares Comedies, Histories, & Tragedies.* It has dedications to the earls of Pembroke and Montgomery, and to "the great Variety of Readers," both dedications signed by two of Shakespeare's "fellows" at the Globe, John Heminge and Henry Condell; and commendatory verses by Ben Jonson, Hugh Holland, Leonard Digges, and an unidentified "I. M." The Droeshout engraving (see *Portraits of Shakespeare*, below) forms part of the title page. Except for *Pericles* (*see* below), all the plays ordinarily printed in modern editions of Shakespeare's works are included in the First Folio. Of these 18 were there published for the first time. The other 18 had already appeared in one or more separate editions, known as the Quartos. The following list gives the date of the First Quarto (Q1) of each such play, and also that of any later Quarto which differs materially from it.

Titus Andronicus (1594)
2 *Henry VI* (1594)
3 *Henry VI* (1595)
Richard II (1597, 1608)
Richard III (1597)
Romeo and Juliet (1597, 1599)
Love's Labour's Lost (1598)
1 *Henry IV* (1598)
2 *Henry IV* (1600)
Henry V (1600)
A Midsummer Night's Dream (1600)
The Merchant of Venice (1600)
Much Ado About Nothing (1600)
The Merry Wives of Windsor (1602)
Hamlet (1603, 1604)
King Lear (1608)
Troilus and Cressida (1609)
Othello (1622)

Entries in the register of copyrights kept by the Company of Stationers (the Stationers' Register) indicate that editions of *As You Like It* and *Antony and Cleopatra* were contemplated, but not published, in 1600 and 1608 respectively.



BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

TABLE OF CONTENTS OF THE FIRST FOLIO, 1623

The Quartos vary in quality. Some contain texts practically identical with those of the First Folio; others show variations so material as to suggest that some alteration, generally by way of shortening for stage purposes, took place. A group of First Quartos are generally known as "bad" Quartos, being first so-called in A. W. Pollard's *Shakespeare Folios and Quartos . . . 1594–1685* (1909), in which he pointed out that the reference in Heminge and Condell's preface to the First Folio to "diverse stolne and surreptitious copies, maimed, and deformed by frauds . . . of injurious impostors" applied not to all the Quartos (as had previously been thought), but to those of which the texts are unusually corrupt —2, 3 *Henry VI*, *Romeo and Juliet*, *Henry V*, *Hamlet*, and *The Merry Wives of Windsor*. To these, later textual researches have added *King Lear* and *Richard III;* there was probably also a "bad" Quarto, now lost, of *Love's Labour's Lost.* The problems presented by "bad" Quartos differ from one play to another: all show signs of a textual corruption far greater than is explicable by bad transcription or printing. By comparison with better texts, it is possible to see that some passages have been omitted (so that "bad" Quartos tend to be shorter than First Folio texts), and that others are confused by interpolations of passages from other parts of the same play and even from other plays (including some not by Shakespeare), and by inversion of word and sentence order, by printing prose as verse and verse as prose, and by use of loose synonyms or verbal equivalents for the words intended. It is generally thought that these Quartos have a "surreptitious" origin in having been reconstructed from memory by an actor or actors, or a prompter, in the company. Earlier theories that they represent early versions of the plays have been abandoned, since it was pointed out in 1923 by W. W. Greg and R. C. Rhodes that the smaller parts are less corrupt than the others, and that this suggests "memorial reconstruction" by actors who were accustomed to playing the minor roles (*see* also *Textual Criticism*, below). Most of them must have been printed without the consent of the theatrical companies owning the plays; this is also observed in connection with plays by other authors.

B. Other Plays Attributed to Shakespeare

Shakespeare's name or initials were printed on the title pages of *Locrine* (1595), *Sir John Oldcastle* (1600), *Thomas Lord Cromwell* (1602), *The London Prodigal* (1605), *The Puritan* (1607), *A Yorkshire Tragedy* (1608), and *Pericles* (1609). It is not likely that (except for *Pericles*) he wrote any part of these plays, some of which were not even produced by his company. They were not included in the First Folio, or in a reprint of it, with minor corrections, in 1632 (the Second Folio; F2); but all seven were appended to the second issue (1664) of the Third Folio (F3; 1663), and to the Fourth Folio (F4; 1685). Shakespeare is named as joint author with John Fletcher (*see* below; also Beaumont and Fletcher) on the title page of *The Two Noble Kinsmen* (published 1634) and with William Rowley (*q.v.*) on that of *The Birth of Merlin* (1662); there is no reason for rejecting the former ascription, or for accepting the latter. Late entries in the Stationers' Register assign to him *Cardenio* (registered 1653; with Fletcher), *Henry I* and *Henry II* (1653; with Robert Davenport), *King Stephen*, *Duke Humphrey*, and *Iphis and Ianthe* (all registered in 1660); none of these is extant. Conjecture has attempted to trace his hand in other plays, of which *Arden of Feversham* (1592), *Edward III* (1596), *Mucedorus* (1598), and *The Merry Devil of Edmonton* (1608) are the most important; he may possibly have had a share in *Edward III.*

More interesting than any of these, however, is a play of which the manuscript, in a poor state of preservation, is now in the British Museum, London. Called, on the vellum cover of the manuscript, "The Booke of Sir Thomas Moore," this was first published as *Sir Thomas More* (by which title it is generally known) in 1844, and since 1911, when W. W. Greg published a type facsimile edition of it, it has been the subject of much scholarly research.

The greater part of the manuscript is written in a handwriting identified as that of Anthony Munday (*q.v.*), but to this interpolations of varying lengths have been added, apparently by five other hands. These have been identified as those of Henry Chettle, perhaps Thomas Heywood, Thomas Dekker, a professional scribe whose handwriting is known from another manuscript play, and Shakespeare. The section thought to be in Shakespeare's handwriting consists of three pages (147 lines of blank verse) dealing with More's pacification of the 1517 May Day riots (*see* More, Sir Thomas), and scholars have based their theory that it is by Shakespeare on comparisons of style and political philosophy as well as of orthography and paleography. (*See* also *Textual Criticism*, below.)

An attempt to publish a collected edition of the plays seems to have been made in 1619. It was apparently planned to bring out ten plays—*The Merchant of Venice* and *A Midsummer Night's Dream;* the "bad" Quarto versions of the *Henry VI* plays, the *Merry Wives*, *King Lear*, and *Henry V;* two of the spuriously attributed plays; and *Pericles.* There is a possibility that extant undated Quartos of *Romeo and Juliet* and *Hamlet* were connected with the venture. Official intervention hampered the enterprise, and the first ten plays came, separately, surreptitiously, and, some of them, with false dates and imprints, from the press of William Jaggard; the other two were printed without dates by William Stansby for the owner of the copyright, John Smethwick, later a First Folio shareholder.

C. Chronology of the Plays

The First Folio, unfortunately, does not give the dates at which the plays in it were written or produced; the endeavour to supply

this deficiency has been one of the main preoccupations of Shakespearean scholarship, since Edmond Malone's "An Attempt to Ascertain the Order in Which the Plays of Shakespeare Were Written" (1778). The investigation is not a mere piece of barren antiquarianism, for on it depends the possibility of appreciating the work of the world's greatest dramatic poet, not as if it were an articulated whole like a philosophical system, but in its true aspect as the reflection of a vital and constantly developing personality. A starting point is afforded by the dates of the Quartos and the entries in the Stationers' Register which refer to them, and by the list of plays inserted by Francis Meres in his *Palladis Tamia* (1598). This includes *The Two Gentlemen of Verona, The Comedy of Errors, Love's Labour's Lost, A Midsummer Night's Dream, The Merchant of Venice, Richard II, Richard III, Henry IV, King John, Titus Andronicus,* and *Romeo and Juliet,* as well as a mysterious *Love's Labour's Won,* which may be a play once printed in Quarto and no longer surviving, or an alternative title for one of the other comedies. (In this case, an earlier dating would have to be assigned to whichever one was thought to be represented by this title. *Much Ado, All's Well that Ends Well,* and *Troilus and Cressida* have been suggested.) There is a mass of supplementary evidence, drawn partly from notices in other writings, or in diaries, letters, account books, and similar records; partly from allusions to contemporary persons and events in the plays themselves; partly from parallels of thought and expression between each play and those nearest to it in time; and partly from considerations of style, including the so-called "metrical tests" (or "verse tests") which depend upon a statistical analysis of the metrical structure of Shakespeare's verse, and the development of his feeling for rhythm. The result of any such investigation is certainly not a demonstration, but, in the logical sense, a hypothesis, which serves to colligate the facts and is consistent with itself and with the known events of Shakespeare's external life.

The following list is an attempt to arrange the dates of production of the plays according to the theatrical seasons, from autumn to autumn, in which they may have fallen. It is framed on the assumption that Shakespeare ordinarily wrote two plays a year (as John Ward, vicar of Stratford 1662–81, records in his notebooks); but some slackening of production in the later years seems probable. This chronology is taken from Sir Edmund Chambers' *William Shakespeare* (1930). Neither the order in which the plays are given nor their distribution lays claim to more than approximate accuracy. Later scholarship has tended to push back the dates of the earlier plays, to date 1 *Henry VI* before 2 and 3, and to argue for numerous minor alterations. However, as Chambers' book remains the standard scholarly life of Shakespeare, it is convenient to retain here his order and chronology:

1590–91
(1, 2) 2, 3 *Henry VI*
1591–92
(3) 1 *Henry VI*
1592–93
(4) *Richard III*
(5) *Comedy of Errors*
1593–94
(6) *Titus Andronicus*
(7) *Taming of the Shrew*
1594–95
(8) *Two Gentlemen of Verona*
(9) *Love's Labour's Lost*
(10) *Romeo and Juliet*
1595–96
(11) *Richard II*
(12) *Midsummer Night's Dream*
1596–97
(13) *King John*
(14) *Merchant of Venice*
1597–98
(15, 16) 1, 2 *Henry IV*
1598–99
(17) *Much Ado About Nothing*
(18) *Henry V*
1599–1600
(19) *Julius Caesar*
(20) *As You Like It*
(21) *Twelfth Night*
1600–01
(22) *Hamlet*
(23) *Merry Wives of Windsor*
1601–02
(24) *Troilus and Cressida*
1602–03
(25) *All's Well That Ends Well*
1604–05
(26) *Measure for Measure*
(27) *Othello*
1605–06
(28) *King Lear*
(29) *Macbeth*
1606–07
(30) *Antony and Cleopatra*
1607–08
(31) *Coriolanus*
(32) *Timon of Athens*
1608–09
(33) *Pericles*
1609–10
(34) *Cymbeline*
1610–11
(35) *Winter's Tale*
1611–12
(36) *Tempest*
1612–13
(37) *Henry VIII*
(38) *Two Noble Kinsmen*

D. Composition and Sources

A more detailed account of the individual plays may now be attempted. The figures prefixed correspond to those given in the table above.

1, 2. The First Quarto of the play called in the First Folio *The second part of Henry the Sixt . . .* is entitled *The First part of the Contention betweene the two Famous Houses, Lancaster and Yorke . . .*, and the Third Quarto (1619), consists of *The First part of the Contention . . .*, together with *The true tragedie of Richard Duke of York* (a play first published in 1595); the Third Quarto title is *The whole Contention betweene . . . Lancaster and Yorke.* The relation of *The Contention* with 2, 3 *Henry VI,* and the extent of Shakespeare's responsibility for either or both, have long been subjects of controversy. The opposite extremes of critical opinion are to be found in two theories, one regarding Shakespeare as the sole author of 2, 3 *Henry VI,* and *The Contention* as a shortened and "surreptitious version" of the original plays, the other regarding *The Contention* as written in collaboration by Christopher Marlowe, Robert Greene, and, possibly, George Peele, and 2, 3 *Henry VI* as a revision of *The Contention,* written, also in collaboration, by Marlowe and Shakespeare. A comparison of the two texts leaves it hardly possible to doubt that the differences between them are to be explained by reporting rather than revision; scholarly opinion has tended more and more to regard the plays as coming from the unaided pen of Shakespeare. Greene's parody, in the "Shake-scene" passage of his *Groatsworth of Wit* (*see* above), of a line which occurs both in *The Contention* and in 3 *Henry VI,* while it clearly suggests Shakespeare's connection with the plays, is evidence neither for nor against the participation of other men, and no sufficient criterion exists for distinguishing stylistically between Shakespeare's earliest writing and that of possible collaborators. But the blank verse of 2, 3 *Henry VI* may quite well be an earlier stage of that found more fully developed in *Richard III,* and it is difficult to assign to anyone other than Shakespeare the humour of the Jack Cade scenes. *Henry VI* is not in Meres's list, but its inclusion in the First Folio is an almost certain ground for assigning to Shakespeare some share in the work. There is no record of contemporary performance of either part.

3. It has been argued that such variety and poor quality of style are found in 1 *Henry VI* that it is difficult to regard Shakespeare as its sole author, even at the earliest stage of his development. It is unlikely that the text of the First Folio represents that of its original performance. The scene in the Temple garden, where the plucking of white and red roses foreshadows the beginning of the Wars of the Roses (Act II, sc. 4) which is the one most obviously Shakespeare's, was probably a later addition. Thomas Nashe, in his *Pierce Penilesse His supplication to the Divell* (1592), refers to a play in which Lord Talbot—the leader of the English against the French—appears on the stage, and this must have been 1 *Henry VI;* it is probably also to be identified with the "Harey the vi" recorded in the Diary of Philip Henslowe (*q.v.*) as having been acted on March 3, 1592, by Strange's Men, probably at the Rose Theatre, a playhouse on Bankside built by Henslowe in 1587–88.

For all three parts of *Henry VI,* based on incidents before and during the Wars of the Roses, the chief historical sources are the *Chronicles* of Edward Hall, Raphael Holinshed, and Robert Fabyan (*qq.v.*).

4. The *Henry VI* series leads directly as historical narrative to *Richard III,* and this, together with its style as compared with that of the plays of 1594–96, suggests the short winter season of 1592–93 as the most likely time for the production of *Richard III.* It is not included in Henslowe's list of the plays acted by Strange's Men during that season, but it may well have been produced by the only other company to appear at court during the Christmas festivities, Lord Pembroke's. The fact that Shakespeare wrote a play, or more than one play, for Strange's Men during 1592–94 does not prove that he never wrote for any other company during the same period; and indeed there is room for guesswork as to the relations between Strange's and Pembroke's Men. The latter are not known to have existed before late 1592, and many difficulties would be solved by the assumption that they originated out of a

division of Strange's, who had amalgamated with the Admiral's, and may have found their numbers too much inflated to enable them to undertake as a whole the autumn tour of that year. If this is so, Pembroke's probably took over the *Henry VI* series of plays (since the part of *The Contention* entitled *The true tragedie of Richard Duke of York* was published as having been performed by them), and completed it with *Richard III* at Christmas. It will be necessary to return to this theory in connection with the discussion of *Titus Andronicus* and *The Taming of the Shrew*. For *Richard III*, as for most of the later historical plays, the second edition (1587) of Holinshed's *Chronicle* is the chief source.

5. To the winter of 1592–93 may also be assigned with fair probability Shakespeare's first experimental comedy, *The Comedy of Errors*. The play contains a reference to the wars of succession in France that would fit any date between 1589 and 1594. The plot is taken from the *Menaechmi*, and to a smaller extent, from the *Amphitruo*, of Plautus (*q.v.*). The *Menaechmi*, translated into English by William Warner, was published in 1595, but there is no known contemporary translation of the *Amphitruo*. A performance of *The Comedy of Errors* by "a company of base and common fellows" (including Shakespeare?) is recorded in the *Gesta Grayorum* (the records of Gray's Inn) as having taken place in Gray's Inn Hall on Dec. 28, 1594.

6. Many scholars have been unwilling to see the hand of Shakespeare in *Titus Andronicus;* the double testimony of its inclusion in Meres's list and in the First Folio makes its total rejection hazardous. A stage tradition was recorded by Edward Ravenscroft, a Restoration dramatist who adapted the play in 1678 (published 1687), that Shakespeare did no more than give a few "master touches" to the work of a "private author." However, abandonment of what are known as "disintegrationist" theories (*i.e.*, theories that any material in Shakespeare's plays markedly inferior to what may be accepted as the normal high standard must be by other dramatists; see *Textual Criticism*, below) has led to greater willingness to accept the play as wholly Shakespeare's. It still remains difficult to produce a convincing and coherent hypothesis about its early stage history. The play was entered in the Stationers' Register on Feb. 6, 1594, and was published in the same year with a title page setting out that it had been acted by the companies of Lords Derby (*i.e.*, Strange, who had succeeded to his father's title on Sept. 25, 1593), Pembroke, and Sussex. It is natural to take this list as indicating the order in which the three companies were concerned with it. Henslowe records the production by Sussex' company of *Titus Andronicus* as a new play on Jan. 23, 1594, only a few days before the theatres were closed by plague, but for the purposes of Henslowe's financial arrangements with the company a rewritten play may have been classed as new. Two years earlier he had also described as new a play called *Tittus and Vespacia* (the title used in Henslowe's Diary; more generally called *Titus and Vespasian*), produced by Strange's Men on April 11, 1592, and although this title suggests a piece about the emperors Titus and Vespasian, there are grounds for supposing this play to have been an early version of *Titus Andronicus*. It is difficult to explain the company names on the title page unless there had been some version earlier than that of 1594. Pembroke's Men are known to have been ruined by August 1593, and it is to be suspected that Sussex' Men, who first appeared in London at Christmas 1593, acquired their stock of plays, and transferred these to the Chamberlain's Men when the companies were reconstituted in the summer of 1594. The Chamberlain's Men were apparently playing *Andronicus* in June. The stock of Pembroke's Men probably included, as well as *Titus and Vespasian*, *Henry VI* and *Richard III*, which also thus passed to the Chamberlain's Men.

The source of the plot is unknown; there are only slight hints for it in Byzantine chronicles. The relationship between the play and a chapbook version of the story, extant only in an 18th-century printing, remains obscure. It is possible that the chapbook represents Shakespeare's source material.

7. An anonymous play called *The Taming of a Shrew*, which can be traced back as far as 1589, was published in 1594 as having been acted by Pembroke's Men. In June 1594 it was being acted by the Chamberlain's. There is little agreement concerning the relationship between *A Shrew* and Shakespeare's *The Shrew*. There are considerable differences between the texts. One theory is that Shakespeare's *The Shrew* is his adaptation of *A Shrew;* another is that *A Shrew* is a "bad" Quarto of *The Shrew* (it shows all the signs of being a "bad" Quarto of something); supporters of the latter view are forced to date Shakespeare's *The Shrew* extremely early. The origins of the play, which is a farce rather than a comedy, are to be found ultimately in widely distributed folktales, and more immediately in the translation (as *Supposes*, 1566?) by George Gascoigne (*q.v.*) of Ludovico Ariosto's *I Suppositi*. Those who regard *The Shrew* as an adaptation of *A Shrew* suggest that it may have been Shakespeare's first task for the newly established Chamberlain's company of 1594 to refurbish the old farce.

8. No very definite evidence exists for the date of *The Two Gentlemen of Verona*, other than the mention of it by Meres in 1598. It is evidently a more rudimentary essay in the genre of romantic comedy than *The Merchant of Venice*, with which it has affinities in its Italian colouring and in the theme of the interrelations of love and friendship; and it may be roughly assigned to the winter of 1594–95. The plot is drawn from various examples of contemporary fiction, especially from the story of the shepherdess Filismena in the Spanish pastoral romance *Diana* (1559?) by Jorge Montemayor (*q.v.*). A play of *Felix and Philiomena* (also based on *Diana*) had already been given at court in 1585, and Shakespeare may have used it.

9. *Love's Labour's Lost*—the first of the plays to be published with his name on the title page—used to be regarded as the first of Shakespeare's plays, and has sometimes been placed as early as 1589. So early a date is improbable. The characters are evidently suggested by Henry of Navarre (later Henry IV of France) and his supporters, the ducs de Biron, de Longueville, and d'Aumont; the last has probably been confused with Charles, duc de Mayenne, to produce the unhistorical character "Dumain." These names would have been familiar to Shakespeare at any time after *c.* 1589, but events from earlier Navarrese history have also been drawn upon, and the channel of transmission of these to Shakespeare is unknown. The absence of the play from the lists in Henslowe's Diary makes it unlikely, though not impossible, that it preceded the formation of the Chamberlain's company; and its lyric character perhaps justifies its being grouped with the other lyric plays of 1594–96. No entry of the play is found in the Stationers' Register; it is most likely that the First Quarto of 1598 replaces a lost "bad" Quarto. The title page describes the play as "corrected and augmented," and as given at Christmas 1597. It was revived for that of 1605.

10. *Romeo and Juliet*, published in 1597 as "often (with great applause)" played by Lord Hunsdon's Men, was probably produced somewhat before *A Midsummer Night's Dream*, as its incidents seem to have suggested the Pyramus and Thisbe parody in the latter play. The text of the First Quarto is "surreptitious," and was "corrected, augmented and amended" in the Second Quarto of 1599. There had been an earlier play on the subject, but the immediate source was *The Tragicall Historye of Romeus and Juliet* (1562) by Arthur Broke (*q.v.*).

11. *Richard II* can be dated with some accuracy by a comparison of two editions of Samuel Daniel's narrative poem *The Civile Warres Between the Two Houses of Lancaster and Yorke*, both dated 1595 and therefore issued between March 25, 1595, and March 24, 1596 (N.S.); the second edition, but not the first contains some close parallels to the play. It is possible that the play was performed on Dec. 9, 1595, at Sir Robert Cecil's house. From the first three Quartos, one in 1597 and two in 1598, the scene of Richard's deposition was omitted, although it was clearly part of the original structure of the play, and its removal leaves an obvious mutilation in the text. There is reason to suppose that this was due to a popular tendency to draw seditious parallels between Richard and Elizabeth I, and it became one of the charges against the earl of Essex and his fellow conspirators in the abortive outbreak of February 1601 that they had produced a performance of a play on Richard's fate to stimulate their followers on the eve of the rising. As the actors of this play were the Lord Chamberlain's Men, it was almost certainly Shakespeare's. The deposition scene

was not printed until after Elizabeth's death, in the Third Quarto of 1608.

12. *A Midsummer Night's Dream*, with its masquelike fairy scenes and the epithalamium at its close, has all the air of having been written for some courtly wedding, and the compliment paid by Oberon to the "fair vestal throned by the west" makes it possible that it was a wedding at which the queen was present. Many occasions have been suggested. The wedding of Mary, countess of Southampton, and Sir Thomas Heneage on May 2, 1594, would fit the setting of the plot; but a widowed countess hardly answers to the "little western flower" of the allegory, and there are allusions in the play to later events and in particular to the rainy weather of 1594–95. The wedding of William Stanley, earl of Derby, brother of the Lord Strange for whose players Shakespeare had written, and Elizabeth Vere, daughter of the earl of Oxford, at Greenwich on Jan. 26, 1595, would meet the conditions, but that of Thomas Berkeley and Elizabeth Carey, granddaughter of the company's patron, Lord Hunsdon, on Feb. 19, 1596, is at least as likely. There appears to be no special source for the play beyond Chaucer's *Knight's Tale* and European fairy lore.

13. *King John* has no very clear indications of date, but stylistically 1596 seems likely, in spite of the improbability of a play on an independent subject drawn from English history's being interpolated in the middle of the Lancastrian series. It would seem that Shakespeare had before him an anonymous earlier play called *The Troublesome Raigne of John King of England*. This was published in 1591, and again, with "W. Sh." on the title page, in 1611. For copyright purposes *King John* appears to have been regarded as a revision of *The Troublesome Raigne*, and in fact the succession of incidents in the two plays is much the same. Shakespeare's dialogue, however, owes little or nothing to that of his predecessor. There is no record of performance before a revival in 1737, and, perhaps because it was regarded as a revision rather than a new play, it was not entered on the Stationers' Register. It was first published in the First Folio.

14. *The Merchant of Venice* is certainly earlier than July 22, 1598, when it was entered in the Stationers' Register. It may have been inspired by the poison plots against the queen and Don Antonio of Portugal by Roderigo Lopez, Elizabeth I's Jewish physician, who was executed in 1594. It shows considerable advance in comic and melodramatic power over earlier plays, and is assigned by most scholars to *c.* 1596. The various stories of which its plot is compounded are based upon themes common to folklore and Italian *novelle*. It has been suggested that Shakespeare had before him a lost play, *The Jew*, of which there are traces as early as 1579, and in which stories illustrating "the greedinesse of worldly chusers" and the "bloody mindes of usurers" seem to have been already combined. *The Jew of Malta* (performed *c.* 1589) by Christopher Marlowe (*q.v.*) may also have contributed something.

15, 16. The first part of *Henry IV* was published in 1598; the entry in the Stationers' Register (Feb. 25) included in the title and description a reference to "the humorous conceits of Sir John Falstaff." The second part was not published until 1600, but it is clear that both parts must have been written by 1598, because, although in Part 1 as published, Shakespeare had already changed the name of the comic character Sir John Falstaff—who in earlier versions had been called Sir John Oldcastle (*q.v.*), after the famous Lollard leader and martyr—in Part 2 there are indications that when it was written, he was still using the name Oldcastle. Thus, in the First Quarto of Part 2 (registered Aug. 23, 1600) the speech-prefix "Oldcastle" is found against Falstaff's speeches; in the epilogue there is the phrase "for Oldcastle died a martyr"; and in Act III, sc. 2, Falstaff is described as having been "page to Thomas Mowbray," which was true of Oldcastle, but not of Sir John Fastolf (*q.v.*), Falstaff's eventual prototype, who had already been represented as a coward in *Henry VI*.

The fact that Shakespeare substituted Fastolf for Oldcastle, with the consequent change of name, and the reasons for it, are referred to in an *Epistle*, written in 1625, by the scholar and librarian Richard James, who explains it as the result of complaints by Oldcastle's descendants at having their ancestor represented as a comic and cowardly buffoon. Nicholas Rowe ascribes the substitution to the queen's intervention. There is also an allusion to the wrong done to Oldcastle in the epilogue to *2 Henry VI*. Presumably Fastolf had no descendants alive in 1598 to protect his reputation.

Shakespeare's source for Oldcastle, and for other details, was probably the anonymous play called *The Famous Victories of Henry V*, registered in 1594 and extant in an edition of 1598, but written much earlier, as there is a record of performance by the Queen's Men in, or before, 1588.

17. A note in the Stationers' Register during August 1600 shows that *Much Ado About Nothing* was already in existence, although its publication was directed to be "stayed." It may plausibly be regarded as the earliest play not included in Meres's list (1598). In 1613 it was revived before James I with the title of *Benedick and Beatrice* (the names of the hero and heroine). The character of the comic constable, Dogberry, is said by Aubrey to have been taken from a constable at Grendon in Buckinghamshire, on Shakespeare's road from London to Stratford. There is no very definite literary source for the play, although some of its incidents are to be found in Ariosto's *Orlando Furioso* and Matteo Bandello's *novelle*.

18. The completion of the Lancastrian series of histories with *Henry V* can be placed in *c.* 1599, since there is an allusion in one of the choruses to the military operations in Ireland in that year of the earl of Essex. The First Quarto (which, although this play was "stayed" at the same time as *Much Ado*, was published in 1600) is a "surreptitious" text, and does not include the choruses. The authentic text was first published in the First Folio.

19. That *Julius Caesar* also belongs to 1599 is shown by an allusion to it in John Weever's *The Mirror of Martyrs: or, the Life and Death of Sir John Oldcastle*, a work written two years before publication in 1601; and by a reference to its performance at the Globe on Sept. 21, 1599, by Thomas Platter of Basel in his account of a visit to London. It was the first of Shakespeare's Roman plays, and, like those that followed, was based on Sir Thomas North's translation (published 1579) of the lives of Brutus, Caesar, and Mark Antony, by Plutarch (*q.v.*).

20. *As You Like It* was one of the plays "stayed" from publication in 1600, and cannot therefore be later than that year. Some trifling evidence suggests that it is not earlier than 1599. The plot is based on *Rosalynde* (1590), a prose romance by Thomas Lodge (*q.v.*), and this in turn was based in part on the pseudo-Chaucerian *Tale of Gamelyn* (*c.* 1350; *see* ROMANCE).

21. *Twelfth Night* may be dated *c.* 1600–01. It quotes part of a song included in Robert Jones's *First Booke of Songes & Ayres* . . . (1600); has been suggested as having been the court entertainment provided on Jan. 6, 1601, before the queen and Virginio Orsini, duke of Bracciano; and is mentioned in the diary of the barrister John Manningham as having been performed at a feast at Middle Temple Hall on Feb. 2, 1602. The principal source of the plot was the "History of Apolonius and Silla," in *Riche His Farewell to Militarie Profession* (1581), by Barnabe Rich (*q.v.*).

22. A play of *Hamlet* (known to scholars as the *Ur-Hamlet*) was performed, probably by the Chamberlain's Men, for Henslowe at Newington Butts on June 11, 1594. There are references to it as a revenge play, and it seems to have been in existence in some form as early as 1589. It was doubtless on the basis of this that Shakespeare constructed his tragedy. There is an allusion in his *Hamlet* to the rivalry between the ordinary stage and the private plays given by boy actors which points to a date after the revival of the acting of plays by the Children of Paul's, a company of boy actors from St. Paul's Choir School who performed both at court and at the cathedral, probably in 1599, and the play is mentioned in a manuscript note by Gabriel Harvey, probably written before the death of Essex in 1601. The play was entered in the Stationers' Register on July 26, 1602. The First Quarto was printed in 1603 and the Second in 1604. The differences of these texts from each other and from that of the First Folio constitute one of the more difficult Shakespearean problems. The First Quarto is certainly "surreptitious." Its title page records performances at the universities of Oxford and Cambridge and elsewhere, as well as in London.

The source of the plot is to be found in legends preserved in the *Gesta Danorum* of Saxo Grammaticus (*q.v.*) and transmitted to Shakespeare or the author of the *Ur-Hamlet* (perhaps Thomas Kyd, *q.v.*) through the version of a novella by Matteo Bandello, in vol. v of François de Belleforest's *Histoires tragiques* (1570). (*See* also HAMLET.)

23. It is reported by John Dennis, in the preface to *The Comical Gallant: or the Amours of Sir John Falstaffe* (1702), his adaptation of Shakespeare's play, that *The Merry Wives of Windsor* was written at the request of Elizabeth I, who wished to see Falstaff in love; and that it was finished by Shakespeare within a fortnight. A date at the end of 1599 or the beginning of 1600, shortly after the completion of the historical Falstaff plays, would seem to be the most natural for this enterprise, and the stylistic evidence agrees with such a date. The play was entered in the Stationers' Register on Jan. 18, 1602. The First Quarto of the same year gives a "surreptitious" text, which was replaced by that of the First Folio. The Windsor setting makes it possible that *The Merry Wives* was produced in the castle there, perhaps with the Children of Windsor Chapel in the fairy parts. The plot has analogies in incidents in Italian *novelle* and in English adaptations of them.

24. Few of the plays present as many difficulties as *Troilus and Cressida*. A play of this name, "as yt is acted by my Lord Chamberlens Men," was entered in the Stationers' Register on Feb. 7, 1603, with a note that "sufficient authority" must be got by the publisher, James Roberts, before he printed it. This can hardly be any other than Shakespeare's play, but it must have been "stayed," for the First Quarto did not appear until 1609, when, on Jan. 28, a new entry for it was made in the Register by another publisher. The text of the Quarto differs from that of the Folio, but not to a greater extent than use of two different copies of the original manuscript might explain. Two alternative title pages are found in copies of the Quarto. On the earlier is a statement that the play was printed "as it was acted by the Kings Maiesties seruants at the Globe"; from the other these words are omitted, and a preface is appended which hints that the "grand possessors" of the play had made difficulties about its publication, and describes it as "never staled with the Stage."

Attempts have been made, mainly on stylistic grounds, to find another hand than Shakespeare's in the prologue and the closing scenes, and even to assign widely different dates to different parts of what is ascribed to Shakespeare. But the evidence does not really bear out these theories, and the style of the whole can be regarded as consistent with a date in 1602. It has been thought that the description of Ajax and his "humours" in Act I, sc. 2, is Shakespeare's "purge" to Jonson in reply to the *Poetaster* (performed *c.* 1601), alluded to in the Parnassus play acted at Cambridge, probably at Christmas 1601–02 (see *Life: Prosperity and Fame*, above). It is tempting to conjecture that *Troilus and Cressida* may have been played, like *Hamlet*, by the Chamberlain's Men at Cambridge, but never taken to London, and in this sense "never staled with the Stage." The only difficulty of a date in 1601 is that a parody of a play on the story of Troilus and Cressida is introduced into an anonymous play called *Histrio-Mastix. Or, the Player Whipt*, written *c.* 1599, but not printed until 1610, in which Troilus "shakes his furious Speare": it has been suggested that this is a punning reference to Shakespeare, and that his *Troilus and Cressida* must therefore have been written before 1599. But Henslowe had produced another play with the same title by Thomas Dekker and Henry Chettle, in 1599, and possibly, therefore, no allusion to Shakespeare is intended.

The material for *Troilus and Cressida* was taken from Chaucer's *Troilus and Criseyde* (*see* also ROMANCE: *Romans d'antiquité*), William Caxton's *Recuyell of the Histories of Troye* (1475), and George Chapman's translation of Homer.

25. Stylistic arguments have led to the dating of *All's Well That Ends Well* in 1602–03 although it has been urged as in the case of *Troilus and Cressida* (and with as little justification) that parts of the play are of a considerably earlier date. The story is derived from one of the tales in Giovanni Boccaccio's *Decameron*, as translated by William Painter in his collection called *The Palace of Pleasure . . .* (1566).

26. A list of plays in the account book of the Revels Office (the office responsible for supervision and payment of special entertainments at court) records that *Measure for Measure* was performed at court on Dec. 26, 1604. It was registered on Nov. 8, 1623, and first published in the First Folio. The fact that the Folio text is not altogether satisfactory has led to theories that the play is a reworking by Shakespeare of an earlier text, and that the two versions of the play (that performed at court in 1604, and that performed later on the public stage) may have differed considerably. It is further suggested that the text finally published in the First Folio contains some passages not by Shakespeare, and that it is, in any case, incomplete.

The plot is taken from a story already used by George Whetstone (*q.v.*) who borrowed it from Giambattista Giraldi Cinthio's *Hecatommithi* (1565).

27. A performance of *Othello* at court on Nov. 1, 1604, is noted in the same list as that recording *Measure for Measure*, and the play may be reasonably assigned to the same year. The play was revived in 1610, when it was seen by the prince of Württemberg at the Globe on April 30. It was entered in the Stationers' Register on Oct. 6, 1621, and a First Quarto was published in 1622. The text of this is less satisfactory than that of the First Folio; it omits a good many lines found in the Folio and almost certainly belonging to the play as originally written. It also contains profane expressions, modified in the Folio, and this points to a date for the original production earlier than the Act to Restrain Abuses of Players of May 1606, which made it an offense to swear oaths on the stage. The plot, like that of *Measure for Measure*, comes from the *Hecatommithi*.

28. The entry of *King Lear* in the Stationers' Register on Nov. 26, 1607, records its performance at court on Dec. 26, 1606. This suggests 1605 or 1606 as the date of first production; and this is confirmed by the publication in 1605 of the anonymous *True Chronicle History of King Leir . . .* (performed 1594, 1605), which Shakespeare used as his source. Two Quartos of *King Lear* were published, in 1608 and 1619; they contain a text rather longer, but in other respects less accurate, than that of the First Folio. The main material of the play consists of a story from Celtic mythology, contained in the *Historia regum Britanniae* of Geoffrey of Monmouth (*q.v.*). It was accessible to Shakespeare in Holinshed and in Edmund Spenser's *Faerie Queene*, as well as in the earlier play.

29. In view of its allusions to topics of particular interest to James I—*e.g.*, witchcraft, in which James had long been interested; events from early Scottish history, and especially the introduction of his legendary ancestor Banquo; and references to recent events connected with the king—*Macbeth* cannot be dated before March 1603, when James VI of Scotland acceded to the English throne. The style and some minor topical allusions suggest *c.* 1605 or 1606, and a hint for the theme of the three witches may have been given by Matthew Gwinne's entertainment of the *Tres Sibyllae*, with which James was welcomed to Oxford on Aug. 27, 1605. The play was revived in 1611, when a physician and astrologer called Simon Forman tells us in his manuscript "Booke of Plaies" that he saw it at the Globe on April 20. The only extant text, that of the First Folio, bears traces of shortening and has been interpolated with additional rhymed dialogues for the witches, possibly by Thomas Middleton (*q.v.*), but the extent of Middleton's contribution has been exaggerated. For sources it is not likely that Shakespeare had consulted any Scottish history other than that included in Holinshed's *Chronicle*; he may have gathered witch lore from Reginald Scot's *Discoverie of Witchcraft . . .* (1584), or from James I's *Daemonologie* (1597).

30. It is not quite clear whether *Antony and Cleopatra* was the play of that name entered in the Stationers' Register on May 20, 1608, for no Quarto is extant, and a fresh entry was made in the Register before the issue of the First Folio. Apart from the 1608 entry there is little external evidence to fix the date of the play, but it is in Shakespeare's later, although not his last, manner and may very well belong to 1607. The second edition (1607 or ...) of Samuel Daniel's *The Tragedie of Cleopatra* shows alterations from the first (1594) that seem likely to have been inspired by Shake-

speare's play. Like *Julius Caesar, Antony and Cleopatra* is derived mainly from Plutarch's life of Mark Antony. There is no record of a 17th-century production, and the play was later eclipsed by *All For Love* (1677), a recasting of the story in the manner of neoclassical tragedy, by John Dryden (*q.v.*). Shakespeare's version was first revived in 1849.

31. For *Coriolanus* the external evidence is even scantier, and all that can be said is that its closest affinities are to *Antony and Cleopatra,* which in all probability it directly followed.

It is based on North's translation of Plutarch's life of Coriolanus. There is no record of a contemporary performance.

32. There is no external evidence as to the date of *Timon of Athens,* but it may safely be grouped by its internal characteristics with *Antony and Cleopatra* and *Coriolanus,* and there is a clear gulf between it and those that follow. It may be placed provisionally in 1607, although some critics put it next after (or before) *Lear,* which, in grandeur of poetic conception, the poetic quality of some passages, and in treating the theme of ingratitude, it closely resembles—although the dullness and lack of humour of the comic scenes are more reminiscent of *All's Well that Ends Well.* The incoherencies of its action and the inequalities of its style have prevented some scholars from accepting it as a finished production of Shakespeare, but there agreement ceases. It is sometimes, and perhaps most reasonably, regarded as an incomplete draft; sometimes as a Shakespearean fragment worked over by a second hand, either for the stage or for printing in the First Folio; sometimes, but not very plausibly, as an old play by an inferior writer partly remodeled by Shakespeare. It seems to have no direct connection with an extant academic play of *Timon,* written *c.* 1585, which remained in manuscript until 1842. The sources are partly in Plutarch's lives of Mark Antony and Alcibiades—where the story is told briefly—on which the version in Painter's *Palace of Pleasure* is based; and partly in Lucian's dialogue *Timon,* which tells the story in greater detail. Lucian was the source, direct or indirect, for the academic *Timon,* and Shakespeare's play has some points that are in Lucian but not in Plutarch, and some that appear only in the academic play. Lucian may have been the source for some unknown version of the story on which both the academic play and Shakespeare's are based.

33. Similar difficulties, equally unsolved, cling to *Pericles, Prince of Tyre.* It was entered in the Stationers' Register on May 20, 1608, and published in 1609 as "the Late and much admired Play" acted by the King's Men at the Globe. The title page bears Shakespeare's name, but the play was not included in the First Folio, and was added to Shakespeare's collected works only in the Third Folio (1663), with others which, although they also had been printed under his name or initials in Quarto, are certainly not his (*see* above). In 1608 a prose story was published, *The Painfull Adventures of Pericles Prince of Tyre . . .* This claims to be the history of the play as it was presented by the King's Men, and is described by its author, George Wilkins—a pamphleteer and novelist to whom other plays, or a share in them, have been attributed —in his "Argument" (preface) to it, as "a poore infant of my braine." The production of the play is therefore to be put in 1608 or a little earlier. It can hardly be doubted on internal evidence that Shakespeare is the author of the verse scenes in the last three acts, with the possible exception of the doggerel choruses. It is probable, although it has been doubted, that he was also the author of the prose scenes in these acts. The poverty of the first two acts has been fairly convincingly accounted for by the theory that the play is a "bad" Quarto, a memorial reconstruction by two reporters, the reporter of the last three acts having an efficiency greater than that of his colleague. It seems reasonable to allow Wilkins some connection with the play: other plays by him were produced *c.* 1607. The writing of the play cannot much have preceded publication, for the close resemblances of the style to that of Shakespeare's last plays argue for its lateness. Unless there was an earlier Shakespearean version now lost, Dryden's statement that "Shakespeare's own Muse her Pericles first bore" (*i.e.,* that *Pericles* was Shakespeare's first play) must be regarded as an error.

The story is an ancient one which exists in many versions. In all except the play, the hero's name is Apollonius of Tyre (*q.v.*). The play is directly based upon a version in John Gower's *Confessio Amantis* (1390, 1393), and Shakespeare's use of Gower as a "presenter" (or chorus) is thereby explained. But another version of the story, in Laurence Twine's *The Patterne of Painefull Adventures* (*c.* 1576, reprinted 1607), may also have been consulted.

34. *Cymbeline* shows a further development in the direction of Shakespeare's final style, and can hardly have come earlier. Simon Forman (*see* above) describes a performance of it, but without giving the date. Forman died in September 1611, and gives the dates of the other performances he describes (of *Macbeth* and *The Winter's Tale*) as 1611, but these were not necessarily first productions—and, in the case of *Macbeth,* what he saw was clearly a revival. On internal evidence, *Cymbeline* may be conjecturally assigned to 1609. It was first registered in 1623, for publication in the First Folio. The masquelike dream in Act V, sc. 4, has been thought to be an interpolation by another hand.

This play also is based upon a widespread story (*see* ROMANCE: *Other Themes*), probably known to Shakespeare in Boccaccio's *Decameron.* The historical part is, as usual, from Holinshed.

35. *The Winter's Tale* was seen by Forman on May 15, 1611, and, as it clearly belongs stylistically and thematically to the last group of plays, it may well have been first produced in that or the preceding year. A document among the Revels Accounts gives Nov. 5, 1611, as the date of a performance at court. The plot is taken from *Pandosto* (1588; reprinted 1607), a romance by Robert Greene (*q.v.*).

36. The wedding masque in Act IV of *The Tempest* has suggested the possibility that the play may have been composed to celebrate the marriage of Princess Elizabeth and Frederick V, the elector Palatine, on Feb. 14, 1613, but the Revels Accounts gives Nov. 1, 1611, as the date of a performance at court. Silvester Jourdan's *A Discovery of the Barmudas . . . ,* containing an account of the shipwreck on the Bermudas (*q.v.*) of Sir George Somers and others in 1609, was published in October 1610, and this and other contemporary narratives of Virginian colonization probably furnished the idea for the plot.

37. Critical opinion has veered away from the theory that *Henry VIII* shows the hand of an author other than Shakespeare. John Fletcher has been named as the collaborator; some have even excluded Shakespeare entirely and allotted the authorship to Fletcher (*see* BEAUMONT AND FLETCHER). The inclusion of the play in the First Folio provides strong argument for Shakespeare's having had some share in the writing. He has been considered to be responsible only for Act I, sc. 1, 2; Act II, sc. 3, 4; Act III, sc. 2, lines 1–203; and Act V, sc. 1, but by some it is thought that the Fletcherian passages merely show Shakespeare writing in the Fletcherian style. The play was probably first produced in 1613, and originally bore the alternative title, *All Is True.* It was being performed at the Globe on June 29, 1613 (perhaps the first performance), when the thatch caught fire and the theatre was burned (*see* above). The principal source was Holinshed, but Hall's *Chronicle,* John Foxe's *Book of Martyrs* (1563, enlarged 1570), and perhaps *When You See Me, You Know Me . . .* (published 1605) by Samuel Rowley (*see* ROWLEY, WILLIAM), seem also to have contributed something.

38. *The Two Noble Kinsmen.*—Of plays not generally included in modern collected editions of Shakespeare's works, the one in which he may be thought to have had most share is *The Two Noble Kinsmen,* published in a Quarto in 1634, and registered as having been performed at the Blackfriars by the King's Men, and as being written by "the memorable Worthies of their time; Mr. John Fletcher, and Mr. William Shakespeare, Gent." An analysis of the play leaves no reason to doubt this title page ascription, although some scholars have rejected it, partly because it was included in 1679 in a Folio collection of plays by Beaumont and Fletcher. The play seems to have been the result of ordinary collaboration, not of later interpolation or reworking by either. There is enough resemblance between Fletcher's and Shakespeare's styles to make it difficult to divide the play between them, but Act I, sc. 1–4; Act II, sc. 1; Act III, sc. 1, 2; and Act V, sc. 1, 3, 4, may probably be assigned to Shakespeare: the great invoca-

tions to Mars, Venus, and Diana in Act V, sc. 1, could scarcely have been written by any other poet.

The best evidence for the dating of the play is provided by the fact that the Morris dance in Act III, sc. 5 (by Fletcher), is borrowed from a similar scene in Beaumont's *Masque of the Inner Temple and Grayes Inne . . .*, given at Whitehall on Feb. 20, 1613; the play must have been written after the masque, but probably not long after, and may therefore be dated 1613. There was probably a performance of it at court *c.* 1619.

The source for the main plot of the two noble kinsmen (Palamon and Arcite) is Chaucer's *Knight's Tale;* the comic but horrifying subplot, not found in Chaucer and unlike anything else of Shakespeare's, is generally assigned to Fletcher.

III. THE POEMS

In volume, Shakespeare's nondramatic work is not considerable: it consists—apart from the songs scattered throughout the plays—of two narrative poems, slight contributions to two collections, and the sonnets.

1. The Narrative Poems.—*Venus and Adonis* was entered on the Stationers' Register on April 18, 1593, and *The Rape of Lucrece* on May 9, 1594. (The latter, when published in a Quarto in the same year, was entitled simply *Lucrece,* and is often known by this title.) Both were dedicated to the earl of Southampton (see *Life: In London,* above).

It used to be thought that Shakespeare arrived in London from Stratford with the manuscript of *Venus and Adonis* in his pocket; this is partly because, in his dedication to Southampton, he refers to it as "the first heire of my invention," but this may well mean the first of his works to be published, rather than the first to be written. It has also been noticed that the poem contains much imagery drawn from country life and rural pursuits, and this has suggested that it was written in Stratford. This type of imagery is characteristic of all Shakespeare's works, however, and particularly of the early ones. It seems likely that, like *Lucrece, Venus and Adonis* was written after Shakespeare went to London, and both poems probably owe their origin to the comparative leisure afforded to playwrights and actors by the closing of the theatres from June 1592 to April 1594, after he had already acquired some reputation as a playwright (see *Life: In London,* above).

Both poems are in long stanza forms, *Venus and Adonis* in sesta rima (a quatrain followed by a couplet), and *Lucrece* in rime royal (*q.v.*), and in both Shakespeare rivals Spenser (who had used sesta rima for his pastoral elegy *Astrophel,* published in 1586) as a master of prosody. He may have borrowed the metre for *Venus and Adonis* from Thomas Lodge's *Scillaes Metamorphosis* (1589), and it is possible that he had read Marlowe's *Hero and Leander* in manuscript. Both poems are typical in subject and style of the heroic narrative poetry of the period: the stories, taken from Ovid's *Metamorphoses* and *Fasti* respectively, occur frequently in Renaissance literature (*see* ENGLISH LITERATURE: *The Renaissance: Sonnets and Mythological Poetry*). But in *Lucrece* especially, there are flashes of the emotional power and poetic genius that were to come to full fruition in the plays. That both poems were very popular is proved by the number of editions: at least 17 are known of *Venus and Adonis* between 1593 and 1675, and 9 are extant of *Lucrece* (1594–1655).

2. Minor Poems.—In 1599 (the year of the 5th and 6th editions of *Venus and Adonis*) William Jaggard published a poetic miscellany called *The Passionate Pilgrim,* attributing it on the title page to Shakespeare. Of the 20 (or 21) short poems it contains, however, only five are certainly Shakespeare's (versions of two of the *Sonnets,* and of three sonnets from *Love's Labour's Lost*). The rest are either anonymous poems found also in other collections (where they are not attributed to Shakespeare), or poems known to be by other authors (Marlowe, Ralegh, Richard Barnfield, Bartholomew Griffin).

More interesting is the poem called *The Phoenix and the Turtle,* an elegy on an unknown pair of wedded lovers, published in 1601 in a volume by Sir Robert Chester called *Loves Martyr: Or, Rosalins Complaint,* and described on the title page as "Allegorically shadowing the truth of Love, in the constant Fate of the Phoenix and the Turtle." To Chester's poem, written to celebrate the love of his patron, Sir John Salisbury, and his wife, symbolically identified as the turtledove (Constancy), and the phoenix (Love), "are added" (to quote the title page again) "some new compositions of severall moderne writers . . . upon . . . the Phoenix and Turtle." These "Diverse Poeticall Essaies" include two attributed to Ben Jonson, one each by John Marston, Chapman, and Shakespeare, and nine of which the authors are unknown. Shakespeare's poem, untitled, but always known as *The Phoenix and the Turtle,* describes a philosophy of fully mutual love, summed up in words spoken by Reason (the poem's principal speaker): "Love hath reason, reason none," thus recognizing, rationally, a power beyond reason, and expressing, for once, what may be taken as Shakespeare's own view.

Finally, before coming to the sonnets, there are the songs from the plays. They belong to the very heart of the Elizabethan Golden Age of lyrical poetry: their variety of mood and metre, freshness of tone, beauty of imagery, and verbal felicity make them some of Shakespeare's best-loved poetry.

3. The Sonnets.—Without doubt, however, the most outstanding of Shakespeare's nondramatic work is contained in the volume registered on May 20, 1609, by the stationer Thomas Thorpe, and published in the same year with the title page inscription, *Shake-speares Sonnets. Never Before Imprinted.* In addition to 154 sonnets, the volume contains the elegiac poem in rime royal called *A Lover's Complaint.* If this is Shakespeare's, which is doubtful, it probably dates from the *Venus and Adonis* period.

In 1640 the sonnets, with poems from *The Passionate Pilgrim* and elsewhere, many of them not Shakespeare's, were republished by John Benson in *Poems: Written by Wil. Shake-speare, Gent.* Here the sonnets are arranged in a different order from that of 1609. Some are omitted, and they are altered, and given titles, and are declared by the publisher to "appeare of the same purity, the Authour himselfe then living avouched."

No other Shakespearean controversy has received so much attention as that concerning the date, character, and literary history of the sonnets; and this has been extended to a controversy about the nature of the poet's relationship with the two people to whom they are addressed—especially the "lovely boy," who is perhaps to be equated with "Mr. W. H." (*see* below). This is intelligible enough, since upon the issues raised depends the question whether or not these poems give a glimpse into the intimate depths of a personality which otherwise is at most only imperfectly revealed through the plays. On the whole, the balance of authority is in favour of regarding them as in a very considerable measure autobiographical. This view has undergone the fires of much destructive argument. The authenticity of the order in which the sonnets were printed in 1609, and even Shakespeare's authorship of some of them, have been doubted; and their subject matter has been variously explained as being in the nature of a philosophical allegory, an effort of the dramatic imagination, or a mere impersonal exercise in the forms of the Petrarchan convention. This last theory rests upon the assumption that the use of literary conventions is inconsistent with the expression of unfeigned emotions (*see* ROMANTICISM); and it is hardly to be set against the conviction which the sonnets themselves convey to the reader of the strength and sincerity of the emotional and spiritual experience out of which they were wrought. This conviction makes due allowance for the inevitable heightening of feeling in the act of poetic composition; and it certainly does not carry with it a belief that all the external events underlying the emotional development revealed in the sonnets are capable at this distance of time of inferential reconstruction. But it does accept the sonnets as reflecting a part of Shakespeare's life during the years in which they were written, and as revealing at least the outlines of a drama that played itself out, for once, not in his imagination, but in his actual conduct in the world of men and women. Exactly what this drama was—whether his relationship with the "lovely boy" was, as some maintain, openly homosexual, or merely expressed in the heightened language of an emotional immaturity—is yet one more of the Shakespearean mysteries that may never be entirely solved.

There is no advantage in rearranging the order of the 1609 volume, even if there were any basis other than that of whim on which to do so. Many of the sonnets are obviously linked to those following or preceding them; and although a few may be misplaced, the order as a whole does not jar against the reader's sense of emotional continuity—the only test that can be satisfactorily applied. The last two sonnets (153 and 154), however, are merely alternative versions of a Greek epigram on Cupid (possibly not by Shakespeare), and it would be rash to assume that the rest all have a common subject matter. On this assumption, however, they have generally been interpreted somewhat as follows. There are two series, which are more probably parallel than successive.

The shorter series (which comes after the longer in the 1609 arrangement), sonnets 127–152, appears to be the record of the poet's relations with a mistress, a dark woman with raven brows and "mourning eyes." In the earlier sonnets of this series he undertakes the half-playful defense of black beauty against the fair Elizabethan ideal; but the greater number are in a more serious vein, and are filled with a deep consciousness of the bitterness of lustful passion and of the slavery of the soul to the body. The woman is a wanton. She has broken her bed vow for Shakespeare, who is himself forsworn in loving her; and she is "twice forsworn" in proving faithless to him with other men. His reason condemns her, but his heart has not the power to throw off her tyranny. Her particular offense is that she, "a woman colour'd ill," has cast her snares not only upon him, but upon his friend, "a man right fair," who is his "better angel," and that thus his loss is double, in love and friendship. In this series occur the so-called "Will" sonnets (135–136), in which the word "will" occurs 18 times, and, in the 1609 edition, is generally capitalized and in italics. As well as being interpreted as a pun on Shakespeare's name, this has also led to a theory that the "lovely boy," Shakespeare's rival for the "dark woman's" affections, was called "Will"—a theory perhaps supported by the opening of sonnet 135

> Whoever hath her wish, thou hast thy *Will*,
> And *Will* to boot, and *Will* in over-plus;

The longer series (1–126) is written to a man, appears to extend over a considerable period of time, and covers a wide range of sentiment. The person addressed is younger than Shakespeare, and of higher rank. He is a "lovely boy," the son of a lovely mother, and has hair like the auburn buds of marjoram. This series falls into several groups, rarely separated by sharp lines of demarcation. Perhaps the first group (1–17) is the most distinct. The sonnets of this group are a prolonged exhortation by Shakespeare to his friend to marry and beget children. The friend is at the peak of his youth, and should make haste, before the rose of beauty dies, to secure himself in his descendants against "this bloody tyrant, Time."

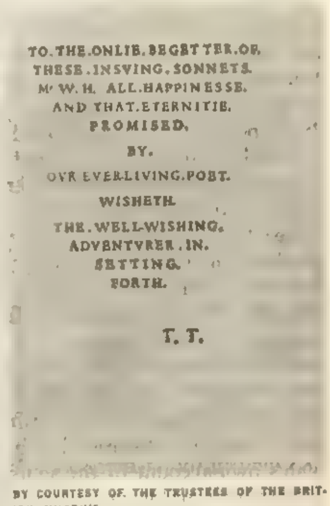
In the next group (18–25) a more personal note is struck, and the writer assumes the attitudes at once of the poet whose genius is to be devoted to making the beauty and honour of his patron eternal, and of the friend whose absorbing affection is always on the point of assuming an emotional colour indistinguishable from that of love. The consciousness of advancing years (22) and of a fate which bars the triumph of public honour and proud titles (25), alike find their consolation in this affection. A period of absence follows (26–32), during which the thought of friendship comes to remedy the sorrows of a life "in disgrace with fortune and men's eyes."

Then (33–42) comes an estrangement. The friend has committed a "sensual fault," which is at the same time a sin against friendship. He has been wooed by a woman loved by the poet, who deeply resents the treachery but in the end forgives it and bids the friend "Take all my loves, my love, yea take them all!" for all are included in the love he has already freely given him. It is difficult to escape the suggestion that this conflict between love and friendship is the same as that which inspired some of the "dark woman" sonnets.

Another journey is filled with thoughts of the friend (43–52). This is followed by three sonnets (53–55) in which the friend's beauty and the immortality already conferred on it by the poet's verse are especially dwelt upon. Once more there is a parting (56–61), and the poet awaits as patiently as may be his friend's return to him. Again (62–65) he looks to his verse to give the friend immortality. He is tired of the world, but his friend redeems it (66–68). Then rumours of some scandal against his friend (69–70) reach him, and he falls (71–74) into thoughts of coming death, bidding his friend not to mourn long for him "for I love you so/That I in your sweet thoughts would be forgot/If thinking on me then should make you woe." In the next three sonnets (75–77), he continues to write of his love for his friend: "O know, sweet love, I always write of you,/And you and love are still my argument." But in sonnets 78 to 86, a new theme is introduced—that of the rival poet who, inspired by Shakespeare's example, has addressed verses to his patron. This rival, whose identification has exercised critics and commentators (*see* below), is described by Shakespeare as "a better spirit" (80), "that able spirit" (85), "by spirits taught to write/Above a mortal pitch. . . ," and as being nightly gulled with intelligence by an "affable familiar ghost" (86). He, with "the proud full sail of his great verse" has inhearsed Shakespeare's "ripe thoughts" in his brain, has "struck me [Shakespeare] dead"; but Shakespeare describes himself and his rival as "both your poets" (83)—suggesting, as well as jealousy, that the rivalry is in some way between equals, even friends. There is another estrangement (87–90), and the poet, already crossed with "the spite of fortune" is ready not only to acquiesce in the loss of friendship, but to find the fault in himself. The friend returns to him, but the relation is still clouded by doubts of his fidelity (91–93) and by public rumours of his "wantonness" (94–96). For a third time the poet is absent from his friend (97–99). Then comes an apparent interval, after which a love already three years old is renewed (100–104) with even richer praises (105–108). It is now the poet's turn to offer apologies (109–112) for offenses against friendship, and for some "brand" upon his name apparently due to the conditions of his profession of player. He is again absent (113), and again renews his protestations of the imperishability of love (114–116), and of his own unworthiness (117–121), for which his only excuse is in the fact that the friend was once unkind. If the friend has suffered as Shakespeare suffered, he has "pass'd a hell of time." The series closes with a group (122–125) in which love is pitted against time; and an *envoi* (126) warns the "lovely boy" that in the end Nature must render up her treasure.

4. The Mystery of "Mr. W. H."—Such an analysis of the sonnets' content can give no adequate idea of their quality, whereby the appeal of universal poetry is built up on a basis of intimate self-revelation. The human document is so legible, and at the same time so incomplete, that it is easy to understand the efforts made to throw further light upon it by tracing the identities of the "lovely boy" and the "dark woman" through his relationships with whom the poet was brought to so fiery an ordeal of soul, and even to the borders of self-abasement.

The search has generally started from the terms of a somewhat mysterious dedication prefixed by Thomas Thorpe to the first edition of 1609 (*see* illustration). The most usual interpretation of this is that the "begetter"—or inspirer—of the sonnets bore the initials W. H.; and history has accordingly been ransacked to find a W. H. whose age and circumstances might conceivably fit the conditions of the problem which the sonnets present. It is perhaps a want of historical perspective that has led to the centring of controversy on two names belonging to the highest ranks of the Elizabethan nobility, those of Henry Wriothesley, 3rd earl of Southampton, and William Herbert, 3rd earl of Pembroke.


TO.THE.ONLIE.BEGETTER.OF.
THESE.INSVING.SONNETS.
Mr.W.H. ALL.HAPPINESSE.
AND.THAT.ETERNITIE.
PROMISED.
BY.
OVR.EVER-LIVING.POET.
WISHETH.
THE.WELL-WISHING.
ADVENTVRER.IN.
SETTING.
FORTH.

T. T.

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

DEDICATION BY THE PUBLISHER THOMAS THORPE OF THE FIRST EDITION OF THE "SONNETS," 1609

There is some evidence to connect Shakespeare with both. To Southampton he dedicated *Venus and Adonis* and *Lucrece*, and the story that he received a gift of no less than £1,000 from him is recorded by Rowe (see *Life*, above). His acquaintance with Pembroke can only be inferred from the statement of Heminge and Condell, in their dedication to them of the 1623 First Folio, that Pembroke and his brother, the earl of Montgomery, had "prosequuted both them and their Authour living, with so much favour." The personal beauty of the rival claimants and of their mothers, their amours, the attempts of their families to persuade them to marry, their relations to poets and actors, and all other points in their biographies which do or do not fit in with the indications of the sonnets have been canvassed with spirit and erudition —but with no very conclusive result. It is in Pembroke's favour that his initials were actually W. H., whereas Southampton's can be turned into W. H. only by a process of metathesis (*i.e.*, transposing the initials of "Henry Wriothesley"); and his champions have been more successful than Southampton's in producing a woman, a certain Mary Fitton (*q.v.*), who was a mistress of Pembroke's and was in consequence dismissed in disgrace from her post of maid of honour to Elizabeth I. But careful investigation of the sonnets, as regards their style and their relation to the plays, renders it almost impossible on chronological grounds that Pembroke can have been their subject. He was born on April 9, 1580, and was much younger than Southampton, who was born on Oct. 6, 1573. The earliest sonnets postulate a marriageable youth, certainly not younger than 18, an age which Southampton reached in the autumn of 1591 and Pembroke in the spring of 1598. The writing of the sonnets may have extended over several years, but as a whole they clearly belong to the years 1593–98 (or even earlier) rather than to 1598–1603. There is indeed not much external evidence available. Meres, in his *Palladis Tamia* (1598), mentions Shakespeare's "Sugred sonnets among his private friends" (*i.e.*, circulated in manuscript), but this allusion might come as well near the beginning as at the end of the series; and the fact that two, not of the latest, sonnets are included in *The Passionate Pilgrim* is equally inconclusive.

The only reference to an external event in the sonnets themselves that might at first sight seem useful is in the following lines (107):

> The mortal moon hath her eclipse endur'd,
> And the sad augurs mock their own presage;
> Incertainties now crown themselves assur'd,
> And peace proclaims olives of endless age.

This has been variously interpreted as referring to the death of Elizabeth I and the accession of James I in 1603, to the relief caused by the death of Philip II of Spain in 1598, and to the illness of Elizabeth and the threatened Spanish invasion in 1596. The "mortal moon" is generally considered to be Queen Elizabeth, but though "eclipse" may well mean "death," it is not clear whether to "endure eclipse" is to die, or to escape death. Arguments that the "mortal moon" was the hostile battle array of the Spanish Armada in 1588 (a reference to the Spanish ships being drawn up in the shape of a crescent), and that all the sonnets are to be dated before 1590, have found little critical acceptance.

Allusions, in sonnets 78–86, to the poet who rivaled Shakespeare in dedicating verses to the friend and patron have been variously interpreted. On critical grounds, "the proud full sail of his great verse" (86) can perhaps most satisfactorily be applied to Spenser, Marlowe, or Chapman, and more doubtfully, to Peele, Daniel, or Drayton. Marlowe has been rejected by those who date the sonnets after 1593—the year of his death—although it is possible to interpret the description of the rival poet as "by spirits taught to write" (86) as a reference to the success of his play *Dr. Faustus*, and to Greene's description of him as having prophetical spirits. The "affable familiar ghost" who is said to school the rival poet by night has been interpreted as a rather obscure reference to Chapman's own account of the sources of his inspiration in the dedication to one of his poems in 1594. Daniel inscribed a poem to Southampton in 1603, but none of the other poets named is known to have written either for Southampton or for Pembroke, or for any other W. H. or H. W., in any year during which the sonnets could have been written. Two minor poets, Barnabe Barnes and Gervase Markham, addressed sonnets to Southampton in 1593 and 1595 respectively, and Thomas Nashe composed improper verses for his delectation.

But even if external guidance fails, the internal evidence for 1593–98 as approximately the sonnet period in Shakespeare's life is strong. It was worked out in detail by two German scholars, Hermann Conrad (formerly Isaac), in an article in the *Shakespeare-Jahrbuch*, XIX (1884), and Gregor Sarrazin in *William Shakespeares Lehrjahre* (1897) and *Aus Shakespeares Meisterwerkstatt* (1906). Conrad's work, in particular, has hardly received enough attention from English scholars, probably because he makes the mistake of taking the sonnets in the order in which they are arranged in the German translation (1862) by Friedrich Martin von Bodenstedt, and of beginning his whole chronology several years too early in order to gratify a fantastic identification of W. H. with the earl of Essex. This, however, does not affect the main force of an argument by which the affinities of the great bulk of the sonnets are shown, on the ground of stylistic similarities, and of parallelisms of expression and theme, to be far closer to the poems and the range of plays from *Love's Labour's Lost* to *Henry IV* (*i.e.*, 1594–98) than to any earlier or later section of Shakespeare's work. This dating has the advantage of putting Shakespeare's sonnets in the full tide of Elizabethan production, which began with the publication of Sir Philip Sidney's *Astrophel and Stella* in 1591, and Daniel's *Delia* and Henry Constable's *Diana* in 1592, rather than later, during years for which this particular kind of poetry had already ceased to be modish. It is to these three volumes that the influence on Shakespeare of his predecessors can most clearly be traced; while he seems in his turn to have served as a model for Drayton, whose *Ideas Mirrour* was published in 1594. It does not of course follow that if the sonnets belong to 1593–98 W. H. is to be identified with Southampton. It is, in fact, unlikely that the publisher of the sonnets would address a great earl as "Master"; and W. H., even if above Shakespeare's rank, seems more likely not to have been of noble birth.

There is a possibility that there is an allusion to Shakespeare's romance in a poem called *Willobie his Avisa*, published in 1594, and written by an undergraduate called Henry Willoughby, of West Knoyle, Wiltshire. In this Willoughby, enamoured of an innkeeper's wife ("his Avisa"), apparently at Sherborne, takes counsel with "his familiar friend W. S. who not long before had tryed the curtesy of the like passion, and was now newly recovered of the like infection." But there is nothing outside the poem to connect Shakespeare with a family of Willoughbys, or with the neighbourhood of West Knoyle or Sherborne—although it is of interest that the commendatory verses prefixed to *Willobie his Avisa* contain the first literary references to Shakespeare by name, as the author of *Lucrece*.

Various other identifications of W. H. have been suggested, and much ingenuity has been shown by historians, critics, and literary detectives in hunting down references, searching through archives, and identifying portraits. One of the more interesting theories is Leslie Hotson's, persuasively argued in *Mr. W. H.* (1964), that the "lovely boy" was a certain William Hatcliffe (1568–1631), of Hatcliffe, Lincolnshire, who in 1587–88 was "prince of Purpoole" (*i.e.*, lord of misrule, *q.v.*), during the celebrations at Gray's Inn probably beginning at Hallowe'en and ending on the day after Candlemas (Oct. 31–Feb. 3). Hotson, dating the sonnets very early, identifies the "rival poet" as Marlowe, the "affable familiar ghost" as Greene, and the "dark woman" as Luce Morgan, one of the queen's gentlewomen during 1579–81, who later became a harlot. In "proving" his theory, Hotson throws much light on Elizabethan manners and customs, use of symbols, and concepts of love and friendship, but, like those of the many others who claim to have "solved the mystery," his explanations are almost too neat to be wholly convincing.

Another school of thought holds that Mr. W. H. was not the "friend" of the sonnets at all, but someone who might have procured the manuscript of them for Thorpe. Sir Sidney Lee identified him with one William Hall, himself a printer, and this explanation has also been applied, rather more plausibly, to Sir William Harvey

(Hervey), the third husband of Southampton's mother, who might have procured the sonnets for Thorpe. (Hotson, assuming William Hatcliff to be the sonnet's "begetter" in the sense of "inspirer," has the best of both worlds by concluding that he also gave the manuscript to Thorpe, thus doubly "begetting" them.) The stumbling-block to acceptance of this theory, however, whoever may be identified as the "procurer" of the sonnets for press, is the intended sense of Thorpe's dedication: it seems certain that the person to whom Thorpe "wishes" eternity and he to whom the poet "promised" eternity are one and the same. The mystery of Mr. W. H. will probably never be solved to everybody's satisfaction, and some may agree with the view of one critic, that its solution would "deprive us of one of the most enjoyable of literary pastimes." (E. K. C.; J. Cw.; X.)

IV. SHAKESPEAREAN STUDIES

A. Literary Criticism

Attempts to interpret Shakespeare's work have, in important ways, controlled and modified literary criticism in English. The plays have continued to be enjoyed whether whole, cut, or adapted, and the response of each age has governed its attitude to literary and aesthetic principles.

Ben Jonson was the first who had to account both for his own and his generation's appreciation of Shakespeare, and for his critical impatience, which resulted from a literary genius and an attitude to literature, especially to drama, very unlike that of Shakespeare (*see* also *Life: Prosperity and Fame*, above). Jonson, a classical scholar of keen, sardonic intellect, a conscious artist and satirist, was concerned both to correct the extravagance of the Elizabethans, and to castigate and reform the morals and conduct of his own age. Both he and, later, Dryden, the supreme poet and dramatist of a neoclassical age, spoke with two voices about Shakespeare, mingling praise for his achievement with disapproval of his aims and methods: their classicism (or neoclassicism) was checked by their instinctive appreciation of Shakespeare's mastery, and this caused them to reexamine their critical principles. Jonson saw bombast and lack of art in Shakespeare the dramatist, and, in his famous appraisal of him in his *Timber, or Discoveries* (written *c.* 1630; *see* above), clearly distinguished himself from those who praised Shakespeare uncritically, honouring his memory "on this side (*i.e.*, below) idolatry." Yet, in his verses for the First Folio, he could, himself "idolatrously," put him above the Greeks and Romans, not only (to borrow words used later by Milton) as a warbler of "native woodnotes wild," but as an artist.

This half line from Milton's "L'Allegro," with its implication that Nature had guided the unscholarly pen of the poet, dogged criticism for two centuries. So the author of the prologue to Shakespeare's *Julius Caesar* (printed, without the author's name, but early ascribed to Dryden, in *Covent Garden Drolery* . . . , 1672), wrote:

> His Excellencies came and were not sought,
> His words like casual Atoms made a thought:
> Drew up themselves in Rank and File, and writ,
> He wondring how the Devil it were such wit.

This may or may not be Dryden's verse, but it does express one of the opinions also expressed in his *Essay of Dramatick Poesie* (1668):

> All the Images of Nature were still present to him, and he drew them not laboriously, but luckily . . . Those who accuse him to have wanted learning, give him the greater commendation: he was naturally learn'd; he needed not the spectacles of Books to read Nature; he looked inwards, and found her there.

Like Jonson, and annoyed by the same elements in Shakespeare, Dryden yet recognized in him the "largest and most comprehensive soul" among "all Modern, and perhaps Ancient Poets," and Pope's preface to his edition of the plays in 1725 follows Dryden in this.

The easily accepted Platonic notion of a poet possessed by the muse continued to militate against a recognition of Shakespeare's conscious artistry, and a succession of works on his education, from Richard Farmer's *Essay on the Learning of Shakespeare* (1767) to T. W. Baldwin's *William Shakespeare's "Small Latine and Lesse Greeke"* (1944), speak for the liveliness of this interest. So, inferentially, do the Baconian and other theories (see *Anti-Shakespearean Theories*, below).

Meanwhile, the 18th-century editors and critics accumulated information on the writers of Shakespeare's time and on his sources, stimulated by Farmer's essay, which had used these for its own purposes. They wrote about Shakespeare's failure to observe the classical unities, his lack of rhetorical "decorum," his quibbles (puns), indecencies, and anachronisms; blemishes were indicated and beauties marked. Samuel Johnson's preface to his 1765 edition, a commanding work surveying the plays and summing up and transcending the issues of the age, while "placing" Shakespeare securely, decided the controversy about the unities, said wise things about the relation of the tragic to the comic, but still balked at the Shakespearean quibble: "A quibble was to him the fatal Cleopatra for which he lost the world, and was content to lose it." The "world" lost is fidelity to nature, and the approbation of an age that recognized in Shakespeare this fidelity above all else: as Johnson himself said, earlier in his preface, "Shakespeare is above all writers, at least above all modern writers, the poet of nature; the poet that holds up to his readers a faithful mirrour of manners and of life."

Walter Whiter defended the quibble in *A Specimen of a Commentary on Shakspeare* . . . (1794), unappreciated in its own day but recognized as important in the 1930s when his points had been made all over again. Stressing the unconscious and associative origin of many puns, he touched on Shakespeare's creative processes, isolated image clusters, and hinted at an awareness of what Caroline Spurgeon (*see* below) came to call "iterative imagery." Shakespeare is, however, no more an artist for being exhibited as a slave of the associative process, and Whiter's apologetic tone is true to his period which, important as it is for its scholarly work, is more concerned with adjudication than interpretation. In surveying 18th-century Shakespearean scholarship and criticism as a whole one sees chiefly promise and prophecy: a clearing of the way.

One work must be excepted: Maurice Morgann's *An Essay on the Dramatic Character of Sir John Falstaff* (1777). This book, important and influential historically, is as wise and broad in its argument as it is vital in its expression. There had been essays in the periodicals of the 1750s on the Shakespearean characters; Johnson had made pertinent comments on them in the notes to his edition; and William Richardson's *A Philosophical Analysis and Illustration of some of Shakespeare's Remarkable Characters*, more intent on indicating casebooks of experience for the benefit of the reader than on psychological analysis, had appeared in 1774; but Morgann's essay transcends these (and also its immediate issue: the courage or cowardice of Falstaff) in its large, liberal, and refreshing response to the plays. Shakespeare, he says: "boldly makes a character act and speak from those parts of the composition, which are *inferred* only, and not distinctly shewn. This . . . seems to carry us beyond the poet to nature itself. . . ." This capacity in Shakespeare for seeing his characters as "*whole*, and as it were original," not imitated, "rather as Historic and Dramatic beings," was central to his purpose, and it remains true despite the critical perversions which have stemmed from it. With Thomas Whately's detailed comparison of *Richard III* and *Macbeth* (*Remarks on some of the Characters of Shakespeare*, 1785) it became clear, that "the distinction and preservation of *character*" were of first importance to 18th-century critics, and this insistence was to prevail throughout the 19th century and into the 20th. It informs the works of William Hazlitt, S. T. Coleridge, F. J. Furnivall, Edward Dowden, A. C. Bradley, Harley Granville-Barker, and John Dover Wilson.

William Hazlitt, admiring Whately's analysis, published his *Characters of Shakespear's Plays* in 1817. It is an excellent handbook; it has an air of novelty and adventure, its author being (like all the Romantics) more conscious of differences from his predecessors than of a tradition. Enthusiastic, reverential, intuitively perceptive, he has a flair for apt quotation and, coming to Shakespeare as a dramatic critic, he is saved from some of the pitfalls of a purely literary study of character.

Coleridge's lectures on Shakespeare had been delivered before

Hazlitt's book appeared, but though they share a certain interest in character, the poet's critical range is far beyond that of his contemporaries and releases us from the 18th-century modes of thought. He was concerned to recognize Shakespeare's powers of judgment and to see him as a conscious artist. He claimed to have argued in a lecture course in 1802 that "Shakespeare's judgment was . . . still more wonderful than his genius, or rather, that the contradistinction itself between judgment and genius rested on an utterly false theory" (*The Canterbury Magazine*, September 1834). This recognition, and the further one that every work of art must be judged by its own organic laws, put an end to the balancing of faults against beauties. Coleridge directed attention to the poems, used before (when read at all) only as quarries for parallel passages, and he analyzed *Venus and Adonis* in chapter 15 of his *Biographia Literaria* (1817). Examples from Shakespeare illuminated his distinction between the imagination and fancy, and both he and Keats, as great poets themselves, revealed much about Shakespeare's creative processes.

The 20th century has seen, indeed, deep affinities in the creative minds of Keats and Shakespeare. Keats's letters are permeated with an awareness of the dramatist, and his attribution to Shakespeare of that envied "negative capability"—"when a man is capable of being in uncertainties, mysteries, doubts, without any irritable reaching after fact and reason" (*Letter to G. and T. Keats*, Dec. 21, 1817)—has proved fruitful.

Less fruitful have been the disputes about precedence and originality in the Romantic movement, and it is more relevant to recognize that a philosophic temper stimulated, in Germany and England, certain approaches and presumptions. G. E. Lessing (*q.v.*) recognized Shakespeare's closer affinity with the Germanic spirit and literature than with the French rules; the prose translations of 22 of the plays by C. M. Wieland (*q.v.*), published during 1762–66; and those in verse (1797–1810, completed by others, 1821–23), of A. W. von Schlegel (*q.v.*), were enthusiastically seized on, and the latter's *Über dramatische Kunst und Literatur* (1809–11; English translation, 1815) explored the Coleridgean world. The philosophical significance of character became central, Shakespeare's profundity as a formal artist was acclaimed, and a leading idea was seen to inform his work as a whole. When this root idea was analyzed in particular plays (for example, in Hermann Ulrici's analyses in *Shakespeares dramatische Kunst*, 1839; Eng. trans. 1876) the interpretation was often profitable, but there was in German criticism of this period the kind of insensitivity to language that made Schlegel accept *A Yorkshire Tragedy* as Shakespeare's without qualms, and a lack of knowledge—more serious, even, than that among English writers—about the circumstances in which the plays were written and their stage origins. Schlegel has, however, fine things to say on particular plays.

In the Victorian period no major critic wrote on Shakespeare. Victor Hugo and Swinburne enthused eloquently, but the first original interpretative work was Walter Pater's, in three essays published in his *Appreciations* (1889). Scholarship, industry, and scientific method were to the fore. Shakespeare societies were begun in England (1840) and Germany (1865), and their published proceedings embody the virtues and faults of the time. Edmond Malone's work on the order of the plays (see *Chronology of the Plays*, above) was supplemented, and interest in Shakespeare's development as a poet and dramatist through the four periods established by Furnivall's verse tests was stimulated: the reach of the whole was surveyed. The dominant work here is Dowden's *Shakspere: a Critical Study of His Mind and Art* (1875). Investigation of Shakespeare's spiritual biography as revealed in the plays persisted into the early 20th century, but the great book then was A. C. Bradley's *Shakespearean Tragedy* (1904). Here is the culmination of all character analysis. Bradley's very greatness, and the finality with which some of his statements were received, have irked later critics, and they have overinsisted on his obsession with character, neglect of the poetry, and unawareness of Elizabethan stage conventions. He irritated most when he pursued characters far beyond the circumstances of the play, or subordinated plot and action to psychological study, or, preoccupied with the moral order at the root of tragedy, sentimentalized some of the issues. But his chapters remain as a necessary counterbalance to the works of the interpreters who reacted against him.

This reaction typifies 20th-century criticism. Sir E. K. Chambers, whose *William Shakespeare: a Study of Facts and Problems* (two volumes, 1930), remains a standard work, brought to the study of Shakespeare's life and plays an unrivaled knowledge of the medieval and Elizabethan stage, and a study of the dramatic conventions of that stage inspired a new school of critics. L. L. Schücking's *Die Characterprobleme bei Shakespeare* (1919; Eng. trans., 1922) and E. E. Stoll's *Art and Artifice in Shakespeare* (1933) are representative works. Dramatic realism, they argue, rightly, is not that of actual life; and inconsistencies of character are best explained in terms of Elizabethan practice and convention. Much of this was salutary for a generation later accused of putting its faith in dramatic realism. It is even necessary as a corrective to the criticism of Granville-Barker (*q.v.*; *Prefaces to Shakespeare*, 1927–48), which was healthy in that it came from a dramatist-actor-producer who had illuminated Shakespeare's art in stage presentation, but limited in its reliance on Bradley and in an underlying predilection for the naturalistic theatre.

A further aspect of historical research was the investigation of Elizabethan physiological psychology. R. L. Anderson's *Elizabethan Psychology and Shakespeare's Plays* (1927), L. B. Campbell's *Shakespeare's Tragic Heroes: Slaves of Passion* (1930), and L. Babb's *The Elizabethan Malady: a Study of Melancholia in English Literature from 1580 to 1642* (1951) demonstrate the range of studies in this kind; studies that are either stimulating and corrective or stultifying and peripheral according to the author's application of his thesis.

One may recognize as a counterblast to historical criticism the contribution of the psychoanalysts. Far from wishing to explain away inconsistencies of characterization, these critics positively welcomed them. J. I. M. Stewart's *Character and Motive in Shakespeare* (1949) discussed the general critical issues arising, and the same year brought Ernest Jones's *Hamlet and Oedipus* (a revision of chapter one of *Essays in Applied Psycho-Analysis*, 1921), a development of ideas prompted by Freud. Specifically Freudian inspiration apart, there is no doubt that the popularization of theories of the unconscious and the mysteries involved in motivation brought a sense of freedom to the interpretation of literature in general and Shakespeare in particular. A wider freedom is claimed by Lascelles Abercrombie, who says that "anything which may be found" in Shakespeare's art, "even if it is only the modern reader who can find it there, may legitimately be taken as its meaning" (*A Plea for the Liberty of Interpreting*, British Academy annual Shakespeare lecture, 1930). His lecture answered the so-called "disintegrators" who, led by J. M. Robertson (1856–1933), had apportioned out plays or parts of plays, deemed to fall below some elusive, subjective standard, to other dramatists. Those elements which Shakespeare adopted and absorbed, Abercrombie argued, became his and were open to interpretation with the rest.

The freedom demanded by Abercrombie had already been exercised. L. C. Knights's article "How Many Children Had Lady Macbeth?" (1933; reprinted in *Explorations*, 1946) reacted against Bradley, and focused a movement which had begun with the early works of John Middleton Murry and George Wilson Knight, and in the researches of Caroline Spurgeon. These writers, conscious that character analysis was not enough, in their various ways redirected attention to the poetry and symbolism. L. C. Knights argued that "we start with so many lines of verse on a printed page," and that we have to "allow full weight to each word, exploring its 'tentacular roots'." (This direction should be seen against the critical and poetical background of the 1920s and 1930s, with its emphasis on the image and on ambiguity, its re-interpretation of the metaphysical poets, and its response to the poetry and criticism of T. S. Eliot. A critical discipline can be seen evolving, sympathetic to Knights's aims and molded by the thought of I. A. Richards, T. S. Eliot, William Empson's *Seven Types of Ambiguity* [1930], and F. R. Leavis; *see* CRITICISM: *The New Criticism.*) Beginning indeed with the "printed page," Caroline Spurgeon card-indexed Shakespeare's images and used them

interpretatively. The most brilliant and influential discovery in her work was that of the "iterative image," and in *Shakespeare's Imagery and what it tells us* (1935) she pointed to prevalent groups of images which characterize the individual plays and unconsciously fashion our response to them. Like Whiter, she indicated clusters of images peculiar to Shakespeare, and these were used to throw light on his creative processes and to guide the attribution of scenes and plays to him. W. H. Clemen, in *Shakespeares Bilder: Ihre Entwicklung und ihre Funktionen im dramatischen Werk* (1936; translated and enlarged as *The Development of Shakespeare's Imagery*, 1951), studied with insight Shakespeare's dramatic use of metaphor.

The impressive and original body of work produced by G. Wilson Knight is historically linked with this school of criticism, but is more free ranging. Actor, producer, and (later) dramatist, he regards each play "as a visionary unit bound to obey none but its own self-imposed laws," and his examination relates each "given incident or speech" either to "the time sequence of story or the peculiar atmosphere, intellectual or imaginative, which binds the play" (*The Wheel of Fire*, 1930). The "symbolic overtones" are revealed. His method radically redirected the attention of the student and freshly interpreted in particular, perhaps, *Measure for Measure*, the Roman plays, *Timon of Athens*, and the last romances. Emphasis on the image has made the business of interpretation more laborious. A book on each play has become common; among the best have been two by Roy Walker on *Hamlet* and *Macbeth* (*The Time is Out of Joint*, 1948; *The Time is Free*, 1949), and Robert Heilman's on "Image and Structure in King Lear" (*This Great Stage*, 1948).

The so-called "new criticism," characterized by close linguistic analysis, has had its brilliant successes. Its excesses, too, have been flagrant, and opposition has taken the form of a move nearer to Bradley, or another kind of historical criticism. Rosemond Tuve (in *Elizabethan and Metaphysical Imagery*, 1947) attacked the Caroline Spurgeon approach; Sister Miriam Joseph (in *Shakespeare's Use of the Arts of Language*, 1947) pointed to the poet's studied use of the rhetorical figures; and John Holloway, in a number of critical articles culminating in the introduction to his *The Story of the Night* (1961), expressed disquiet about the "current coin of Shakespeare criticism," "its vocabulary of metaphor," and its refusal to examine the fundamentals of its method, with its implication that there is only one valid approach to dramatic writing. Moreover. a wider knowledge of the thought and beliefs in his day, demonstrated in such works as Hardin Craig's *The Enchanted Glass* (1936) and E. M. W. Tillyard's *The Elizabethan World Picture* (1943), valuably checked and controlled any tendency to think peculiar in Shakespeare what was common to his age. Scholarly work on the sources of the plays has also intensified, again with the implication that the ardent critic may overstress the liberty of interpreting. Scholarship and interpretation have, however, happily come more nearly into alliance than at any time since Johnson's day. New editions of the plays are in the hands of enlightened critics. An interpretative work, however, roving over all the relevant fields of knowledge and research and surveying (as Johnson could) Shakespeare's achievement, is unlikely to appear. One is thrown back on Heminge and Condell's First Folio preface: ". . . his wit can no more lie hid, then it could be lost. Read him, therefore; and againe, and againe . . ."

(G. A. O.; X.)

B. Textual Criticism

The textual critic of Shakespeare needs to know about the status of the various texts, and their value as evidence; the language of Shakespeare's time; Shakespeare's idiosyncrasies as a writer; and the barrier which an Elizabethan, or Jacobean, printer might be likely to erect between the manuscript "copy" and the printed text.

1. Texts.—Modern knowledge of the status of the Folio and Quarto texts, and some of the problems involved—for example, that of the "bad" Quartos—have already been discussed under *The First Folio and the Quartos*, above. The last collected Folio edition of the plays (F4) was published in 1685; thereafter editors aimed at making Shakespeare more accessible to the reader by printing the works in smaller format, in several volumes. They also aimed, from the 18th century onward, at elucidation and amendment of the text, by comparison of earlier editions (the Folios and Quartos), and by a study of Shakespeare's period, and of the works of his contemporaries. Important collected editions of the works have been published as follows:

1709	Nicholas Rowe: 6 vol. Based largely on F4; notable for adding stage directions, completing lists of *dramatis personae* for every play, and as containing the first life of Shakespeare.
1725	Alexander Pope: 6 vol. Based on Rowe, collated with some Quarto texts.
1733	Lewis Theobald: 7 vol. The first scholarly edition.
1744	Sir Thomas Hanmer: 6 vol. Based on Theobald.
1747	William Warburton: 8 vol. Based on Theobald.
1765	Samuel Johnson: 8 vol. Based on Warburton, with valuable preface (see *Literary Criticism*, above).
1767–68	Edward Capell: 10 vol. Based on the collection and collation of Quarto texts, and research on Elizabethan literature.
1773	George Steevens: Johnson's text, with quotation from Shakespeare's contemporaries; greatly enlarged in the 2nd edition, 10 vol. (1778).
1790	Edmond Malone: 10 vol. Important for its "Attempt to ascertain the order in which Shakespeare's plays were written," and the first use of the Revels Accounts, Henslowe's Diary, and other contemporary records.
1795–96	Samuel Johnson (Philadelphia): 8 vol. (1st U.S. edition).
1803	First Variorum: 21 vol. The 5th edition of the Johnson-Steevens text of 1773, edited after Steevens' death by Isaac Reed, and containing three volumes of introductory matter: prefaces by earlier editors, Malone's account of the English stage, extracts from Henslowe's Diary, etc.
1813	Second Variorum: 21 vol. A reprint of the First Variorum.
1818	Thomas Bowdler: 10 vol. The so-called *Family Edition*, suitably "bowdlerized."
1821	Third Variorum, also known as "Boswell's Malone": 21 vol. An edition planned by Malone, and finished after his death by James Boswell the Younger. Incorporates the prolegomena to the first two Variorum editions, with Malone's life of Shakespeare, The "Attempt . . . ," and an essay on Shakespeare's phraseology and metre. Unlike Variorum I and II, and earlier editions, it contains the *Poems*.
1842–44	J. P. Collier: 8 vol.
1857	Alexander Dyce: 6 vol. 3rd edition (1895–1901): 10 vol.
1853–65	J. O. Halliwell-Phillipps: 16 vol. A scholarly Folio edition.
1863–66	W. G. Clark, W. Aldis Wright, and J. Glover: 9 vol. The Cambridge edition. Revised, 1891–93.
1864	Clark and Wright: 1 vol. The Globe edition. Based on the Cambridge.
1871–	New Variorum, edited by H. H. Furness *et al.* The most comprehensive edition; supplementary volumes bring early volumes up to date.
1880–81	H. N. Hudson: 20 vol. Harvard edition.
1886–1906	A. Morgan *et al.*: 22 vol. Bankside edition. Parallel texts.
1888–90	Henry Irving edition: 8 vol. with full stage history of each play.
1899–1924	General editors: W. J. Craig, R. H. Case *et al.* Arden edition; 1 vol. for each play. 37 vol.
1918–28	General editors: W. L. Cross, C. F. T. Brooke *et al.* Yale edition: 1 vol. for each play. 40 vol.
1921–63	General editors: Sir Arthur Quiller-Couch, J. Dover Wilson *et al.* New Cambridge edition. 1 vol. for each play.
1937–	G. B. Harrison: Penguin Paperback edition. 1 vol. for each play.
1938–	General editors: R. E. C. Houghton *et al.* New Clarendon edition. 1 vol. for each play.
1951	Peter Alexander: 1 vol.
1951–	General editors: Una Ellis-Fermor (until her death in 1958), H. F. Brooks, H. Jenkins *et al.* New Arden edition. 1 vol. for each play.
1951	Hardin Craig: 1 vol.
1954–	General editors: H. Kökeritz, C. T. Prouty *et al.* Revised Yale edition. 1 vol. for each play.
1956–	General editor: A. Herbage. Pelican Paperback edition.
1958	P. Alexander: 3 vol. Heritage Shakespeare.
1960–	General editor: L. B. Wright. Folger Library Paperback edition.

New methods of printing in the 19th century led to a new development: the production of facsimiles of the early editions. Important facsimiles include:

Of the Quartos:

1862–76	Edited by J. O. Halliwell-Phillipps, 48 vol. Lithographic.
1880–89	The Griggs-Praetorius Quarto Facsimiles, general editor, F. J. Furnivall, 43 vol. Photolithographic.
1939 *et seq.*	By W. W. Greg, for the Shakespeare Association. Collotype.

Of the First Folio:
1862–64 Edited by L. Booth, 3 vol. Type facsimile.
1876 Edited by Halliwell-Phillipps. Reduced facsimile. Photo-lithographic.
1902 The Oxford Facsimile, ed. by Sir Sidney Lee. Collotype.
1955 The Yale Facsimile, ed. by H. Kökeritz and C. T. Prouty. Reduced facsimile. Photographic.
Of the Four Folios:
1904–10 Published by Methuen. Photozincographic.
Of the Poems:
1905 Edited by Sir Sidney Lee. Collotype.

Of the editors who preceded him, W. A. Wright wrote in 1893, almost certainly with some excess of charity:

> And it is a study of great interest to follow them as they exercise their varied talents on the noblest field which the literature of their country afforded: Rowe, himself a dramatist of no mean skill; Pope, with his deep poetic insight; Theobald, with his fine tact and marvellous ingenuity; Hanmer, whose guesses, however they may pass the sober limits of criticism, are sometimes brilliant, often instructive, and never foolish; Warburton, audacious and arrogant, but now and then singularly happy; Johnson, with his masculine common sense; Capell, the most useful of them all, whose conscientious diligence is untiring, whose minute accuracy is scarcely ever at fault; Steevens, Malone, . . . Boswell, . . . with all their varied learning; together with their successors of the present generation in England, Germany and America, who have devoted themselves to the illustration of Shakespeare as to a labour of love.

It would be unfair to suggest that the earlier editors, from Rowe onward, did not interest themselves in the establishment of a correct text; but by modern standards they worked by hit-or-miss systems, and sometimes substituted their own (often excellent) tastes for critical principles. They tended to assemble as many various readings, from various editions, as possible, and to use their own judgment as to which was the best. Considering their methods their success was striking. (*See* also articles on many of the individual editors.)

The editors of the Cambridge Shakespeare seem to have been the first to make a reasonable attempt to list the early editions and to sort out the relationships between them. Their critical apparatus of variant readings and emendations remains unsurpassed. The Globe edition provided a standard text which modern scholars find impossible to dislodge: this is the text generally used for quoting line numbers. The almost impeccable Booth type facsimile of F1 placed much necessary information before the public, and remains the handiest edition. The Griggs-Praetorius Quarto facsimiles should have been even more reliable, but Aldis Wright has demonstrated that they had undergone remarkable tinkering (one "facsimile" of a 1595 edition contains words foisted in from a 1619 edition), and that no confidence can be placed in their punctuation. The critical introductions to some of these Quarto facsimiles, however—notably those by P. A. Daniel—are of great value. The Sidney Lee F1 facsimile, though less portable than the Booth, may be said to replace it on the score of reliability. The Kökeritz and Prouty facsimile of F1 has been faulted, in that the photographic reproduction sometimes omits punctuation. The collotype facsimiles of the Quartos under the general editorship of W. W. Greg are trustworthy.

The path of the Shakespearean textual critic used to contain the major stumbling-block of the problem of the early Quartos (the so-called "bad" Quartos; see *The First Folio and the Quartos*, above) of certain plays: *Romeo and Juliet, Henry V, Hamlet, The Merry Wives of Windsor*, 2 and 3 *Henry VI, Richard III*, and *King Lear*, in which the disparity between the earlier and later texts is remarkable. Rival explanatory theories long disputed the field: either that here were early versions later revised by Shakespeare (the *Henry VI* plays being, it was suggested, the work of earlier, cruder dramatists), or that the corrupt texts were versions of stage performances taken down in shorthand. These theories were eventually routed by W. W. Greg, A. W. Pollard, J. Dover Wilson, and Peter Alexander (see *Bibliography*), who established that the "bad" Quartos are the result of reconstruction from memory by actors of texts to add to the repertories of near-derelict traveling theatrical companies, not, as had been earlier suggested, that the shorthand writers worked at the behest of fraudulent, stealthy, and injurious publishers. Philip Edwards (in "An Approach to the Problem of *Pericles*," *Shakespeare Survey*, number v, 1952) convincingly suggested that *Pericles* should be added to the "bad" Quartos: it differs from the others in that it is the work [illegible] two reporters of widely different skill, and in that no good vers[illegible] has ever been found to replace the existing "bad" Quarto [illegible] The shorthand writer theory was given its deathblow by [illegible] Duthie's *Elizabethan Shorthand and the First Quarto of King Lear* (1950). The relative quality of the Quartos of *Richard III* and *King Lear* is so good that no satisfactory explanation of them has been offered, but there is general agreement that the former is a memorial reconstruction, whilst the *Lear* text may be basically a "surreptitious" transcript (*see* A. Walker, *Textual Problems of the First Folio*, 1953).

General attacks on the editorial problem began in three of the annual Shakespeare lectures of the British Academy: A. W. Pollard, *The Foundations of Shakespeare's Texts* (1923), W. W. Greg, *Principles of Emendation in Shakespeare* (1928), and R. B. McKerrow, *The Treatment of Shakespeare's Text by His Earlier Editors, 1709–1768* (1933). McKerrow's *Prolegomena for the Oxford Shakespeare: a Study in Editorial Method* (1939) is a full-dress discussion of the problems; the author's death prevented publication of any of the projected edition. Greg's *The Editorial Problem in Shakespeare* (1942) carried the discussion further; in it, and in *The Shakespeare First Folio: Its Bibliographical and Textual History* (1955), he laid before the scholar convincing arguments to support a theory of the text for every one of the plays. There is no starkness of simplicity in his findings. For no two plays, it seems, is the history of the text the same. F. T. Bowers *On Editing Shakespeare and the Elizabethan Dramatists* (1955) is a valuable manifesto of analytical bibliography.

Two of the less complicated text histories may be given as examples. For *Romeo and Juliet*, the first edition (1597; Q1) is a "bad" Quarto. Q2 (1599) is an authoritative text printed, it is thought, from a Shakespeare autograph. Q3 (1609) is printed from Q2. Q4 (undated, but probably 1619) is printed, with [illegible]ous but apparently unauthoritative changes, from Q3. F1 is printed from Q3. Later researches have led to a theory that a pen-corrected example of Q1 may have been used as copy for [illegible] which complicates a hardly pellucid situation with theories th[illegible] good text was contaminated by a bad one. For *Titus Andron[illegible]* Q1 (1594) seems to have been set from a Shakespeare autogr[illegible] Q2 (1600) from Q1, Q3 (1611) from Q2. F1 was printed from [illegible] with the addition of Act III, sc. 2, which is absent in all [illegible] Quartos, and must derive from an authoritative manuscript [illegible]plications arise from the fact that printer's copy for any of [illegible] plays may have been any one of the following: Shakespeare's [illegible] copy, his rougher "foul papers," fair copies apparently made [illegible] known scribe) especially for use in the printing house, theatr[illegible] prompt copies, earlier editions corrected by pen and ink [illegible] manuscripts of various status, etc.

An important development for Shakespearean textual criticism was the general acceptance of the view that three pages in [illegible] manuscript "Booke of Sir Thomas Moore" (British Mus[illegible] Harleian MS. 7368) are in Shakespeare's handwriting—the [illegible] certainly known example of it except for the six legal signa[illegible] (see *Life*, above). This view, suggested in the 1870s, was [illegible] thoroughly discussed by A. W. Pollard, W. W. Greg, E. [illegible] Thompson, J. Dover Wilson, R. W. Chambers, and others [illegible] *Shakespeare's Hand in the Play of Sir Thomas More* [illegible] which brought forward paleographic, orthographic, linguistic, [illegible]listic, and psychological evidence to support the identification [illegible] despite much controversy, and a frenzy of attempts during [illegible] 1920s and '30s to solve all textual problems by an appeal to Sha[illegible]speare's calligraphic style as here revealed, it has proved help[illegible] the elucidation of disputed readings, and is thus a valuable aid [illegible] modern editors of the plays.

Under the influence of Pollard, Greg, and McKerrow, a number of new editions of single plays have been published. Peter Alexander's one-volume edition (1951) is probably the most satis[illegible] complete text. The New Cambridge Shakespeare, begun [illegible] J. Dover Wilson and Sir Arthur Quiller-Couch with *The Tem[illegible]* (1921), and carried on by Dover Wilson and others after Q[illegible] Couch's death in 1944, has provided a number of stimulating [illegible]

controversial texts. Some of the New Arden (1951 *et seq.*) editions of single plays, each by a different editor, have evidently been influenced by the new bibliographical methods; others have not.

2. The Language of Shakespeare's Time.—The 18th-century editors of Shakespeare, particularly Theobald, Johnson, Capell, Steevens, and Malone, were aware that much of Shakespeare's language was obsolete, and so not easily intelligible to 18th-century readers, and as a result they made considerable study of the work of Shakespeare's contemporaries. Several glossaries and illustrative notes on Shakespeare's language were published. Knowledge of Elizabethan English was greatly advanced by *A Glossary: or, Collection of Words, Phrases, Names . . . in the Works of English Authors, particularly Shakespeare and his Contemporaries* (1822), by the philologist, and keeper of manuscripts at the British Museum, Robert Nares.

By the mid-19th century, various scholarly societies (*e.g.*, the Camden Society, founded 1838; the Shakespeare Society, 1840; and the Percy Society, 1840) had begun the reprinting of a considerable body of Elizabethan literature. Among the leaders in this were J. P. Collier (*q.v.*) and J. O. Halliwell (later Halliwell-Phillipps), two of the founders of the Shakespeare Society, but Collier was led, by zeal for scholarship or some other motive, to extensive forgeries to support his theories and textual emendations, and Halliwell was suspected of the theft of manuscripts from the library of Trinity College, Cambridge. These lapses, although they resulted in the winding-up of the Shakespeare Society in 1853, do not prevent modern Shakespearean scholars from acknowledging a considerable debt to both. The reprinting of Elizabethan literature was continued by the New Shakespeare Society, founded by F. J. Furnivall (*q.v.*) in 1873; it lasted until 1894.

Other 19th-century works on Shakespeare's language were E. A. Abbott's *A Shakespearian Grammar* (1869), and Alexander Schmidt's *Shakespeare-Lexicon* (1874–75; revised by G. Sarrazin, 1902). In 1874, too, began the publication of the *New English Dictionary* (N.E.D.; later the *Oxford English Dictionary*, O.E.D., completed in 1928), which owed much to the labours of the reprinters who had disseminated earlier English writings, and which, in its turn, is an invaluable aid to Shakespearean scholars. Based on the N.E.D. was C. T. Onions' *Shakespeare Glossary* (1911, revised 1919), and this, with the N.E.D., and Schmidt's *Lexicon*, laid the foundations for later work on Shakespeare's language. New directions, with implications for the text, were explored in H. Kökeritz' *Shakespeare's Pronunciation* (1953).

3. Shakespeare's Idiosyncrasies.—Knowledge of those matters in which Shakespeare differed from his contemporaries—versification, language, methods of thought, methods of composition, etc.—is necessary for the establishment of his text. Here, however, agreement among scholars is not easy, and subjective thinking and personal taste—or prejudice—play a large part. Scholarly aids of the highest value can be found in the great *Concordance to Shakespeare* (1894) of John Bartlett (keyed to the Globe text), and its still useful pioneer predecessor (*Complete Concordance to Shakespeare*, 1844–45) by Mrs. Cowden Clarke. Studies of Shakespeare's education, methods of writing, use of images, and use of sources, continue to be made (see *Literary Criticism*, above). It is impossible to sum up the findings of the researchers; as yet no clearly recognizable pattern in the Shakespearean carpet emerges.

4. Shakespeare and His Printers.—The modern scholar, presented with a text of Shakespeare, wants to know what relationship it bears to the author's manuscript, and hopes to be told how he can reconstruct that manuscript from it. Much of the effort made to satisfy him has been described above. Considerable activity however, continues among those who, like Greg and many other scholars mentioned, are interested in what happened in an Elizabethan printing house. McKerrow's *Introduction to Bibliography for Literary Students* (1927) remains the starting-point for such investigations.

The first serious examinations of the First Folio came from Sir Sidney Lee, in his introduction to the Oxford Facsimile (1902), and other writings; and from Pollard and Greg, largely in their attempts to traverse his theories. Appreciable headway began to be made with E. E. Willoughby's *The Printing of the First Folio of Shakespeare* (1933). Using a variety of new techniques, Willoughby was able to reconstruct many of the day-to-day events in Jaggard's printing house while F1 was being printed. The later work of Charlton K. Hinman (*see* below) and J. W. Shroeder (published in *The Great Folio of 1623*, 1956) destroyed the validity of some of Willoughby's work, but his achievement remains of value.

The key question would once have seemed to be "What did printers do to texts?" Willoughby showed that this question must be narrowed to the particular printers dealing at a particular time with a particular text. Hinman saw that any given line of type in F1 was set up by an individual compositor, possibly a different man from the one who composed a given line on the next page, and by a variety of techniques (including use of an instrument for the mechanical collation of two copies of a single book—known as the Hinman collating machine—and the elaboration of methods for distinguishing between the work of different compositors) he reached a stage of knowing far more of what went on in Jaggard's shop in 1623 than Jaggard himself could ever have known.

Hinman's findings began to be made available in the 1950s, and his methods were put to good use by other scholars, notably Fredson T. Bowers and Alice Walker. The latter, concentrating on the evidence provided by Hinman's methods that made it possible to identify the various compositors working on the First Folio, established that two (known as A and B), distinguishable by their spelling, were liable to different sorts of error in setting type from copy. Finally, with the publication of Hinman's *The Printing and Proof-Reading of the First Folio of Shakespeare* (2 vol., 1963), it became possible also to identify three other compositors—C, D, and E—and to conclude that E, whose hand is seen in the section of the Folio containing the tragedies, was a raw apprentice, not regarded as capable of setting type from a manuscript copy, but only from one previously printed.

Furthermore, Hinman's new and scientific methods, used to collate the 79 copies of the First Folio collected at the Folger Shakespeare Library, Washington, D.C., made it possible to reconstruct the details of Jaggard's printing and proofreading procedures, and revealed that, by modern standards, the methods used when producing the First Folio were haphazard in the extreme. For instance, the method of setting up the type not by successive pages, but in sets of 3 sheets, doubled to form 6 leaves and 12 pages, with the consequent need to estimate from copy where each page would begin and end, led to miscalculations of length, especially if the compositors were working from annotated copy, so that they were obliged to split lines of verse to fill up gaps, to print verse as prose to save space, and to omit or insert stage directions.

But perhaps the most important discoveries relate to proofreading. Hinman demonstrated that the pages were read and corrected while the printing was in progress, and that to form any one copy of the Folio, both corrected and uncorrected sets of three sheets, not necessarily in order of printing, were collected and bound together. Thus, it seems likely that no two copies of the First Folio were identical. From those available it is clear that few proof corrections were made—only on 134 of the 900 pages of the First Folio are any to be found, and on those only an average of fewer than four corrections a page were made. The proofreader was clearly chiefly interested in typographical niceties, and usually applied his wit to the correcting of substantive errors, rather than consulting copy; in fact, only two indubitable cases of the printer's having consulted copy in making a correction were claimed by Hinman.

Thus, it can be seen that the results of his work for textual criticism were twofold: the opening up of new possibilities of research into the idiosyncrasies of compositors; and the development of a new, and justifiable, eclecticism in dealing with disputed readings.

V. ANTI-SHAKESPEAREAN THEORIES

The idea that Shakespeare's plays and poems were not actually written by William Shakespeare of Stratford has been the subject

of many books and is widely regarded as at least an interesting possibility. The source of all doubts about the authorship of the plays lies in the disparity between the greatness of Shakespeare's literary achievement and his comparatively humble origin, the supposed inadequacy of his education, and the obscurity of his life. In Shakespeare's writings men have claimed to discover a familiarity with languages and literature, with such subjects as law, history, politics, and geography, and with the manners and speech of courts, which they regard as inconceivable in a common player, the son of a provincial tradesman. This range of knowledge, it is said, is to be expected at that period only in a man of extensive education, one who was familiar with such royal and noble personages as figure largely in Shakespeare's plays. And the dearth of contemporary records has been regarded as incompatible with Shakespeare's eminence, and as therefore suggestive of mystery. That none of his manuscripts has survived has been taken as evidence that they were destroyed to conceal the identity of their author.

1. The Claims Put Forward for Bacon.—Though earlier writers had expressed skepticism about Shakespeare, the first identification of the author as Francis Bacon, viscount St. Albans, seems to have been made *c.* 1785 by the Rev. James Wilmot, rector of Barton-on-the-Heath, Warwickshire. Wilmot communicated his views to J. C. Cowell of Ipswich, whose account of the matter is contained in a manuscript in the University of London library (*see* Allardyce Nicoll, "The First Baconian," in the *Times Literary Supplement*, Feb. 25, 1932). A letter from Sir Toby Matthew to Bacon contains the statement, "The most prodigious wit that ever I knew . . . is of your Lordship's name, though he be known by another," and this has been regarded as proof of Bacon's pseudonymous writings. Their similar treatment of references from the Bible, the law, and the classics has made Bacon a candidate for the authorship of the plays. The first published statement of this theory was W. H. Smith's *Was Lord Bacon the Author of Shakespeare's Plays?* (1856), which set on foot the Bacon-Shakespeare controversy. W. S. Melsome (*The Bacon-Shakespeare Anatomy*, 1945) found references in Shakespeare's plays to works by Bacon not published during Shakespeare's lifetime, which therefore he could not have seen; from this Melsome concluded that, as only Bacon could have known of these unpublished works, he must have written the plays containing references to them. Another method of investigation was used by I. Donnelly who, in *The Great Cryptogram: Francis Bacon's Cypher in the so-called Shakespeare Plays* (1887), found ciphered messages embedded throughout the plays that proved Bacon's authorship. Sir E. Durning-Lawrence (*Bacon is Shakespeare*, 1910) discovered in *Love's Labour's Lost* the word "honorificabilitudinitatibus," which forms the anagram *Hi ludi F. Baconis nati tuiti orbi* ("These plays, the offspring of F. Bacon, are preserved for the world"). Others carried this cryptographic line of investigation further, but the professional cryptologists W. F. and E. S. Friedman examined all the Baconian ciphers and rejected them as invalid (*Shakespearean Ciphers Examined*, 1957). Several societies and periodicals devoted to the Baconian claims were founded at the end of the 19th century, but interest later declined.

2. Other Candidates.—For several years after the appearance of his *Shakespeare Identified in Edward de Vere, the 17th Earl of Oxford* (1920), the views of J. T. Looney attracted enough adherents to lead to the formation of Oxfordian societies in London and New York. Looney argued from the biographical similarity of Bertram (in *All's Well That Ends Well*) and of Hamlet to Oxford himself, and also from the resemblance of Oxford's poems to Shakespeare's early work. Oxford's interest in the drama extended beyond noble patronage, for he himself wrote some plays, although there are no known examples extant. His 22 acknowledged poems were written in youth, and, as he was born in 1550, Looney argued that they were the prelude to his mature work, and that this began in 1593 with *Venus and Adonis*. This theory is supported by the coincidence that Oxford's poems apparently ceased just before Shakespeare's work began to appear. It is claimed that Oxford assumed a pseudonym in order to protect his family from the social stigma attached to the stage, and also because extravagance had brought him into disrepute at court. (*See* also OXFORD, EDWARD DE VERE, 17th Earl of.)

Another candidate is William Stanley, 6th earl of Derby, whose claims were advanced by A. W. Titherley (*Shakespeare's Identity: William Stanley, 6th Earl of Derby*, 1952). He was keenly interested in the drama and became the patron of his own company of actors; and his elder brother, Ferdinando (the 5th earl), was, as Lord Strange, patron of the company to which Shakespeare belonged (see *Life: The Missing Years*, above). Several poems written early in the 1580s exhibit signs of an immature Shakespearean style, yet cannot well have been written by Shakespeare. One is in Derby's handwriting, and three ("Of silver pure thy penne is made, dipte in the Muses well"; "What shall I say of Gold, more than tis Gold?"; "My thoughts are winged with hopes, my hopes with love") are signed W. S.; these initials have been thought to conceal Derby's identity, and to have been expanded later into "William Shakespeare." Derby's motive for this self-concealment would have been, like Oxford's, to avoid the association of his family name with the stage, then regarded as low in the social scale.

A theory put forward by C. Hoffman in *The Man Who Was "Shakespeare"* (1955; U.S. title, *The Murder of the Man Who Was "Shakespeare"*) identifies Shakespeare with Marlowe. Marlowe's influence on Shakespeare's early work has long been recognized, and it has even been suggested that he collaborated with Shakespeare in some of the plays, but Hoffman has discovered stylistic and linguistic similarities throughout the Shakespeare canon. He has also noticed that Shakespeare emerged from obscurity with the publication of *Venus and Adonis* in September 1593, less than four months after Marlowe was stabbed to death at Deptford. Because Marlowe's avowed "atheism" and probable homosexuality are known to have incurred religious and political disapproval Hoffman argues that when the privy council ordered Marlowe's arrest, his friend and patron Sir Thomas Walsingham (cousin of Sir Francis Walsingham) evolved an elaborate plan for his escape. Two of Walsingham's servants murdered an unknown man and alleged the body to be that of Marlowe, while Marlowe fled to France and thence to Italy. He continued to write in exile, his work being sent to Walsingham, who arranged for it to be transcribed and paid Shakespeare to "father" it. The original manuscripts were thought to be buried with Sir Thomas Walsingham at Chislehurst, Kent, but when the tomb was opened in 1956 they were not found. (*See* also MARLOWE, CHRISTOPHER.)

Sir G. G. Greenwood (*The Shakespeare Problem Restated*, 1908) and others, although not venturing to identify the "true" author, have nevertheless vigorously disputed Shakespeare's claim to the title.

3. The Case for Shakespeare.—It can hardly be maintained that no difficulty exists. In spite of recorded allusions to Shakespeare as the author of many plays in the canon, made by about 50 men during his lifetime, it is arguable that his greatness was not as clearly recognized in his own time as might be expected. But, on the other hand, the difficulties are not so great as disbelievers have held, and their proposals have raised larger problems than they have resolved. The difficulty, inasmuch as it exists, can only be viewed profitably in the light of extensive knowledge of the conditions under which plays were written, produced, and preserved, and of the prevailing attitudes toward authorship in general, in the period *c.* 1590 to *c.* 1612.

Shakespeare's origin raises no difficulties, for poets and men of genius arose in his age, as in others, from all social levels. Of his education more must be said. Stratford possessed a grammar school of better than average quality, which was free for the sons of burgesses, and this included Shakespeare (*see* above). That we can point to no documentary evidence of his attendance there would be significant only if his name were missing from existing records of attendance, but no such records are known, and it is doubtful that they ever existed. Though we have no specific knowledge of the curriculum of the Stratford school, we can infer it with some confidence from that of other schools, and from the degree of uniformity found among these. The assumption is, then, that Shakespeare spent some years in a good school, and that he

acquired (there and elsewhere) some competence in Latin, some acquaintance with Ovid, Virgil, Cicero, Horace, Livy, Quintilian, and other classical authors, and a familiarity with parts of the Bible and with the Book of Common Prayer, together with a thorough grounding in Christian ethics. As Virgil K. Whitaker has shown (*Shakespeare's Use of Learning*, 1953), the Latin authors whom we should expect a boy from a good grammar school *c.* 1575 to know are just those with whom Shakespeare shows some familiarity in his earliest plays and poems. Neither in these nor in his later works does he reveal knowledge of a large number of books, Latin or English, or a scholar's knowledge of those relatively few books more or less well known to him. Likewise, his knowledge of the law as reflected in his plays has been shown (by P. S. Clarkson and C. T. Warren, *The Law of Property in Shakespeare and the Elizabethan Drama*, 1942) to be neither so broad nor so accurate as that of "about half of Shakespeare's fellows." Similarly, his acquaintance with history was not extensive; his knowledge of English history appears to have been limited to what he found in those portions of Holinshed, Hall, and other works which he mined for the purposes of his own history plays. On his ignorance of geography (or carelessness about it), as shown by his famous reference in *Twelfth Night* to the "sea-coast of Bohemia," he was twitted by Ben Jonson. As for the accuracy of the reproduction in the plays of the speech and manners of courtiers, which has been thought to provide evidence of their author's high birth, this would be difficult to challenge in the absence of authentic records of conversation and comportment at the courts of Elizabeth I and James I: we form our impressions of how such persons talked at least in part from the plays of the time. How Shakespeare attained his unparalleled insight into human nature must remain a mystery, and it will scarcely be maintained that university education or noble birth have in general been of much service in this direction. Everyone agrees that the author of Shakespeare's plays was a man of extraordinary genius, endowed with brilliant powers of observation and an unexampled ability to synthesize what he experienced (directly or in books), and to give to it artistic shape.

The often repeated assertion that we know almost nothing about Shakespeare's life must be tested historically. We are entitled to suspect the presence of a mystery in the dearth of knowledge of Shakespeare's daily life, personality, and opinions (as compared with those of, say, William Wordsworth or Robert Browning) only if we find that he is unusual in this respect among dramatists of his day. But this is not the case. It is true that we know more about Ben Jonson, but this is clearly because, beside being a dramatist and actor, Jonson was a scholar, a literary critic, the leader of a poetic school, a writer of court and city masques, a man often involved in troubles which became matters of record: indeed, a quarrelsome and colourful public figure. With the exception of Jonson, we know more about Shakespeare than about any other dramatist of the time. Stage plays were not regarded as serious literature by the Elizabethans, and their writers were not lionized. Much of our comparative ignorance about the dramatists must be attributed to the fact that it was not an age of biography or casual letter writing, of journalism or advertising.

These facts have bearing upon the disappearance of Shakespeare's manuscripts. Once a play was in print, the manuscript possessed no value except for the acting company that owned it: by whom it would be altered and amended, roughly handled, and probably in the end, destroyed. There was no custom of collecting the autograph manuscripts of particular writers (*i.e.*, those written in their own hand). Moreover, because paper was a costly, imported commodity, manuscripts were in demand for many practical uses, all in themselves destructive. It is not surprising, therefore, that we have no manuscript in the handwriting of Shakespeare or any one of his fellow dramatists, of any successful professional public stage play printed before 1640: such manuscripts as have survived, even in the writing of copyists, are of "academic" plays; *i.e.*, plays produced in universities and schools.

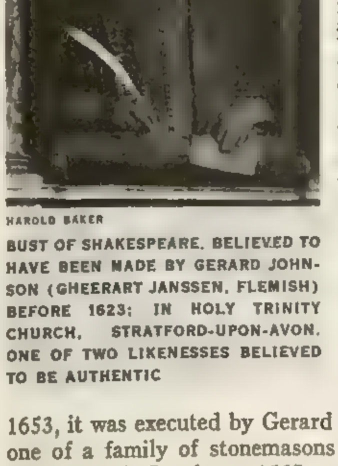
From 1598 onward, Shakespeare is named as author on the title pages of the Quarto editions of the separate plays and poems generally attributed to him, and of the Folio editions of the collected plays. Shakespeare's contemporaries wrote of him unequivocally as the author of the plays. Ben Jonson, who knew him well, contributed verses to the First Folio of 1623, where (as elsewhere) he criticizes and praises Shakespeare as the author. John Heminge and Henry Condell, fellow actors and theatre owners with Shakespeare, signed the dedication and a foreword to the First Folio, and described their methods as editors. In his own day, therefore, he was accepted as the author of the plays. Throughout his lifetime, and for 150 years after, no person is known to have questioned his authorship. In an age that loved gossip and mystery as much as any, is it conceivable that Jonson and Shakespeare's theatrical associates shared the secret of a gigantic literary hoax without a single leak, or that they could have been imposed upon without suspicion? Unsupported assertions that the author of the plays was a man of great learning and that Shakespeare of Stratford was an illiterate rustic no longer carry weight, and only when a believer in Bacon or Oxford or Marlowe produces sound evidence will scholars pay close attention to it and to him. (G. E. D.)

VI. PORTRAITS OF SHAKESPEARE

The problems surrounding the details of Shakespeare's life and work extend also to his portraiture. Only two of the many supposed representations of him can be regarded as authoritative, and neither is a portrayal direct from life. This, and the fact that they differ in some essential features, has encouraged acceptance of numerous portraits of wholly different types. The result has been a series of portraits which may be classed as: (1) genuine portraits of Shakespeare's period of persons not Shakespeare but like various conceptions of him; (2) memorial portraits based on originals believed to be genuine; (3) portraits of persons, both known and unknown, deliberately faked to resemble Shakespeare, or some conception of what he looked like; (4) complete fabrications and forgeries.

The two portraits (neither of them of great artistic merit) that can be accepted as authentic are the bust (really a half-length statue) on the wall monument in Holy Trinity Church, Stratford-upon-Avon; and the copperplate engraved by Martin Droeshout (1601–*c.* 1650), whose father, a Flemish engraver, settled in London *c.* 1590, as the frontispiece for the First Folio (1623).

1. The Stratford Bust and Monument.—These must have been erected on the north wall of the chancel between 1616, when Shakespeare died, and 1623, for, in his verses for the First Folio, Leonard Digges refers to the "Stratford *Moniment*." The monument, mainly of white marble, inlaid with black panels, consists of a central arch, in which the bust stands, flanked by two Corinthian pillars of black marble with gilt capitals and bases, supporting a cornice. The cornice holds a square block containing Shakespeare's arms, helm, and crest; above is a skull, and on each side, a cherub. One cherub represents Rest, the other Labour. Below the arch is a tablet, on which is the inscription. The bust, showing Shakespeare full face, with his hands holding pen and paper and resting on a tasselled cushion, is of soft Cotswold limestone.



HAROLD BAKER

BUST OF SHAKESPEARE. BELIEVED TO HAVE BEEN MADE BY GERARD JOHNSON (GHEERART JANSSEN, FLEMISH) BEFORE 1623: IN HOLY TRINITY CHURCH, STRATFORD-UPON-AVON. ONE OF TWO LIKENESSES BELIEVED TO BE AUTHENTIC

In design the monument is typical of the period: according to an entry by the antiquarian Sir William Dugdale in a diary for 1653, it was executed by Gerard Johnson—*i.e.*, Gheerart Janssen, one of a family of stonemasons from Amsterdam, whose father had settled in London *c.* 1567, and who had a workshop on Bankside. The bust is said to have been commissioned by John Hall,

Shakespeare's son-in-law, and he must, presumably, have approved it; Shakespeare's widow must also have seen it. It has been suggested that it was modeled from a death mask; it is certainly a portrait, not merely a conventional generalization.

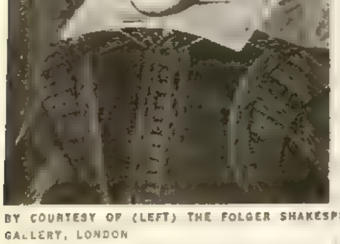
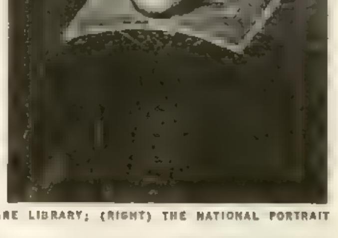
As was usual, especially in the memorial work of Flemish sculptors of the Jacobean period, the bust was originally coloured, and this probably explains the lack of detail in the carving; the detail (*e.g.*, of eyebrows, which are scarcely indicated by the chisel; and of teeth, represented by a solid surface) was supplied by the paintbrush. The high colouring, especially of the face, and of the dark hazel eyes, auburn beard and hair; scarlet doublet; black gown; and crimson and green cushion is responsible for the bust's somewhat wooden, artificial appearance.

The bust was repaired and restored in 1748, and in 1793 Edmond Malone, the Shakespearean scholar and editor, brought it back to what he thought to have been its original state by having it painted "a good stone colour." In 1861, this was removed, and the original colouring restored. There is no sign of any breakage and repair of the nose, although an apparently unusually long upper lip has been attributed to such a breakage; actually, this appearance is mainly an optical illusion, resulting from the smallness of the nose and the thinness of the moustache.

Theories that the monument and bust are not those originally set up have been brought forward to explain its many dissimilarities from the engraving of it in Dugdale's *Antiquities of Warwickshire* (1656), but many other engravings in this book are demonstrably inaccurate, and there is no reason to doubt that the bust and monument now in the church are those set up between 1616 and 1623.

2. The Droeshout Print.—The importance of the portrait by Droeshout is that it was commissioned by John Heminge and Henry Condell (Shakespeare's fellow-actors and first editors), as the frontispiece to the first collected edition of his plays (the First Folio, 1623). The volume itself was to be his real "monument," but as the whole work was regarded as a memorial undertaken by Shakespeare's friends in a spirit of pious veneration, it may be assumed that the engraving was thought to be a satisfactory likeness. Anne Shakespeare must have seen it; Ben Jonson extolled it—though his verses may be regarded as merely conventional praise.

The original from which the print was made can scarcely have been drawn from life by Droeshout, but he probably worked from a lost line drawing by an unknown artist, showing Shakespeare as a young man. It has been deduced from the strong highlights and shadows of the head, and from the fact that the body is out of scale with it, and the dress out of perspective, that this original was probably a shadowless outline drawing of the head only, and that the chiaroscuro, body, and dress were unsuccessfully added by the young and inexperienced Droeshout. This theory (first propounded by Sir George Scharf in *On The Principal Portraits of Shakespeare*, 1864) is supported by the existence of

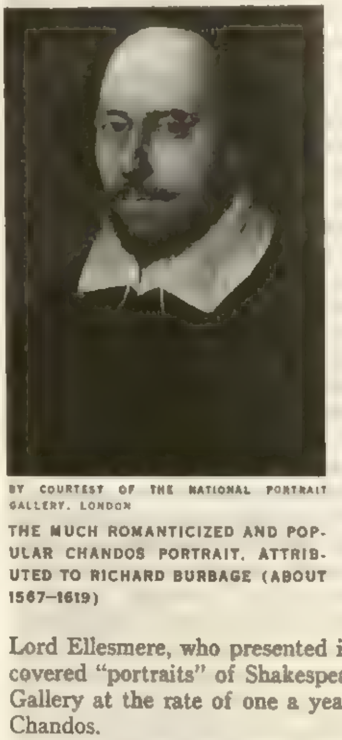
BY COURTESY OF (LEFT) THE FOLGER SHAKESPEARE LIBRARY; (RIGHT) THE NATIONAL PORTRAIT GALLERY, LONDON

(LEFT) FIRST PROOF AND (RIGHT) PRINTED VERSION OF THE ENGRAVED PORTRAIT OF SHAKESPEARE MADE BY MARTIN DROESHOUT AS THE FRONTISPIECE TO THE FIRST FOLIO, IN 1623. THE FIRST PROOF IS THOUGHT TO BE THE BETTER LIKENESS

copies of the first proof of the engraving. The first such proof to be discovered was acquired by J. O. Halliwell-Phillipps in the 1870s, and, with another, is now in the Folger Shakespeare Library, Washington, D.C. There is another in the British Museum, London, and one in the Bodleian Library, Oxford. In the proof—thought to be probably the best extant likeness of Shakespeare—the head is far more human than in the First Folio frontispiece as printed; the bone structure of the other features corresponds to that of the nose (which is longer than in the Stratford bust); and the moustache is thin and wiry.

3. Other Portraits.—The existence of the first proof of the Droeshout engraving has undermined the claim that the Flower portrait (presented by a Mrs. Charles Flower to the Shakespeare Museum, Stratford, in 1895, and now in the Stratford Memorial Theatre) is the original from which Droeshout worked: the Flower portrait resembles the second, printed, stage of the engraving, not the proof. It was probably painted from the Droeshout print, perhaps in the first half of the 17th century, and may be the earliest painted portrait of Shakespeare.

The portrait that has made the greatest appeal, because it suits popular ideas of what Shakespeare ought to look like, is the Chandos, in the National Portrait Gallery, London. The romantic, Italian-looking head differs greatly in appearance from that of the Stratford bust and the Droeshout print, and the facial measurements do not agree. According to a doubtful tradition, it was painted by Richard Burbage, who gave it to his fellow-actor Joseph Taylor, who bequeathed it to Sir William Davenant, Shakespeare's godson; but this story is strongly suspect. The esteem in which it was held within 70 years of Shakespeare's death is proved by the fact that Sir Godfrey Kneller made a copy of it to give to Dryden, acknowledged in Dryden's *Fourteenth Epistle* (1694). In Malone's time it was in the possession of the earl of Fitzwilliam, and, inherited by the duke of Buckingham and Chandos, in 1847 it was sold to Lord Ellesmere, who presented it to the nation. The newly discovered "portraits" of Shakespeare sent to the National Portrait Gallery at the rate of one a year generally tend to resemble the Chandos.



BY COURTESY OF THE NATIONAL PORTRAIT GALLERY, LONDON

THE MUCH ROMANTICIZED AND POPULAR CHANDOS PORTRAIT, ATTRIBUTED TO RICHARD BURBAGE (ABOUT 1567–1619)

After the Stratford bust, there seems to have been no statuary portrait of Shakespeare until 1740, when the statue in the Poets' Corner, Westminster Abbey, was set up by public subscription organized by the earl of Burlington and Alexander Pope, among others. Designed by William Kent, and executed by the popular Flemish sculptor, Peter Scheemakers (15 of whose monuments are in the abbey), it is based on the Chandos portrait, as is the statue by L. F. Roubillac, commissioned by David Garrick in 1758, and left by him to the British Museum. (M. H. S.; X.)

VII. SHAKESPEARE THE DRAMATIST

Shakespeare's powers as a dramatist have been demonstrated beyond all doubt by the success of his plays in performance. His plays are performed throughout the world; his birthday is honoured by many nations; and his quatercentenary in 1964 was the occasion of worldwide celebrations. He is the one universal playwright, accepted in countries with diverse ideologies, religions, and artistic conventions and standards: in Canada, Ceylon, China, and Czechoslovakia; in India, Israel, Italy, and Japan; in the U.S.S.R. and the United States.

Shakespeare did not achieve this universal appeal by promulgating ideas. *Coriolanus* and *Julius Caesar* may sometimes have

given political offense, but to both sides of a political debate. The history plays may be seen as supporting a kind of monarchical government now found only in very few countries, and as giving more importance to "providence" in human affairs than to other factors—economic, ideological, or racial—on which many have placed more importance. But ideas are not dominant, nor is "meaning" unequivocal. Shakespeare's plays appeal to thinking men, but do not obviously direct their thoughts; in this they differ from the works of, for example, Aeschylus, Molière, Ibsen, Shaw, and many other dramatists.

The most obvious and general appeal of the plays derives from Shakespeare's power to create intrinsically interesting situations and plots. These enable a wide range of characters to be opposed or contrasted; and the action displays their antagonisms and attractions so that a full gamut of reactions is expressed, and the deeper instincts of human nature exposed. There are opportunities for sober argument and fantastic wit, for dreamlike and for waking reality. Moreover, the action is so displayed that it holds attention. One proof of this power of plot and situation lies in the many adaptations made from Shakespeare's plays into other art forms: into prose narratives, operas, musical dramas, into ballets without words, and radio plays or recorded readings without visual effects, into films and television dramas. The situations, plots, and characters seem endlessly interesting. They have yielded new plays too: innumerable rewritings of *Romeo and Juliet;* Bertolt Brecht's version of *Richard III* in *Arturo Ui;* thinly-disguised borrowings from *Hamlet* in Anton Chekhov's *The Seagull.* Sometimes they have encouraged formal experiment, as in W. H. Auden's *The Sea and the Mirror,* a poetic "commentary on Shakespeare's *The Tempest,*" written as spoken by the various *personae* from the drama. With the immediate and localized interests of situation and plot larger issues are involved, so that Shakespeare's plays can be described in archetypal or symbolistic terms: it can be said that the trial scene in *The Merchant of Venice* opposes mercy and justice; that *Hamlet* places an individual against society, a son against father and mother, a moment against destiny; that *A Midsummer Night's Dream* juxtaposes innocence and experience, elemental beings and mortals, imagination and reason. Shakespeare chose basic stories of man's search for assurance, or knowledge, or integrity; of purgation, reconciliation, communion.

The intrinsic interest of the action of Shakespeare's plays is most readily discovered in their diverse *dramatis personae,* some of whom have become proverbial—as, for instance, a Shylock, a Hamlet, a Falstaff. But the settings in which the figures appear are also arresting. The Elizabethan theatre had no stage scenery in the modern manner (*see* THEATRES [STRUCTURES]) but by changes of costume, by groupings, behaviour, and verbal implication, Shakespeare ensured that in performance his plays would realize a series of particular backgrounds. The movements from Venice to Cyprus, and from public places to private, in *Othello;* from castles to heath, seashore, and battlefield in *Lear;* from the court of Sicilia to the countryside of Bohemia in *The Winter's Tale;* or, in *Hamlet,* from the battlements to the interiors of the castle at Elsinore, to the open country, to the graveyard, and back again to the court, are all large dramatic devices with an appeal that is instinctive and powerful because it relies on sense-impressions rather than intellectual or psychological perceptions. The Elizabethans called their plays "shows" for good reason: at their best, in Shakespeare's works, they not only spoke, but also demonstrated or displayed; and they presented not merely individual human beings in action but large views as well. Proof of this achievement lies in the way in which Shakespeare's plays have appealed to those 19th- and 20th-century theatrical managers who have sought to make "spectacles" of the poetic drama: entrances and exits are often elaborate and potentially meaningful; intimate scenes and soliloquies alternate with crowd scenes involving general movement; and casual, informal, scenes with ceremonies and rituals. Such producers as Henry Kemble, Charles Kean, Max Reinhardt (*q.v.*), Sir Tyrone Guthrie, and Peter Brook have amply demonstrated the power of these devices; so that sometimes in their productions such elements seem to overshadow others. But even in the barest or most inexperienced production the sweeping effects of movement and setting speak for Shakespeare.

The radio and phonograph have demonstrated the opposite pole of the plays' attractions—the astonishing vitality of the words. Shakespeare's verse form gives continuity and control of tempo. Allied to his use of rhetorical devices, it can sustain a moment, hold the drama still in verbal and almost musical elaboration so that its implications and power may grow in the minds of the audience. In verse and prose, Shakespeare weighted single syllables, or set one rhythm against another. And with words of complex meaning and images of wide or precise associations he displayed thematic, psychological, or "atmospheric" subtleties. He mixed formal and deliberate speech with the quick and colloquial, and so familiarized the remote and lent power and significance to the familiar. In the 20th century the astonishing "poetic texture" of the dialogue has often drawn attention from other aspects of Shakespeare's dramaturgy, so that the silent study or the classroom has sometimes seemed the proper place for admiring his works.

Perhaps the most astonishing fact in considering Shakespeare's dramatic art is that a writer who has so satisfied the subtle literary critics has also satisfied the most realist actors, and often with the same element of his art, the dialogue. For his "lines" are eminently actable; in general, they can both be spoken as poetry and also serve as one element in a full realization of a role in performance. Verbal complexity is used to show conscious and subconscious thought at one and the same time; so, clearly, in *Twelfth Night,* Viola speaks for herself as well as for Cesario (the name she has assumed when disguised as a man, and servant of the duke, Orsino, with whom she is in love) when she talks to Orsino of the sister who died for love (Act II, sc. 4); more subtly, Macbeth reveals his own sense of guilt in asking if the murderers are "so gospelled" that they will kill Banquo (Act III, sc. 1). The actor can present two levels of consciousness: Hamlet, in talking (Act II, sc. 2) of the Tragedians of the City who are coming to Elsinore, may also voice his own deeper desire when he says that the "lover may not sigh gratis," that "the humorous man shall end his part in peace . . . and the lady shall say her mind freely . . ."; and in his repeating of the words "mobled queen" from the player's speech describing the death of Priam and the anguish of his wife Hecuba, the words are revalued, so that Hecuba may become his mother Gertrude to the audience, as she is to Hamlet in his deepest thoughts. Such opportunities are numerous; if all were taken a performance would be overfilled, and the imagination of the actor overstretched. So it is that Shakespeare's major roles can be interpreted in many ways; there is such a wealth of opportunity that each actor may select what best suits his physique and inspiration. Malvolio (in *Twelfth Night*) may be comic, or dignified, or pathetic; Hamlet philosophical or active, neurotic or resolute. Even a comparatively small-scale character like Olivia in *Twelfth Night* can be interpreted as the conventional "widow" that a member of an early audience saw, as the elegant, sad lady of numerous productions, as an impetuous minx, or as a rather frightened girl; and such a full-scale character as Shylock can satisfy the widely differing powers of a Macklin, a Kean, or an Irving, or in our own time, of a Donald Wolfit, an Alec Guinness, a John Gielgud, or a Peter O'Toole.

The scope of a role does not depend on words alone, for there are moments for which Shakespeare has chosen gesture or mere physical presence as the primary means of impressing character-in-action. The most famous example is the stage direction in *Coriolanus* (Act. V, sc. 3), *"He holds her by the hand, silent,"* indicating the simple movement that answers the power of Volumnia's many words—and, indeed, her physical presence. When Macbeth first sees the witches (Act I, sc. 3) his reaction is silence, and it is Banquo who speaks to them; and, again, when they have spoken, he is silent, and continues so after Banquo (with the words "Good sir, why do you start and seem to fear/Things that do sound so fair?") has identified the physical response that dominates the stage as one of fear. At the end of the tragedy, in the penultimate scene, Macbeth's last contribution is the wordless fight with Macduff (indicated by the stage direction "Exeunt

fighting . . . re-enter fighting and exeunt. Macbeth exit falling"); and, beyond this, the final impression he makes comes at the very end of the play, when his "cursed head"—a painted stage-property—is held aloft by Macduff. Shakespeare "realized" the characters of his dramas at several levels of consciousness, in their physical and instinctive as well as in their mental or emotional reactions, and in the effect of their mere presence or image on the other characters, and audience. The progress of the play's action reveals character with progressive clarity and penetration; one of the major structural devices is a progressive revelation of psychological truth, a casting away of masks, a growing perception—for the audience if not for the character himself—of basic and seldom-named desires.

In many ways Shakespeare might seem limited by the primitive nature of his instrument. To Elizabethans, the wooden "O" (the circular auditorium and central, uncurtained stage) had to stand for the fields of France, and a boy actor for Juliet. But time has proved that he was not so restricted. Obviously, he turned to good account the speed and chameleon qualities of his stage, and the agility and clarity of his boy actors, but large scenic resources and the full sensibility and power of a great actress do not cause his plays to seem crippled; they extend themselves easily into wider areas of realization. The secret of this is perhaps the final secret of Shakespeare's art as a dramatist, the element we can readily identify, but only with greatest difficulty analyze. His plays are not confined by his theatrical instruments, nor by the thought and manners of his age, nor even by a particular form of dramatic construction, because while mastering the techniques of his art Shakespeare seems to have been constantly concerned to reflect the world in which he lived, to make the variously and richly attractive characters-in-action a "mirror held up to nature." Johnson put this magisterially:

> Shakespeare's plays are not in the rigorous and critical sense either tragedies or comedies, but compositions of a distinct kind; exhibiting the real state of sublunary nature, which partakes of good and evil, joy and sorrow, mingled with endless variety of proportion and innumerable modes of combination. (Preface to Shakespeare's *Works*, 1765.)

See also references under "Shakespeare, William," in the Index.

(J. R. Bn.)

BIBLIOGRAPHY.—Many important books have been referred to in the text; *see* especially the sections *Literary Criticism* and *Textual Criticism*, the latter containing lists of editions and facsimiles. For historical and other background studies, *see* bibliographies to the articles ENGLISH HISTORY; ENGLISH LITERATURE; DRAMA; THEATRE.

Bibliographies, Reference Books, etc.: W. Jaggard, *Shakespeare Bibliography* (1911; rev. ed. 1965); W. Ebisch and L. L. Schücking, *A Shakespeare Bibliography* (1931), *Supplement . . . 1930–35* (1937); G. R. Smith, *A Classified Shakespeare Bibliography, 1936–1958*, 4 vol. (1963–); W. W. Greg, *A Bibliography of the English Printed Drama to the Restoration*, 4 vol. (1939–59). *See* also the *Shakespeare-Jahrbuch* (1865–); *Shakespeare Quarterly* (1950–); *PMLA* (1884–); *The Year's Work in English Studies* (1920–); bibliographical supplements to *British Book News*, 58, 129, 138, 140, 143, 146 (1955 *et seq.*); *Shakespeare Survey* (1948–). On early editions, etc. *see* H. C. Bartlett, *Mr. William Shakespeare, Original and Early Editions of His Quartos and Folios, His Source Books and Those Containing Contemporary Notices* (1922); H. C. Bartlett and A. W. Pollard (eds.), *A Census of Shakespeare's Plays in Quarto, 1594–1709* (1916; rev. ed. 1939). Useful handbooks are H. Granville-Barker and G. B. Harrison (eds.), *A Companion to Shakespeare Studies* (1934); F. E. Halliday, *A Shakespeare Companion, 1550–1950* (1952; revised 1964); L. Marder, *His Exits and His Entrances: the Story of Shakespeare's Reputation* (1963).

Editions: As well as those listed under *Textual Criticism* above, the following are useful: The Stratford Town Edition, by A. H. Bullen, 10 vol. (1907); the Caxton Edition, by Sir Sidney Lee, 20 vol. (1910); the Tudor Edition, by W. A. Neilson and A. H. Thorndike, 39 vol. (1911–13); the Nonesuch Edition, by H. Farjeon, 7 vol. (1929–33; reprinted in 4 vol., with new introduction by Ivor Brown, 1953); the New Temple Shakespeare, ed. by M. R. Ridley, 40 vol. (1934–36); *The London Shakespeare*, ed. by J. J. Munro (1957). One-volume editions include *The Players' Shakespeare*, ed. by H. Granville-Barker (1923); and the editions by G. L. Kittredge (1936), and C. J. Sisson (1955). There are editions of the poems by G. Wyndham (1898), and H. E. Rollins (1938; the New Variorum edition); of the poems and songs by Edward Hubler (1959); and of the sonnets by E. Dowden (1881), R. M. Alden (1916), T. G. Tucker (1924), and F. T. Prince (1960; Arden). Ascribed plays are in *The Shakespeare Apocrypha* (1908), ed. by C. F. T. Brooke. *Sir Thomas More* was ed. by F. J. Shirley (1938) and by H. Jenkins (1954); *Arden of Feversham* by H. Macdonald, for the Malone Society (1947).

Life, Personality, Interests, etc.: General: The standard scholarly life, including documents and bibliographical introductions to each chapter, is still Sir E. K. Chambers, *William Shakespeare: a Study of Facts and Problems*, 2 vol. (1930), supplemented by his *Shakespearean Gleanings* (1944), and *Sources for a Biography of Shakespeare* (1946). Earlier lives still of value are J. O. Halliwell-Phillipps, *Outlines of the Life of Shakespeare* (1881; 7th rev. ed. 1887); Sir S. Lee, *A Life of William Shakespeare* (1898; 14th ed. 1929); Sir Walter Raleigh, *Shakespeare*, "English Men of Letters" Series (1907). J. Q. Adams, *A Life of William Shakespeare* (1923); C. F. T. Brooke, *Shakespeare of Stratford* (1926); L. Hotson, *Shakespeare Versus Shallow* (1931), *I, William Shakespeare, do Appoint Thomas Russell, Esquire, . . .* (1937); J. Dover Wilson, *The Essential Shakespeare* (1932); G. B. Harrison, *Shakespeare at Work, 1592–1603* (1933); E. I. Fripp, *Shakespeare, Man and Artist*, 2 vol. (1938; new ed. 1964); P. Alexander, *Shakespeare's Life and Art* (1939), *Shakespeare* (1964); Ivor Brown, *Shakespeare* (1949), *How Shakespeare Spent the Day* (1964); M. M. Reese, *Shakespeare: his World and his Work* (1953); F. E. Halliday, *The Life of Shakespeare* (1961); G. E. Bentley, *Shakespeare: a Biographical Handbook* (1961); A. L. Rowse, *William Shakespeare* (1963); P. Quennell, *Shakespeare: the Poet and his Background* (1963).

Special Aspects: A. Acheson, *Shakespeare's Lost Years in London, 1586–1592* (1920); M. H. Spielmann, "Shakespeare's Portraiture," in *Studies in the First Folio*, ed. for the Shakespeare Association by Sir I. Gollancz (1924); G. B. Harrison, *Willobie His Avisa* (1926); F. Marcham (ed.), *William Shakespeare and His Daughter Susannah* (1931); R. S. H. Noble, *Shakespeare's Biblical Knowledge and Use of the Book of Common Prayer* (1935); K. Muir and S. O'Loughlin, *The Voyage to Illyria* (1937); T. W. Baldwin, *William Shakespere's Petty School* (1943); C. M. Mitchell, *The Shakespeare Circle: a Life of Dr. John Hall, Shakespeare's Son-in-Law* (1947); Sir A. D. Cooper, *Sergeant Shakespeare* (1949); C. W. Scott-Giles, *Shakespeare's Heraldry* (1950).

Background Studies: Sir W. Raleigh, Sir S. Lee, and C. T. Onions (eds.), *Shakespeare's England* (1916); J. Dover Wilson (ed.), *Life in Shakespeare's England* (1911); M. St. C. Byrne, *Elizabethan Life in Town and Country* (1926; 8th rev. ed. 1961); Sir E. K. Chambers, *The Elizabethan Stage*, 4 vol. (1923); W. W. Greg, *Shakespeare and the Theatre* (1927), (ed.), *Dramatic Documents from the Elizabethan Playhouses*, 2 vol. (1931); W. R. Davies, *Shakespeare's Boy Actors* (1939); J. C. Adams, *The Globe Playhouse* (1942; 2nd ed. 1961); T. Spencer, *Shakespeare and the Nature of Man* (1942); B. L. Joseph, *Elizabethan Acting* (1951), *Acting Shakespeare* (1960); I. Smith, *Shakespeare's Globe Playhouse* (1956); L. Hotson, *Shakespeare's Wooden O* (1959); I. Brown, *Shakespeare in his Time* (1960); M. Holmes, *Shakespeare's Public* (1960); R. A. Foakes and R. T. Rickert (eds.), *Henslowe's Diary* (1961); B. Beckerman, *Shakespeare at the Globe, 1599–1609* (1962); M. C. Bradbrook, *The Rise of the Common Player* (1962); A. Nicoll (ed.), *Shakespeare in His Own Age* (*Shakespeare Survey*, vol. xvii, 1964).

Shakespeare's Sources: T. Spencer (ed.), *Shakespeare's Plutarch* (1964); M. W. MacCallum, *Shakespeare's Roman Plays and Their Background* (1910); G. C. Taylor, *Shakespeare's Debt to Montaigne* (1925); S. Guttman, *The Foreign Sources of Shakespeare's Work* (1947); K. Muir, *Shakespeare's Sources*, vol. i (1957); G. Bullough (ed.), *Narrative and Dramatic Sources of Shakespeare*, 7 vol. (1957–); C. G. Smith, *Shakespeare's Proverb Lore* (1963).

Textual Studies: A. W. Pollard, *Shakespeare Folios and Quartos . . . 1594–1685* (1909), *Shakespeare's Fight with the Pirates and the Problems of the Transmission of His Text* (1917), *The Foundations of Shakespeare's Texts* (British Academy lecture, 1923); W. W. Greg, *The Editorial Problem in Shakespeare* (1942; rev. ed. 1951, 1955), A. Walker, *Textual Problems of the First Folio* (1953); L. Kirschbaum, *Shakespeare and the Stationers* (1955); J. Dover Wilson, *The New Way with Shakespeare's Texts*, in *Shakespeare Survey*, vol. vii, viii ix (1954–56); H. Craig, *A New Look at Shakespeare's Quartos* (1961).

Criticism and Interpretation: History: A. J. Ralli, *A History of Shakespearian Criticism*, 2 vol. (1932); D. N. Smith (ed.), *Shakespeare Criticism* (1916); A. Bradby (ed.), *Shakespeare Criticism, 1919–1935* (1936); F. E. Halliday, *Shakespeare and His Critics* (1949), *The Cult of Shakespeare* (1957); B. M. Wagner (ed.), *The Appreciation of Shakespeare: a Collection of Criticism . . . 18th, 19th, and 20th Centuries* (1949); L. D. Lerner (ed.), *Shakespeare's Tragedies: an Anthology of Modern Criticism* (1964); P. Alexander (ed.), *Studies in Shakespeare* (1964).

General Studies: Sir A. T. Quiller-Couch, *Shakespeare's Workmanship* (1918); Sir E. K. Chambers, *Shakespeare: a Survey* (1926; new ed. 1964); G. Wilson Knight, *The Imperial Theme* (1931; 3rd ed. 1951), *The Shakespearian Tempest* (1932); M. C. Bradbrook, *Themes and Conventions of Elizabethan Tragedy* (1935); W. E. Farnham, *The Medieval Heritage of Elizabethan Tragedy* (1936), *Shakespeare's Tragic Frontier* (1950); J. Middleton Murry, *Shakespeare* (1936); H. B. Charlton, *Shakespearian Comedy* (1938), *Shakespearian Tragedy* (1948); D. A. Traversi, *Approach to Shakespeare* (1938); M. van Doren, *Shakespeare* (1939); S. L. Bethell, *Shakespeare and the Popular Dramatic Tradition* (1944); J. L. Palmer, *Political Characters of Shakespeare* (1945), *Comic Characters of Shakespeare* (1946); E. A.

Armstrong, *Shakespeare's Imagination* (1946); E. C. Pettet, *Shakespeare and the Romance Tradition* (1950); C. Leech, *Shakespeare's Tragedies, and Other Studies in 17th Century Drama* (1950); G. I. Duthie, *Shakespeare* (1951); E. T. Sehrt, *Vergebung und Gnade bei Shakespeare* (1952); H. Fluchère, *Shakespeare*, Eng. trans. (1953); P. Cruttwell, *The Shakespearean Moment and its Place in the Poetry of the 17th Century* (1954); M. H. Parker, *The Slave of Life* (1955); P. N. Siegel, *Shakespearean Tragedy and the Elizabethan Compromise* (1957); H. S. Wilson, *On the Design of Shakespearian Tragedy* (1957); J. R. Brown, *Shakespeare and His Comedies* (1957); M. M. Mahood, *Shakespeare's Wordplay* (1957); L. C. Knights, *Shakespeare's Politics* (1958), *Some Shakespearean Themes* (1959); A. B. Kernan, *The Cankered Muse* (1959); J. R. Brown and B. Harris (eds.), "Stratford-upon-Avon Studies": I *Jacobean Theatre* (1960); B. Evans, *Shakespeare's Comedies* (1960); J. Lawlor, *The Tragic Sense in Shakespeare* (1960); K. Muir, *Shakespeare as Collaborator* (1960); I. Ribner, *Patterns in Shakespearian Tragedy* (1960); W. Rosen, *Shakespeare and the Craft of Tragedy* (1960); U. M. Ellis-Fermor, *Shakespeare the Dramatist, and Other Papers*, ed. by K. Muir (1961); A. P. Rossiter, *Angel with Horns, and Other Shakespeare Lectures*, ed. by G. Storey (1961); H. M. V. Matthews, *Character and Symbol in Shakespeare's Plays* (1962); A. Righter, *Shakespeare and the Idea of the Play* (1962); C. J. Sisson, *Shakespeare's Tragic Justice* (1962); R. M. Frye, *Shakespeare and Christian Doctrine* (1963); N. Frye, *A Natural Perspective: the Development of Shakespearean Comedy and Romance* (1965).

On Particular Plays and Groups of Plays: The Early Comedies: F. A. Yates, *A Study of Love's Labour's Lost* (1936); L. Hotson, *The First Night of Twelfth Night* (1954); D. A. Traversi, *William Shakespeare: the Early Comedies* (1960); J. R. Brown and B. Harris (eds.), "Stratford-upon-Avon Studies": III *Early Shakespeare* (1961); J. Dover Wilson, *Shakespeare's Happy Comedies* (1962); E. W. Talbert, *Elizabethan Drama and Shakespeare's Early Plays* (1963).

The Histories: A. C. Bradley, *Oxford Lectures on Poetry* (1909); J. Dover Wilson, *The Fortunes of Falstaff* (1943); E. M. W. Tillyard, *Shakespeare's History Plays* (1944); W. Clemen, *Kommentar zu Shakespeares Richard III* (1957); I. Ribner, *The English History Play in the Age of Shakespeare* (1957); D. A. Traversi, *Shakespeare from Richard II to Henry V* (1957); M. M. Reese, *The Cease of Majesty* (1961); C. Leech, *William Shakespeare: the Chronicles* (1962); S. C. Sen Gupta, *Shakespeare's Historical Plays* (1964).

The Problem Plays: W. W. Lawrence, *Shakespeare's Problem Comedies* (1931); O. J. Campbell, *Comicall Satyre and Shakespeare's Troilus and Cressida* (1938); E. M. W. Tillyard, *Shakespeare's Problem Plays* (1949); M. M. Lascelles, *Shakespeare's Measure for Measure* (1953); P. Ure, *Shakespeare, the Problem Plays* (1961); E. Schanzer, *The Problem Plays of Shakespeare* (1963).

The Roman Plays: A. C. Bradley, *Coriolanus* (British Academy lecture, 1912); M. M. Charney, *Shakespeare's Roman Plays* (1961).

The Main Tragedies: J. Dover Wilson, *What Happens in Hamlet* (1935; 3rd ed. 1951); W. C. Curry, *Shakespeare's Philosophical Patterns* (1937); C. Brooks, *The Well Wrought Urn* (1947); R. B. Heilman, *This Great Stage* (1948), *Magic in the Web: Action and Language in Othello* (1956); J. F. Danby, *Shakespeare's Doctrine of Nature: a Study of King Lear* (1949); P. Alexander, *Hamlet: Father and Son* (1955); J. Bayley, *The Characters of Love* (1960); L. C. Knights, *An Approach to "Hamlet"* (1960); K. Muir, *William Shakespeare: the Great Tragedies* (1961).

The Last Plays, or Romances: G. L. Strachey, *Books and Characters, French and English* (1922); E. M. W. Tillyard, *Shakespeare's Last Plays* (1938; new ed. 1951); S. L. Bethell, *The Winter's Tale: a Study* (1947); F. R. Leavis, *The Common Pursuit* (1952); D. A. Traversi, *Shakespeare: the Last Phase* (1954); G. K. Hunter, *William Shakespeare, the Late Comedies* (1962).

On the Sonnets and Poems: L. Hotson, *Shakespeare's Sonnets Dated, and Other Essays* (1949); G. Wilson Knight, *The Mutual Flame* (1955); J. W. Lever, *The Elizabethan Love Sonnet* (1956); J. B. Leishman, *Themes and Variations in Shakespeare's Sonnets* (1961); J. Dover Wilson, *An Introduction to Shakespeare's Sonnets for the Use of Historians and Others* (1964); M. Krieger, *A Window to Criticism: Shakespeare's Sonnets and Modern Poetics* (1964); H. J. Landry, *Interpretations in Shakespeare's Sonnets* (1963). (G. A. O.; X.)

SHAKHTY (formerly ALEXSANDROV-GRUSHEVSKI), a town in Rostov Oblast' of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the upper Grushevka (a small stream flowing by way of the Tuzlov into the Don), 40 mi. (64 km.) NE of Rostov-na-Donu (Rostov-on-Don). Pop. (1959) 196,190. As its name ("Pits") implies, the town is engaged in coal mining. It is the main centre of the eastern end of the Donets Basin (Donbass), with about 30 pits in the immediate vicinity of the town. The coal is of high quality and anthracite forms about 85% of the production. Coking coal is also found. Other industries include iron casting, canning, brewing, and the manufacture of footwear and clothing. There is a coal-burning power station. In its southern suburb of Kamenolomni there are stone quarries and metal works. The main lines of communication for Shakhty are the railway and highway from Rostov to the Donbass. Shakhty has a pedagogic institute and technical schools for coal mining. (R. A. F.)

SHAKTISM (ŚAKTISM), an ensemble of Indian philosophical systems and religious cults which consider the divinity as Supreme Power or Energy, personified as the Mother Goddess. Together with Vishnuism and Shivaism (*qq.v.*), it is one of the main forms of modern Hinduism, especially popular in northeast India, particularly Bengal and Assam. Shakti (Śakti) is conceived either as the sovereign deity or as the consort of a male deity, generally Shiva. Indian speculation regards the male principle as quiescent (transcendence being understood as repose) and the female principle as active and creative; both are complementary aspects of the ultimate monism. This conception of Shiva and Shakti further coalesced with the Sankhya idea of *purusha-prakriti* and with the popular Vedantic idea of Brahma-*maya* (for these, *see* INDIAN PHILOSOPHY: *Sankhya* and *Vedanta*).

Spiritually minded men worship Shakti as the Divine Will, the Divine Mother who calls for absolute surrender. Yogis consider this Shakti as the power, lying dormant within the body as a coiled serpent, which must be aroused and realized to reach liberation. Shaktism is inseparably related to Tantrism (*q.v.*), a system of practices for the purification of both mind and body. In popular worship the goddess has many names, standing for the forms in which she has appeared in the world. Thus she is worshiped as Durga, with both terrifying and gracious aspects; as the four-armed destroyer of Kali; as the gracious Lakshmi and Sarasvati, consorts of Vishnu; as Shashthi, protector of children; and as Shitala, the personification of smallpox. *See* also HINDUISM: *Gods*.

See Sri Aurobindo, *The Mother* (1928, often reprinted); Sir John Woodroffe, *Shakti and Shakta* (1951). (S. B. D. G.)

SHALE is a fine-grained, earthy sedimentary rock that is a somewhat indurated clay and is characterized by a thinly layered or laminated structure, by which it is differentiated from clay. It is usually harder than clay and has less tendency to slake in water. It is usually gray, yellow, green, or red.

Shales are sediments that accumulated in past geologic time in lakes, lagoons, oceans, etc. Their properties and composition depend upon the character of those sediments; the uniformity of the source material from day to day and from year to year; variations within the area of accumulation; and changes that took place during and after its deposition in the area of accumulation (*see* SEDIMENTARY ROCKS).

Some shales contain components other than clay minerals (*see* CLAY AND CLAY MINERALS), and varietal names are given on this basis; *e.g.*, sandy shale, limy shale, bituminous shale, etc. Shales find extended use in the manufacture of structural clay products, such as brick, tile, and portland cement. *See* also FUELS: *Carbonisation;* and references under "Shale" in the Index. (R. E. GM.)

SHALE, OIL: *see* FUELS: *Carbonization.*

SHALLOT, a hardy onionlike perennial (*Allium ascalonicum*), of the lily family (Liliaceae), with small, elongated, somewhat angular bulbs that develop in clusters on a common base. The leaves are short, small, cylindrical, and hollow. The flowers of the compact umbel are lilac or reddish in colour. The shallot is probably of Asiatic origin. The common so-called shallot that is marketed extensively as green spring onions is in fact a form of the onion. (V. R. B.)

SHALMANESER (Assyrian SHULMANU-ASHARIDU, "the god Shulman is chief"), the name of five Assyrian kings. Of the reigns of Shalmaneser II (1030–19 or, perhaps 1019–08 B.C.) and Shalmaneser IV (782–772 B.C.) little is known.

SHALMANESER I, son of Adad-nirari I, ruled 1275–45 B.C. While the Hittites warred with Egypt, the Assyrians invaded Cappadocia (in eastern Asia Minor) and founded an Assyrian colony at Luha. By the defeat of Shattuara of Hani and his Hittite allies and by raids on Carchemish, Irridi, and the Kashiari Hills (all in the region around the borders of northern Iraq, northeastern Syria, and southeastern Turkey), Shalmaneser reopened the main northwestern trade routes. At his capital, Ashur, he built a palace and restored one of the temples. His buildings at Nineveh and at Kalakh (or Kalhu; modern Nimrud), the latter of which he

founded, have been partially recovered.

SHALMANESER III, son of Ashurnasirpal II, reigned 859–824 B.C. His vigorous policy of military expansion was opposed on all frontiers. Babylonia was quiet except for a rebellion by the usurper Marduk-bel-usate which was quashed in 851 B.C. On the eastern frontiers (near the borders of northeastern Iraq and northwestern Iran), the rulers of Musasir, Namri, Zamua, and Kirhi were made vassals and Parsua was raided. Although Shalmaneser twice led his army as far as Lakes Van and Urmia and campaigned in Tabal and Cappadocia (836 B.C.), Sardur (Sarduris) I of Urartu remained undefeated. The main military effort was devoted to the conquest of North Syria. Ahuni of Bit-Adini was captured only after four campaigns, and progress farther west was barred by a coalition led by Irhuleni of Hamath and Bar-Hadad I (Hebrew, Ben-Hadad; throne name, Hadadezer; Akkadian, Adad-idri) of Damascus. In 853 B.C. Shalmaneser fought these allies, whose forces numbered 63,000 infantry, 2,000 cavalry, 4,000 chariots, and 1,000 camels. Of these, 10,000 men and half the chariots were provided by Ahab (Ahabbu) of Israel according to the first reference to Israelites in the Assyrian annals. This battle, at Karkar, was indecisive and, apart from two skirmishes, Shalmaneser did not appear in the west until the coalition had broken up.

In 841 B.C. he defeated Hazael and, failing to capture Damascus itself, marched to the Mediterranean coast where he received tribute from Tyre, Sidon, and Samaria. The submission of the latter is shown on the "Black Obelisk" (from Nimrud, now in the British Museum) where "Jehu, son of Omri" bows before Shalmaneser. By 832 Cilicia had been invaded, Tarsus captured, and the region made an Assyrian dependency. The remaining campaigns of Shalmaneser's reign were led by the army commander against Sardur I and the Mannai. Before the king died in 824 B.C. civil war broke out between a son, Ashur-danin-apal, and his heir, Shamshi-Adad V. Shalmaneser rebuilt a palace and ziggurat at Nimrud. His wars were commemorated both on palace reliefs there and on the gates of the temple at Balawat.

SHALMANESER V reigned from 727 to 722 B.C. None of his historical records survive, but the King-List of Babylon, where he ruled as Ululai, links him with Tiglath-pileser III, whose son he may have been. When Hoshea of Israel rebelled (II Kings 17), he marched via Bit-Adini to besiege Samaria and attack Tyre. For three years he laid siege until "he broke the resistance of Shamara'in" (Samaria). However, he died in 722 B.C. before the capture of the city, which was claimed by his successor, Sargon II.

See also BABYLONIA AND ASSYRIA: *History*.

See *Cambridge Ancient History*, vol. ii (1926) and vol. iii (1925); D. Winton Thomas (ed.), *Documents from Old Testament Times* (1958). (D. J. WI.)

SHAMANISM is a religious phenomenon characteristic of Siberian and Ural-Altaic peoples. The word shaman itself is of Tungus origin (*saman*) and it has passed, by way of Russian, into European scientific terminology. But shamanism, although its most complete expression is found in the arctic and central Asian regions, must not be considered as limited to those countries. It is encountered, for example, in southeast Asia, Oceania and among many North American aboriginal tribes. A distinction is to be made, however, between the religions dominated by a shamanistic ideology and by shamanistic techniques (as is the case with Siberian and Indonesian religions) and those in which shamanism constitutes rather a secondary phenomenon.

The shaman is medicine man, priest and psychopompos; that is to say, he cures sicknesses, he directs the communal sacrifices and he escorts the souls of the dead to the other world. He is able to do all this by virtue of his techniques of ecstasy; *i.e.*, by his power to leave his body at will. In Siberia and in northeast Asia a person becomes shaman by hereditary transmission of the shamanistic profession or by spontaneous vocation or "election." More rarely a person can become shaman by his own decision or upon request of the clan, but the self-made shamans are regarded as weaker than those who inherit the profession or who are "elected" by the supernatural beings. In North America, on the other hand, the voluntary "quest" for the powers constitutes the principal method. No matter how the selection takes place, a shaman is recognized as such only after a series of initiatory trials after receiving instruction from qualified masters.

Asia.—In Asia as a rule the trials take place during an indefinite period of time in which the future shaman is sick and stays in his tent or wanders in the wilderness, behaving in such an eccentric way that it could be mistaken for madness. Several authors went so far as to explain arctic and Siberian shamanism as the ritualized expression of a psycho-mental disease, especially of arctic hysteria. But the "chosen" one becomes shaman only if he can interpret his pathological crisis as a religious experience and succeeds in curing himself. The serious crises that sometimes accompany the "election" of the future shaman are to be regarded as initiatory trials (*see* PASSAGE RITES). Every initiation involves the symbolic death and resurrection of the neophyte. In the dreams and hallucinations of the future shaman may be found the classical pattern of the initiation: he is tortured by demons, his body is cut in pieces, he descends to the netherworld or ascends to heaven and is finally resuscitated. That is to say, he acquires a new mode of being which allows him to have relations with the supernatural worlds. The shaman is now enabled to "see" the spirits, and he himself behaves like a spirit; he is able to leave his body and to travel in ecstasy in all cosmic regions. However, the ecstatic experience alone is not sufficient to make a shaman. The neophyte must be instructed by masters in the religious traditions of the tribe, and he is taught to recognize the various diseases and to cure them.

BY COURTESY OF THE ARCHIV DES MUSEUMS FÜR VÖLKERKUNDE, HAMBURG

SIBERIAN SHAMAN, EARLY 20TH CENTURY

Among certain Siberian peoples the consecration of the shaman is a public event. Among the Buryats, for example, the neophyte climbs a birch, a symbol of the world tree, and in doing this he is thought to ascend to heaven. The ascension to heaven is one of the specific characteristics of Siberian and central Asian shamanism. At the occasion of the horse sacrifice, the Altaic shaman ascends to heaven in ecstasy in order to offer to the celestial god the soul of the sacrificed horse. He realizes this ascension by climbing the birch trunk, which has nine notches, each symbolizing a specific heaven.

The most important function of the shaman is healing. Since sickness is thought of as a loss of the soul, the shaman has to find out first whether the soul of the sick man has strayed off far from the village or has been stolen by demons and is imprisoned in the other world. In the former case the healing is not too difficult: the shaman captures the soul and reintegrates it in the body of the sick person. In the latter case he has to descend to the nether world, and this is a complicated and dangerous enterprise. Equally stirring is the voyage of the shaman to the other world to escort the soul of the deceased to its new abode; the shaman narrates to those present all the vicissitudes of the voyage as it goes on.

Another type of descent is represented by the journey of the Eskimo shaman (*see* below) to the bottom of the ocean, to the Mother of the Seals; he undertakes this submarine voyage when the seals become scarce and the population is menaced by a famine.

As a rule the shaman has several auxiliary spirits at his disposal, but he is not possessed by them. The spirits help him to find the soul of the sick person, and sometimes they accompany

him in his ecstatic voyages. In certain regions the shaman ends by being really possessed by his auxiliary spirits, and he becomes their mouthpiece; but the possession is a secondary phenomenon found especially in those regions where shamanism is mingled with other magico-religious conceptions and techniques.

Shamanism represents the mystical experience characteristic of the primitive and archaic religions. The "flight" of the shaman to heaven can be regarded as the most ancient expression of mystical experience known to mankind. But the shaman is not only a mystic. He is just as much the guardian (and largely the creator) of the traditional lore of the tribe, he is the sage and even the poet of primitive societies. The narrations of his adventurous descents to the netherworld and of his ascents to heaven constitute the material of popular epic poetry in central Asia and in Polynesia.

Eskimo.—Shamanism predominates in the religious life of the Eskimos. The chief prerogatives of the Eskimo shaman (*angakok;* pl., *angákut*) are healing, the ecstatic underwater journey to the Mother of Animals for the purpose of assuring an abundance of game, and the aid he brings to barren women. Sickness is brought on by the violation of tabus or results from the capture of the soul by a ghost. In the first case, the shaman strives to drive out the impurity by collective confessions; in the second case, he undertakes the ecstatic journey to heaven or to the depths of the sea to retrieve the sick person's soul and restore it to his body. The *angakok* is also a specialist in magic flight. Some shamans are reputed to have visited the moon; others claim to have flown around the earth. The *angákut* also know the future, make prophecies, predict changes in the weather and excel at magic feats.

American Indians.—Among many North American tribes shamanism constitutes the most important aspect of the religious life. The shaman is characterized by the supernatural power he acquires as the result of a direct personal experience. This power is obtained either spontaneously or after a voluntary quest, but in either case the future shaman has to undergo certain initiatory trials. In general, the power is utilized in such a way as to affect the whole society. The shaman's principal function is healing, but he also plays an important role in other magico-religious rites such as, for example, communal hunting and, where they exist, secret societies (*see* SECRET SOCIETIES, PRIMITIVE) or mystical movements (Ghost Dance [*q.v.*] religion type). North American shamans, like all their fellows, claim to control the weather (bring on or stop the rain, etc.), know future events, expose the perpetrators of thefts, etc. Furthermore, they defend men against sorcerers. But the magico-medical powers held by North American shamans do not exhaust their ecstatic abilities. There is every reason to suppose that modern secret societies and mystical movements among the Indians have appropriated in large part the ecstatic activity that once characterized shamanism.

In the tribes of South America the shaman enjoys considerable prestige and authority. Not only is he the healer *par excellence* and, in certain regions, the guide of souls of the dead to their new abode; he is also the intermediary between men and the gods or spirits, substituting himself for the priest at times. He guarantees the respect for ritual observances, defends the tribe against evil spirits, points out places for fruitful hunting and fishing, increases the wild life, controls the weather, eases childbirth, reveals future events, etc. Of course, the South American shaman can also fill the role of sorcerer; he can, for example, change himself into an animal and drink the blood of his enemies. Yet it is rather to his ecstatic abilities that the South American shaman owes his magico-religious position and social authority.

It is probable that a certain form of shamanism was diffused on the two American continents with the first waves of immigrants from Asia; later contacts between northern Asia and North America made Asian influence possible well after the penetration of the first immigrants.

Africa.—Shamanism does not play a role of the first order in Africa.

Southeast Asia and Oceania.—Shamanism is prevalent in the Malay peninsula and in Oceania. Among the Negritos of the Malay peninsula, the shaman heals with the help of celestial spirits or by using crystals of quartz. But the influence of Indo-Malayan beliefs is noticeable, too (the shaman changing into a tiger, trance achieved by dancing, etc.). In the Andaman Islands, the shaman gets his power from contact with spirits. The commonest method is to "die" and return to life, the traditional pattern of shamanic initiation. The shamans gain their reputation through their acts of healing and their meteorological magic (they are thought to bring on storms).

The distinctive marks of Malayan shamanism are the calling forth of the tiger's spirit and the achievement of the trance (*lupa*), during which the spirits seize the shaman, possess him and reply to questions asked by the audience. Mediumship is also characteristic of different forms of shamanism in Sumatra, Borneo and Celebes. Among the Ngadju-Dayak of Borneo there even exists a special class of shamans, the *basirs* (literally, "incapable of procreation"), hermaphrodites who dress and act like women. These are considered to be intermediaries between heaven and earth because they unite in their own person the feminine element (earth) and the masculine element (heaven).

Possession by gods or spirits is a peculiarity of Polynesian ecstatic religion. The extreme frequency of possession in that region has made possible a proliferation of healers. Priests, inspired persons, medicine men and sorcerers may all perform magical cures. For this reason it is not possible to speak of shamanism *stricto sensu* in Polynesia.

In Australia, a person becomes a medicine man through a ritual of initiatory death, followed by a resurrection to a new and superhuman condition. But the initiatory death of the Australian medicine man, like that of the Siberian shaman, has two specific marks not found elsewhere in combination: first, a series of operations performed on the candidate's body (opening of the abdomen, renewal of the organs, washing and drying of the bones, insertion of magical substances); second, an ascent to heaven, sometimes followed by other ecstatic journeys into the other world. The revelations concerning the secret techniques of the medicine men are obtained in a trance, a dream or in the waking state before, during or after the initiatory ritual proper.

See also references under "Shamanism" in the Index.

See Mircea Eliade, *Shamanism: Archaic Techniques of Ecstasy*, Eng. trans. by Willard R. Trask (1964); M. A. Czaplicka, *Aboriginal Siberia* (1914). (M. EL.)

SHAMASH, the name of the sun-god of the Semitic pantheon. The Akkadian Semites (*c.* 2300 B.C.) identified their sun-god Shamash with the local Sumerian sun-god Utu, whose temple was called Ebabbar, and whose picture sign was a disk rising from two mountains in the east to indicate the rising sun. In processions of gods, Shamash is pictured seated on a horse, and in pre-Islamic Arabia the horse is a symbol of the sun. In the Old Testament it is said of King Josiah that "he removed the horses that the kings of Judah had dedicated to the sun, at the entrance to the house of the Lord . . . and he burned the chariots of the sun with fire" (II Kings 23:11).

The early chief attribute of the Sumerian god Utu was fruitfulness. But by about 2130 B.C., the ruler of Lagash, Gudea, is saying that "Utu causes Righteousness to come forth; as for Unrighteousness Utu treads its head underfoot." King Ur-Nammu, who established the 3rd dynasty of Ur (*c.* 2130 B.C.), and of whose Sumerian law code fragments have been found, speaks of the just laws of Utu. The last king of Isin (*c.* 1820 B.C.) dedicates to the Semitic Shamash a temple named "House of the Judge of the Land." In Akkadian Semitic thought, the sons of Shamash are Justice (*kettu*) and Right (*mesharu*). The code of Hammurabi, which is concerned precisely with these, is eloquent in praise of "the great judge of heaven and earth" to whom the king in the text attributes the laws. Before this god the king is shown standing in reverence on the relief which forms part of the stele on which the laws are inscribed.

The chief centres of the cult of the sun-god were at Zararim (Akkadian Larsa) in Sumer and at Sippar in Akkad. At the former city Ur-Nammu built a temple to Utu, but this is the only mention on the surviving tablets of a temple to this god, though there

is frequent mention of a shrine (*bara*) and of offerings to him. The Akkadian king Naramsin (*c.* 2230 B.C.) founded the temple, at Sippar, whose foundations Nabonidus, the last king of Babylon (556–539), claims to have discovered. Here the cult of the sun-god was revived after a long period of neglect. Inscriptions of Assyrian kings from the 16th century B.C. onward testify to his cult, and there were temples in his honour at Ashur and Nineveh, the ancient and later capitals.

Shamash ranked below Sin (*q.v.*), the moon god, whose son he was. With his father and his sister, Ishtar, he forms the secondary cosmic triad of the pantheon. Already among the Sumerians he was a warrior god, and later Akkadian sources represent him as conqueror of darkness in the heavens and on earth. To mortals he is the giver of light, the searcher of hearts, the source of laws of righteousness, the giver of life, protector of the oppressed and the terror of the wicked, whom his rays, like a net, catch and hold fast. Augurers have recourse to him, and oaths are sworn before him.

His consort was Aya, a Sumerian goddess, later absorbed by Ishtar.

See also BABYLONIA AND ASSYRIA: *Religion;* SUN WORSHIP.

BIBLIOGRAPHY.—A. Deimel, *Pantheon Babylonicum* (1914); A. Falkenstein and W. von Soden, *Sumerische und akkadische Hymnen und Gebete* (1953); E. (Paul) Dhorme, *Les Religions de Babylonie et d'Assyrie* (1949). (T. FR.)

SHAMIL (1798?–1871), the leader and unifier of the Muslim mountaineers of Dagestan and the Chechen country against the Russians. He was born at Gimry, the son of a free landlord. He studied grammar, logic, rhetoric, and Arabic and soon acquired the prestige of a learned man. In 1830, under Gazi Mohammed, the first elected imam of Dagestan, Shamil became one of the exponents of Muridism, a movement which demanded holy war against the Russians and the maintenance of the sacred law of the *Sheriat* against the *Adat* (a less orthodox interpretation of the Koran).

In 1832 Gazi Mohammed was killed by the Russians; two years later his successor Gamzat Bek was assassinated by his own supporters. Then, in 1834, Shamil was elected third imam of Dagestan. He conducted the war against the Russian invaders with personal courage and resourcefulness. His extermination of khans and beks who submitted to Russia, his liberation of their serfs, and, especially, his success in eluding capture made him a legendary hero. The struggle was so intense and passionate that the peasant peoples of northeastern Caucasia surmounted their tribal and linguistic particularisms and became united.

In 1859 the Russians decided to put an end to Shamil's rule. An army of 200,000, with 200 guns, under General A. I. Baryatinski started operations from all sides. Resistance was gradually overcome; and on Sept. 6 Shamil himself, besieged with a handful of Murids at Gunib, was taken prisoner. Incarcerated first in St. Petersburg, he was later sent to reside at Kaluga. In 1870 he made a pilgrimage to Mecca; he died, probably at Medina, in March 1871. His romantic fight had earned him a European reputation.

See K. von Seeger, *Imam Schamil: Prophet und Feldherr* (1937).
(A. GU.)

SHAMMAI, along with Hillel (*q.v.*) the leading Palestinian Jewish sage at the end of the last century B.C. and the early part of the 1st century A.D. In his time, on more than a few occasions, his views and those of his school (the House of Shammai) found readier acceptance in Pharisaic intellectual circles than they did later, after the Hillelite influence triumphed. But as late as the 2nd century, some talmudic scholars still reflected and favoured Shammaite tendencies, essentially conservative, adhering closely to earlier, transmitted teachings and strict in the application of legal injunctions. Shammai, like other Pharisaic and later rabbinic scholars, was no fundamentalist, and he engaged in biblical exegesis as well as development of the Oral Law; but his interpretations reveal a literalness of mind which contrasts sharply with the greater imaginativeness and plasticity encouraged by the Hillelite school. (*See* also TALMUD: *Oral Law: Shammaites and Hillelites.*)

To say seriously what Shammai was like as a person is impossible, for in most of the (few) narratives about him he is used by the storytellers as a foil, to highlight the virtues of Hillel. The effect is chiaroscuro, neither history nor biography. The stories were probably popular exempla of the late 1st and 2nd century intended for the amusement of audiences and the instruction of children. His few sayings and decisions, and the views of the school over which he presided, suggest that he was a product of the older and more conservative generation, that he regarded with suspicion compromises with tradition, that he was noteworthy for his impatience with verbosity or, despite his insistence that all men be received cheerfully, with strangers who made grotesque demands upon him. There is good reason to believe that he felt strong sympathy for those groups that encouraged revolt against the Roman authorities. (JU. G.)

SHAMROCK. According to an Irish legend, St. Patrick first chose the shamrock as a symbol of the trinity of the Christian church because of its three leaflets. With a shamrock he is said to have driven the snakes of Ireland into the sea. The name is applied to several plants of the pea family (Leguminosae), among them wood sorrel (*Oxalis acetosella*); white clover (*Trifolium repens*); suckling clover (*Trifolium dubium*); and black medic (*Medicago lupulina*). Wood sorrel is shipped from Ireland to London and other places in great quantity for St. Patrick's Day. For that holiday in the United States the plants most commonly used as shamrock are white clover and species of *Oxalis* (*q.v.*).

The shamrock pea is *Parochetus communis*, a creeping legume with bicoloured blue and pink flowers; it is used as a pot plant and for hanging baskets.

SHAN, the Burmese term for speakers of Thai (Tai) languages other than Siamese. Thai (Tai) principalities (*möng*) of varying size exist in Assam, Burma, Thailand, Laos, Vietnam, Cambodia, and throughout southwest China. In Burma the Shan numbered more than 1,250,000 in the 1960s (*see* THAI PEOPLES).

Although holding mainly high upland territory in Burma, the Shan themselves live only in the valleys and in stretches of plain. The surrounding hill country is occupied by relatively primitive hill peoples living in economic symbiosis with the Shan, but there is no clear genetic distinction between the Shan and their neighbours. With minor exceptions modern Shan communities are Buddhist. Most Shan are cultivators of rice; they have a characteristic species of feudal political structure. Shan culture probably diffused southward and westward from west-central China from Han times (*c.* 200 B.C.) onward.

In a Shan *möng* the power of the prince (*saohpa*, anglicized *tsawbwa*) is theoretically absolute but is really exercised by appointed ministers (*amat*). The prince has semidivine attributes and, unlike the commonality who are monogamous, is expected to maintain an extensive harem. In some areas he should marry a half-sister as chief wife. Emphasis on patrilineal descent is confined to the royal lineages. Rank status is hereditary but there is no rigid caste endogamy. Thai counterparts of the Shan are known locally by a great variety of names; *e.g.*, Ahom, Khamti, Pa-yi, Lu, and Lao. *See* also AHOM; PA-YI; SHAN LANGUAGE; SHAN STATE; and references under "Shan" in the Index.

BIBLIOGRAPHY.—W. W. Cochrane, *The Shans* (1915); E. R. Leach, *Political Systems of Highland Burma* (1954); F. M. Le Bar *et al.*, *Ethnic Groups of Mainland Southeast Asia* (1963). (E. R. L.)

SHANGHAI (SHANG-HAI), a commercial metropolis of China and one of the world's largest seaports, occupying an extraordinarily favourable geographical situation for both internal and external trade. Geographic conditions have concentrated China's foreign trade into a single gateway to a degree unparalleled in any other great country. One of the original treaty ports (opened by the Treaty of Nanking, 1842), it owed its phenomenal growth to Occidental commercial enterprise.

The magnificent natural waterways united in the drainage of the Yangtze river give to the port at its outlet to the Pacific a vast commercial hinterland producing a great range of commodities and a large consumer market of about 200,000,000 inhabitants. In addition, it serves as the master port for the lesser ports of the China coast from Fukien to Shantung province. Ex-

ternally, it is the nearest port to Japan, and it lies directly on the great shipping lane from San Francisco to Singapore. Locally, Shanghai occupies one apex of the delta land south of the Yangtze, containing 20,000 sq.mi. (51,800 sq.km.) with more than 40,000,000 people, of whom about 70% are farmers who supply the city with food and with cotton and silk for its industries.

Climatically, Shanghai has the monsoon characteristics common to most of the east coast region, with its concentration of high rainfall in the spring and summer. The annual average is 45 in. of precipitation. Weather variability is provided by the frequent cyclonic storms moving eastward along the Yangtze valley, especially in spring. Shanghai's latitude is almost as far south as that of New Orleans, so that summers are uncomfortably hot and humid. Summer temperatures are kept below 32° C (90° F) by rains for part of July and August, and they seldom rise above 37° C (98° F), but humidity averages 84% for this period. In winter humidity drops to about 30% and absolute minimum temperature to −10° C (13° F).

The Harbour.—Shanghai does not escape the problems besetting delta cities. It lies about 14 mi. (23 km.) above the mouth of the Whangpoo River (Huang-p'u Chiang), a small tributary of the Yangtze near its estuary. Daily, rising tide water from the Pacific shunts Yangtze River water up the Whangpoo. The slow current at this stage leads to rapid deposition of sediments. In early times this also resulted in the silting up of Soochow Creek (Wu-sung Chiang), which flows from Soochow into the Whangpoo. The Whangpoo during the mid-19th century scoured out a channel to the Yangtze estuary at least 15 ft. (5 m.) deep and was able to accommodate the sailing vessels of the time. This led the British to select Shanghai as one of the five treaty ports opened to trade as a result of their success in the war with China (1839–1842). The Whangpoo's initial advantages were rapidly reduced by further silting, however, and by 1905 the channel depth was only 10 ft., leading to gloomy forecasts of Shanghai's future as a port. The problem, however, was largely solved by the work of the Whangpoo Conservancy Board. Its vigorous harbour policy resulted in the development of a good shipway with a minimum depth of 24–26 ft. (7.3–7.9 m.) at the lowest tides and a high-water depth of 30–42 ft. (9.1–12.8 m.). Requiring much greater efforts were the main bars of the Yangtze itself at the so-called Fairy Flats (Shen T'an). Large ships could enter the river only at high tide and through the south channel, while tides and currents made the approach to Shanghai somewhat difficult. Neglect during World War II and after the Communists took over resulted in a less commodious and useful harbour in the mid-20th century than during the 1930s.

Port Communications.—Numerous canals, some navigable to 80-ton barges and steamboats, lead from the Whangpoo to the surrounding countryside. These include a 36-mi.-long (58 km.) canal for navigation and flood control from T'ai Hu, a large lake in the west, to the Whangpoo, constructed in 1958. Shanghai's two chief railroad connections are with Nanking in the northwest and Hangchow in the south, 193 mi. (311 km.) and 118 mi. (190 km.) distant respectively. Through these cities Shanghai has rail connections with the north and south China networks. The role of railroads in Shanghai's growth has been minor, and in the 1930s her two main rail lines carried only the moderate amount of 1,000,000 tons annually per mile. The withdrawal of most of China's shipping by the Nationalist government of China to Formosa in 1949 left water transport at Shanghai at a low ebb. From June 1949 to May 1950 the Nationalist blockade brought the relative shares in total freight of railways and watercraft of all kinds to an approximate ratio of 3.2 to 1. The heavy reliance that was put on carriage by rail contributed in part to the piling up of large quantities of cargo here as in other parts of the country in the late 1950s when there were transport bottlenecks. Although several hard-surfaced roads usable to motor transport lead from Shanghai to the hinterland, the high cost of motor transport resulted in a low volume of trucking from the port.

Port Development.—When the British first realized the port potentials of Shanghai during naval operations in the Anglo-Chinese war, Shanghai was a small unimportant walled city on the Whangpoo about a mile above its junction with the Soochow Creek. In 1843 the British settlement was established between the Soochow Creek and the Yang-ching Canal a short distance north of the walled city. Later this strip between the canal and the walls became the site of the French concession, while the American settlement (Hongkew) grew up on the northern side of the Soochow Creek. The American and British sectors were joined in 1863 to form the International settlement. The foreign zones were administered separately until World War II. The settlement area once was marshy paddy land often flooded, but through careful drainage and filling the area was converted to sites suitable for buildings. However, construction faced the difficulty of finding stable foundations, since the unconsolidated sand and silt was about 1,000 ft. deep. Uneven settling of large buildings made it dangerous to construct skyscrapers until the problem was solved for buildings up to about 20 stories by floating them on concrete rafts. Many of the canals and open sewers were culverted over, although some were used for small boat and barge navigation. Pontoon jetties anchored along the bank provided landing facilities. From the walled city down river on the left bank, and for a longer distance up and down the river on the right bank (an area known as Pootung or P'u-tung) there were in the 1930s less than 10 mi. of riverside shipping frontages with a minimum depth of 12 ft. The congestion at Shanghai was great. Subsequent work by the Chinese municipality, which was continued after 1937 by the Japanese, resulted in additional wharfage just below the Pootung sector of the waterfront. Postwar construction by the Communist regime includes the K'ai-p'ing wharf at Jih-hui Chiang, which for the first time provided railroad connection directly to a wharf, and the new wharf at Pei-p'iao begun in June 1956.

RENE BURRI – MAGNUM

(LEFT) BARGES LOADING MACHINERY AND IRON ORE ON THE WHANGPOO RIVER; (RIGHT) THREE CORNER MARKET (SANG CHOW TE MARKET) IN THE CENTRE OF SHANGHAI

Trade.—Shortly after being opened to foreign commerce in the 18th century the port of Shanghai began to dominate China's trade and by the 1870s it was handling about 62% of the total. By 1920 this had dropped to 45% as northern ports developed. In the post-World War II period before the Nationalists took refuge on Formosa, Shanghai accounted for 75% of the imports and 60% of the exports of China. Its importance declined abruptly after 1949 because of the Nationalist blockade, and by the mid-1950s Shanghai was in second place after Tientsin which had become the leading port for Communist China. However, by

the 1960s, Shanghai's trade had rapidly and progressively increased.

The export of tea and silk and the import of opium and cotton textiles dominated the greater part of the trade period from 1850 to 1937. During the 20th century, however, especially during the 1920s and 1930s, imports of machinery, metals, and raw materials rose to account for more than 65% of the import trade. In exports, the position of raw silk and tea, although not the quantity, diminished sharply, while bulk commodities such as raw materials, vegetable oils, hides and skins, and some manufactured products became increasingly important. Tung oil and pig bristles were two important raw materials marketed predominantly through Shanghai. In 1939 the value of exports from Shanghai was about the equivalent of U.S. $126,000,000, while imports reached about $94,000,000. These figures, somewhat incomplete because of the abnormal war situation, give some idea of the trade volume. Wartime trade from 1941 to 1945 was greatly reduced and toward the end amounted to little. The 1946 figures for China's exports as a whole (and Shanghai handled about 60% of them) were very small absolutely and relatively to their prewar ratio to imports. Large quantities of imports were UNRRA and American aid materials. After 1949, mainland China's trade was predominantly with the Soviet Union and Communist-dominated countries. There was a marked decline in the importance of consumer goods and a corresponding increase in the import of capital goods. Machinery, metals, chemical raw materials, scientific instruments, drugs, and fibres, especially cotton, were leading imports. Food and raw materials continued to furnish the bulk of the exports, but manufactured goods and machinery were also of some importance.

The number of foreign ships visiting the port, not unexpectedly, has declined since the 1930s. While the People's Republic of China claimed in the late 1950s that Shanghai had regained its importance as a leading seaport, this was not borne out by observations of foreign travelers nor by official shipping statistics. It was reported in 1958 that Shanghai harbour handled a record volume of 2,220,000 tons of cargo, the highest in the history of the harbour. Hence, it is assumed that much of this pertained to domestic freight movements. However, the food crisis in China after 1960 caused about 5,000,000 tons of grain to be shipped from foreign countries, much of it imported through Shanghai.

Industry.—Shanghai not only is one of China's leading ports, but also a leading manufacturing city. Before 1937 it accounted for at least half of the country's large-scale Western-type industrial production. Of this industry, 40–50% of the value of output and 50–60% of the labour of the city was in textile manufacturing. This was led by cotton-spinning followed in order by silk-reeling, cotton weaving, wool spinning and wool weaving. Food industries came second with 30–35% of the value of production. Among these, flour milling and cigarette manufacturing were most important. Other industries in order of importance were leather goods, rubber goods, paper and publishing, chemicals, and machine manufacturing. While cotton and silk industries found their raw materials in Shanghai's immediate hinterland in the Yangtze delta, flour mills depended upon wheat shipped from abroad, while tobacco came from the north China plain. The metals- and rubber-using industries were farthest removed from their sources of supply. Shanghai's coal was imported from north China, Japan, and northern Vietnam prior to 1949. The Shanghai power company, using coal as fuel, produced 70% of the city's power, and as coal could be brought in by cheap ocean transportation the city's power rates were among the cheapest in the world. Industries developed along both banks of the Whangpoo except along the Bund (a waterfront thoroughfare), but the east bank was less important because of location factors.

Shanghai emerged from World War II in a precariously depressed condition. Most of the textile plants and spindles as well as other factories were destroyed or inoperative. Rehabilitation began gradually. In 1949 Shanghai metal industries turned out 5,000 tons of steel and about 13,000 tons of steel products. Following the initial take-over, the Communist industrial policy was to keep privately owned plants in operation with the government absorbing the output but at fixed prices. The second step was the formation of joint government-private companies which by the beginning of 1956 took in all industrial ventures. This smoothed the transition to full government ownership and operation, since the old managerial staffs were largely retained on a salary basis. Following the Nationalist air bombardment of February 1950, there was a temporary industrial exodus and a move to decentralize Shanghai's industries by removing them to the interior. In 1953 the city had only 20% of the total value of mainland China's industrial output, but this was officially claimed to be 368% higher than that of 1950.

In the mid-20th century Shanghai had about one third of China's cotton spindles and produced about 2,000,000 bales of cotton yarn. China's first rayon plant went into production in 1958. In 1950 the city's flour mills had 20% of China's flour-making capacity of approximately 140,000,000 bags per year. This was above the consumption requirements of Shanghai, so it was decided to move some of the plant equipment to wheat-growing areas in the interior. Rice-milling production, however, was insufficient and was increased by the construction of new mills.

Next to Tientsin, Shanghai in the 1960s was a chief centre of chemical industry in China. The largest chemical factory was the Yung-li plant which produces ammonium sulfate and pure soda. The Wu-ching fertilizer plant built in 1960 went into full operation in 1963 with a 100,000-ton capacity for producing ammonium sulfate. One of China's chief cement plants is at Chiang-nan. A new policy to build up Shanghai as a heavy-industry and machine-manufacturing centre was well underway in the 1960s. In 1958, the year of the "great leap," Shanghai claimed to have achieved its goal of 1,222,000 tons of steel, 1.4 times the production of 1957. Output was scheduled to be further increased in subsequent years. The heavy industry was also pledged to make steel manufacturing equipment for other parts of China, electric power generators, and tractors. Among the major plants built or enlarged during the first five-year plan (1953–58) were shipyards at Chiang-nan and Hu-tung, a steam turbine works, a boiler works, and a lathe plant. In 1964 a 6,000-ton ocean freighter was launched at the Chiang-nan shipyard. Eight steel rolling mills for making seamless tube, welded pipe, alloy steel, and sheet iron were also included in the plan. In 1964 the first tinplate workshop was completed at the steel mill, and cylindrical grinders were being produced. Heavy industrial development increased considerably in the 1950s but on the other hand there was a marked decline in light industry. In all, 168 industrial plants in 17 industries were enlarged and 43 new plants built between 1949 and 1960. By 1961 Shanghai was the main centre of much heavy machine industry in China, and of the production of precision lathes, paper, glass, plastics, and watches.

Population.—According to the Nationalist census bureau Shanghai's population was 4,630,385 in 1948. The distress in the rural areas following the Communist cooperative farm program grain collections, together with natural calamities, brought a rapid urban influx of peasants which by 1953 had boosted Shanghai's population to 6,204,417. Efforts were made to reduce this, by forcible evacuation to the country, but many of these deportees probably managed to drift back. In 1957 the population was estimated at 6,900,000. By the mid-1960s reports indicated that the Shanghai urban area had ten million inhabitants. Shanghai's natural population increase since 1949 was estimated at an average rate of 200,000 per year. In 1964 the rate of population increase at Shanghai was 1.7%, much lower than the 2.2% estimated for the country at large. The city's industrial labour force in the 1960s was more than 1,000,000. The male to female sex ratio of the population in the 1953 census was 123 to 100.

Social Change and Culture.—Shanghai's social character has greatly changed. In the 1930s there were between 30,000 and 60,000 foreign residents, but now few foreigners are to be seen. In a city known once as a centre of gangsterism and of beggars and prostitutes, it is claimed that now not a single night club exists. About 10,000 prostitutes were put into reformatories for labour reform and rehabilitation to Communist society. Gambling was banned along with prostitution, and the famous racecourse and

dog race stadium were converted into a recreation park and museum. Numerous regulations were applied to urban inhabitants of Shanghai after 1950 and individual movement about the city was greatly restricted. Under the so-called Democratic Reform movement, begun in December 1952, former independent labour leaders were liquidated, and labour organizations were disbanded. Wholesale arrests of suspected Nationalist sympathizers were made. That the regime's problems were not all solved was indicated by the drive in 1958 to eradicate theft, gambling, vagrancy, census evasions, delinquencies, political sabotage, and illegal capitalistic enterprises. Shanghai was charged in 1957 with having over 9,000 clandestine factories with more than 30,000 workers.

In conformity with the national policy of making office workers and officials do manual labour, thousands of cadres were reported to have been organized and sent to rural areas for work. The city is being used as a training centre for clerks, skilled labour, and technicians for the new plants being established in the interior regions. In 1964, for instance, 4,000 workers were sent from Shanghai to Lan-chou to aid the development of light industry. In spite of censorship and reorganization, Shanghai is a major publishing centre, and also maintains its position as a leading educational and cultural centre. It is the site of Futan and Tungchi universities, and there are several medical colleges, museums, and libraries. Institutes of Foreign Languages, Technology, Drama, Law, Finance and Economics, and Marine Products and the Shanghai Conservatory of Music are located there.

Municipal Administration.—Shanghai is located in, but is politically independent of, Kiangsu province. Shanghai municipality, along with Peking, is treated on a par with provinces. The Communist municipal government has a mayor, 11 elected deputy mayors, 44 members of a peoples' council, and a peoples' congress consisting of 737 delegates. The latter is a number equal to that of any provincial congress. The area of Shanghai in 1958 was 772 sq.mi. (1,999 sq.km.). When the Communists first took over in 1949, they went into such detail of organization that in each of Shanghai's 11,185 streets a street committee was set up whose work was divided into six departments: health, education and culture, defense and security, social welfare, conciliation, and relief. This was furthered by the establishment in 1951 of the Shanghai Peoples' Control Commission. The control function was most likely performed by the Discipline Inspection Committee already established in all the administrative units of Shanghai's municipal government. Moreover, the municipal peoples' court of Shanghai set up in 1949 was placed under the control of the local Military Control Committee, thus reducing the court to a combination of military and administrative tribunal. The close tie with the party organization is shown in the position of Jao Shu-shih, the first secretary of the East China Party Bureau and of the Shanghai Municipal Party Committee, who in 1950 was also political commissioner to the 3rd Field Army under Ch'en I.

BIBLIOGRAPHY.—Lucian Taire, *Shanghai Episode* (1957); Rhoads Murphey, *Shanghai, Key to Modern China* (1953); E. M. Hinder, *Life and Labour in Shanghai* (1944); R. W. Barnett, *Economic Shanghai: Hostage to Politics* (1941); W. C. Johnstone, *The Shanghai Problem* (1937); G. E. Miller, *Shanghai, Paradise of Adventurers* (1937).

(H. J. Ws.)

SHANKARA (780–820), Indian philosopher, the leading exponent of Advaita Vedanta, nondualism, and author of commentaries on the Upanishads, the Bhagavad Gita, and the Brahma Sutra which are known for their profundity of spirit and subtlety of thought. Shankara attempts to build a spiritual view of life on rational foundations: by reasoning we find that the world is not self-maintaining and must have a cause which is transcendent to it. In spiritual experience we apprehend the reality of this Absolute Spirit, which is called Brahman (*see* BRAHMA AND BRAHMAN). The differences of knower, known, and knowledge disappear when the Supreme Reality is known.

The world does not exist of itself but is derived from and dependent on Brahman and so is less real than Brahman. It is not, however, an illusion. The individual self which feels, suffers, and is affected by the experiences of the world is a phenomenon while the truth is Brahman. To recognize the highest truth as Brahman is to attain release. Meditation, worship, ritual are intended for a lower class of aspirants, and *jnana* (wisdom, knowledge of God) is the path to be pursued by the higher class of aspirants who have no desire for earthly prosperity or heavenly joy. There are two kinds of release: *sadyo-mukti*, or instantaneous liberation, and *krama-mukti*, or gradual liberation. The former is the result of *jnana*, the latter of *upasana*, or worship and prayer. Shankara makes out that the identity with the Highest Self is not destruction of soul. It is the perfection of the soul which has no more specific cognition or objective knowledge. *See* also INDIAN PHILOSOPHY: *Vedanta*. (S. RA.)

SHANKLIN: *see* SANDOWN-SHANKLIN.

SHAN LANGUAGE. Shan is a language of the Tai linguistic group spoken in the Shan and Kayah states of northern Burma. It is a noninflecting, basically monosyllabic, tonal language bearing considerable phonological and grammatical resemblance to other Tai languages such as Lao and Siamese (Thai). Compared with these, Shan exhibits a simpler syllabic structure, but greater differences exist in lexicon. Long contact with the Burmese language has influenced Shan, which contains many Burmese loan words.

The Shan script was developed from the Burmese. The traditional writing system is inefficient and causes grave reading difficulties even to Shans. In particular, it lacks sufficient vowel symbols and does not mark tone. These deficiencies were remedied in a series of reforms culminating in that of 1955, and the modern Shan script is an excellent phonetically based writing system.

Dialectal differences exist within Shan, but these are relatively minor. Dialects to be grouped with Shan are spoken by some Tai peoples in southwest China (*e.g.*, the Tai Mao people) and on the borders of Assam (*e.g.*, the Khamti people). These groups employ scripts which, though generally similar to that of Shan, exhibit marked stylistic variations.

Within the Shan state some peoples, notably the Khu'n of the formerly autonomous Kengtung state, speak Tai dialects which must be grouped with those of north Thailand rather than with Shan. *See* SHAN; SHAN STATE.

BIBLIOGRAPHY.—J. N. Cushing, *Grammar of the Shan Language*, 2nd ed. (1887), *A Shan and English Dictionary*, 2nd ed. (1914); H. W. Mix, *English-Shan Dictionary* (1920); S. Egerod, "Essentials of Shan Phonology and Script," *Acad. Sinica* (Taiwan), vol. 29, pp. 121–129 (1957). (E. H. S. S.)

SHANNON, CHARLES HASLEWOOD (1863–1937), English artist who was notable for his lithographs, was born at Quarrington, Lincolnshire, April 26, 1863, the son of a clergyman. He was apprenticed to a trade wood engraver, in whose workshop he presently met Charles Ricketts, with whom he lived and worked until Ricketts' death in 1931, although their last direct collaboration was the periodical *The Dial* (1889–97).

Shannon's earliest designs were for Harry Quilter's *Universal Review*; but in 1893 he and Ricketts produced *Daphnis and Chloë*, a book in which text, illustrations and paper of choice quality formed a unique harmony. The first designs savour of early Renaissance book illustration, but the lithographs undertaken by Shannon from 1889 to 1898 are more personal, austere, delicate, and full of a silvery chiaroscuro. Often their subject matter is of a recondite mythological character, but among them are valuable portraits such as those of Alphonse Legros, George Moore, and Lucien Pissarro. Shannon as a painter was influenced by Titian, Tintoretto and, in allegorical subjects such as the "Vanity and Sanctity" (Diploma gallery, Royal Academy, London), by G. F. Watts; but here again he produced notable portraits such as the "Miss Bruce, Sculptress" in the Luxembourg, Paris. He was elected associate of the Royal Academy in 1911, full member in 1921. Between 1904 and 1919 appeared 46 more lithographs, fuller in form and richer in chiaroscuro than before. Shannon died at Kew, Surrey, on March 18, 1937, after eight years' disablement following a fall. The Ricketts and Shannon collection of prints, famous for its careful and original selection, is now in the British Museum, London.

See E. B. George, *C. H. Shannon* (1924). (D. C. T. T.)

SHANNON, the longest river in the Republic of Ireland, rises in the northwest of County Cavan and flows for about 161 mi. (260 km.) in a southerly direction to enter the Atlantic Ocean

via a 70-mi.-long (113-km.) estuary below Limerick. It drains an area of 6,060 sq.mi. (15,700 sq.km.). As the main river draining the Irish Central Lowland it is surrounded by marshes and bogs for much of its course, and widens out at various points into lakes, many with numerous islands.

The Shannon source is generally considered to be the pools at the foot of Tiltinbane Mountain (1,881 ft. [573 m.]), in County Cavan, but it is fed by a number of mountain streams in this area. After a few miles it enters Lough Allen (164 ft. above sea level, 7 mi. long, and 1 to 3 mi. wide) and then flows south through a wide belt of marshes and water meadows, locally called callows. In dry years these provide excellent natural pastures for cattle but in wet years the river becomes what the poet Edmund Spenser called "the spacious Shenan spreading like a sea." It is crossed by a bridge at Leitrim and also at Carrick-on-Shannon (*q.v.*), above which it is joined by the Boyle from the west. From north of Carrick south to Roosky, it flows through a landscape dominated by drumlins (low glacial hills), interspersed with bogs and marshes. Both the main stream and its tributaries, the Mountean from the west and the Rinn from the east, broaden out into lakes, such as Boderg and Bofin, and smaller sheets of water. Around Roosky the drumlins disappear and the river flows through ground moraine country or meadows and bogs, and crossings are infrequent.

Near Termonbarry (9 mi. below Roosky), at Richmond Harbour, the river is joined by the Royal Canal, and at Lanesborough it enters Lough Ree (nearly 18 mi. long and 1 to 6 mi. wide) which separates Longford and Westmeath counties on the east from Roscommon. The River Inny joins the lake on the east. From Athlone, a mile south of Lough Ree and originally built on an esker (or gravel ridge), the barrier nature of the Shannon becomes even more marked. The bridge at Athlone is 15 mi. from the next at Shannonbridge, though there is one ford at Clonmacnoise (*q.v.*). The Suck, the largest tributary, joins the Shannon at Shannonbridge and 2 mi. north of Banagher, at Shannon Harbour, the river is joined by the Grand Canal. About 15 mi. south it reaches Lough Derg at Portumna. This lake extends for over 20 mi. south to Killaloe, and in one section is 10 mi. broad. All of its waters are now dammed for the headrace, in effect a canal 7½ mi. long which feeds the Ardnacrusha hydroelectric station (1925–29). This uses the fall of 109 ft. from the lake to the sea level near Limerick, and concentrates it in one spot. The tailrace connects Ardnacrusha with the Shannon riverbed. Though power is generated, the river has proved less reliable than was first hoped owing to its fluctuations in level. A fish ladder has been constructed at the weir at the outlet of Lough Derg to allow salmon to pass up river at spawning time. From Limerick the river enters its long estuary where shipping has dwindled since the 19th century and where some of the large reclaimed areas have been used for the Shannon airport, located near the estuary of the Fergus River in County Clare.

In the middle reaches the Shannon divides the provinces of Leinster and Connaught and for most of its length it is a county boundary with contrasting landscape and land use on either bank. Clare and Galway on the west are poorer counties (with smaller farms) than Limerick and Tipperary on the east (though Leitrim in the north is an exception, being one of the poorest counties in Ireland). One reason for the economic divide is the lack of east-west communications.

In the early part of the 19th century the Shannon was a vital link in the waterways of the country, though before the days of steam transport navigation was difficult. From 1755 the Grand Canal was built from Dublin across the Central Lowland (though the Shannon was reached only in 1804) and until the coming of the railways was an important means of communication. From 1789 to 1817 the Royal Canal was constructed from the north side of Dublin through Mullingar to the Shannon at Richmond Harbour but, running through sparsely populated country, was never as successful as the Grand Canal and in 1846 was sold to the new railway company. At this period steamers used the Shannon between Killaloe and Richmond Harbour, and horsedrawn boats used the canal from Killaloe to Limerick. There were passenger services along the canals from 1780 but the canal services ceased in the 1850s because of railway competition. The Shannon steamboats survived until the coming of the automobile and were a tourist attraction until World War I. Since the 1950s a steamer service has been provided as a holiday attraction. (*See* also the separate articles on the counties mentioned above.)

See N. T. H. Delany, "The development of the river Shannon navigation," *J. Transport Hist.*, vol. 3, no. 4 (1958); T. W. Freeman, *Ireland: a General and Regional Geography*, 3rd ed. (1965).

(T. W. Fr.)

SHANSI (Shan-hsi Sheng), a province of north China, pop. (1957 est.) 15,960,000, occupies part of the loess plateau and is bounded by the Yellow River in the west and south, the Great Wall in the north, and Wu-t'ai and T'ai-hang mountains in the east where they drop off toward the North China plain. Area 60,656 sq.mi. (157,099 sq.km.). A northeast-southwest series of river basins and valleys roughly divides the eastern plateau from the western plateau into nearly equal halves. The Wu-t'ai and T'ai-hang ranges rise above the general 5,000–6,500 ft. levels and reach 9,974 ft. (3,040 m.) in a Wu-t'ai peak. In the west the Lu-liang is the chief range of the dissected plateau and reaches a peak elevation of 9,186 ft. (2,800 m.). Most of the valley land ranges between 2,300–2,600 ft., but southwest of An-i drops to below 1,300 ft. (400 m.). In the north, tributaries of the Yung-ting and Hai Ho have created the Ta-t'ung and Hsin Hsien basins, the latter dividing the Wu-t'ai from the T'ai-hang range. Southward lies the T'ai-yuan Basin and provincial seat on the upper Fen River, which runs thence southward and again widens into the lower Fen plain ending at the great lower bend of the Yellow River. The dry climate exhibits a decrease in precipitation from southeast to northwest, from 20 in. (510 mm.) to less than 10 in. yearly, much of the unreliable fall occurring in July and August.

A severe drought in 1957, resulting in a drop in production of 883,000 short tons of grain, was typical of the farming hazards. The growing season ranges from 150 days in the north to 250 days in the south. The rivers, which provide intermittent navigation in summer to flat-bottomed junks, are frozen as well as much constricted in winter. Millets, spring wheat in the north and winter wheat in the south, kaoliang, and corn are the chief crops. Some potatoes and oats are grown. Industrial crops include cotton, tobacco, hemp, and peanuts, and there are fruits such as the pear, apricot, persimmon, and grape, as well as walnut and the jujube date.

Animal husbandry is an important sideline, especially for the hill farmers, who raise chickens, sheep, pigs, horses, and donkeys and produce large quantities of eggs, wool, sheepskins, and pig bristles. The fertile but easily eroded loess soil blankets most of the hill slopes and permits extensive dry-terrace cultivation. The agricultural heart of the province lies in the Fen Valley which is part of the cotton belt that extends southwest into the Wei Valley of adjoining Shensi Province. The landscape generally is barren of trees except in the higher mountains, where over scattered areas forest covers an estimated 10,000,000 mou (1,500,000 ac.). Another 60,000,000 mou were reported suitable for afforestation.

Shansi has numerous scattered iron ore deposits, copper ores in the An-i region in the southwest, where one of China's salt sources also is found in Chieh-chih Lake (600,000-ton annual yield), and zinc and silver deposits in the north near Ta-t'ung. However, the province is most noted for being China's richest coal region, with about 47% of China's total coal or an estimated 400,000,000,000 tons, including coking and anthracite coals. The Ta-t'ung deposits lead among the six major fields in the province. Production is not as proportionately large. The more important iron mines are near Ch'ang-chih in the southeast, and at Yang-ch'uan, about 100 km. E of T'ai-yuan. Railroad communications connect T'ai-yuan with Ta-t'ung and Kalgan in the north, Shih-men in Hopeh Province in the east, and the Yellow River bend in the south. Ta-t'ung has rail connections with the industrial centre at Pao-t'ou in Inner Mongolia, and itself is the second most important industrial centre in Shansi, producing cement, locomotives, and mining machinery as well as coal. T'ai-yuan is the leading industrial centre (pop. [1957 est.] 1,021,000) and has one of the largest steel plants

China, as well as numerous chemical, machine, and iron smelting plants. Of the newer steel plants the largest is the Hsiang-feng, approximately equaling the T'ai-yuan plant in capacity. A phosphorus fertilizer plant with an annual capacity of 100,000 tons was completed in 1957. A high-voltage power line carries power to Yang-ch'uan along the railway eastward, while another line strung in 1957 carries high-voltage power to Fen-hsi in the Fen Valley to the south. Highway connections are poor and limited, but a major highway leads from T'ai-yuan to Ta-t'ung in the north, and other roads connect the Fen Valley with Ch'ang-chih in the east and the Yellow River in the west.

The provincial area has been part of the earliest heartland of Chinese civilization and was part of the state of Chin under the Chou dynasty. Its name originated during the Yüan period. From 1911 to 1949 it was governed for the most part by Yen Hsi-shan. Between 1937 and 1945 the Japanese occupied parts of the province, while Communist guerrillas harassed the Japanese from the T'ai-hang hills. (H. J. Ws.)

SHAN STATE in the Union of Burma consists of a group of formerly semi-independent states on the eastern frontier, which were formed into an administrative unit under the constitution of independent Burma introduced on Jan. 4, 1948. Pop. (1962 est.) 1,972,924. The group comprised the Shan states, inhabited by the Shan or Thai and other races, which in 1941 had a population of 1,699,000 and an area of 57,816 sq.mi. (149,744 sq.km.), and also the Wa states, with a population in 1941 of 82,614. Before Burma became independent the Shan states were governed by their hereditary chiefs, called sawbwas by the Burmese, under the control of the governor of Burma, but in 1922 they were federated for certain common purposes. Under the constitution of independent Burma the Shan State has a minister of the Burmese government at its head, aided by a state council comprising all the members of the Burmese Parliament representing the area; these members are elected by universal adult suffrage in the case of the lower house of Parliament and by the sawbwas in the case of the upper house. The state was given governmental authority in matters of agriculture, fisheries, public works, communications, police, administration of justice, education, public health, and local government, thus enjoying powers similar to those of a county council in the United Kingdom. The sawbwas retained their former administrative and judicial powers, subject to the minister and state council, until in October 1952 they were induced by the government to surrender their authority, retaining only their right to elect to the upper house of Parliament.

The Shan State is divided into three administrative areas, Southern, Northern, and Eastern, each under the charge of a resident, but the original states survive as minor administrative units. Taunggyi, 100 mi. (161 km.) SE of Mandalay, is the administrative centre of the state and the meeting place of the state council. Most other towns of the state are really large villages, formerly state capitals and now minor administrative centres, though they become very much alive on the periodic market days.

The former autonomous states are now mainly of historic interest. The 25 in the southern group ranged from Mong Nai (3,152 sq.mi. [8,164 sq.km.]) and Mong Pan (2,988 sq.mi. [7,739 sq.km.]) down to Kyong with a mere 24 sq.mi. (62 sq.km.) The northern group included the large states of North Hsenwi (capital Hsenwi), South Hsenwi (capital Mong Yai), and Hsipaw, or Thibaw, with considerable stretches of cultivated plain. The eastern group, east of the Salween River, includes Keng Tung (the capital Keng Tung can be reached by road east from Taunggyi but only in favourable weather), Manglun, Kokang, and the Wa country. The Wa are the most primitive of the peoples of the Shan State and until recently regarded headhunting as essential in their fertility rites, at plowing time, when rows of skulls were placed in avenues near villages. The Wa area is still only loosely administered.

Physical Features.—The Shan State borders Burma proper on the west, China on the north and east, Laos on the southeast, and Thailand on the south. The shape of the state is roughly that of a triangle, with its base near the plains of Burma and its apex on the Mekong River. The Shan Plateau is properly only the country between the deep trench of the Salween and the steep edge overlooking the Upper Irrawaddy and Sittang rivers. The average height of the plateau is between 2,500 and 4,000 ft. (750–1,200 m.), but it is seamed and ribbed by mountain ranges, which split up and run into one another. On the north the Shan State is barred by ranges following the line of the Myitnge River (Namtu). The huge mass of Loi Leng, 6,146 ft. (1,873 m.), projects south from this, and from each side of it and to the south is the wide plain extending to Mong Nai. The highest peaks are in the north and the south. Loi Leng is the highest point west of the Salween and in Kokang and other parts of North Hsenwi there are many peaks above 7,000 ft. (2,135 m.). The majority of the intermediate parallel ranges have an average height of between 4,000 and 5,000 ft. with peaks rising to over 6,000 ft. The country beyond the Salween is a mass of broken hills, ranging in the south from 2,000 to 3,000 ft., while in the north toward the Wa states they average from 5,000 to 7,000 ft. From December to March it is cool everywhere, and over five degrees (C) of frost is experienced on the open downs. The hot season temperature is 27 to 30° C (80–90° F), rising to about 38° C (100° F) in the Salween Valley. The annual rainfall varies from 60 in. (1,500 mm.) in the broader valleys to 100 in. (2,500 mm.) on the higher mountains.

The hill ranges rising from the plateau are usually forested. On the margins with Burma proper are important forest reserves with much teak but the Shan forests are best described as mountain forests dominated by evergreen oaks and, more locally, by pines such as *Pinus khasya*. The plains and valleys have much grassland and so, with their rolling character, are aptly described as downs. Over very large areas shifting cultivation is practised to the detriment of the forests; settled cultivation with rice is carried on when adequate water is available. Elsewhere cultivation with cereals including wheat and vegetables is limited to a few of the most fertile tracts.

Race and Language.—Shan State contains about two-thirds of all the Shans in Burma. The Thai or Tai, as they call themselves, were first known to the Burmese as Taroks or Tarets. The original home of the Thai people was China. Their first settlement in Burma proper was probably in the Shweli Valley, and from this centre they radiated at a comparatively recent date north, west, and southeast, through Upper Burma into Assam and into Thailand and Indochina. It is supposed that the group includes not only the Shans proper, the Laos, and the Thais or Siamese, but also the Muongs of Vietnam, the Hakkas of south China, and the Li, the inhabitants of the interior of the far eastern island of Hainan.

The Thai language may be divided into two subgroups, the north and the south. The south includes Siamese, Lao, Lu, and Hkun; the north, the three forms of Shan, namely North-Burmese Shan, South-Burmese Shan, and Chinese Shan with Hkamti and Ahom. The vernacular of the people who are known in Burma as Shan is South-Burmese Shan. This language is isolating and polytonic. It possesses five tones.

The Economy.—The Shans are a peaceful race, fond of trading. During the 20th century the trade with Burma increased considerably, especially after the construction of the Lashio (1903) and Shwenyaung (1928) railways. The huge silver-lead-zinc mines of the former Burma Corporation Ltd. are at Bawdwin in the northern part of the Shan State. Production declined for a time because of destruction suffered during the Japanese invasion and the frequent disruption of road and rail communications in the civil disturbances which often prevailed in Burma from 1948 on. The cultivation of wheat and potatoes in the southern areas for export to Burma proper is normally profitable but similarly suffered from interrupted communications.

Two railways penetrate into Shan State from the main Rangoon-Mandalay line. One from Thazi junction climbs the western edge of the plateau to the hill station of Kalaw, and across an extensive plain north of the large shallow Inle Lake (famed for its leg-rowers) to Heho and Shwenyaung at the foot of the ridge on which stands Taunggyi the capital. Farther north a branch line from near Mandalay climbs the scarp by a very scenic route

through Maymyo (*q.v.*) across a deep limestone gorge by the famed Gokteik viaduct to Lashio, chief town of the north. A narrow-gauge branch line runs to the Bawdwin mines through the smelting town of Namtu. Motorable roads follow along these two railways and extend eastward: through Taunggyi almost to the eastern limit of the state, with a ferry across the Salween River; from Lashio the famous Burma Road (*q.v.*) extends into China. There are also north-south roads, notably from Bhamo through Lashio and so southward to Loi-lem and the Thai border. Air services offer ready communication with Rangoon.

(B. R. P.; L. D. S.)

SHANTUNG (SHAN-TUNG SHENG), a densely populated coastal province of China between the Yellow Sea and the Gulf of Chihli, is bounded by Hopeh in the north, Honan in the west, and Kiangsu and Anhwei in the south. Area 59,189 sq.mi. (153,300 sq.km.); pop. (1957 est.) 54,030,000.

In early geological times its inland and peninsular hill masses comprised at least two large islands, while an inland sea or strait stood between them and the T'ai-hang hills of Shansi. Heavy sedimentation from the Yellow and Huai rivers filled this in to form the low-lying Yellow plain around the central hill mass. The province thus forms three physiographic regions: (1) the central hills, a triangular mass culminating in T'ai Shan (Mt. T'ai), elevation 5,069 ft. (1,545 m.); most of the hill lands rise to only about 1,300 ft., but several peaks reach over 3,200 ft. (2) The peninsular hills, with most elevations around 600 ft., but rising to 3,839 ft. (1,170 m.) in the Lao Shan northeast of Chiao-chou Bay and Tsingtao city. A submarine ridge connecting Shantung with Liaotung protrudes above the sea as the Ch'ang-shan archipelago. The islands and the indented peninsula favoured the development of sheltered harbours for fishing and commerce. (3) The plain which from the west slopes eastward to a northwest-southeast oriented trough at the western base of the central hills, and which in the north and south continues to the sea in the alluvial deposits of the Yellow River. The trough is occupied by several elongated lakes utilized in part by the Grand Canal cutting through western Shantung. The Yellow River in its present course flows past the northern foothills. Between the central hills and the peninsular hills is the corridor plain formed by sediments of the Pei-chiao and Wei rivers, over 62 mi. (100 km.) long and from 18 to 30 mi. wide.

History.—Western Shantung formed the easternmost parts of the earliest known organized state of China, the Shang (1520–1030 B.C.). Even earlier, an indigenous culture had developed there, as shown by a unique type of polished black pottery excavated at Ch'eng-tzu-yai. Southern Shantung as the state of Lu gave birth to Confucius and Mencius, honoured by a temple on T'ai Shan. The Lu kingdom, together with the Ch'i kingdom in northern Shantung, formed parts of the feudal domains of the Chou dynasty which lasted until the 3rd century B.C. Thus Shantung has from the beginning been part of the core area of the Chinese state. In more recent history Shantung symbolized China's humiliation by Japan and by European powers. It was occupied by Japanese forces during the Chinese-Japanese War of 1894–95. In 1898 Germany forced a lease of Tsingtao from which to exploit the province, while Great Britain similarly obtained a lease of Wei-hai for a naval base. The Boxer Rebellion developed out of Chinese resentment against foreigners in Shantung in 1899 and spread to other provinces in 1900. After World War I Japan tried to take over the German sphere of interest in Shantung and garrisoned troops there off and on between 1915 and 1929. When Japan attacked China in 1937, it was at T'ai-erh-chuang in southern Shantung that China won its first major victory over Japanese troops. In the Nationalist campaign against the Communists after World War II, the Communists were able to hold most of the province, and gained their first decisive victory in east China with the capture of the provincial capital of Tsinan in September 1948.

Climate and Agriculture.—In climate the peninsula is more moderate than the inland mass which shares the continental dryness and severity of north China winters. The northern part of the province is especially exposed to the northwest winds from Mongolia, but these winds pick up some moisture from the Gulf of Chihli and deposit some snow and rain on Shantung. The harbours freeze over at Chefoo and Lung-k'ou, and the January average for Chefoo is −4° C (24° F) while at Tsingtao it is 0° C (32° F). In summer when the inland plains have 35° C (95° F) temperatures, Tsingtao and Wei-hai only reach about 25° C (77° F). The precipitation has the usual summer peak, with south-facing slopes of the hill masses receiving over 30 in. (760 mm.) and the interior plains only about 20 in. (500 mm.). Some 1,200,000 wells (1951 estimate) provided the water for 85% of the land that is irrigated, although river water also is used. Expanded irrigation projects in 1956 provided water for some 17,930,000 mou (about 2,720,000 ac.) of additional irrigated land. The aggregate grain production (largely wheat, kaoliang, millets, and corn) increased greatly after 1949. Other crops include cotton, tobacco, and peanuts. Hogs are raised in large number.

The hills are capable of growing forests, but have been denuded during past centuries. Forests were being restored through afforestation in the second half of the 20th century, however, and along the coasts sand drift was being reduced by sand-fixing plants and trees. The convergence of warm and cool currents in the sea off Shantung is conducive to fish production. Tsingtao and Chefoo are the leading fishing centres.

Communications and Industry.—Communication facilities in the mid-20th century included more than 5,650 mi. (9,100 km.) of highways connecting all the 104 hsien (districts) of the province, but only about 930 mi. (1,500 km.) were all-weather roads. One major railroad runs from Tsingtao westward to Tsinan, the provincial seat, where it meets the north-south line from Tientsin to P'u-k'ou on the Yang-tzu. A rail line also leads from Tsingtao, the chief port, to Chefoo which is an embarkation port for travelers to Manchuria. Spur lines lead to the coal mining centres of Tzu-pu, Hsin-t'ai, and Tsao-chuang in the north-central, central, and southern hill areas respectively. The mining development at Tsao-chuang is one of the largest and most mechanized in China. There are numerous gold mines in the peninsular hills, the most important being in Chao-yuan district. The more noted iron ore mines are at Chin-tien-chen and Li-kuo-i, and important aluminum ore mines are at Nan-ting. The coastal salt industry is centred at Chiao Hsien.

Important industries of the province include cotton textiles, cigarette and match making, flour milling and oil pressing, brewing and distilling, machinery and chemicals; about 90% of the country's peanut oil was produced there in the mid-20th century.

(H. J. Ws.)

SHANTY (sometimes spelled CHANTY, a reflection of its putative derivation from the French *chanter*, "to sing"), in folk music, is a work song, usually of leader-chorus pattern. A group engages in a collective task, such as canoe paddling, net hauling, making sail, capstan turning, laying railway tracks, and road building. The group leader intones a line of song, and the group responds in chorus, usually singing a fixed refrain, heaving at a given point in the melody. Then follows another solo phrase, and again the concerted refrain-and-heave. By means of verbal improvisation and repetition, the shanty may be spun out for as long as the task endures; thus shanty texts are far more fluid than published versions would indicate. The function of such songs may be twofold: to synchronize the group's muscular effort, and to afford diversion from the wearisome nature of the task. The practice is ancient and widespread, being found in nearly all parts of the world, except where mechanical hoists, motor-driven winches, and such have made it superfluous. The Far East and Bantu Africa are notably rich in this kind of song.

In Britain and North America the term "shanty" is applied particularly to the songs used to accompany hauling, pumping, and capstan work aboard sailing ships. In English vessels, shanty singing rather fell into abeyance during the 17th century when merchant ships came, more or less, under naval discipline, but the practice revived vigorously in the 19th century when the newly formed shipping lines, each anxious to complete voyages more rapidly than its rivals, found that "a good shantyman was worth four extra hands on the rope." Thus, despite their primitive-seeming cast of melody and poetry, most of the well-known English-language shanties were evolved between 1825 and 1875, by

which time sail was giving way to steam and group work aboard ship was diminishing. (A. L. Ll.)

SHAPING MACHINE: *see* Machine Tools; Lathe; Woodworking Machinery.

SHAPLEY, HARLOW (1885–), U.S. astronomer, pioneer in studies of the structure of the Milky Way and of the universe beyond it, was born in Nashville, Mo., on Nov. 2, 1885. After studying at the University of Missouri and at Princeton, he joined the staff of the Mount Wilson Observatory in 1914, and later became professor of astronomy at Harvard University. From 1921–52 he was director of the Harvard College Observatory.

Shapley's most important early work concerned eclipsing binary stars. With H. N. Russell, in 1913, he devised practical methods for deducing the dimensions of the component stars of these systems from measurement of the light variation during eclipse. These methods, in essentials, remained the standard procedure for more than 30 years. Shapley showed that Cepheid variables cannot be eclipsing binaries, and was the first to advocate the theoretically important view that they are pulsating stars.

He studied the colours and magnitudes of stars in clusters, especially of those in the globular clusters, and was a pioneer in the use of variable stars (Cepheids and RR Lyrae stars) as distance indicators. He made the bold hypothesis, then revolutionary but later well established, that the globular clusters form a fairly symmetrical swarm about the centre of the Milky Way and that our sun must therefore be located in the outer parts of the Milky Way system, some tens of thousands of light years from the centre.

With the aid of large numbers of photographs of the sky, made both at Harvard and at the Harvard southern station in South Africa, Shapley studied the structure of the Milky Way and of nearby galaxies, especially the Magellanic Clouds, and the distribution of galaxies in space. He was one of the first to demonstrate that galaxies tend to occur in clusters and that the Milky Way is a member of a local cluster.

Shapley's work, characterized by bold imagination and drive, was mainly of the nature of first surveys. As he himself pointed out in the late 1950s, much further detailed investigation was required.

Shapley's works included *A Source Book in Astronomy*, with H. E. Howarth (1929), *Star Clusters* (1930), *Flights from Chaos* (1930), *Galaxies* (1943), *Inner Metagalaxy* (1957), *Of Stars and Men . . .* (1958). He edited *Climatic Change* (1954); *A Treasury of Science*, 3rd rev. ed., with S. Rapport and H. Wright (1954). He also wrote many technical papers, especially in the *Astrophysical Journal* and in the *Proceedings of the National Academy of Sciences*. (R. O. R.)

SHAPUR (Pahlavi Shahpuhr, "son of the king"; Greek Sapores; commonly Sapor), the name of three Sasanian kings of Persia.

Shapur I (according to T. Nöldeke, reigned A.D. 241–272; to W. B. Henning, 239–270, and to S. H. Taqizadeh, 241–273), son of Ardashir I and second in line of the Sasanian kings. The Persian legend that makes him the son of an Arsacid princess is not historical. Ardashir, toward the end of his reign, had renewed the wars against Rome inherited from the Arsacids. On his death Shapur continued the war, conquering Nisibis (Nusaybin) and Carrhae (Harran) and advancing deep into Syria. Defeated at Resaena in 243, he was yet able to conclude an advantageous peace in 244 because of the murder of the emperor Gordian III. In 256 he took advantage of the internal chaos within the Roman Empire to renew his attacks. Invading Syria, Asia Minor, and Armenia, he sacked Antioch but was repulsed by the emperor Valerian. However, in 260 Shapur not only defeated Valerian at Edessa but captured him and kept him a prisoner for the rest of his life. The capture of Valerian is a favourite subject of Sasanian rock carvings. In attacking the eastern Roman provinces, Shapur does not appear to have aimed at a permanent occupation. He merely carried off enormous booty both in treasure and men. The captives from Antioch were set to work to build the city of Gunde-Shapur (called in an inscription *Vahy-Andiok-Shapuhr*, "better than Antioch [has] Shapur [built]"), later to become famous as a centre of learning. Using the same captives, who excelled the Persians in technical skill, he built the dam at Shushtar known to this day as the Band-i-Qaisar, "the dam of Caesar."

Shapur was no longer content to describe himself as "king of kings of Persia" as his father had done: he styled himself "king of kings of Persia and Aniran (Aneran)," that is to say, of non-Persian territories as well. Possibly in connection with these secular imperialist ambitions Shapur appears to have tried to find a religion that would be suitable to all his subjects, whether of Persian stock or not. He showed marked favour to Mani, the founder of Manichaeism (*q.v.*), and at least one of his brothers adopted the Manichaean faith. He himself does not seem to have taken this final step, for the inscriptions show that he founded Zoroastrian fire temples both on his own behalf and on behalf of his wife and sons. He sought, however, to broaden the basis of the newly revived Zoroastrian religion by the addition of material derived from both Greek and Indian sources; there is evidence that he favoured the heretical "Zurvanite" wing of Zoroastrianism which derived both Ahura Mazda and Ahriman, God and the devil, from the principle of Infinite Time.

If the credit of the founding of the Sasanian empire belongs to Ardashir I, it was nevertheless Shapur who consolidated and expanded it. He thought of the new empire as a multinational state in which Manichaeism, drawing on Christian and Buddhist as well as on Zoroastrian sources, might serve as a unifying factor.

Shapur II (310–379), usually called "the Great," ninth of the Sasanian kings. At the death of Ormizd II, the Persian nobles killed his eldest son, blinded the second, and imprisoned the third. The throne was reserved for the unborn child of one of the wives of Ormizd. This was Shapur, who was thus born king.

With the adoption of Christianity by Constantine the Great and the subsequent christianization of the Roman Empire, Persia found itself saddled with what it considered to be a powerful Christian "fifth column." Thus Shapur resumed the policy of fierce intolerance toward non-Zoroastrians which had been initiated by the high priest Kartir in the reign of Bahram I, and singled out the Christians for persecution. He summoned a conference of all creeds, and decided in favour of one Adhurbadh, son of Mahraspand, a dualist Zoroastrian of strictly orthodox views. Thus Shapur II sought to unite his empire by enforcing Zoroastrian orthodoxy on frequently unwilling subjects.

His external policy was one of war and expansion; during his reign the insane and semipermanent struggle between the rival empires of Rome and Persia swayed in favour of the latter. In 337 Shapur broke the peace which had been concluded 40 years before between Narses and Diocletian and embarked on a 26-year war. Shapur attempted with varying success to reduce the great Mesopotamian fortresses of Singara, Nisibis, and Amida (Diyarbakir). Though the Roman emperor Constantius II was invariably defeated on the field, Shapur made little real progress. He was not strong enough to occupy permanently the conquered provinces because his eastern frontier was continually raided by central Asian tribes, prominent among which were the Chionites (*see* Hephthalites). After a prolonged struggle Shapur finally forced their king Grumbates to conclude a peace, whereupon the latter allied himself to Shapur in a renewed onslaught on the Roman power. With Grumbates' aid Shapur at last reduced both Amida and Singara (359–360). In 363 the emperor Julian advanced against him in force but was killed in battle. His successor, Jovian, was defeated and made to sign an ignominious peace by which he ceded the frontier marches on the Tigris and Nisibis and gave the Persians a free hand in Armenia. This victory is celebrated in the rock carvings near Shapur, not far from the modern Shiraz.

Shapur then turned his attention to the subjugation of Armenia. He captured the pro-Roman Arsaces III and imprisoned him in Susiana (Elam), where he committed suicide. His attempt to force Zoroastrianism on an Armenia which had only just accepted the Christian faith met with fierce and successful resistance. Arsaces' son Pap was smuggled out into Pontus and the Armenian resistance remained unbroken. The emperor Valens succeeded in restoring Pap in violation of the treaty concluded between Jovian

and Shapur, but Pap repaid him by poisoning the patriarch, Nerses, the very symbol of national resistance, and executed Cylax and Artaban, the leaders of the anti-Persian party. In so doing he may have been trying to preserve a precarious balance between Persia and Rome in Armenia. Be this as it may, he was treacherously murdered by the Romans in or about 374.

For five years Armenia was progressively weakened by an internecine struggle among its nobles. Shapur's last act was to win over Armenia by a discreet distribution of gifts. The country's autonomy under Persian suzerainty was recognized and the Christian religion respected. Thus Shapur II left Persia with its territory considerably augmented: he had pacified the eastern frontier and gained suzerainty over Armenia. The Persian empire had never been stronger; but war with Rome became a natural state of affairs, and its continuance did more than anything to undermine the strength of the Sasanian empire.

SHAPUR III (383–388), son of Shapur II, was raised to the throne by the nobles against his uncle Ardashir II, and killed by them after a reign of five years. He concluded a treaty with Theodosius the Great. (R. C. Z.; X.)

SHARECROPPING, though it is usually thought of as a form of crop-share renting, is essentially a labour contract involving payment in kind. The landowner furnishes the land, sources of power (such as a mule), usually a house, and sometimes food, clothing, and medical expenses. The sharecropper furnishes his labour, usually working under the close supervision of the landowner, who retains the right of possession and control of the land and crops. In most jurisdictions, the sharecropper is legally a labourer with a statutory lien for his labour upon the property of his employer.

Ordinarily the crops, usually cotton and tobacco, are shared 50–50, but they may be divided in other proportions. The cropper's share of production costs and advances for living expenses are deducted from his share of the crop.

With the introduction of modern machinery and technology, sharecropping has declined in importance. The furnishing of living expenses has largely disappeared, as have quasi-peonage and other questionable practices.

See METAYAGE SYSTEM; FARM TENANCY; SOUTH, THE: *Sharecropping.* (M. D. H.)

SHARI (CHARI), the principal river feeding Lake Chad in north-central Africa. It flows through the Republic of Chad and the Central African Republic. It is formed by a series of smaller rivers: the Bamingui, considered as the true headstream of the Shari, but in the dry season carrying little water, drains the region of Dar Kouti; the Gribingui rises south of Fort-Crampel and is a river of considerable size when it joins the Bamingui at Irena; the Ouham, or Bahr Sara, a long river coming from the Yade massif which at times flows through a great marshy valley, and in places is interrupted by rapids, particularly upstream from Batangafo at its confluence with the Fafa. The Ouham brings the greatest volume of water to the Shari. Near Fort-Archambault the Shari is joined on its right bank by the Aouk, the Bahr Kéita, and the Bahr Salamat, the waters of whose parallel courses mingle in an immense floodplain. The Bahr Salamat is the longest; its upper branch, the Bahr Azoum, rises in the volcanic hills of Darfur. Other rivers from the slopes of Hadjer Telfân and Massif de Guéra also join the Salamat. In its middle course the waters of Lake Iro drain into it and the river divides into numerous branches which spread into a delta and end in the Shari. The Aouk, which forms the boundary between the republics of Chad and Central Africa, also rises in Darfur in the Sudan and drains sluggishly an immense marshy plain. In the dry season it is sometimes reduced to a succession of stagnant ponds; it empties its flood waters into the Shari in an extensive delta.

Downstream from Fort-Archambault the Shari crosses the Gay Rapids, at the high ground near Niellim. From there on it is a large slow-running river, 3 to 4 mi. wide, which may reach 6 mi. in time of flood, and its lower reaches divide into many channels. It pushes west across the level plain toward the Logone, which for many miles flows parallel to it; the two floodplains are linked by the Ba Illi. At Fort-Lamy, where the Shari is no longer tortuous and frequently overflows its crumbling banks, the two rivers unite and flow on into Lake Chad (*see* CHAD, LAKE) through several distributaries including the Serbewel, the largest, the Ebeji-Ngada, and the Taftaf. Total length of the Shari between Fort-Lamy and Fort-Archambault is 869 mi. (1,398 km.). The Shari basin covers 231,600 sq.mi. (600,000 sq.km.) and the river unites the drainage of the southern part of the Chad inland drainage basin.

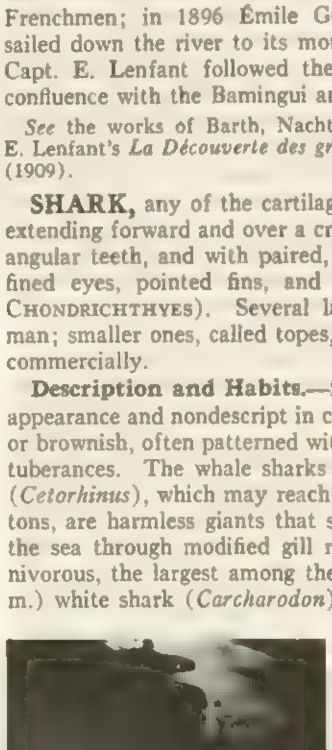
Steam navigation is possible for more than 680 mi. (1,100 km.) on the main rivers in the wet season, as far as Baïbokoum on the western Logone (Mbere), Goré on the eastern Logone (Pendé), Batangafo on the Ouham, and Irena on the Gribingui. The average annual rate of flow at Fort-Lamy is 1,230 cu.m. per sec., but it may exceed 4,500 cu.m. per sec. in time of flood.

The existence of the Shari was made known by Dixon Denham, Hugh Clapperton (*qq.v.*), and Walter Oudney, the first Europeans to reach Lake Chad (1823). In 1852 Heinrich Barth (*q.v.*) spent some time in the region of the lower Shari and Logone, and in 1872–73 Gustav Nachtigal (*q.v.*) studied the hydrographical system of these rivers and explored the Gribingui, which he called the Bahr el Ardhe. The most prominent explorers have been Frenchmen; in 1896 Émile Gentil reached the Bamingui and sailed down the river to its mouth; in 1904 an expedition under Capt. E. Lenfant followed the Ouham from its source to the confluence with the Bamingui and discovered the eastern Logone.

See the works of Barth, Nachtigal, and other travelers, especially E. Lenfant's *La Découverte des grandes sources du centre de l'Afrique* (1909). (J. D.)

SHARK, any of the cartilaginous fishes with a pointed snout extending forward and over a crescentic mouth set with sharp triangular teeth, and with paired, largely lateral, gill slits, well defined eyes, pointed fins, and a muscular, upturned tail (*see* CHONDRICHTHYES). Several larger kinds can be dangerous to man; smaller ones, called topes, hounds, and dogfishes, are fished commercially.

Description and Habits.—Shark species are often similar in appearance and nondescript in colour, varying from gray to bluish or brownish, often patterned with spots, bands, marblings, or protuberances. The whale sharks (*Rhincodon*) and basking sharks (*Cetorhinus*), which may reach 50 ft. (15 m.) and weigh several tons, are harmless giants that subsist on plankton strained from the sea through modified gill rakers. All other sharks are carnivorous, the largest among them being the voracious 40-ft. (12 m.) white shark (*Carcharodon*), which attacks seals, sea turtles, large fish, and, on more than one occasion, man (see *Hazards to Man*, below). The sluggish Greenland shark (*Somniosus*) of cold, deep waters has been known to feed on seals, reindeer, large fish, and whales. Normally sharks feed on fish, often attacking in schools; open-ocean species such as the mackerel, mako, and thresher sharks frequently feed near the surface, and are much sought after by rod and reel sportsmen as prime game fish. Beautifully streamlined and powerful swimmers, these open-ocean sharks are adept at feeding on fast tuna, marlin, etc. Bottom-feeding sharks are stout, blunt-headed forms, tending to flattened bodies with well developed spiracles and larger eyes; the shellfish eaters have coarse, pavementlike crushing teeth. By far the oddest looking sharks are the hammerheads (*Sphyrna*), whose heads resemble double-headed hammers



BY COURTESY OF NEW YORK ZOOLOGICAL SOCIETY

SAND SHARK (CARCHARIAS TAURUS)

and have an eye on each stalk. So characteristic are shark teeth that it is possible to identify fossil and modern species from one or two teeth.

The origin of sharks is obscure, but their geological record goes back to pre-Devonian times (more than 320,000,000 years ago). Their present diversity of forms and numbers, estimated at some 300 species, equals their past abundance; many recently found species have long been known from the fossil record.

The living sharks (order Pleurotremi) fall into three suborders. The first is the Notidanoidea, a small group of primitive sharks distinguished by six or seven gill clefts and one dorsal fin; examples are the frilled (*Chlamydoselachus*) and six-gilled (*Hexanchus*) sharks. These are deepwater fishes; the former are eel-shaped and may reach over 6 ft. (2 m.); the latter attain a length of over 25 ft. (8 m.). The suborder Squaloidea is characterized by five, rarely six, gill clefts and two dorsal fins preceded by spines. This group includes the bullheaded, or horn, shark (*Heterodontus*); the spiny dogfish (*Squalus*), a small shark of temperate seas, which travels in schools and is a pest to commercial fisheries; the bramble shark (*Echinorhinus*), a bulky shark with spiny tubercles on the skin; and the Greenland shark, which grows to 20 ft. (6 m.). The specialized saw sharks (*Pristiophorus*) and the bottom-dwelling angel sharks (*Squatinidae*) with elongated bodies and muscular tails are also of this suborder. The third suborder, the Galeoidea, are typified by five gill clefts, two spineless dorsal and one anal fin. The bulk of the sharks, and many aggressive representatives of the requiem sharks (Carcharinidae), feared because of their attacks on man, are associated with the galeoids. This suborder also includes the families Odontaspidae (the sand and elfin sharks), Lamnidae (man-eater, porbeagle, basking, mackerel, and thresher sharks), Carchariidae (topes, hounds, blue, and hammerhead sharks), Scyliorhinidae (dogfishes), and Orectolobidae (carpet, tiger, nurse, and whale sharks).

Although chiefly marine, sharks often enter fresh water; requiem sharks of the genus *Carcharhinus* occur commonly in the Zambezi, Tigris, and Ganges as well as other rivers in subtropical and tropical areas. The Lake Nicaragua shark is a landlocked form found in Central American lakes. *Carcharhinus* species are small to medium sized, but are aggressive, persistent, and generally feared.

Hazards to Man.—In Australia, South Africa, and elsewhere along coasts exposed to the shark nuisance, public beaches are often provided with lookout towers, bells and sirens, or nets—proof of public concern toward the shark hazard. The most feared species is the great white shark, whose erratic presence in United States coastal waters has given rise to particularly distressing attacks in Buzzard's Bay, Mass., off the New Jersey shore and, with most frequency, along the California coast. Other sharks involved in attacks on humans are the tiger (*Galeocerdo*), blue (*Prionace*), gray nurse (*Carcharias*), and hammerhead.

Of course, the larger the shark the more formidable the attack, but several small specimens can be equally hazardous, a fact well attested to by wartime sea survivors. The zone of greatest hazard lies approximately between latitudes 40° N and 40° S and coincides with the warmer months, when average surface sea temperatures vary between 60° and 70° F (16° and 21° C). Fatalities are greatest in late afternoon; men have been attacked more frequently than women. The majority of attacks are on bathers in shallow water (between 2 and 3 ft. [0.6 to 1 m.]); however, participants in underwater activities also risk injury from sharks in warm seas. (The above information, however, may reflect the habits of bathers as well as the habits of sharks; bathers are most numerous under the conditions listed.) Injury is greatest on the lower limbs and buttocks and, secondarily, on the forearms and hands. Mortality is as high as 70%; this is largely due to hemorrhage and shock, controllable through first-aid measures. Increasing public interest in seashores for recreation and sport emphasizes the need for uniform safety measures to prevent loss of life. The incidence of attack by sharks is no greater than that for land carnivores, but the probability of a rise in shark attacks is greater since increasingly greater numbers of people, with little realization of marine hazards, are entering the sea during the season when sharks are closest to the shore. Shark repellents are constantly being sought; several, like "Shark Chaser," consisting of copper acetate and a black marker dye, nigrosine, in a water-soluble wax, appear to be some help in repelling some species.

Economic Importance.—Shark fishing is a relatively minor industry of local importance in a few regions. Shark-liver oil, high in vitamin A, has been used for pharmaceutical and proprietary medicines when supplies of cod-liver oil, formerly a chief source of the vitamin, have been reduced. Lower-grade shark oils have industrial application in tanning and other industries. Shark skin is the source of shagreen, a very durable leather for which there is only a very limited demand. Spiny (*Squalus*) and smooth (*Mustelus*) dogfishes serve as objects of anatomical dissection in many school laboratories of zoology. Past attempts in the United States to popularize spiny dogfish (*Squalus acanthias*) flesh as edible "gray fish" proved unsuccessful. However, the use of shark flesh elsewhere is more widespread; large quantities are sold in Europe, South America, and South Africa. In Latin-American countries sun-dried or salted shark flesh is shipped inland.

The school shark (*Galeorhinus australia*), originally fished for vitamin A in Australia, is now processed into fresh shark fillets. Several million pounds annually are sold as "flake," the principal source of fish in "fish and chips"; the Australian government has been concerned with the dwindling number of this commercially valuable species in its territorial waters. (Strangely enough, regardless of where shark flesh is sold, it rarely is described as a shark product.) Sharks are also used as bait for lobsters and crabs, and considerable amounts of shark meat are processed for fertilizer and for livestock feed.

See also references under "Shark" in the Index.

BIBLIOGRAPHY.—V. M. Coppleson, *Shark Attack* (1958); G. A. Llano, *Airmen Against the Sea*, Air University USAF (1955); "Sharks *v.* Men," *Scientific American* 196(6):54–61 (1957); J. R. Thompson and S. Springer, "Sharks, Skates, Rays and Chimaeras," U.S. Fish and Wildlife Service *Circular 119* (1961). (G. A. LL.)

SHARP, CECIL JAMES (1859–1924), English musician whose work as a collector of English folk song and dance had an important influence on 20th-century British music, especially on Gustav Holst and Ralph Vaughan Williams, was born in London on Nov. 22, 1859. Educated at Uppingham and Cambridge University, he was self-taught as a musician. In 1882 he emigrated to Australia, where, after some false starts, he became assistant organist of Adelaide Cathedral and co-director of the Adelaide College of Music. In 1892 he returned to England and obtained posts as music master at Ludgrove School (1893–1910) and principal of the Hampstead Conservatoire (1896–1905).

In 1903 Sharp discovered that an unsuspected wealth of native folk song survived in England. Although work in this field had already been begun, the publication of Sharp's collection *Folk Songs from Somerset* (1904–09) and of his study *English Folk Song: Some Conclusions* (1907; 3rd ed. rev. by M. Karpeles, 1954) led to a new and widespread interest in English folk music. He was invited to lecture and to teach the songs, and later the dances, which he had begun to collect in 1905; and his life's purpose became the collection and performance of folk music. In 1911 he founded the English Folk Dance Society (later to be amalgamated with the Folk Song Society); he demonstrated and taught annually at the Vacation School of Folk Song and Dance at Stratford-upon-Avon; and he initiated the teaching of folk song and dance in English schools.

Between 1916 and 1918 Sharp three times visited the United States to teach and to collect songs of English origin. His reputation in the U.S. was established by the publication, with Olive Dame Campbell, of *English Folk Songs from the Southern Appalachians* (1917; enlarged and edited by M. Karpeles, 1932; rev. ed., 1960). In 1923 he received the honorary degree of Mus.M. at Cambridge. He died in London on June 23, 1924.

Sharp's other published works include *English Folk Songs* (1921; reprinted 1959) and teaching handbooks of morris, sword, and country dances.

See A. H. Fox-Strangways and M. Karpeles, *Cecil Sharp* (1933; rev. ed., 1955). (M. J. D.-S.)

SHARP, JAMES (1618–1679), archbishop of St. Andrews from 1661, who was largely responsible for the reintroduction of

episcopacy in Scotland after the restoration of Charles II, was born on May 4, 1618. A native of Banffshire, he grew up in the episcopalian and royalist tradition, and after the presbyterian triumph and the abolition of episcopacy in Scotland in 1638, went to England. Returning in 1643, he taught at St. Andrews University, later becoming minister of Crail, in Fife (1649). The presbyterians were divided into two factions, and Sharp sided with the more moderate and royalist "Resolutioners" against the extreme "Protestors." In 1657, as the spokesman of the Resolutioners before Oliver Cromwell in London, he showed ability that won him much esteem.

Sharp took part in the negotiations leading to the restoration of Charles II, and in 1660 again represented his party in London. He soon saw that the drift in England toward the restoration of episcopacy would affect Scotland. Being leader of the moderate party, he was consecrated archbishop of St. Andrews on Dec. 15, 1661. Most of the Resolutioners acquiesced in episcopacy, but to the Protestors and all who adhered to presbyterianism Sharp was a scheming traitor. He therefore bore the odium for many unpopular government measures (*see* SCOTLAND: *History*), but the truth is that his determination to retain his position made him merely compliant with the Privy Council, whatever its policy. Thus, although Sharp disapproved of the Act of Supremacy of 1669, giving the crown supreme power in all church affairs in Scotland, he did not imitate Archbishop Alexander Burnet of Glasgow, who was deprived because of his opposition.

An attempt was made to assassinate Sharp in Edinburgh on July 11, 1668. A presbyterian, James Mitchell, arrested in 1674, confessed to the crime on a promise of his life, but was nevertheless executed in 1678. This affair intensified antipathy to the archbishop. On May 3, 1679, while driving to St. Andrews with his daughter, he was murdered on Magus Muir, 3 mi. W of St. Andrews, by a band of fanatical presbyterians. (GN. D.)

SHARP (IN MUSIC): *see* ACCIDENTALS.

SHARPEY-SCHÄFER, SIR EDWARD ALBERT (1850–1935), British physiologist and inventor of the prone-pressure method of artificial respiration (*q.v.*), was born at Hornsey, Middlesex, on June 2, 1850. He was educated at University College, London, where he met William Sharpey, the professor of general anatomy and physiology, and in 1918 he prefixed Sharpey's name to his own in order to perpetuate the name of his teacher. He became Jodrell professor at University College (1883), Fullerian professor at the Royal Institution (1878–81), and professor of physiology at Edinburgh University (1899–1933). Sharpey-Schäfer made important contributions to the knowledge of cerebral localization, of the heart sounds, and of muscular contraction. In 1894 (with George Oliver) he demonstrated the existence in the adrenal glands of a substance (adrenaline) which raised the blood pressure, and this stimulated research on the internal secretions (hormones) of the ductless glands. Sharpey-Schäfer's method of artificial respiration was adopted by the Royal Life Saving Society, and he received many honours, including the presidency of the British Association (1912) and a knighthood (1913). He died at North Berwick, East Lothian, on March 29, 1935.

(W. J. BP.)

SHARQIYAH, ASH, a *muhafaza* (governorate) of Lower Egypt (United Arab Republic) in the east of the Nile Delta, reaches the Mediterranean Sea in the northeast and is bounded by the Suez Canal Zone in the east, by the Eastern Desert (As Sahra' ash Sharqiyah) in the south, and by the *muhafazat* of Al Qalyubiyah in the southwest, and Ad Daqahliyah in the west. Area 1,815 sq.mi. (4,700 sq.km.); pop. (1960) 1,819,798, giving a density of 1,003 persons per square mile. The capital and only large town is Az Zaqaziq (Zagazig).

Much of the area's development dates from the mid-19th century with the spread of cotton growing in the Nile Delta. New canals, especially the Ismailia or Sweet Water Canal (1863), carrying fresh water from the Nile to the Suez Canal Zone, converted the eastern part of the territory from desert into fertile farmland. However, to the northeast contiguous with Lake Manzilah there lies a large area of saline marsh. Reclamation of this marsh and the lake area proceeded slowly in the 1960s, with rice as the main crop of the reclaimed farmland. The principal crops are maize (corn), barley, cotton, wheat, and rice. Peanuts and sesame are important in the sandy soils to the east. About 65% of the occupied population is engaged in agriculture. The principal water for irrigation is obtained from the Ismailia and Muways canals. Fisheries are important in Lake Manzilah. The principal industries are related to the agricultural resources. There are no mineral resources. (A. B. M.)

SHASI (SHA-SHIH), a commercial city and water transportation centre on the Yangtze in central Hupeh Province, China. Pop. (1953) 85,800. Called "Little Hankow" after the most important commercial port of central China about 120 mi. (190 km.) eastward, Shasi is the centre for numerous canal connections, one of which shortens to a third the circuitous Yangtze River distance to the Wu-han metropolitan area. Across the Yangtze and southward, the T'ai-p'ing and Ou-ch'ih canals lead to Tung-t'ing Lake in Hunan. Between these two canals a flood-detention basin covering 350 sq.mi. (900 sq.km.) was constructed in 1952–53. Because of its situation in the cotton-growing region of Hupeh, cotton and cotton seed, as well as other produce such as rape seed, are exported through Shasi. The city also is a distribution point for upriver Szechwan salt. Cotton textile and flour mills west of the city form two of its most important industries. Shasi has no rail connections but a motor highway runs south from across the Yangtze around Tung-t'ing Lake to Ch'ang-sha, and another highway runs north to Hsiang-yang on the Han River.

Shasi was flourishing as early as the T'ang dynasty, and superseded in the 1870s contiguous Chiang-ling (formerly Ching-chou, which dates far earlier) as the local commercial centre. Shasi was one of the treaty ports opened in 1896. (H. J. WS.)

SHASTA, a Hokan-speaking group of North American Indians, lived on and south of the Klamath River in California, east of Mt. Shasta. Culturally they stood intermediate between the northwestern tribes, such as the Yurok and Hupa (*qq.v.*), and the Maidu (*q.v.*) and other central Californians. When first encountered by whites (early 19th century) the Shasta were sedentary villagers who collected roots and seeds, relying largely on salmon for animal food. Houses were of wooden planking and built about half below ground level. They used dugout canoes, spears, bow and arrow, weirs, and nets. Numbering an estimated 2,000 in 1770, they dwindled to a few dozen scattered households in the second half of the 20th century; two Shasta communities with a combined population of 130 were reported from northern California in the 1960s. *See* also HOKAN.

See J. R. Swanton, *The Indian Tribes of North America* (1953); F. W. Hodge (ed.), *Handbook of American Indians North of Mexico* (1959).

SHASTRI, LAL BAHADUR (1904–1966), Indian leader, prime minister of India in succession to Jawaharlal Nehru, was born on Oct. 2, 1904, at Mughalsarai in the United Provinces (later Uttar Pradesh). He went to school in Benares (Varanasi) until the age of 17, when he joined the noncooperation movement and was imprisoned for a short time. On his release he studied in the Kashi Vidyapith, a school specializing in religious and nationalist instruction, where he was awarded the title of "Shastri" (learned in the scriptures). He then returned to politics, was imprisoned several times, and attained influential positions in the local Congress Party hierarchy. He was elected to the United Provinces legislature in 1937 and again in 1946. After Indian independence Shastri gained experience as minister for home and transport in Uttar Pradesh. He was elected to the central Indian legislature in 1952 and became union minister for railways and transport. He resigned in 1956 because he felt that he was constitutionally responsible for a serious railway accident, but returned to office in 1957. He gained a reputation as a skilful mediator after his appointment to the influential post of minister for home affairs in 1961, but he resigned in 1963 under a plan whereby ministers should voluntarily leave office to devote themselves to the reform of the Congress Party. In January 1964, on Jawaharlal Nehru's illness, he was recalled to the cabinet as minister without portfolio, and after Nehru's death he became prime minister (June 9, 1964). Before the end of that month, he suf-

fered a heart attack. He was of slight, almost frail, physique, and there were fears that he might not be able to stand the strain of his office. He was criticized for failing to deal effectively with economic problems but won great popularity for his firmness on the outbreak of hostilities with Pakistan (1965). He died of a heart attack at Tashkent, in Soviet Central Asia, on Jan. 11, 1966, after signing a "no-war" agreement with Pres. Ayub Khan of Pakistan.

Shastri was unassuming and unpretentious in manner, and devoid of doctrinaire views. His strength lay in the general recognition of his administrative competence and in the respect for his upright character. *See* also INDIA: *History*.

(KE. A. B.)

SHAVUOTH (SHABUOTH): *see* JEWISH HOLIDAYS.

SHAW, GEORGE BERNARD (1856–1950), the most important British comic dramatist since Congreve, was born in Dublin on July 26, 1856. His ancestry lay among the Protestant landed gentry, but his father was first a minor civil servant and later an unsuccessful wholesale merchant, so that he grew up in an atmosphere of polite urban indigence. He largely rejected the indifferent education offered by the schools he attended (beginning in 1867 with the Wesleyan Connexional School—later Wesley College), but developed a wide knowledge of music, art, and literature, as a result of home influence and visits to the National Gallery of Ireland. By the time he was 16, he was working as a land agent's clerk. In 1872 his mother virtually abandoned her husband, and took her two daughters to London, following George Vandaleur Lee, her singing teacher (with whom the Shaws had shared a house in Dublin and a cottage at Dalkey since *c.* 1866), upon whom her hopes of a musical career depended.

Early Years in London.—In 1876 Shaw resolved to become a writer, and joined her and his elder sister (the younger having died). He thus left Ireland for 29 years, returning then only for a visit. The emotional temperature of the family seems to have been low, and Shaw and his mother took little interest in each other. But they lived together until he married in 1898, the household making do on her earnings as a music teacher, and on £1 a week from her husband in Dublin, augmented, after 1877, by Shaw's share of a family bequest.

Shaw wrote five novels in steady succession: *Immaturity* (written 1879, publ. 1930); *The Irrational Knot* (serialized 1885–87; as a book, 1905); *Love Among the Artists* (1887–88; as a book, Chicago, 1900); *Cashel Byron's Profession* (1886, several times revised; a dramatic version, *The Admirable Bashville*, being appended to the 1901 edition); and *An Unsocial Socialist*, the last novel completed, and the first to be printed (serially, 1884; as a book, 1887). A fragment of a sixth novel, begun in 1887, was published in 1958. It is now possible to find much that is agreeably "Shavian" in these works. But they had little success, and for a decade Shaw's industry earned him less than 10*s.* a year. He read in the British Museum, became a vegetarian, joined the recently founded Fabian Society (*q.v.*) in 1884, and began his long career as a public speaker.

THE GRANGER COLLECTION

GEORGE BERNARD SHAW, PHOTOGRAPHED IN HIS GARDEN AT AYOT ST. LAWRENCE, ENGLAND

Eventually, in 1885, William Archer (*q.v.*), whom he had met in the British Museum reading room, found him employment as a book reviewer on the *Pall Mall Gazette* and art critic on the *World* (1886–89); and then as a writer on music in the *Star* (1888–90) and the *World* (1890–94). He had a good acquaintance with music, particularly opera, and eked out his knowledge with a brilliance of digression that makes many of his notices permanently delightful. His first real mark, however, was made as drama critic on the *Saturday Review*, to which he was recruited in 1895 under the editorship of Frank Harris.

Marriage and Success.—By this time Shaw was a playwright, and his energies began to concentrate themselves on drama. In 1898, after an illness caused by overwork, he married Charlotte Payne-Townshend (d. 1943), an Irishwoman of good family who combined progressive political views with considerable fortune. He had met her when staying with his friends Sidney and Beatrice Webb, and she had nursed him during his illness. Shaw was not in the common sense a passionate man, and what may be called the two love affairs of his life, with Ellen Terry and Mrs. Patrick Campbell (*qq.v.*), seem to have been mainly conducted by letter.

His marriage established him in a settled mode of living. Although he remained an active Socialist, he was henceforward essentially a dramatist. He was fond of declaring that he was "a social reformer and doctrinaire first, last, and all the time." But it was the theatre that he loved, although he insisted on surrounding his work with manifestos, and on importing into it a strong polemical tone. He was in every sense a showman, and kept himself in the public eye by a sort of intellectual clowning. When he died at 94 (at Ayot St. Lawrence, Hertfordshire, on Nov. 2, 1950), he had long been the greatest living man of letters in the English-speaking world.

In 1925 he had been awarded the Nobel Prize for literature (using the money to establish the Anglo-Swedish Literary Foundation for translation of Swedish literature into English); his plays, often popular in the United States before they were appreciated in England, were performed in most European countries; his political writings—notably in *Fabian Essays in Socialism*, which he edited for the Fabian Society (1889); *The Common Sense of Municipal Trading* (1904; new ed. 1908), a result of his experience as vestryman and borough councilor of St. Pancras, London (1897–1903); *Common Sense About the War*, published (November 1914) as a supplement to the *New Statesman*, which he had helped to found in 1913; works on Irish nationalism (1917 and 1920), on the League of Nations (1929), and *What I Really Wrote About the War* (1931)—were widely read and influential; and his personality had made an unusually wide impact. He conducted a large correspondence—often on postcards—and, although burdened by a marked constitutional shyness, made many friendships with notable contemporaries (*e.g.*, T. E. Lawrence). His will revealed the breadth of his interests. What remained of his fortune of more than £300,000 after death duties were paid was left in trust for research on a "Proposed British Alphabet" of at least 40 letters; the residue to be divided between the National Gallery of Ireland, the British Museum, and the Royal Academy of Dramatic Art. Shaw's Corner, at Ayot St. Lawrence, his home since 1906, was left to the National Trust.

The Plays.—*"Unpleasant and Pleasant."*—When Shaw began writing for the English stage its most prominent dramatists were Sir A. W. Pinero and H. A. Jones (*qq.v.*). Both were concerned to develop a modern realistic drama; neither had the power to break away from the type of threadbare artificial plot expected by theatregoers. (*See* also DRAMA: *Modern Drama.*) The poverty of this sort of drama had become apparent in 1889, when Ibsen's *A Doll's House* was played in London; his *Ghosts* followed in 1891, and the possibility of a new freedom and seriousness on the English stage was introduced. Shaw, about to publish *The Quintessence of Ibsenism* (1891; completed, 1913, to Ibsen's death), rapidly refurbished an abortive comedy, *Widowers' Houses*, as a play recognizably "Ibsenite" in tone, making it turn on the notorious scandal of slum landlordism in London. The result (performed 1892), in both its strength and its limitations, is characteristic of much that was to follow. A well-intentioned young Englishman falls in love, and then discovers that his prospective father-in-law's fortune and his own private income derive from exploitation of the poor. Potentially this is a tragic situation, but Shaw seems

to have been always determined to avoid tragedy. The unamiable lovers do not attract sympathy; it is a social evil and not a romantic predicament on which attention is concentrated; and the action is kept well within the key of ironic comedy. The same dramatic predispositions control *Mrs. Warren's Profession,* written in 1893 but not performed until 1902, as the lord chamberlain refused it a licence. Its subject is organized prostitution, and its action turns on the discovery by a well-educated young woman that her mother has graduated through the "profession" to become part-proprietor of brothels throughout Europe. Again, the economic determinants of the situation are emphasized, and, although there is real horror for Vivie Warren in what is revealed to her, she is endowed with a certain hardness and coldness which ensure that her private dilemma shall not engage us at the expense of the public issues involved. The play is, within limits, a drama of ideas; the vehicle by which these are presented is essentially one of high comedy; here is something at once horrible and steadily entertaining.

Shaw called these plays "unpleasant," because "their dramatic power is used to force the spectator to face unpleasant facts"; he followed them with four "pleasant" ones. Both groups were revised and published in *Plays Pleasant and Unpleasant* (1898). The first of the second group, *Arms and the Man* (1894), has a Balkan setting and makes lighthearted, though sometimes mordant, fun of romantic falsifications of both love and warfare. The second, *Candida* (performed 1895), has more substance, and was important for English theatrical history, for its successful production at the Royal Court Theatre in 1904 encouraged Harley Granville-Barker (*q.v.*) and J. E. Vedrenne to form a partnership which resulted in a series of brilliant productions there. The play represents its heroine as forced to choose between her clerical husband —a worthy but obtuse Christian Socialist—and a young poet who has fallen wildly in love with her. She chooses her confident-seeming husband because she discerns that he is actually the weaker. The poet is immature and hysterical, but, as an artist, has a capacity to renounce personal happiness in the interest of some large creative purpose. This is a significant theme for Shaw; it leads on to that of the conflict between man as spiritual creator and woman as guardian of the biological continuity of the race which is basic to *Man and Superman.* In *Candida* such speculative issues are only lightly touched on, and this is true also of *You Never Can Tell* (performed 1899), which ends in a romp and is often farcical in tone. But when the hero and heroine, who believe themselves to be respectively an accomplished amorist and an impregnably rational New Woman, find themselves in the grip of a vital force which takes little account of these notions, we are again brought directly into contact with Shaw's emerging philosophy. This play is a light Gilbertian comedy given a novel twist.

"The Devil's Disciple."—The Devil's Disciple (performed New York, 1897; publ., in *Three Plays for Puritans,* with *Caesar and Cleopatra* and *Captain Brassbound's Conversion,* 1901) is melodrama similarly refreshed. It is set in New Hampshire during the American Revolution, and its hero, Dick Dudgeon, the black sheep of his family, is represented as heroically taking the place of a minister whom the English propose to hang. His motive is not (as the minister's wife supposes) romantic, but public and impersonal. Moreover, he acts spontaneously and at the command of some unquestioned imperative, so that his virtue is seen less as an achievement than as a profession: he is thus, surprisingly, distantly related to the characters in Corneille's dramas. Despite its serious implications, the play abounds in wit. The English General Burgoyne murmurs as the time of execution approaches: "We must not detain Mr. Dudgeon." And when the minister rushes in at the first stroke of 12, thanking God that he is in time, Burgoyne glances at his watch and says he would never dream of hanging any gentleman by an American clock.

"Caesar and Cleopatra," "Captain Brassbound," "Man and Superman."—Caesar and Cleopatra (performed 1906), Shaw's first great play, is a study in magnanimity, and outstandingly successful because the eminence upon which Caesar is set is never allowed to become a pedestal. When the great library of Alexandria goes up in flames and an agonized scholar cries out: "What is burning there is the memory of mankind," this lonely and austere man replies: "A shameful memory. Let it burn." But he does his best to educate Cleopatra, a spoiled and vicious child—and, moreover, he moves easily (and here consummate dramatic tact is at work) amid a crowd of minor characters who are simply and triumphantly amusing. In *Captain Brassbound's Conversion* (1900), by an extension of this technique, the voice of reason and good sense is set in the mouth of an aristocratic Englishwoman who is nevertheless a delightful figure of fun. The play is a sermon against various sorts of folly masquerading as duty and justice, but there is scarcely a speech that does not exhibit character or motive in an entertaining light. And this is true—except for a long and excrescent third act—of *Man and Superman* (publ. 1903, performed 1905). In outline it is merely the story of a ruthless and resourceful girl "out to get her man," but this is made the medium of sustained and scintillating debate upon the relationship of the sexes as it might variously appear in the light of sundry persuasions—moral and scientific, traditional and experimental—actually at play in the speculative climate of the early 20th century. Its hero, Jack Tanner, is the most intellectually resilient of Shaw's characters. At the same time—and particularly in the discursive third act—we may feel that he is being employed to explore the operatic rather than the intellectual boundaries of drama. Shaw himself declared, indeed, that opera had taught him to shape his plays into recitatives, ensemble finales, and the like, so as "to display the technical accomplishments of the executants." Behind *Man and Superman* there is still a Shaw belonging to the world of Karl Marx and Henry George, Sidney Webb and H. G. Wells, who believes that the human lot is to be ameliorated, and, indeed, the destiny of the human race to be determined, by purposeful activity at the level of enlightened legislation. But already there are doubts. The inverted "marriage by capture" with which the piece ends, when Ann claims Jack Tanner, scarcely suggests that controlled scientific mating is just round the corner. And further thought on the nature of evolution was to lead Shaw to very different conclusions.

"John Bull's Other Island" and "Major Barbara."—John Bull's Other Island (1904; publ., with *Major Barbara* and *How He Lied to Her Husband,* 1907) was intended for the Dublin Abbey Theatre, and its main aim was to exhibit to an Irish audience a comic Englishman as absurd as the many comic Irishmen on the London stage. Its background in Anglo-Irish relations has necessarily proved ephemeral, and, although it is often wildly amusing, it is not well-constructed. It has, however, a point of growth in the minor character of Father Keegan, in the portrayal of whom Shaw's interest in the religious consciousness emerges strongly. This is developed in the far richer *Major Barbara* (1905). Here the heroine, who has joined the Salvation Army, finds that her father, an armaments manufacturer, holds tenets which are in a way themselves religious. Barbara ranks next to Joan of Arc among Shaw's women, and some scenes are more moving than any that Shaw had hitherto allowed himself to write. On analysis the play may lack intellectual coherence, but it gains in dramatic vitality from the freshness of the doubts that lie behind it.

"The Doctor's Dilemma" to "Pygmalion."—Shaw now entered on a period of somewhat relaxed achievement. *The Doctor's Dilemma* (1906, publ., with *Getting Married* and *The Shewing-up of Blanco Posnet,* 1911) is a successful light satire on the medical profession, knit to a less convincing representation, in the consumptive painter, Dubedat, of the strengths and frailties of the artistic temperament. Shaw, himself a great creative artist, is seldom at his best in depicting artists. *The Shewing-up of Blanco Posnet* (1909) is a one-act play on a theme similar to that of *The Devil's Disciple. Getting Married* (1908) and *Misalliance* (1910, publ., with *Fanny's First Play* and *The Dark Lady of the Sonnets,* 1914) are dramatized debates on marriage and on parents and children, respectively. *Fanny's First Play* (1911) and *Overruled* (1912) are light exercises on familiar Shavian themes, rapidly run up upon tenuous artificial situations uncommon in Shaw. *Androcles and the Lion* (1913; publ., with *Overruled* and *Pygmalion,* 1916) returns, uncertainly, to exploration of religious sentiment by way of a group of early Christians condemned to the

arena. It is slight and in some danger of appearing frivolous. *Pygmalion* (1913) is another matter. This story of a flower girl trained by a phonetician to pass as a lady, although a most effective satire upon the English class system, is less a play of ideas than is any other of Shaw's major plays. Its richer human content has made it a favourite with a wide public, and it has achieved success in both film and musical versions. It is a masterpiece of comedy. The scene in which Eliza Doolittle appears in society when she has acquired a correct accent but no notion of polite conversation is one of the funniest in English drama.

"Heartbreak House" and "Back to Methuselah."—Heartbreak House occupies a special place in Shaw's work. Begun in 1913, published in 1919 (with minor pieces), and produced in New York in 1920, it ends with casualties and destruction caused by an air raid, but there is no other sign that the action takes place in wartime. Superficially it owes much to Chekhov, its theme being the boredom and frustration of a cultivated and sophisticated upper-middle-class social group conscious of lacking a system of beliefs and a function. The play seems to be the product of a more sombre vision than is readily reconcilable with its familiar comic machinery. The central character, Captain Shotover, is old and mad; and, although he tries to recall the heroine from a cynical despair induced by disillusioned love, his own sense of the plight of the modern world is expressed in his having turned his skill as an inventor to projects of universal annihilation. The play symbolizes the human race brought to an impasse by its subjection to modern scientific determinism.

This leads on naturally to *Back to Methuselah* (publ. 1921; performed in New York, 1922; rev. ed. with postscript "After Twenty-Five Years," 1945), five linked plays in which Shaw expounds his conception of Creative Evolution. "You imagine what you desire; you will what you imagine; and at last you create what you will" (the Serpent to Eve, Act I). The superman is no longer thought of as attainable through eugenic breeding. He will come—and come with a leap—only as the product of profound travail at the unconscious roots of Life itself. And in A.D. 2170 "the thing happens." Two characters whom we have met in the 1920s and judged undistinguished prove to be still alive, and longevity has begun to endow them with wisdom far greater than that which "short livers" can enjoy. Shaw's drawing out of this theme into a long dramatic sequence is an astonishing feat, and its conclusion, in A.D. 31,920, has an impressiveness wholly unexpected in a work of speculative fantasy.

"Saint Joan."—Shaw remained vigorously and variously productive into his 90s. Only one more play, however, is to be counted among his best, and it is generally regarded as his greatest. *Saint Joan* (1923; publ. 1924) achieves what he had always sought (and had most nearly achieved in *Caesar and Cleopatra* and *Major Barbara*): the fusion of high seriousness with high comedy. The life and death of Joan of Arc are made to typify a perennial conflict between human institutions and human inspiration. And Shaw is absolute for the inner voice of the individual as the instrument through which evolutionary change accomplishes itself in the world as we know it. The strength of his conviction about this makes Joan the strongest and most moving of his characters.

Prose.—Shaw was from the first concerned to make his plays readable, believing that they would best make their way on the stage if widely known in published form. For all of them, therefore, he wrote unusually full "stage directions"; and, for nearly all, long polemical prefaces. He was a master alike of dramatic and nondramatic prose, and the prefaces are inferior to the plays only as a good parliamentary speech must necessarily be inferior to an equally good parliamentary debate. On the stage Shaw contrives the effect of seeing fair play all round; characters relegated to upholding the most absurd intellectual positions are endowed with pungent powers of repartee. The incidence of what he declared to be his master formula—"Find the right thing to say, and then say it with the utmost levity"—sometimes bears too heavily upon the prefaces to make for good persuasive writing on serious topics. But some of them are masterly. And *The Intelligent Woman's Guide to Socialism and Capitalism* (1928; new ed., 2 vol., 1937, with chapters on Sovietism and Fascism), his most sustained nondramatic work, is a vigorous, lucid, and in places eloquent book, such as could have been written only by a mature master of English prose.

See also references under "Shaw, George Bernard" in the Index.

BIBLIOGRAPHY.—*Editions, Selections, etc.:* Collected editions of the plays include those in 33 vol. (1930–38), and 36 vol. (1931–). The *Complete Plays*, 1 vol. (1931; enlarged 1936, 1938). *Prefaces*, with additional notes (1934; enlarged 1938; new enlarged ed. 1965). *Short Stories, Scraps and Shavings* (1934) includes *The Adventures of the Black Girl in her Search for God* (1932). Collected reviews, etc., include *London Music in 1888–89* (1937); *Our Theatres in the Nineties*, 3 vol. (1932); *Bernard Shaw and Karl Marx. A Symposium, 1884–1889* (1930). *Selected Prose*, ed. by D. Russell (1952), emphasizes the Socialist writings; other selections are: *Selected Novels*, ed. by A. Zeiger (1946); *Plays and Players: Selected Essays*, ed. by A. C. Ward (1952); *Bernard Shaw: A Prose Anthology*, ed. by H. M. Burton (1959); *Shaw on Theatre*, ed. by E. J. West (1958); *Platform and Pulpit*, ed. by Dan H. Laurence (1961); *Shaw on Shakespeare*, ed. by E. Wilson (1961); and *The Matter with Ireland*, ed. by D. H. Greene and Dan H. Laurence (1962), which collects early journalistic writings.

Correspondence.—Collected Letters, ed. by Dan H. Laurence, 4 vol. (1964–); *Ellen Terry and Bernard Shaw. A Correspondence*, ed. by C. St. John (1949); *Correspondence between Shaw and Mrs Patrick Campbell*, ed. by A. Dent (1952); *Bernard Shaw's Letters to Granville-Barker*, ed. by C. B. Purdom (1956).

Biography and Criticism.—Early works include: H. L. Mencken, *G. B. Shaw: his Plays* (1905); G. K. Chesterton, *George Bernard Shaw* (1909; augmented 1935); P. P. Howe, *Bernard Shaw: a Critical Study* (1915); *Frank Harris on Bernard Shaw* (1931). Among later works are: H. Pearson, *Bernard Shaw: His Life and Personality* (1942), and *G. B. S. A Postscript* (1951), combined, with additions (1961); S. Winsten (ed.), *G.B.S. 90: Aspects of Shaw's Life and Works* (1949); W. Irvine, *The Universe of G.B.S.* (1949); C. E. M. Joad, *Shaw* (1949), mainly on Shaw's philosophy; D. MacCarthy, *Shaw* (1951); Max Beerbohm, *Around Theatres* (1953); A. Henderson, *G. B. Shaw: Man of the Century* (1956); J. B. Kaye, *Bernard Shaw and the 19th-Century Tradition* (1958); M. Meisel, *Shaw and the 19th-Century Theatre* (1963).

Special Aspects.—E. Strauss, *Shaw: Art and Socialism* (1942); A. West, *A Good Man Fallen Among Fabians* (1950); R. Mander and J. Mitchenson, *A Theatrical Companion to Shaw* (1955), mainly a pictorial record of productions. F. E. Loewenstein (ed.), *Bernard Shaw Through the Camera* (1948), and *Bernard Shaw* (1952); and M. Shenfield, *Bernard Shaw: A Pictorial Biography* (1962), are pictorial records. *See* also F. E. Loewenstein, *Rehearsal Copies of Shaw's Plays: A Bibliographical Study* (1950). For Shaw's interest in language, see *On Language: G.B.S.*, ed. by A. Tauber (1963); I. L. Currell, *Shavio Phoenetica: A System of Phonetic English Based on Shaw's Instructions* (1946). (J. I. M. S.)

SHAW, HENRY WHEELER: *see* BILLINGS, JOSH.

SHAW, LEMUEL (1781–1861), chief justice of the supreme judicial court of Massachusetts from 1830 to 1860, who left an indelible mark on the law of the commonwealth and significantly contributed to the structure of American law. Born on Jan. 9, 1781 at Barnstable, Mass., he was educated at Harvard, studied law with David Everett, and was admitted to the bar in 1804. While a practitioner and Federalist officeholder in Boston he had extensive experience with public affairs, and was counselor to many ascendant industrial interests.

Respectful of the legislative will when statutes made it clear, Shaw did not hesitate to mold the common law to meet the needs of a changing society. He made the fellow-servant rule a principle of American law (*Farwell* v. *Railroad*, 4 Metcalf); he freed unions from the abusive application of the law of conspiracy (*Commonwealth* v. *Hunt*, 4 Metcalf); and, though opposed to slavery, he denied that racial segregation in public schools created unconstitutional inequalities (*Roberts* v. *Boston*, 5 Cushing). He died in Boston on March 30, 1861.

Shaw was closely associated with the family of Herman Melville, whose father-in-law he became. At the age of 21 Shaw was engaged to Nancy Melville, the author's aunt, but she died before they could be married. He maintained his relationship with her family after his marriage to Elizabeth Knapp in 1818; the Shaws' only daughter, Elizabeth, was married to Herman Melville in 1847.

See F. H. Chase, *Lemuel Shaw . . .* (1918); L. W. Levy, *The Law of the Commonwealth and Chief Justice Shaw* (1957). (M. DEW. H.)

SHAW, SIR (WILLIAM) NAPIER (1854–1945), English meteorologist, whose scientific contributions bridge the transition from the 19th-century era to the rejuvenated state of meteorology resulting from the ideas of the Norwegian School, as represented

by the work of V. and J. Bjerknes and their colleagues, after World War I, and from the direct analysis of the upper air. Born March 4, 1854, at Birmingham, he began his career teaching physics at Cambridge, where he had been a student, a task he performed with distinction for more than 20 years. Elected a fellow of the Royal Society in 1891, he was called from Cambridge in 1900 to become secretary of the Meteorological Council, which was the governing body of the Meteorological Office. His influence led to the reorganization, with noteworthy improvements, of the Meteorological Office. In 1905 he became director of the office, serving until his retirement in 1920. He then was named the first professor of meteorology at the Royal College of Science of the Imperial College of Science and Technology, retiring in 1924. Subsequently he devoted his attention to international meteorological affairs and to completing his four-volume *Manual of Meteorology* (1926–31). He died in London, March 23, 1945.

(H. R. B.)

SHAW, RICHARD NORMAN (1831–1912), British architect who played an important role in the Domestic Revival of the second half of the 19th century in England, was born in Edinburgh, Scot., on May 7, 1831. After an apprenticeship to William Burn, Shaw attended the architectural schools of the Royal Academy, winning a traveling scholarship to Germany, Italy, and France. He subsequently entered the office of G. E. Street. His early works were in collaboration with W. E. Nesfield. In his reaction against High Victorian Gothic, Shaw adopted divergent manners (often simultaneously), including neo-Baroque (Piccadilly Hotel, London, of 1905–08), but it was in the sphere of town and country houses that his influence was strongest and his style appears most characteristically: the former rely primarily on Shaw's very individual adaptation of the Queen Anne style; the latter derive from a study of regional developments in the English manorial style of the 16th century, and are carried out with a marked respect for the differing nature of different materials. The publication of Shaw's designs carried his influence outside England and was an element in the development of the American Shingle style.

Shaw's country houses are widespread. Characteristic examples are Wispers in Sussex (1875) and Adcote in Shropshire (1877). Town houses are best seen in London at Kensington, Chelsea, and Hampstead. In the field of town planning, the "Garden Suburb" laid out by Shaw in 1876 at Bedford Park (on the western side of London) was an important pointer to future developments of this nature. Shaw died in London on Nov. 17, 1912.

Bibliography.—Sir R. Blomfield, *Richard Norman Shaw, Architect* (1940); N. Pevsner, "Richard Norman Shaw" in *Architectural Review*, lxxxix (1941); H. R. Hitchcock, *Architecture: Nineteenth and Twentieth Centuries* (1958).

SHAWINIGAN, previously called Shawinigan Falls, is an industrial city in St. Maurice County, Quebec, Can., on the right bank of the St. Maurice River, 20 mi. above Trois-Rivières. Pop. (1961) 32,169. The falls near the city drop 165 ft. and 416,000 h.p. have been developed in two plants, supplying power both locally and to Montreal and Quebec. In 1901 an aluminum smelter, and a pulp and paper mill were installed in the village of Shawinigan Falls, forming the basis for its considerable industrial expansion. Other industries produce chemicals, abrasives, and textiles. The city is divided into two residential areas by the industrial belt and railway lines of Canadian Pacific and Canadian National railways. In 1961, 96% of the population was French-Canadian and 98% Roman Catholic. Shawinigan South, on the left bank of the St. Maurice River, had 12,683 persons in 1961. Shawinigan was incorporated as a village in 1901, a town in 1920, and a city in 1921. The name is the Algonkin word for "crest," used as a place name because the portage beside the falls of the river led over a crest of rocks. (M. C. Ba.)

SHAWL, a term applied to a square, oblong, or triangular article of dress worn as a protective or ornamental covering over the shoulders, neck, or head. It is found as clothing in most parts of the world. The word is of Oriental origin deriving from the Persian *shal.* Records seem to indicate Bukhara as the birthplace of the shawl-weaving industry. Shawls have been worn since the earliest times but the 19th century is known as the "shawl period" because then the shawl reached such a height of popularity that it became a necessity in the wardrobe of every woman in western Europe and America.

The first of the handsome Oriental shawls appeared in Europe after the Egyptian campaign in 1798. The most beautiful import was the Cashmere shawl produced in the valley of Kashmir in India. There the Tibetan goat produced a long fine fleece called pashm, which was softer than any wool known. The wool was woven into shawls by expert weavers trained for generations to a sensitive feeling for design and glowing colour. There are two types of Cashmere shawls, the woven shawl and those embroidered in imitation of weaving. In shape they were both square and oblong. The characteristic design was always the "cone" pattern inspired by the jeweled ornament in the turban of the Mughal emperors.

The Paisley shawl, machine produced in Scotland to meet the demand for the Cashmere, was an adaptation rather than an imitation of the Oriental shawl. Scarves, boas, mufflers, stoles, and smaller shawls are forms of the shawl popular as articles of dress in the 20th century.

Bibliography.—Katherine Morris Lester and Bess Viola Oerke, *Illustrated History of Those Frills and Furbelows of Fashion Which Have Come To Be Known as Accessories of Dress* (1940); Alice Morse Earle, *Two Centuries of Costume in America,* vol. i (1903); Millia Davenport, *The Book of Costume,* vol. ii (1948). (M. B. K.)

SHAWM (from the Latin *calamus,* a reed; Old French *chalemie*), a wind instrument of Near-Eastern origin, a precursor of the oboe. Like the oboe, the shawm is conically bored and sounded with a double reed of cane. Bore, bell, and fingerholes are, however, wider than the oboe's and the tone, intended for the open air, is of great power.

The shawm first appeared in the Near East about the beginning of the Christian era. It was later widely disseminated by Islamic influence, and numerous Oriental varieties are still played in many countries from Morocco to China. At the time of the Crusades the shawm was introduced into Europe and became widely used for dance and ceremonial music. In the 16th century shawms were constructed in various pitches from treble to great bass, the larger sizes being known in French as *bombardes* and in German as *Pommern.* The shawm was rarely used in Europe after the 17th century, but in Spain it survived as the *tiple* (treble) and *tenora* (tenor) which, modernized with complete keywork, still lead the bands for the sardana, the national dance of Catalonia.

(A. C. Ba.)

SHAWN, TED (Edwin M.) (1891–), U.S. dancer, choreographer, and dance educator who significantly increased the stature of the American dance and the respect accorded to it especially as a career for men, was born in Kansas City, Mo., on Oct. 21, 1891. He was originally destined for the ministry, but turned his attention to the dance after a serious illness interrupted his studies. He was just beginning his career when in 1914 he met and married Ruth St. Denis (*q.v.*), with whom he founded the Denishawn School, where every dance technique was taught, and the Denishawn Dancers, a troupe which toured extensively in the United States and visited the Orient in 1925–26.

From 1933 to 1940, after his separation from St. Denis, Shawn devoted himself to building a group of men dancers that presented such impressively masculine works as *Labor Symphony, Olympiad* and *Kinetic Molpai.* He next developed the Jacob's Pillow Dance Festival (Lee, Mass.), which was modestly inaugurated in 1933; under Shawn's leadership it became a dance centre of international importance, with its own school and theatre. Shawn also wrote and lectured extensively.

See Katherine S. Dreier, *Shawn, the Dancer* (1933), and the chapter on Shawn in Walter Terry, *The Dance in America* (1956). (Ln. Me.)

SHAWNEE, an Algonkian-speaking Indian tribe that lived in the central Ohio Valley until the 17th century, when they were driven out by the Iroquois and scattered into widely separated areas. Some settled in what is now Illinois and others in the Cumberland Valley, while one group moved to the southeast, giving its name to the Savannah River. After 1725 the tribe united again in the region of Ohio, where they formed the principal barrier to the advance of U.S. settlers. Following their defeat by

Gen. Anthony Wayne (*q.v.*) at the Battle of the Fallen Timbers, and the failure of Tecumseh (*q.v.*) to unite the Indians of the Ohio Valley, they broke into three independent branches (Absentee, Eastern, and Cherokee Shawnee) that eventually settled in different parts of Oklahoma.

Closely related in language and culture to the Sauk, Fox Indians (*qq.v.*), and Kickapoo, the Shawnee were also influenced by their long association with the Seneca and Delaware Indians (*qq.v.*), and combined eastern woodland and prairie traits. During the summer they lived in bark-covered houses grouped into large villages where women raised corn while men hunted; in winter, small family groups realigned themselves into hunting camps. Each village had a large council house that was also used for such religious ceremonies as the ritual purification of warriors. Other important ceremonies included the spring Bread Dance held when the fields were planted; the Green Corn Dance marking the ripening of crops; and the autumn Bread Dance. The social organization included patrilineal clans and an Omaha type of kinship system based on the lineage principle (*see* KINSHIP TERMINOLOGY). Early Shawnee population was estimated at 2,000 to 3,000; in the 1960s the three main settlements in Oklahoma numbered about 2,250.

See also ALGONKIAN TRIBES.

BIBLIOGRAPHY.—C. C. Trowbridge, *Shawnese Traditions* (1939); E. W. Voegelin, *Mortuary Customs of the Shawnee and Other Eastern Tribes* (1944); C. Klinck (ed.), *Tecumseh: Fact and Fiction in Early Records* (1961). (CH. C.)

SHAWNEE, a prairie city of central Oklahoma, U.S., on the North Canadian River, 38 mi. ESE of Oklahoma City; the seat of Pottawatomie County. It was named for the Shawnee Indians who came west after the defeat of Tecumseh in 1812, joined the Potawatomi, and in 1867 were brought into the Indian Territory. In 1891 their land was opened to white settlement. The city, originally known as Shawnee Town because it was a trading place for Indians, was incorporated in 1894 as Shawnee. Farming, stock raising, oil, and diversified industries are the bases of its economy. Educational facilities include St. Gregory's College, a Roman Catholic institution for men established in 1915, a two-year college with a three-year professional school (noted for its museum and library), and Oklahoma Baptist University, a four-year coeducational institution established in 1911. Places of historical interest include the Quaker Mission Church, established in 1872.

For comparative population figures *see* table in OKLAHOMA: *Population*. (AN. L.)

SHCHERBATSKY, FEDOR IPPOLITOVICH (1866–1942), Russian philologist, whose works on Buddhist logic and philosophy became known as authoritative and valuable, was born in St. Petersburg on Oct. 1 (new style; Sept. 19, old style), 1866, of a family of old nobility. He was educated at Tsarskoe Selo and studied philology at St. Petersburg, Vienna, and Bonn. In 1904 he became professor of Indian literature at St. Petersburg. He spent a few years in Mongolia and India and selected for special study Buddhist logic and metaphysics. He was greatly influenced by Dharmakirti's logical work, of which he published a translation with a detailed commentary. His chief works are *Logic in Ancient India* (1902); *Buddhist Philosopher on Monotheism* (1904); *Theory of Knowledge and Logic According to the Later Buddhists* (part i, *Text Book of Logic* by Dharmakirti; part ii, *Study of Perception and Deduction* [1909]); *Study of the Categorical Imperative Amongst Brahmans* (1918); *The Conception of Buddhist Nirvana* (1927); and *Buddhist Logic*, two volumes (1930–32). He also published many Sanskrit manuscripts in the *Bibliotheca Buddhica* of the Academy of Sciences. He died on March 18, 1942. (S. RA.)

SHEARING MACHINES: *see* MACHINE TOOLS.

SHEARS: *see* SCISSORS AND SHEARS.

SHEARWATER, any of 40 or so medium-sized plainly coloured petrels (*q.v.*), chiefly of the genera *Puffinus* and *Pterodroma*. Most species nest in the southern hemisphere; a few, such as the sooty shearwater (*Puffinus griseus*) and the greater shearwater (*Puffinus gravis*), reach the shores of the northern hemisphere on their migrations. The Manx shearwater (*P. puffinus*) nests on the coasts of Europe. Shearwaters spend the greater part of their lives on the ocean. The single white egg is laid in a hole underground. The young are thickly covered with long down and become extremely fat. In this condition they are considered good eating, and enormous numbers are caught in some localities, especially on islands off Australia, where they are known as muttonbirds. (DN. A.)

SHEATHBILL, a bird so called from the horny case around the base of its bill. There are two species, constituting the family Chionididae. They are found only in the antarctic regions.

The sheathbills are about the size of a pigeon, which they resemble in many ways. The plumage is white. In the white sheathbill (*Chionis alba*) the bill is yellowish at the base, becoming pale pink toward the tip; the skin around the pink-rimmed eyes bears cream-coloured pimples; and the legs are bluish gray. The lesser sheathbill (*Chionis minor*), only slightly smaller than the white sheathbill, has a dark bill and legs and a differently shaped "sheath." The white sheathbill collects its food, consisting of seaweeds and shellfish, on rocks at low tide; it is also known to eat birds' eggs. Though most commonly encountered as a shore bird, it can be seen far out at sea. The eggs of both species are not unlike those of oyster catchers.

SHEBA (SABA): *see* SABAEANS.

SHECHEM (SICHEM), a city of ancient Palestine, 8 mi. SE of Samaria. It is now a large mound called Balatah about 1½ mi. E of Nablus (Roman Neapolis) in western Jordan, in the valley between Mt. Ebal and Mt. Gerizim. A few hundred yards from Balatah toward the adjacent alluvial plain are the traditional sites of Jacob's well and of the tomb of Joseph. Since the mound, despite its extent, had been almost buried by the earth and stones washed down from the mountains, it was not identified with Shechem until shortly before World War I, when Ernst Sellin recognized its significance and led a German expedition to open it up (1913–14 and 1926–34). The excavations of Sellin, followed after 1928 by G. Welter and H. Steckeweh, unearthed a massive pre-Israelite city wall, two monumental city gates, a citadel with a massive Canaanite temple and other important remains. Most of the German work was done before the chronological value of pottery was generally accepted, so the direct historical results were somewhat meagre. In 1956 excavations were resumed by a U.S. expedition headed by G. E. Wright and B. Anderson. By means of modern pottery chronology, the history of the site has been worked out in detail. Combined with the information from Egyptian and cuneiform inscriptions, as well as with biblical and Greco-Roman data, the archaeological finds on the site yield a fairly complete picture from *c.* 2000 B.C. down to the destruction of the old city in the late 2nd century B.C. During the first centuries A.D. a village occupied part of the site.

By the time of the Egyptian middle kingdom (*c.* 2000–1800 B.C.) Shechem was one of the most important towns of the Palestinian hill country, as shown by references to it as a focus of actual or feared rebellion. It must already have been a leading centre for trade and communication because of its position at the southeastern edge of the Nablus pass, which led directly across the watershed ridge of western Palestine. When the Hyksos were ruling in Egypt, Shechem became a stronger fortress than any yet excavated in Palestine. Its importance in Hebrew tradition of the early 2nd millennium B.C. is illustrated by numerous references to it in connection with the stories of Abraham and Jacob (*e.g.*, Gen. 12:6; 33:18). In the 15th–14th centuries B.C. the population was already very mixed, as shown by the personal names in cuneiform tablets from the Shechem area found at Tell el Amarna in Egypt as well as at Balatah itself. Horites (so read in Gen. 34:2, with the Greek translation, instead of "Hivite") and Indo-Aryans lived at Shechem together with Hebrews and other Semites. In the early 14th century B.C. a chieftain with the good Hebrew name of Lab'ay ("lion-man") ruled the Shechem district. In the Amarna tablets the Hebrews (Apiru) are said to have been in control of the town. (In Gen. 34 the house of Jacob is said to have conquered Shechem.)

Shechem was not destroyed at the time of the Israelite occupa-

tion of the rest of the hill country of Palestine, and in the time of Joshua it became (together with Shiloh) the chief meeting place for representatives of the tribal confederacy (Josh. 24:1). In the early 11th century it was briefly the capital of the ephemeral monarchy set up by Abimelech (Judg. 9), and Jeroboam made it the first capital of the northern kingdom (I Kings 12:25) in the 10th century. After the exile it became the religious capital of the Samaritan sect, which replaced Jerusalem by Gerizim as its holy mountain. About 130 B.C. it was occupied by the Hasmonaean king John Hyrcanus I, and it never recovered its importance.

See also references under "Shechem" in the Index.

See G. E. Wright, *Shechem: The Biography of a Biblical City* (1965).
(W. F. A.)

SHEEP. Sheep belong to the family of hollow-horned ruminants or Bovidae (*q.v.*). They pass almost imperceptibly into the goats. Both sexes often possess horns, but those of the females are small. In the males the horns are generally angulated, and marked by fine transverse wrinkles; their colour is greenish or brownish. They are directed outward, and curve in an open spiral with the tips directed outward. Although there may be a fringe of hair on the throat, the males have no beard on the chin; and they also lack the strong odour characteristic of goats. Usually the tail is short; and in all the wild species the outer coat takes the form of hair, though beneath lies a short undercoat of fine wool which has been developed into the fleece of domesticated races. Weight varies from somewhat less than 100 lb. to several hundred pounds.

Like goats, sheep have narrow upper molar teeth, very different from those of the oxen, and narrow hairy muzzles. Between the two middle toes, in most species, is lodged a deep glandular bag, having the form of a retort with a small external orifice, which secretes a strong-smelling, oily substance. This, tainting the herbage or stones over which the animal walks, affords the means by which, through the powerfully developed sense of smell, the neighbourhood of other individuals of the species is recognized. The crumen, or suborbital face gland, which is so largely developed and probably performs the same office in some antelopes and deer, is present, although in a comparatively rudimentary form, in most species, but is absent in others. It may be added that the long tails of most tame breeds are, like wool, in all probability the results of domestication.

Origin and History.—The bones of sheep are found with other evidences of early human habitation at sites ranging from the Middle East to the crannogs of Ireland. Remains at Hassuna in upper Mesopotamia and at Sialk in central Iran are indicated as very early sites. But archaeological evidence presumably accumulated considerably later than domestication took place, for the Neolithic period appears to have opened with a well-established agriculture and a rather full complement of crops and animals which are still basic to farmers. Other evidence, particularly that from religious history, also suggests early domestication—the sheep is the preferred animal for sacrifice in ancient as well as some modern religious ceremonials.

Carl O. Sauer, author of *Agricultural Origins and Dispersals* (1952), associated domestication of sheep not with hunters or nomads but with sedentary seed farming, with southwest Asia indicated as the most probable hearth of domestication. Leopold Adametz, Austrian livestock-research specialist, associated an early domestic form with spiral horns with a variety of the urial (*Ovis vignei*) living from the Salt Range of the Punjab to Baluchistan, whence they passed to the Near East and beyond. In the Copper Age a new breed (*O. aries studeri*), with massive spiral horns, appeared in Europe, and J. U. Duerst, Swiss animal geneticist, showed that it was almost certainly derived in part at least from the wild mouflon (*O. musimon*). In Sardinia, as has been known since the time of Pliny, the mouflon interbreeds freely with domesticated sheep. The bulk of modern breeds are obviously much more closely related to the sheep of the Copper Age than to the earlier type; while it thus seems likely that the mouflon played the major part and the urial a minor part in producing farm flocks, it is by no means certain that other species, such as the argali (*O. ammon*), were not concerned.

VARIETIES AND DISTRIBUTION

Wild Sheep.—Wild sheep attain their maximum development both in number and size, in central Asia. They associate either in large flocks or in family parties, the old males generally keeping apart from the rest. Although essentially mountain animals, sheep generally frequent open, undulating districts rather than the precipitous heights to which goats are partial.

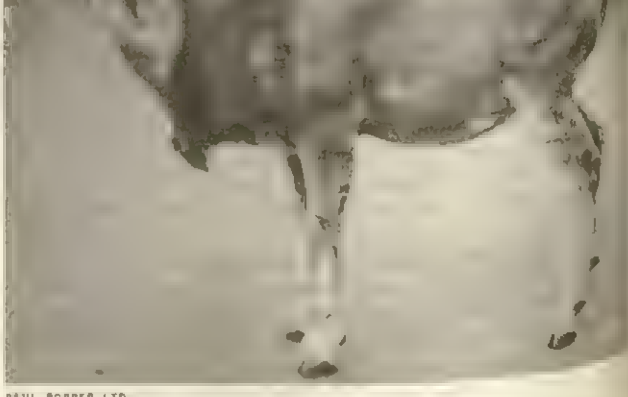
The Pamir Plateau, on the confines of Turkistan, at an elevation of 16,000 ft. (4,900 m.) above sea level, is the home of the magnificent *Ovis poli*, named after the celebrated Venetian traveler Marco Polo, who met with it in the 13th century. It is remarkable for the great size of the horns of the old rams and the wide open sweep of their curve, so that the points stand boldly out on each side, far away from the animal's head, instead of curling round nearly in the same plane, as in most of the allied species. A variety inhabiting the Tien Shan range is known as *O. karelini*, and other forms occur in the mountains and lower ground of Turkistan and in central Asia.

An even larger animal is the argali (*O. ammon*), typically from the Altai range but represented by one race in Ladakh and Tibet, by a second in eastern Mongolia, and by a third in the Gobi Desert. The largest living wild sheep, it may stand four feet high at the shoulders and weigh over 300 lbs. (136 kg.). Although its horns are less extended laterally than those of the Marco Polo sheep, they are grander and more massive. In their short summer coats the old rams of both species are nearly white. Closely allied forms are found in the Sair Mountains and Kulja.

In the Stanovoi Mountains and neighbouring districts of east Siberia and in Kamchatka occur two sheep which have been respectively named *O. borealis* and *O. nivicola*. They are, however, so closely allied to the bighorn (*q.v.*) or mountain sheep of North America that they can scarcely be regarded as more than local races. These bighorns are characterized by the absence of face glands and the comparatively smooth front surface of the horns of the old rams, which are thus very unlike the strongly wrinkled horns of the argali group.

The typical bighorn is the khaki-coloured and white-rumped Rocky Mountain animal; but on the Stikine River in British Columbia there is a nearly black race, with the usual white areas, while this is replaced in Alaska by the nearly pure-white Dall sheep, the gray sheep of the Yukon being perhaps not a distinct form. Other geographical races of the bighorn, distinguished chiefly by the colour of the coat, include the mountain sheep of Mexico, of Lower California, and of the Kenai Peninsula of Alaska.

Returning to Asia, in Ladakh, Afghanistan, and the Punjab



PAUL POPPER LTD.

URIAL OR WILD SHEEP OF THE PUNJAB

ranges is found the sheep whose local races are variously known as urial, urin, and shapo, which has tightly curved horns. It is a smaller animal than the members of the argali group, and approximates to the Persian and Armenian moufion (*O. orientalis*) and the Sardinian and Corsican mouflon (*O. musimon*), and on Cyprus a small race (*O. orientalis ophion*) hunted almost to extinction. Mouflon rams carry good, widely curving horns. In Tibet is the bharal (*Pseudois nahura*), the wild blue sheep of the Himalayas. They live in bands of 10 to 50 in mountain meadows above 10,000 ft. (3,050 m.). The ram is characterized by a black band separating the blue-gray back from the white underparts, and by smooth sigmoid horns sweeping backward in a semicircle. The udad, aoudad, or audad (*Ammotragus lervia*), the Barbary sheep, is the only sheep native to Africa, where it inhabits the mountain ranges of the north, descending eastward far into the Sudan. It is distinguished by long, abundant hair on the throat and forequarters of the rams. Both the bharal and the udad have no face glands, and in this and their smooth horns approximate goats. Additional information on varieties and distribution of wild sheep will be found in the separate articles on various countries and areas. (*See* also HUNTING: *Big Game Hunting*.)

(R. LY.; J. RI.; J. K. R.)

Domestic Sheep.—Domestic sheep differ from their wild progenitors and even among themselves in conformation, quantity, and quality of fleece, colour, size, milk production, and other characteristics. Breeds of sheep have been developed to meet environmental conditions influenced by latitudes and altitudes and to satisfy variable desires of the people for clothing and food. The commodities for which sheep have been bred include wool, fur, meat, and milk.

Of more than 200 breeds of sheep in existence in the world, the majority are of limited interest, except in the localities where they are raised. About 30 of the breeds which are most extensively raised or have otherwise attained prominence provide the principal basis for this discussion. They may be grouped into the following six types according to the kind and quality of commodities which they produce:

Fine-wool type.—Merino and Rambouillet.

Medium-wool type.—White-face sheep—Cheviot, Columbia, Corriedale, Dorset, Ile de France, Le Cotentin, Montadale, No-Tail, Oldenburg White Head, Panama, Romeldale, Ryeland, Targhee, Texel, and Welsh Mountain; dark-face sheep—Hampshire, Kerry Hill, Oxford, Shropshire, Southdale, Southdown, and Suffolk. Except for the Kerry Hill, these dark-face sheep are also called Down sheep, as they originated in the Downs of England and may have been the native sheep of the open-field farmer. The Tunis, whose modern native home was north Africa, has a brown face.

Long-coarse-wool type.—White-face sheep—English Leicester, Border Leicester, Cotswold, Lincoln, Romney, and Wensleydale; dark-face sheep—Blackface Highland of Scotland. The progenitors of these breeds, especially the Lincolns and the Romney Marshes, were probably of Flemish origin.

Fur type.—Karakul and Romanov.

Woolless mutton type.—Blackhead Persian.

Milk type.—East Fresian, La Razza Sarda, Pelvin, Sevlievo, Stara Zagora, and Svishtov.

Brief descriptions of these breeds follow.

Breeds of Sheep.—The Merino, a fine-wool sheep, originated in Spain. It was known as early as the 12th century and may have been a Moorish importation. It was particularly well adapted to semiarid climates and to transhumance. It became a prominent breed in many countries, notably Australia, the United States, the Union of Soviet Socialist Republics, South Africa, Argentina, France, and Germany. This breed has been designated by various names such as Australian Merino in Australia, American Merino and Delaine Merino in the United States, Merino Transhumante in Spain, Merino Volosh, the Soviet Merino, Ascanian, Caucasian, and Altai in the Union of Soviet Socialist Republics, Merino Argentino in Argentina, Merino Precoce in France, and Thuringia Merino-Meat in Germany. Some of these names signify strains of the parent breed that have been modified by selective mating, but all the sheep have rather heavy fleeces of fine wool that is pure white when scoured.

Merinos vary considerably in size, conformation, and extent of skin folds, but the prevailing trend in breeding is to develop sheep of medium size, fair mutton conformation, and freedom from excessive skin folds. The colour of their faces and legs is white. Although they have a considerable growth of wool on their faces, it is seldom extensive enough to cause wool blindness. The fine wool fibres of Merino fleeces are beautifully crimped and the fleeces have a tendency to carry sufficient natural fat to keep the fibres of the inner part of the fleeces soft and pliable. The fat in combination with the dust and soil of natural surroundings forms a dark covering on the outer part of the fleece that protects the beautiful, fine, white wool fibres of the inner portion. Through selective mating and crossing with other breeds, the Merino has served extensively as foundation stock in the creation of many useful breeds and strains of sheep. Of the four Merino types which became widely known and provided the basis for the expansion in apparel wool production in the 19th century, the heavily wrinkled Negretti became the basis for most of the English Merino development. The small, tighter-skinned Escurial appears to have been basic to the Rambouillet and Saxon developments, whereas the Paular and Infantado appear to have participated only slightly. After World War II considerable interest developed in Australia in polled Merino.

The Rambouillet, also a fine-wool sheep, originated in France. It was developed from excellent selections in 1786 and again in 1799 of a few hundred of the best Merino sheep in Spain. The French government made these selections and placed the flock at its national sheepfold, Rambouillet, France, where this breed originated. It is the largest of fine-wool sheep. Breeders in Germany made important contributions to the development of the Rambouillet, especially to its fleshing qualities.

Sheep of this breed were first imported into the United States in 1840. Through selective mating the breed was successfully molded to meet the needs of a large class of U.S. sheep producers. Rambouillets prevail on the western ranges where two-thirds of the sheep of the United States are produced. The face and legs of this breed are white. The face covering of wool is rather heavy, even to the extent of wool blindness in some specimens, but through selective mating breeders have made progress in developing strains free from such excessive face covering. Fleeces of Rambouillet sheep are relatively heavy. The lambs grow rapidly, attaining under good feed conditions satisfactory market weights, and finish at from six to nine months of age. The crossbreeding of Rambouillet ewes with medium-wool and long-coarse-wool rams is practised extensively for the production of choice market lambs and of rugged breeding ewes that produce heavy fleeces of attractive medium wool.

The Cheviot is a medium-wool breed. It is a hardy, white-face, hornless sheep developed in Scotland and Northumberland, Eng. Cheviots have no wool on their heads and ears or on their legs below the knees and hocks. As a consequence they present an attractive and alert appearance. The wool of their fleeces is relatively straight, of moderate length, close set, and free from black fibre. Cheviots are used in crossbreeding, especially with Border Leicesters and Lincolns, for market lambs.

The Columbia is a medium-wool, white-face, hornless breed developed in the United States by the U.S. Department of Agriculture. It was formed by crossbreeding Lincoln rams and Rambouillet ewes and interbreeding the resulting crossbred ewes and rams and their descendants without backcrossing to either parent stock. The initial crossbreeding began at Laramie, Wyo., in 1912, and after 1918 the molding of the breed was carried out at the United States sheep experiment station, Dubois, Ida. The sheep of this breed are large and rugged, and they excel in the production of both wool and lambs. The breed is popular, especially in the western U.S. range region, where private breeders organized the Columbia Sheep Breeders' Association.

The Corriedale is also a medium-wool, white-face, hornless breed. It originated in New Zealand as a result of crossbreeding Lincoln rams and Merino ewes, beginning about 1880. The breed is valued especially in New Zealand, Australia, and the United States. Corriedales are of medium size, and they produce wool and lambs of very good quality. The breed is adapted to both farms and ranges.

MERINO RAM

MERINO EWE

[illegible] sheep. Both ewes and [illegible] called the Dorset Horn. [illegible] heavy milkers and their [illegible] rather short and [illegible] the United States and other [illegible] hothouse lambs as a portion [illegible] They also cross well with [illegible] -bred ewes well suited to hot- [illegible] Both Dorset Horns and polled Dorsets [illegible]

The [illegible] a medium-wool white-face, hornless sheep [illegible] noted [illegible] of both meat and wool in France. It [illegible] early maturity and the rapid growth of [illegible]

The Coter[illegible] maintains [illegible] medium-wool, white-face, hornless sheep [illegible] kers and [illegible] hardy as a grazing sheep. The ewes are [illegible] and their [illegible] lambs well. They have good confor- [illegible] ers. [illegible] quality. They are also good wool [illegible]

The Montadale [illegible] developed [illegible] Cheviot rams. [illegible] -wool, medium-sized mutton sheep, [illegible]

The No-Tail, [illegible] after 1932 from Columbia ewes and [illegible] 1913 in South D[illegible] several fine [illegible] wool breed, has been developed since [illegible]

The Oldenburg [illegible] Siberian, "Fat-rumped" sheep, crossed [illegible] sheep of Germ[illegible] medium-wool breeds. [illegible] prolific and an exc[illegible] a medium-wool, white-face, horn- [illegible]

The Panama is [illegible] a hardy grazing breed. It is very [illegible] the western United [illegible] of both lambs and wool. [illegible] breeding Lincoln ew[illegible] wool, white-face, hornless breed of [illegible] good size, are hardy [illegible] developed in Idaho in 1912 by cross- [illegible] are good milkers and [illegible] Rambouillet rams. These sheep have [illegible] wool are heavy. [illegible] adapted to the range. The ewes [illegible] well. Their fleeces of white [illegible]

The Polwarth is an A[illegible] was developed in Victori[illegible] rather large sheep which [illegible] from crosses of Merino and Lincoln. Less numerous than Merino or Corriedale, it still is important in the Australian sheep industry in localities too cold and wet for Merino.

The Romeldale is a medium-wool, white-face, hornless sheep of the western United States, especially raised in California. It resulted from the crossbreeding of Romney rams and Rambouillet ewes, beginning about 1915. It is well suited to pasturage of alfalfa and heavy forage, as well as being hardy and well suited to range conditions. The Romeldale produces wool and lambs of high quality.

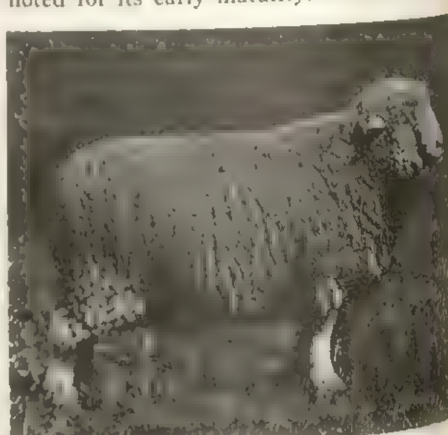
The Ryeland is a medium-wool, white-face, hornless sheep of England. The lambs grow quickly; the wool is good quality, of deep staple, and thickly set on the skin. Sheep of this breed were raised to some extent in the United States, especially in Kentucky.

The Targhee was developed at the U.S. experiment station at Dubois, Ida., after 1926 by vigorous selection and inbreeding of various Rambouillet-Columbia-Corriedale combinations to produce a true-breeding strain intermediate to the smaller fine-wool breeds and the large crossbreeds. Hardy and of medium size, it is well adapted to the range; it is of compact conformation with a moderately low-set, broad, level back, good bones, and straight legs. The rump and leg of mutton are well developed. The fleece is heavy, of desirable fineness (60s and 62s) and staple length. Ewes are good mothers and handle easily; necks are without folds and faces are open.

The Texel is a medium-wool, white-face, hornless sheep of the Netherlands, well adapted to range conditions and very prolific. The lambs grow rapidly. The fleeces of this breed are of good, heavy, white wool. The Texel is raised also in France, Belgium, Denmark, Spain, Indonesia, South America, South Africa, and Mexico.

The Welsh Mountain is a medium-wool, white-face sheep of Wales, and is of importance in other parts of the United Kingdom. The rams have horns and the ewes are hornless. It is noted for its very high-quality mutton. The fleeces are light in weight but of good quality.

The Hampshire is a medium-wool, dark-face, hornless sheep that originated in England. It is large and blocky and, as a superior mutton breed, is noted for its early maturity. It is one of the



TARGHEE RAM

BORDER LEICESTER EWE

PHOTOS COURTESY OF THE U.S. DEPARTMENT OF AGRICULTURE

most popular meat breeds in the United States, where it is raised extensively for market-lamb production in farming regions and for crossing with white-face range ewes in the western range regions for the production of market lambs. The wool of Hampshire fleeces is strong, of medium fineness and length, and desirable for manufacturing purposes except for the occurrence of black fibres in a small percentage of fleeces.

The Kerry Hill is a medium-wool, hornless sheep with markings that consist of black speckles on white faces, ears, and legs. It originated in the British Isles and is a hill sheep that combines desirable features of the Hill and Down types. Its fleece resembles that of the Shropshire but is not so good in quality. It produces good mutton and is excellent for crossbreeding.

The Oxford is a medium-wool, dark-face, hornless sheep, and it is the heaviest of the Down breeds. It was produced by mating Hampshires and Cotswolds in Oxfordshire, Eng. The ewes are good mothers and the lambs grow rapidly. Their fleeces are heavy and the wool is of good length for sheep of the medium-wool type. They are popular on farms that have abundant feed. Considerable numbers are found in England and the Great Lakes region of the United States.

The Shropshire is a medium-wool, dark-face, hornless breed that originated in the Downs of England. It is one of the most popular farm sheep in middle-western United States. It produces good wool and mutton and subsists on sparse pasturage better than breeds such as the Hampshire or Suffolk. For crossbreeding it is better adapted to farms than to range conditions. Its face covering of wool is excessive, which is more easily dealt with on farms.

The Southdale is a breed resulting from crossing the Southdown and the Corriedale.

The Southdown is a medium-wool, dark-face, hornless sheep. It is a Down breed that was raised on the Sussex hills of England. The Southdown is the oldest of all British breeds of sheep, and it has an ideal body conformation for meat production. Its fleece is close and the finest of the British breeds; but although white and of good quality, the wool is short and the fleeces relatively light in weight. The colour of its face and legs is brown to light brown or mouse colour. Southdowns are popular in many parts of the world, especially in Commonwealth of Nations countries and the United States.

The Suffolk is a medium-wool, dark-face, hornless sheep. It was developed in England by mating Norfolk horned ewes with Southdown rams during the years 1800 to 1850. Suffolks are prolific, early maturing sheep with excellent mutton carcasses. They are energetic and the whole carriage is alert, showing stamina and quality. Suffolk lambs gain rapidly and the breed is noted for mutton production. This breed is not desirable for wool production. The fleeces are short in staple and light in weight, and they have black fibres. Suffolks are popular as lamb producers.

The Tunis, a medium-sized to small, polled, coarse- to medium-wool breed, was imported into the U.S. from north Africa as early as 1799. Early grown in the southeast, they are now largely produced in the New York to Indiana section, where they are renowned for mating during all seasons, thus favouring the off-season milk lamb industry.

The Zenith is an Australian-bred sheep, founded at "Trelawney," Donald, Victoria, that came into prominence after World War II. The breed was considered fixed in 1947, when it was named. A dual-purpose comeback sheep (Merino crossed with an English breed and crossed back to Merino), it is rather large, with a dense fleece.

The English Leicester is a long-coarse-wool breed that originated in England through improvement by the noted breeder, Robert Bakewell, about 1755. This breed has white faces and legs. They are hornless, broad-backed, thick-fleshed, and early maturing. Their fleeces are heavy, with a tuft on the forehead and wool hanging in compact locks on the body. Some of the choicest wool of crossbred sheep is produced by offspring of Merino ewes and English Leicester rams.

The Border Leicester is a long-coarse-wool, white-face, hornless sheep that was developed in the border counties of England and Scotland by crossbreeding English Leicester and Cheviot sheep. The head is free from wool, bold and carried high. The whole carriage of this choice mutton body presents a square, alert, stylish appearance. The wool is long, soft, and in small locks, but fairly compact.

The Cotswold is a long-coarse-wool, white-face, hornless sheep that originated in England. The back is broad and flat, but the body is only moderate in depth. The fleeces are heavy and on the surface the wool lies in open curls. This breed is used successfully in crossing with Merinos and Rambouillets to increase size of body and length of wool fibre.

The Lincoln is a long-coarse-wool, white-face, hornless sheep, native of England. It is a large mutton sheep with heavy fleeces of tufted wool, and the poll is covered with wool. Rams of this breed are used extensively in many countries for crossbreeding with fine-wool ewes.

The Romney is a long-coarse-wool, white-face, hardy, polled sheep that originated in Kent, Eng. It is sometimes called the Kent or the Romney Marsh. Sheep of this breed are popular in South America, Australia, New Zealand, and in some western areas of the United States. Their fleeces are relatively dense and on the average are the finest of the long-coarse-wool fleeces. The rams are used rather extensively in the western U.S. for crossbreeding with Rambouillet ewes.

The Wensleydale is a long-coarse-wool, white-face, hornless breed of England. The wool is long, lustrous, and in wavy locks. The skin on the face, ears, and legs, and sometimes on the body, is of a bluish tinge. In England they are crossed with other long-coarse-wool breeds. Their mutton is of good quality, but they mature rather slowly.

The Blackface Highland is a long-coarse-wool, dark-face, horned, mountain sheep, native of Scotland. It is extremely hardy, matures quickly, and produces mutton of excellent quality. The wool of the fleeces of this breed is used extensively in the manufacture of tweeds, and the coarsest of it makes good, strong carpet wool.

The Karakul is a fur sheep, native of central Asia. Black is the prevailing colour of newborn lambs, which in most cases have beautiful tight-curled coats that, when skinned at about one to three days old, provide the commercial lambskin fur for which they are raised. A large percentage of the lambskins are classified by the trade as Persian lambskins. The wool of grown sheep is a mixture of coarse and fine fibres, of colours varying from black to various shades of brown and gray. The prevailing colour of the face and legs is black. The ewes are good mothers, and the lambs grow rapidly and produce good meat, but the breed is primarily intended for lambskin-fur production. Horned and hornless sheep occur in both sexes. Karakul sheep are raised in several countries of Asia, Europe, Africa, and the Americas.

The Romanov is a fur sheep of the Union of Soviet Socialist Republics, and it is found particularly in the Yaroslavl region. It is noted as a fur-producing sheep that excels in prolificacy. Black and white occur on face and legs and in the fibres of the fleeces.

The Blackhead Persian is a woolless fat-rumped sheep that has been highly developed for mutton production in South Africa, where it thrives on lands not well suited to wooled sheep. The head and neck are black and the body is white. This breed is indigenous to North Africa, Somaliland, and Arabia. The name "Persian" is misleading, for none of these sheep have come from Persia.

The East Fresian is a white-face milk sheep, native of Germany. It is considered to be one of the best milk breeds of sheep in Europe. In addition to its heavy milking qualities it is also well fleshed and wooled with heavy fleeces of white medium to coarse wool. The head is bare of wool and polled.

The La Razza Sarda is a milk sheep, native of Sardinia. The face is white with small black spots about the eyes and muzzle. It is noted as a heavy milk producer and also valued as a producer of long, white wool of coarse but good quality.

The Pelvin is a black-face milk sheep of Bulgaria, valued especially for milk production in one of the largest sheep-producing areas. It also produces a good weight of wool that is grayish in colour and medium to coarse in fineness.

MERINO RAM

MERINO EWE

The Dorset is a medium-wool, white-face sheep. Both ewes and rams have horns; the breed is sometimes called the Dorset Horn. It originated in England. The ewes are heavy milkers and their lambs grow rapidly. The fleece is pure white, rather short and open. The breed is valued also in the United States and other countries for the production of fancy hothouse lambs, as a portion of the ewes will breed out of season. They also cross well with Merinos for the production of crossbred ewes well suited to hothouse lamb production. Both Dorset Horns and polled Dorsets are bred in Australia.

The Ile de France is a medium-wool, white-face, hornless sheep that is an excellent producer of both meat and wool in France. It is especially noted for its early maturity and the rapid growth of its lambs.

The Le Cotentin is a medium-wool, white-face, hornless sheep of France maintained primarily as a grazing sheep. The ewes are good milkers and raise their lambs well. They have good conformation and their meat is of first quality. They are also good wool producers.

The Montadale, a medium-wool, medium-sized mutton sheep, was developed in the U.S. after 1932 from Columbia ewes and Cheviot rams.

The No-Tail, a medium-wool breed, has been developed since 1913 in South Dakota from Siberian, "fat-rumped" sheep, crossed with several fine-wool and medium-wool breeds.

The Oldenburg White Head is a medium-wool, white-face, hornless sheep of Germany, noted as a hardy grazing breed. It is very prolific and an excellent producer of both lambs and wool.

The Panama is a medium-wool, white-face, hornless breed of the western United States, originated in Idaho in 1912 by crossbreeding Lincoln ewes and Rambouillet rams. These sheep have good size, are hardy, and are well adapted to the range. The ewes are good milkers and raise their lambs well. Their fleeces of white wool are heavy.

The Polwarth is an Australian breed of rather large sheep which was developed in Victoria in the 1880s from crosses of Merino and Lincoln. Less numerous than Merino or Corriedale, it still is important in the Australian sheep industry in localities too cold and wet for Merino.

The Romeldale is a medium-wool, white-face, hornless sheep of the western United States, especially raised in California. It resulted from the crossbreeding of Romney rams and Rambouillet ewes, beginning about 1915. It is well suited to pasturage of alfalfa and heavy forage, as well as being hardy and well suited to range conditions. The Romeldale produces wool and lambs of high quality.

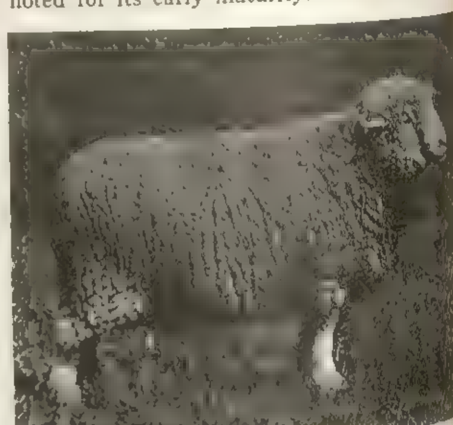
The Ryeland is a medium-wool, white-face, hornless sheep of England. The lambs grow quickly; the wool is good quality, of deep staple, and thickly set on the skin. Sheep of this breed were raised to some extent in the United States, especially in Kentucky.

The Targhee was developed at the U.S. experiment station at Dubois, Ida., after 1926 by vigorous selection and inbreeding of various Rambouillet-Columbia-Corriedale combinations to produce a true-breeding strain intermediate to the smaller fine-wool breeds and the large crossbreeds. Hardy and of medium size, it is well adapted to the range; it is of compact conformation with a moderately low-set, broad, level back, good bones, and straight legs. The rump and leg of mutton are well developed. The fleece is heavy, of desirable fineness (60s and 62s) and staple length. Ewes are good mothers and handle easily; necks are without folds and faces are open.

The Texel is a medium-wool, white-face, hornless sheep of the Netherlands, well adapted to range conditions and very prolific. The lambs grow rapidly. The fleeces of this breed are of good, heavy, white wool. The Texel is raised also in France, Belgium, Denmark, Spain, Indonesia, South America, South Africa, and Mexico.

The Welsh Mountain is a medium-wool, white-face sheep of Wales, and is of importance in other parts of the United Kingdom. The rams have horns and the ewes are hornless. It is noted for its very high-quality mutton. The fleeces are light in weight but of good quality.

The Hampshire is a medium-wool, dark-face, hornless sheep that originated in England. It is large and blocky and, as a superior mutton breed, is noted for its early maturity. It is one of the



TARGHEE RAM

BORDER LEICESTER EWE

PHOTOS COURTESY OF THE U.S. DEPARTMENT OF AGRICULTURE

most popular meat breeds in the United States, where it is raised extensively for market-lamb production in farming regions and for crossing with white-face range ewes in the western range regions for the production of market lambs. The wool of Hampshire fleeces is strong, of medium fineness and length, and desirable for manufacturing purposes except for the occurrence of black fibres in a small percentage of fleeces.

The Kerry Hill is a medium-wool, hornless sheep with markings that consist of black speckles on white faces, ears, and legs. It originated in the British Isles and is a hill sheep that combines desirable features of the Hill and Down types. Its fleece resembles that of the Shropshire but is not so good in quality. It produces good mutton and is excellent for crossbreeding.

The Oxford is a medium-wool, dark-face, hornless sheep, and it is the heaviest of the Down breeds. It was produced by mating Hampshires and Cotswolds in Oxfordshire, Eng. The ewes are good mothers and the lambs grow rapidly. Their fleeces are heavy and the wool is of good length for sheep of the medium-wool type. They are popular on farms that have abundant feed. Considerable numbers are found in England and the Great Lakes region of the United States.

The Shropshire is a medium-wool, dark-face, hornless breed that originated in the Downs of England. It is one of the most popular farm sheep in middle-western United States. It produces good wool and mutton and subsists on sparse pasturage better than breeds such as the Hampshire or Suffolk. For crossbreeding it is better adapted to farms than to range conditions. Its face covering of wool is excessive, which is more easily dealt with on farms.

The Southdale is a breed resulting from crossing the Southdown and the Corriedale.

The Southdown is a medium-wool, dark-face, hornless sheep. It is a Down breed that was raised on the Sussex hills of England. The Southdown is the oldest of all British breeds of sheep, and it has an ideal body conformation for meat production. Its fleece is close and the finest of the British breeds; but although white and of good quality, the wool is short and the fleeces relatively light in weight. The colour of its face and legs is brown to light brown or mouse colour. Southdowns are popular in many parts of the world, especially in Commonwealth of Nations countries and the United States.

The Suffolk is a medium-wool, dark-face, hornless sheep. It was developed in England by mating Norfolk horned ewes with Southdown rams during the years 1800 to 1850. Suffolks are prolific, early maturing sheep with excellent mutton carcasses. They are energetic and the whole carriage is alert, showing stamina and quality. Suffolk lambs gain rapidly and the breed is noted for mutton production. This breed is not desirable for wool production. The fleeces are short in staple and light in weight, and they have black fibres. Suffolks are popular as lamb producers.

The Tunis, a medium-sized to small, polled, coarse- to medium-wool breed, was imported into the U.S. from north Africa as early as 1799. Early grown in the southeast, they are now largely produced in the New York to Indiana section, where they are renowned for mating during all seasons, thus favouring the off-season milk lamb industry.

The Zenith is an Australian-bred sheep, founded at "Trelawney," Donald, Victoria, that came into prominence after World War II. The breed was considered fixed in 1947, when it was named. A dual-purpose comeback sheep (Merino crossed with an English breed and crossed back to Merino), it is rather large, with a dense fleece.

The English Leicester is a long-coarse-wool breed that originated in England through improvement by the noted breeder, Robert Bakewell, about 1755. This breed has white faces and legs. They are hornless, broad-backed, thick-fleshed, and early maturing. Their fleeces are heavy, with a tuft on the forehead and wool hanging in compact locks on the body. Some of the choicest wool of crossbred sheep is produced by offspring of Merino ewes and English Leicester rams.

The Border Leicester is a long-coarse-wool, white-face, hornless sheep that was developed in the border counties of England and Scotland by crossbreeding English Leicester and Cheviot sheep. The head is free from wool, bold and carried high. The whole carriage of this choice mutton body presents a square, alert, stylish appearance. The wool is long, soft, and in small locks, but fairly compact.

The Cotswold is a long-coarse-wool, white-face, hornless sheep that originated in England. The back is broad and flat, but the body is only moderate in depth. The fleeces are heavy and on the surface the wool lies in open curls. This breed is used successfully in crossing with Merinos and Rambouillets to increase size of body and length of wool fibre.

The Lincoln is a long-coarse-wool, white-face, hornless sheep, native of England. It is a large mutton sheep with heavy fleeces of tufted wool, and the poll is covered with wool. Rams of this breed are used extensively in many countries for crossbreeding with fine-wool ewes.

The Romney is a long-coarse-wool, white-face, hardy, polled sheep that originated in Kent, Eng. It is sometimes called the Kent or the Romney Marsh. Sheep of this breed are popular in South America, Australia, New Zealand, and in some western areas of the United States. Their fleeces are relatively dense and on the average are the finest of the long-coarse-wool fleeces. The rams are used rather extensively in the western U.S. for crossbreeding with Rambouillet ewes.

The Wensleydale is a long-coarse-wool, white-face, hornless breed of England. The wool is long, lustrous, and in wavy locks. The skin on the face, ears, and legs, and sometimes on the body, is of a bluish tinge. In England they are crossed with other long-coarse-wool breeds. Their mutton is of good quality, but they mature rather slowly.

The Blackface Highland is a long-coarse-wool, dark-face, horned, mountain sheep, native of Scotland. It is extremely hardy, matures quickly, and produces mutton of excellent quality. The wool of the fleeces of this breed is used extensively in the manufacture of tweeds, and the coarsest of it makes good, strong carpet wool.

The Karakul is a fur sheep, native of central Asia. Black is the prevailing colour of newborn lambs, which in most cases have beautiful tight-curled coats that, when skinned at about one to three days old, provide the commercial lambskin fur for which they are raised. A large percentage of the lambskins are classified by the trade as Persian lambskins. The wool of grown sheep is a mixture of coarse and fine fibres, of colours varying from black to various shades of brown and gray. The prevailing colour of the face and legs is black. The ewes are good mothers, and the lambs grow rapidly and produce good meat, but the breed is primarily intended for lambskin-fur production. Horned and hornless sheep occur in both sexes. Karakul sheep are raised in several countries of Asia, Europe, Africa, and the Americas.

The Romanov is a fur sheep of the Union of Soviet Socialist Republics, and it is found particularly in the Yaroslavl region. It is noted as a fur-producing sheep that excels in prolificacy. Black and white occur on face and legs and in the fibres of the fleeces.

The Blackhead Persian is a woolless fat-rumped sheep that has been highly developed for mutton production in South Africa, where it thrives on lands not well suited to wooled sheep. The head and neck are black and the body is white. This breed is indigenous to North Africa, Somaliland, and Arabia. The name "Persian" is misleading, for none of these sheep have come from Persia.

The East Fresian is a white-face milk sheep, native of Germany. It is considered to be one of the best milk breeds of sheep in Europe. In addition to its heavy milking qualities it is also well fleshed and wooled with heavy fleeces of white medium to coarse wool. The head is bare of wool and polled.

The La Razza Sarda is a milk sheep, native of Sardinia. The face is white with small black spots about the eyes and muzzle. It is noted as a heavy milk producer and also valued as a producer of long, white wool of coarse but good quality.

The Pelvin is a black-face milk sheep of Bulgaria, valued especially for milk production in one of the largest sheep-producing areas. It also produces a good weight of wool that is grayish in colour and medium to coarse in fineness.

BY COURTESY OF THE U.S. DEPARTMENT OF AGRICULTURE

SOUTHDOWN RAM

SOUTHDOWN EWE

RAMBOUILLET RAM

RAMBOUILLET EWE

CHEVIOT RAM

CHEVIOT EWE

The Sevlievo is a dark-face milk sheep of the mountain type, from Bulgaria. It is one of the more abundant milk-producing sheep and a fair producer of pigmented wool of medium fineness.

The Stara Zagora is a large, white-face, horned milk sheep of Bulgaria, developed for an abundance of milk but also valued for its white medium-wool fleece.

The Svishtov is a white-face, horned milk sheep of the region between the Danube River and the Balkan Mountains. It is noted for its abundant production of milk and for its good fleece of long, white, coarse wool. (D. A. S.; J. K. R.)

Improvement of Sheep.—The existence of more than 200 different breeds of sheep, differing in appearance as well as in function, indicates the variety of ideals held by breeders. It also indicates the wide range of genetic material available for experimentation. Unfortunately, most herdsmen in the past made little use of either line breeding or inbreeding to develop their flocks. By mid-20th century those techniques were utilized as well as progeny testing and some selection indexes and performance yardsticks. The use of artificial insemination has become more widespread, and breeding synchronization techniques employing steroid drugs to control the onset of estrus (and ovulation) have been developed. (*See* ANIMAL BREEDING.)

Most sheep are raised for two principal products, meat and wool. In the case of the Merino, the annual wool clip is more valuable to the grower than the yearly crop of lambs sold for meat. In most of the Down breeds the meat, produced chiefly as fat lambs, is more valuable than the yield of wool.

In a few parts of the world, such as in the Roquefort cheese-producing region in France and the Balkan countries, sheep are triple-purpose creatures because of their yield of milk. Even in their production of wool, sheep vary greatly. Some breeds yield long, coarse wool used chiefly in rug- and carpetmaking, like the sheep bred by the Navaho tribe of American Indians, which yield a coarse, carpet-type fleece. Merino wool can be woven into a fabric almost as sheer and resilient as silk, yet possessing greater strength and warmth. Karakul sheep yield a tightly curled, beautifully patterned and lustrous soft dark fur. It was found that the curl size of Karakul lambskin fur was shifted favourably toward a smaller curl by selection in the breeding flock over a few generations.

Environment, distance from market, and economic competition also influenced sheep types. Some breeds, such as the Rambouillet, have a natural herding instinct. They graze by day and bed down at night in closely gathered flocks, sometimes containing more than 1,000 sheep. Sheep which scatter when grazing because of lack of the herding instinct would soon be lost or fall victim to predatory animals.

Fortunately it was possible, by careful breeding, to make notable progress with the Rambouillet as to staple length of wool, mutton type, and freedom from skin folds. Not only are good mutton type and condition of flesh associated with heavier body weights and fewer skin folds, but staple length of wool and the weight of wool, either as clean or grease wool, are positively associated. Freedom from wool blindness is somewhat less closely associated. Progress was far greater from the selection of sires than from the selection of dams.

The semiarid regions of Australia, where an area three-fifths as large as continental United States averages only 15 in. of rainfall a year and offers spare grazing of a weedy, shrubby character, are unsuitable for producing choice fat lambs, but the Merino sheep with its excellent fleece of fine wool thrives there. The long distance of Australia from most of the world's markets is not a seri-

ous factor for a product as imperishable and light in weight as wool. A notable post-World War II development was the increase in "other" breeds, mainly Corriedales and Polwarths. The Merino was found to be uncommonly effective in grading up the common sheep flocks of India and the Middle East.

Shifts in the relative profitability of mutton and wool have been a cause of considerable crossbreeding in sheep production. The Corriedale breed, which originated in New Zealand, is about equally valuable for lambs and wool under range conditions. The same is true of the Columbia. The Corriedale and Columbia possess the herding instinct; their ewes have the mothering qualities so necessary in range sheep and they produce large lamb crops. The Corriedale is adaptable to areas as diverse as Kenya and Tierra del Fuego and gives a high lambing ratio. Flocks were established in most sheep regions, particularly those of the southern hemisphere. In western United States, Corriedale rams were used especially on Rambouillet ewes to give an improved quarter-bred.

Scientific crossing has given good results. Crossbred Hampshire, Shropshire, and Southdowns gave a much better survival ratio to weaning time, plus more rapid weight gains. Crossing of Columbia with Southdown gives wool production of very good staple length, quality, and quantity, as well as producing excellent feeders with a satisfactory market finish at desirable weights. Southdown and Shropshire lambs excel in edible meat production, giving the highest proportion of preferred cuts, including legs, loins, and ribs.

In the United States several state and regional experiment stations, particularly the sheep experiment station at Dubois, Ida., and the southwestern range and sheep breeding laboratory, Fort Wingate, N.M., compare different systems of breeding and develop more effective methods of selection. Various traits of lamb and wool production are analyzed for heritability and economic importance. Comparisons are made between inbred lines formed by selection within lines and by recurrent selection of sires from line-cross and top-cross progeny tests. Non-inbred selected groups are developed in each breed for comparison with inbred lines and their crosses. A genetically stabilized group is also established with selection at random to serve as a control for the selected groups.

(J. K. R.; M. A. D.)

CARE AND MANAGEMENT

This differs considerably from area to area, and more particularly between the small flock of 10 to 100 operated as an incidental part of a general farming operation and the large ranch or station band of one to several thousand head. The latter may involve moving with the season from valley to mountain to valley again (transhumance) over long trails distant from the home base.

Such far-flung operations may require spending the major part of the year on public land, returning to the home base only for wintering and the spring lambing season.

Regardless of the size or nature of the operation, skilled care and attention is required at lambing time, not only to assist with difficult deliveries but especially to make certain that the lamb does not become chilled after birth and that the ewe recognizes the lamb and allows it to nurse. Heat lamps are used to some extent to warm and dry the newborn. Protection and even rescue and perhaps supplementary feeding of the flock may be required during the worst winter months in some areas, particularly if there is deep snow.

It is seldom profitable to feed grain to breeding sheep or suckling lambs as long as they have an abundance of succulent grazing forage. Though they do best on pastures and ranges which provide

SUFFOLK RAM

SUFFOLK EWE

SHROPSHIRE RAM

SHROPSHIRE EWE

CORRIEDALE RAM

CORRIEDALE EWE

BY COURTESY OF THE U.S. DEPARTMENT OF AGRICULTURE

choice grass or legume forage which is short and fine, they will consume considerable quantities of high, coarse, brushy material and weeds.

Hay and other roughage can be used, particularly in winter or drought periods. Grain may be used in small amounts for flushing at breeding time and, of course, for fattening. Fattening on grain and concentrates is carried on mostly between the ranch and the consuming market and in general is a more delicate operation than cattle fattening. In some areas lambs may be sent to market "milk fat" from the ewe at 3 to 5 months of age; the greater part of the annual lamb crop moves to market at the close of the range season at 6 to 9 months of age.

Except on alkali ranges, breeding sheep must be provided with about one-half ounce of salt a day and fattening lambs about half as much. Calcium supplements of about one-fourth ounce a day may be needed if legume forage or hay is not included. Phosphorus, if deficient in the range plants, may be supplemented by adding two parts of bone meal to one part of salt. In areas where newborn lambs are afflicted with goitre, iodized salt in moderate amounts may be fed to the pregnant ewes. Vitamin A appears to be the only one likely to be deficient and that when food is limited to cereal straws, low-quality hay, or cottonseed hulls.

Sheep need as much as $1\frac{1}{2}$ gal. of water a day, though dew on grass or snow, if available, may suffice for much of the requirement. They should not have to travel more than three to four miles between grazing and water in cool weather, or about two miles in warm weather; animals with young should be even closer. Thus the number of animals which may be supported in semiarid areas is rather strictly limited by the availability and spacing of water holes.

Shade during hot weather is desirable as is protection from predatory animals—dogs, wolves, eagles, etc. Sheep should be shorn only when the fleece is dry and settled weather is in prospect, for wet snowstorms or severe cold may cause heavy losses from exposure after shearing unless protection is available. Except in areas of bitter winter, only the simplest shelter or none at all is required.

Approximately one mature ram is kept for each 35 ewes, unless artificial insemination is employed. Range ewes are seldom kept beyond the sixth year though they may be moved to areas with softer food and used as breeding stock for one or more additional years.

There is considerable variation in lamb crop, loss of animals and, of course, in income, from year to year. Under range conditions the annual lamb crop may average less than 75% of the number of breeding ewes maintained; in smaller farm flock, with an effort to save twins and triplets, the lamb crop may exceed the number of breeding ewes. The wool clip under favourable management will average 10 lb. per head for some breeds.

Diseases.—*Bacterial and Viral Infections.*—Prior to the 20th century, sheep owners commonly accepted heavy losses in their flocks as inevitable. After the early 1900s, however, marked progress was made by the veterinary profession in the prevention and control of many of the bacterial and viral infections that were once ovine scourges.

Bacteria of the genus *Clostridium* probably cause greater death losses than any other single class of microorganisms. These spore-bearing organisms are capable, generally, of producing powerful, lethal toxins. Some of the diseases they produce in sheep, and the organisms which cause them, are as follows: enterotoxemia, struck and pulpy kidney disease, caused by *C. perfringens;* braxy or bradsot, *C. septicum;* blackleg, *C. chauvoei;* black disease or infectious necrotic hepatitis, *C. novyi;* tetanus, *C. tetani.* The symptoms effected by the different *Clostridium* infections are frequently very similar; accurate diagnosis can only be made on bacteriological isolation and identification of the causative agent.

Struck, braxy, and black disease are particularly acute and lethal diseases. As in infection by anthrax (*q.v.*), affected sheep may die in a few hours after they first show symptoms, or they may be found dead without symptoms ever having been observed. These diseases are peculiar in that they commonly affect the most vigorous and thrifty sheep in the flock. Fortunately, through proper management and vaccination, losses from these diseases may be largely controlled, if not eliminated.

Struck or lamb dysentery affects the young, as the latter name implies. Unsanitary conditions and close confinement contribute to the danger from this disease, which attacks in from a few hours to a week after birth. The mortality may be high if precautionary measures are not used. In the second half of the 20th century the anthrax problem was not as serious in the western hemisphere as in the old world, where repeated epizootics of anthrax over the decades and centuries resulted in heavy contamination, particularly in Asia, southern Europe, and Africa.

The spore form of the anthrax bacterium (*Bacillus anthracis*) is very resistant to chemical and environmental influences and can survive for years in contaminated soil and in animal products, including wool. Feedstuffs and fertilizers, moreover, may be contaminated. The main source of infection in agricultural workers is from contact with contaminated carcasses, wool, hair, hides, and insufficiently cooked meat.

Many different preparations of anthrax spore vaccines are useful for the prevention of the disease, particularly spore suspensions in glycerin-saline solution for intradermic use and spore suspensions in saponin solution for subcutaneous use. The avirulent, noncapsulated strain of *B. anthracis* used in the spore vaccine developed at the Onderstepoort Veterinary Research Laboratory, Pretoria, S.Af., was used with considerable success in many countries.

Blackleg, tetanus, and malignant edema result from infections of wounds. The freshly ruptured navel, shearing cuts, and the wounds resulting from castration, docking, and earmarking are the usual avenues for entrance. To prevent these diseases, great care should be taken in operations, using clean instruments and techniques. Lambing should be done in cleaned and disinfected buildings or, when weather permits, outside on uninfected land. Infections with pus-forming bacteria also enter through wounds, resulting in arthritis and abscesses in the internal organs, including the brain and spinal cord. An effective means of prevention is to dip the stump of the navel cord in tincture of iodine or other effective antiseptic as soon as possible after birth.

Contagious pustular dermatitis or sore mouth is an especially widespread virus disease of sheep and goats. The disease occurs in England and other parts of Europe, North and South America, and Australia. Aside from the loss of body weight that results, the disease is usually not serious. The animals usually recover, but secondary infections caused by bacteria, especially *Sphaerophorus necrophorus,* may cause death. An effective vaccine was developed as a preventive.

Bluetongue, a little-known virus, has infected thousands of sheep in the southwestern United States and is known in other countries. Studies of scrapie, a disease involving the central nervous system, indicate that its transmission may be by a viral agent linked to certain genetic factors.

Foot rot or foul foot, due to infection with *S. necrophorus*, is more or less prevalent in most sheep-raising countries of the world. Because of extreme pain in the feet, some animals on sparse pastures may starve to death through inability to travel and procure feed. Cure is tedious and painstaking. The infection may be largely avoided by preventing healthy sheep from having contact with infected sheep or soil.

Sheep rarely develop tuberculosis. The few cases observed were mostly of the avian type, contracted from tuberculous fowls. A common disease of sheep the world over is pseudotuberculosis or caseous lymphadenitis. It is a chronic, bacterial disease of the lymph nodes, caused by *Corynebacterium ovis* (*Preisz-Nocard bacillus*). It seldom causes death, except in aged animals, but it requires considerable attention from the standpoint of meat inspection.

In some flocks mastitis (inflammation of the udder, blue bag) is a serious problem. Various bacteria are implicated as causative agents. Sanitation and biological products may be applied as preventives.

Louping ill, an encephalitis caused by a virus, was reported only

from Scotland and other parts of the British Isles. The disease is transmitted by the tick *Ixodes ricinus*. An effective vaccine was developed as a preventive.

Pregnant ewes, especially those carrying twins or triplets, may develop pregnancy toxemia or pregnancy disease if they are underfed, closely confined, or subjected to sudden changes in feed or environment. Treatment is generally ineffectual for this highly fatal metabolic disease. Prevention consists in keeping ewes in a thrifty gaining condition during pregnancy, especially during the last two months.

Other diseases that affect sheep, but are generally less widespread or less serious than those mentioned, include sheep pox, Johne's disease, hemorrhagic septicemia, paratyphoid, rabies, foot-and-mouth disease (*q.v.*), Q fever, and various forms of malnutrition. (M. S. Sn.; J. K. R.)

Parasites.—Sheep probably suffer more from parasitism than any other class of livestock, the damage being greatest among lambs and young animals. The losses from the effects of protozoa, worms, and arthropods are so extensive at times as to make sheep raising an unprofitable undertaking. Fortunately, control measures for most parasites are available and, if followed, the loss can be substantially reduced.

The most injurious of the protozoan parasites are coccidia, of which there are several species. Coccidia are minute animals that develop in and destroy the cells lining the intestinal tract, causing bloody diarrhea and frequently death. Coccidiosis (*q.v.*) is particularly prevalent where lambs are concentrated in feed lots for fattening and finishing for market. It can be controlled to a large extent by sanitation and frequent cleaning of the feed troughs to prevent accumulation in them of manure containing the oöcyst or infective stage of the organisms, which may be swallowed with the feed.

The worm parasites include flukes, tapeworms, and roundworms. The common liver fluke (*Fasciola hepatica*) occurs in most parts of the world and is one of the deadliest parasites of sheep. The flat, leaflike parasite lives in the liver, causing a condition known as liver rot. The symptoms of fluke infestation are principally unthriftiness, anemia, and abdominal dropsy, and differ little from those caused by other worm parasites. An acute, frequently fatal bacterial disease known as black disease is often associated with liver fluke infestation. The life cycle of the fluke is complicated, requiring water snails (*Lymnaea* and related genera) as intermediate hosts. The larval stages develop in the snails and the infective stage, or cercaria, that escapes becomes encysted on aquatic vegetation and is swallowed by sheep while grazing.

Control of liver flukes depends upon destruction of the snails, either by drainage to destroy their breeding places or by the use of chemical poisons, such as copper sulfate, broadcast over the snail-infested areas. Medicinal treatment with small individual doses of carbon tetrachloride or hexachlorethane will destroy the adult flukes in the bile ducts. Young flukes in the liver tissue are not affected by the treatment.

Several species of adult tapeworms parasitize the digestive tract of sheep, but as a general rule they cause little injury. The fringed tapeworm (*Thysanosoma actinioides*) frequently occurs in the bile ducts, rendering the liver unfit for human consumption. So far as known, the larval stage of sheep tapeworms develops in grass mites The larval stages of several dog tapeworms (*Echinococcus granulosus*, *Taenia hydatigena*, and *T. ovis*) occur in sheep; they render the carcass or the organs affected unfit for food. As a preventive measure, all dogs on farms where sheep are kept should be examined periodically for tapeworms by a veterinarian, and the infested animal should be treated for their removal.

The important roundworms of sheep are the stomach worm (*Haemonchus contortus*), nodular worm (*Oesophagostomum columbianum*), and several species of small threadlike worms belonging to the genera *Cooperia*, *Ostertagia*, and *Trichostrongylus*. Infestations are acquired by the animals while grazing. The common symptoms are unthriftiness, anemia, and diarrhea. Death loss from stomach worms and the small roundworms is sometimes very extensive. In addition to the usual bad effects of parasitism caused by roundworms, the nodular worm ruins the small intestines for use as sausage casings, surgical suture material, and for other purposes. In areas where freezing weather prevails for long periods, the infective larvae of roundworm parasites on pastures are destroyed. Treatment of the breeding flock with appropriate doses of thiabenzole or of phenothiazine is used to control infestations.

The arthropod parasites include the sheep tick or ked, nose grub, screwworm, and several kinds of lice and scab mites. The sheep ked is a wingless fly which sucks blood and causes skin irritation; it also stains the wool and reduces its market value. Keds may be controlled by immersing infested sheep in dips containing rotenone or other insecticidal substances and also in an aqueous suspension of 0.03% lindane. Nose grubs, which are larvae of the sheep bot fly *Oestrus ovis*, live in the head sinuses. They may be destroyed by spraying into the nostrils, under pressure, a solution of saponified cresol. The larvae of the screwworm fly (*Callitroga hominivorax*) infest wounds and cause extensive damage. Screwworm infestations are best treated by the application of a smear having diphenylamine as its basic ingredient. Lice suck blood and irritate the skin; they may be controlled by the use of insecticidal dips or applications containing coal tar creosote, arsenic, or nicotine. Scab mites cause skin irritation, falling out of the wool, unthriftiness, and even death; control or eradication is accomplished by dipping in lime sulfur and nicotine dips.

A dip prepared from diazone or from benzene hexachloride (BHC) is highly effective against scale mites and requires only one dipping instead of the usual two needed when the older dips are used. The dipping vat is filled with enough clean, unheated water to cover the sheep and the required amount of diazone or BHC powder added and thoroughly stirred. The head of each sheep is submerged at least twice for an instant so that the wool about the head and face is thoroughly wet.

(E. W. Pe.; J. K. R.; M. A. D.)

WORLD SHEEP INDUSTRY

Major Producing Areas.—Distribution of sheep is, of course, limited by grazing lands and somewhat by climatic conditions. Some cold areas, even if unfavourably wet, are usable, and warm areas, if dry, are used, but sheep are not numerous in the warmer areas of the tropics.

However, northern Australia above the Tropic of Capricorn contained in the 1960s fully 5% of the more than 160,000,000 sheep population of Australia. These were about 75% Merinos and grazed successfully, except for some problems of infertility which are not uncommon in hot arid areas with approximately equal hours of daylight and darkness.

The distribution pattern of sheep differs from other livestock in the emphasis on the southern hemisphere. Entirely a development that occurred after 1850, it appears to be related, first, to the storability and transportability of wool, well adapted to frontier conditions of subhumid climates.

In such subhumid areas the Merino sheep with its fine apparel wool developed under large-scale ranching conditions. Later the development of refrigerated ocean shipping enabled some of the thinly settled areas to increase the production of lamb for the European market.

World sheep numbers in the 1960s stood near 1,000,000,000 head, about one-third more than pre-World War II. Increases in the U.S.S.R. were particularly rapid. Other leading sheep-producing countries were: India, China, Turkey, and Iran in Asia; Australia and New Zealand in Oceania; United Kingdom, Spain, and Italy in Europe; Argentina, Uruguay, and Brazil in South America; South Africa and Morocco in Africa; United States in North America. There were increases in all continents except North America, where sheep numbers decreased by 35%. The decline in the United States is difficult to understand at first glance. One might think that emphasis on soil conservation, grassland farming, new developments in better-adapted and more productive grasses, and the adoption of new ways of curing hay and making and using grass silage should have led to increased sheep production.

The explanation appears to be more in the realm of comparative

economics and cultural trends. Cattle seemingly made more profitable the use of some grazing areas which previously had been devoted to sheep, paying a better return on capital invested. Moreover, labour of the types needed for successful sheep ranching, particularly the excellent Basques, became increasingly difficult to obtain. (M. S. Sn.; J. K. R.)

Products.—For wool and meat production, *see* Wool; Lamb and Mutton. The pelt is also a major by-product of sheep slaughter. The pelts, with the wool removed, known as slats, are tanned for leather which is used for upholstery, bookbinding, gloves, clothing, and shoe uppers. With the wool, the pelts are a basic material for the manufacture of durable and warm outer clothing.

Not only the carcass as such but the liver, heart, kidney, and some other parts are used for human food. Some of the internal glands have pharmaceutical uses. The small intestines are valuable sausage casings and are significant in international trade. They also are used to make surgical and musical catgut. Wool grease or lanolin has important uses in lubricants, ointments, and cosmetics. Sheep tallow has both edible and inedible uses.

Live sheep enter international trade in comparatively small numbers, mostly as select breeding stock.

See also references under "Sheep" in the Index.

Bibliography.—*Wild sheep:* R. E. Trippensee, *Wildlife Management*, ch. xvii (1948). *Domestic sheep:* Ministry of Agriculture and Fisheries, *British Breeds of Live Stock* (1938); Carl O. Sauer, *Agricultural Origins and Dispersals* (1952); W. Kupper, *The Golden Hoof: the Story of Sheep in the Southwest* (1945); H. Munz, *The Australian Wool Industry* (1950); W. G. Kammlade, *Sheep Science*, rev. ed. (1955); M. E. Ensminger, *Sheep and Wool Science*, 3rd ed. (1964); Marion Clawson, *The Western Range Livestock Industry* (1950); A. H. Clark, *The Invasion of New Zealand by People, Plants and Animals: the South Island* (1949); E. T. Baker, *Management and Feeding of Sheep* (1947); S. B. Collins, *Profitable Sheep* (1956); P. V. Ewing, *Karakul Sheep* (1943); J. L. Horlacher and C. Hammonds, *Sheep* (1953); F. S. Hultz and J. A. Hill, *Range Sheep and Wool in the Seventeen Western States* (1931); E. M. Roberts *et al., Sheep and Wool Production in the Soviet Union* (1963); E. M. Juergenson, *Approved Practices in Sheep Production*, 2nd ed. (1963); I. E. Newsom, *Sheep Diseases*, 3rd ed. (1964); J. F. Walker, *Breeds of Sheep* (1942), *Sheep Breeding* (1945); H. B. Austin, *The Merino: Past, Present and Probable*, 3rd ed. (1950); H. G. Belschner, *Sheep Management and Diseases*, 3rd ed. (1953); A. Hawkesworth, *Australasian Sheep and Wool*, 7th ed. (1948); E. H. Pearse, *Sheep Farm and Station Management*, 6th ed. (1946); Victoria Department of Agriculture, *The Sheep Industry*, 9th ed. (1949); New Zealand Department of Agriculture, "The Sheep and Wool Industry of New Zealand," *Bulletin No. 274* (1950); J. R. Gray, "Southwestern Sheep Ranches Organization Costs, and Returns, 1940–54," Agricultural Experiment Station *Research Report 7* (1956); U.S. Department of Agriculture, *1956 Yearbook of Agriculture, Animal Diseases* (1956). (J. K. R.; M. A. D.)

SHEERNESS, a seaport, naval establishment and urban district on the Isle of Sheppey (*q.v.*), in the Faversham parliamentary division of Kent, Eng., on the right bank of the Medway estuary at its junction with the Thames, 48 mi. E. of London. Pop. (1961) 13,691. The town is divided into three sections—naval; residential and business; and recreational. Blue Town, the oldest, contains the naval dockyard built on a site surveyed by Samuel Pepys, secretary to the Navy, in 1665. The fort at Garrison point was built after the taking of Sheppey by the Dutch in 1667. In 1797 Sheerness was involved in the Nore mutiny by the landing in force of the mutineers and it was there that Lord Nelson's body was brought ashore from his flagship, the "Victory." Mile Town, built after Blue Town, comprises shopping and residential streets; while Marine Town, lying along the sandy eastern shore, forms an additional residential area, with pleasure gardens and an open-air swimming pool.

SHEFFIELD, a city, county and parliamentary borough in the West Riding of Yorkshire, Eng., 158 mi. N.N.W. of London, 38 mi. E.S.E. of Manchester and 52 mi. S.S.W. of York by road. Pop. (1961) 494,344. Area 61.9 sq.mi. It is in the extreme south of the county, at the foot of the Pennines and at the junction of the Don with its tributaries the Sheaf, Porter, Rivelin and Loxley.

At the time of the Domesday survey Sheffield (Escafeld) was a subdivision of the large manor of Hallam. At that time the Meersbrook, the Limb Dyke and the Sheaf, which had divided Mercia from Northumbria, were the boundary between Yorkshire and Derbyshire and continued to be so until 1900. After that, extensions of the city boundary carried Yorkshire several miles into Derbyshire.

Early in the 12th century, the Norman William de Lovetot built a castle at the confluence of the Sheaf and the Don, and a parish church. These two focal buildings were in, or near to, the submanor of Sheffield, and this was no doubt why the name of Sheffield soon superseded that of Hallam for the town and manor; though Hallamshire (Sheffield, Ecclesfield and Bradfield) continued to be an area of administration cut off from the rest of Yorkshire by the great moors north of Bradfield.

In 1297 Thomas de Furnival, then lord of Hallamshire, granted to his "Free tenants of the town of Sheffield" certain privileges as burgesses. These free tenants continued to be the body most concerned with local government until in the reign of Edward VI certain property which had been left to them in trust was forfeited to the crown under the Act for the Suppression of Colleges and Chantries. On their petition it was restored by Queen Mary in 1554, and as she preferred to separate the ecclesiastical from the civil duties she created by charter a body of 12 capital burgesses who discharged their functions side by side with the free tenants. Both are now charitable trusts—the Church Burgesses and the Town Trustees. Mary, queen of Scots, spent some time at the Manor lodge during the 14 years of her captivity at Sheffield castle.

Industries.—The district is rich in sources for metalworking. From early times the local iron ore had been smelted by charcoal obtained from the abundant woodlands. Smiths, and probably cutlers, were active in the neighbourhood as early as the 14th century. The millstone grit of the district provided excellent grindstones. During the 15th century the streams which converge on Sheffield began to be used for power for grinding and forging. Though the cutlery industry was for many years a rural one, carried on in a succession of water mills which packed the banks of the rivers from their sources on the moors to the boundary of Rotherham, the importance of Sheffield was established as the market town where the wares were finished and sold. This resulted in its emergence as the main provincial cutlery town and a powerful rival to the cutlery trade of London. By 1700 the London cutlers also were defeated; thereafter Sheffield enjoyed a monopoly of the English cutlery trade. The variety and quality of products reached a very high level in the late 18th and early 19th centuries—spring and table knives, razors, scissors, surgical and mathematical instruments, edge tools, saws, agricultural and joiners' tools, files and rasps of every grade and size. The introduction of steam grinding released the industry from its dependence on the rivers but until very recent times there was little change in the organization of an essentially small-firm industry.

The crisis of Sheffield's industrial history was reached about 1740, when Benjamin Huntsman moved from Doncaster to Handsworth (now within Sheffield) to experiment in the production of a steel harder and more even than could be made by the older methods of smelting and forging. The crown of his experiments was crucible steel, the reliable tool steel which marked the beginning of special steelmaking in England. By 1830 Sheffield was recognized as the centre of high-grade steel manufacture. Shortly afterward the opening of the first local railways made possible the movement, and therefore the manufacture, of really large castings. It followed naturally that when in 1856 Henry Bessemer invented his method of making inexpensive steel in large quantities, he should set up a works in Sheffield. During the remainder of the 19th century the heavy industries grew rapidly, aided by the development, by Sir John Brown (*q.v.*) and Charles Cammell of the manufacture of armour plate for ships, and the discovery by Sir Robert Hadfield (*q.v.*) of a workable manganese steel. Development in alloy steels, including the perfection by Harry Brearley of noncorrosive ("stainless") steel, in 1912, went on apace. The area between Sheffield and Rotherham was quickly filled with the great works which clang and glow so impressively through the night in the Don valley.

About the same time that Huntsman was experimenting Thomas Boulsover, a cutler, discovered the process of plating copper with silver by fusion which resulted in the beautiful ware

known as old Sheffield plate. (See SHEFFIELD PLATE.) Factories for making it were established rapidly and played a large part in bringing to the town wealth which had eluded the older small-scale cutlery manufacturers. The discovery of the method of electroplating killed the more expensive plate trade almost overnight in 1850, and the surviving specimens have become museum pieces eagerly sought by collectors; but the silver and electroplating trades continue, and the city still has one of the few provincial assay offices (established 1773).

The cutlers organized their industry on craft guild lines very early; the first trade-mark known was granted in 1554. In 1624 they were incorporated as the Company of Cutlers in Hallamshire, with jurisdiction for six miles round Hallamshire as well. As a guild, the company relinquished most of its powers in 1812, but has since revived as a powerful influence in upholding the quality of Sheffield wares. By acts of 1883–88, the company has the right of registering trade-marks for all goods composed in whole or in part of any metal, wrought or unwrought, in its ancient area of jurisdiction, and its membership now includes the manufacturers of steel as well as cutlers.

Although the cutlery, steel and auxiliary trades still occupy the greater part of Sheffield's wage earners, there are important minor industries. Haft, handle and case making naturally accompany the cutlery trades. Sheffield has the largest type foundry in England and there are snuff factories, food factories, confectionary works, fruit and vegetable canning, iron and brass founding, paper manufacture, bookbinding and the making of optical instruments, bicycles, brushes, railway fittings, chemicals, paints and varnish.

Communications.—Sheffield lies on no great lowland route or natural highway. To the northwest and south the valleys lead to Pennine dales and lofty moorlands. For many centuries the only ways were along steep and narrow tracks, except beside the Don from Rotherham.

In 1726 improvements were made in the navigation of the Don from Doncaster west to Tinsley, to help the trade of Sheffield; in 1814 a canal from the heart of Sheffield to Tinsley completed the outlet by water. Ten turnpike roads made between 1756 and 1819 brought the town out of its isolation, but the gradients daunted the early railway engineers, and the first line, opened in 1835, only connected Sheffield with Rotherham, on the original Midland line.

The present main line through Chesterfield was not built until 1876, after the tunnelling of the Sheaf-Rother watershed. Long tunnels connect Sheffield by rail with the west and even in the 1950s relatively little of the trunk traffic of England passed through the city.

Later History.—During the period generally known as the Industrial Revolution the population expanded so rapidly that the ancient bodies of local government could not cope with the new problems of urban management. The first interest of the day was in parliamentary reform and in 1832 Sheffield was allotted two members in the reformed parliament. The Municipal Corporations act of 1835, allowing named towns to apply for a charter of incorporation as boroughs of the modern type, was adopted in 1843. In 1893 the borough was created a county borough and city and in 1897 its chief magistrate received the title of lord mayor. In 1911 the census returns showed that Sheffield had taken a place above Leeds as the largest city in Yorkshire. In this period the built-up area spread rapidly along the valleys; during the 20th century pleasant new suburbs were extended over the hillsides.

In 1914 the diocese of Sheffield was created from the see of York, the old parish church becoming the cathedral church of St. Peter and St. Paul. Only the tower and chancel of the 15th-century building (which replaced two earlier ones) remain, but the present building has some fine alabaster monuments and some interesting modern stained glass.

In 1897 the university college was created by the amalgamation of the medical school (founded in 1828), the technical school (1886) and Firth college (1879), a university extension venture for which Mark Firth, a steel manufacturer, had provided the building. In 1905 the college received its charter as the University of Sheffield. Although it provides ample opportunities for a broad cultural education, the university especially developed branches of study and research in glass and fuel technology, metallurgy, mining and engineering. Sheffield was one of the first towns to provide secondary education; at mid-20th century there were eight grammar schools, including King Edward VII school, formed in 1905 by the union of the old Royal Grammar school and two private schools, administered by the municipality. There were also colleges of arts and crafts, and commerce and technology.

BARNABY'S PICTURE LIBRARY

THE TOWN HALL, SHEFFIELD, DESIGNED BY C. W. MOUNTFORD, 1897. THE TOWER IS CROWNED WITH A BRONZE STATUE OF VULCAN

The town hall, opened in 1897, is a good example of Victorian Gothic housing the municipal offices. Twentieth-century buildings include the city hall (1932) for concerts and meetings; the city museum (1937) with excellent collections of cutlery, plate and Derbyshire antiquities, and the central library and Graves art gallery (1934). The library is a fine one, well known for material on ferrous metallurgy and for collections of local archives. In Meersbrook park was the Ruskin museum (closed 1952) containing John Ruskin's collection of works of beauty lent to Sheffield from St. George's guild by his desire.

Sheffield's growth and appearance have been adversely affected by the lack of large commercial undertakings in earlier periods, by the necessarily dirty nature of its major industries and by two heavy air raids sustained in Dec. 1940. Nevertheless, finely sited on a hill-and-valley system of great beauty, Sheffield is a pleasant and in some ways a beautiful city. There are open spaces and tree-lined streets less than a mile from the city centre, and the purple moors of the Pennines and the deep wooded dales of Derbyshire sweep up to the very edge of the residential area.

See A. Gatty, *Sheffield: Past and Present* (1873); *Handbook and guide to Sheffield*, pub. for Brit. Ass. (1910); M. Walton, *Sheffield; Its Story and Its Achievements*, 3rd ed. (1952). (My. W.)

SHEFFIELD PLATE, articles produced from copper to which a coating of silver has been fused. The process was discovered about 1742 in Sheffield, Eng., by Thomas Boulsover (Bolsover), a cutler, and was soon used to make food-serving dishes and the like. Design and workmanship were in time brought to a very high level, and these early pieces were sometimes impressed with hallmarks resembling those used on silver—a practice that was prohibited by an injunction obtained in 1773 by the London silversmiths. By 1774, however, the Sheffield plate makers were again authorized to use marks that bore the name of the maker and a distinctive device not used for silver. The industry was located almost entirely in Sheffield, where there were a number of manufacturers, and to a lesser extent in Birmingham, where the great name was that of Matthew Boulton (*q.v.*). By about 1860 the fusion process had been superseded by electroplating (*q.v.*) and the industry was defunct.

About 30 years after its disappearance as a commercial commodity, the acquisition of "old Sheffield plate" became a cult among collectors. Since the demand soon exceeded the supply, a number of manufacturers began to make pieces reproducing the old designs by electroplating on copper. The sale of such

pieces as "Sheffield plate" provoked the Sheffield Cutlers' Company to take action, and in 1911 it was established in court that the term "Sheffield plate" could only be applied to articles made by the process of plating by fusion. An intermediate category of goods was left, however, of genuine "Sheffield plate" which had been worn through to the copper by hard usage and later electroplated.

The cult of "old Sheffield plate" gradually declined and specialist collectors disappeared, although the more ornamental pieces such as candelabra and tea urns, which can be lacquered and used as decorative adjuncts to antique furniture, continued to be prized by connoisseurs.

BY COURTESY OF SHEFFIELD CITY MUSEUM

SAUCEPAN, BY JOSEPH HANCOCK, MADE ABOUT 1755, SHEFFIELD, ENG.

BIBLIOGRAPHY.—F. Bradbury, *A History of Old Sheffield Plate* (1912) remains the standard authority; S. B. Wylear, *The Book of Sheffield Plate* (1949) includes reproductions of early electroplaters' marks, which serve as a check during the period when firms were changing from fused to electroplating; E. Wenham, *Old Sheffield Plate* (1955) gives a concise account of the subject.

SHEIKH (SHAIKH or SHAYKH), an Arabic title of respect dating from pre-Islamic antiquity, strictly means a venerable man, of more than 50 years of age. It is specially borne by heads of religious orders, heads of colleges (*e.g.*, Al Azhar in Cairo), chiefs of tribes, and headmen of villages and of separate quarters of towns. It is also applied to learned men, especially members of the class of *'ulama'* (*q.v.*), and has been applied to anyone who had memorized the whole Koran, however young he might be.

Combined with other terms the word is used in various titles: *shaikh al-balad*, mayor of a town; *shaikh al-jabal* ("the mountain chief"), a popular term for the head of the Assassins mistranslated by the crusaders as "the Old Man of the Mountain" (*see* ASSASSIN). By far the most important of these is the title *shaikh al-Islam*, which by the 11th century A.D. was being given to eminent *'ulama'* and mystics and by the 15th century could be claimed by any outstanding mufti (*q.v.*). In the Ottoman Empire this title was restricted by Suleiman I (1520–66) to the mufti of Istanbul, who was equal in rank to the grand vizier and was head of the religious institutions which controlled law, justice, religion, and education. Because of his right to issue legally binding *fatwas* (opinions) this official came to wield great power. In 1924, under the Turkish Republic, the last vestiges of the institution were abolished. (W. M. WT.)

SHEIKH OTHMAN, a town of Aden state, part of the British colony of Aden, in the Federation of South Arabia, lies on the sandy mainland part of Aden, about 6 mi. N of the port. Pop. (1960 est.) 35,000. Purchased from the sultan of Lahej in 1882 as a settlement for the surplus population of the Aden Peninsula, it proved too unhealthy until sanitary measures completely changed the outlook. Governed municipally by the Settlement Committee when Aden formed part of British India, in 1945 Sheikh Othman (including the villages of Hiswah, Al Burayqah (Bureika), Bi'r Fuqum, and Little Aden) was given its own township authority. In 1955 Little Aden and the fishing villages of Al Qaissa (Al-Khaisa) and Bi'r Fuqum were granted separate status as the Little Aden Township Authority. (W. H. IS.)

SHEKINAH, a Hebrew noun (derived from the verb *shakan*, "dwell") meaning "dwelling" or "presence" with specific reference to the presence of God. It first appears in the Aramaic form *shekinta* in the paraphrastic Aramaic translations of the Old Testament known as Targums (*see* TARGUM), and is used frequently in Talmud and Midrash and other postbiblical Jewish writings. In the Targums it is used, like *memra* ("word") and *yeqara* ("glory"), as a substitute for "God" in passages where the anthropomorphism of the original Hebrew seemed likely to mislead. Thus, in Ex. 25:8, "that I may dwell in their midst" becomes in the Targum of Onkelos "and I will cause my Shekinah to rest among them." By this usage belief in the omnipresence and transcendence of God was safeguarded; and in many passages Shekinah is simply a reverential substitute for the divine name.

In rabbinic literature the Shekinah is associated with several other religious and theological terms. There is appropriateness in the thought that the Shekinah descended on the tabernacle (Hebrew *mishkan*, from the same root as Shekinah) and on Solomon's Temple, though it is said that it was one of five things which were lacking from the Second Temple (the other four being the ark, holy fire, spirit of prophecy, Urim and Thummim; *see* TEMPLE, JEWISH). The glory of God which filled the tabernacle (Ex. 40:34) was thought of as a bright radiance; and the Shekinah is sometimes similarly conceived. This idea appears in John 1:14, where the Greek verb "dwelt" or "tabernacled" (*eskenosen*) recalls the form as well as the sense of the Hebrew verb *shakan*. There is also an affinity between the Shekinah and the Holy Spirit, though the two are not identical. Both are associated with prophecy; both may be lost because of sin; and both are connected with the study of the Torah. Two later developments of the idea are that the Shekinah was a created entity, distinct from God (Maimonides), and that Shekinah was a female element in God: the bride and daughter of God (Cabala).

BIBLIOGRAPHY.—J. Abelson, *The Immanence of God in Rabbinical Literature* (1912); G. F. Moore, *Judaism in the First Centuries of the Christian Era*, vol. i, pp. 434–438 (1927); G. G. Scholem, *Major Trends in Jewish Mysticism*, 3rd ed. reprinted with enlarged bibliography (1961). (G. W. AN.)

SHELLAC: *see* LAC; RESINS: *Natural Resins.*

SHELLEY, MARY WOLLSTONECRAFT (1797–1851), English writer, who has a place in literary history as the author of *Frankenstein*, and as the daughter of William Godwin and Mary Wollstonecraft Godwin (*qq.v.*) and the second wife of Percy Bysshe Shelley, was born in London on Aug. 30, 1797.

BY COURTESY OF THE NATIONAL PORTRAIT GALLERY, LONDON

MARY WOLLSTONECRAFT SHELLEY, PORTRAIT BY R. ROTHWELL, 1841

Mary came as near as any woman could to meeting Shelley's requirements for the partner of his life: "one who can feel poetry and understand philosophy." After his death in 1822 she published his *Posthumous Poems* (1824), edited his *Poetical Works* (1839), with long and invaluable notes, and his prose works (*Essays, Letters from Abroad*, etc., 1840). Her *Journal* is the backbone of Shelley biography and her letters are an indispensable adjunct.

Frankenstein, or the Modern Prometheus (1818) is perhaps the most widely known pseudoscientific novel. Its style is immature, but the basic idea is fascinating: the power of the scientist to form and give life to a human being and the dreadful consequences which this act produces. The treatment of psychological and social problems is extraordinary, the more so in that the book was completed before Mary's 20th birthday. Her second, historical, novel, *Valperga: or, the Life and Adventures of Castruccio, Prince of Lucca* (1823), is in literary style her best work; it was produced under Shelley's direct stimulus. But in general esteem *The Last Man* (1826) ranks as her best novel. In this and later novels—*The Fortunes of Perkin Warbeck* (1830) excepted—she made use of her rich early life by portraying Shelley, Byron, Emilia Viviani (the inspiration of Shelley's *Epipsychidion*), E. J. Trelawny (*q.v.*), and real episodes from her experience. *Lodore* (1835) and *Falkner* (1837) are inferior to the earlier novels. To Dionysius Lardner's *Cabinet Cyclopaedia* (1829–46) she contributed *Lives of the Most Eminent Literary and Scientific Men of Italy, Spain, and Portugal*

(1835–37) and *France* (1838–39). Of more permanent value are her travel books: *History of a Six Weeks' Tour* (1817), which recounted the continental tour following her elopement with Shelley in July 1814 and their summer near Geneva in 1816; and her eminently readable *Rambles in Germany and Italy* (1844).

After Shelley's death her life was devoted to educating her only surviving child, Percy Florence, and to assisting her aged father (d. 1836). Shelley's father, Sir Timothy, made only a small allowance for Percy's education, and when, in 1844, Mary achieved monetary freedom following Sir Timothy's death at the age of 90, she was by that time too ill to derive much pleasure from her relative affluence. She died in London on Feb. 1, 1851, and was buried at Bournemouth.

BIBLIOGRAPHY.—R. Glynn Grylls, *Mary Shelley* (1938); F. L. Jones (ed.), *The Letters of Mary W. Shelley* (1944) and *Mary Shelley's Journal* (1947); Elizabeth Nitchie, *Mary Shelley* (1953). *See* also SHELLEY, PERCY BYSSHE, especially the *Bibliography*. (F. L. J.)

SHELLEY, PERCY BYSSHE (1792–1822), English Romantic poet, was born on Aug. 4, 1792, at Field Place, near Horsham, Sussex. He was the son and heir of Timothy Shelley, Conservative Member of Parliament for Shoreham, and the grandson of Sir Bysshe Shelley, bart. Even as a child he was remarkable for the subtle feeling and originality of ideas which, combined with his later mastery of language, were to give him his place in literature. His short life, dedicated to the promotion of peace and love between individuals and among humanity at large, was, paradoxically, to involve both himself and others in much suffering; this was because his powers and his ideals were often in advance of his experience. Yet out of suffering came a strength which passed into his poetry.

Early Years.—Shelley's childhood was coloured by family affections, a comfortable home, the Sussex countryside, and a world of imagination derived from books, among them the fashionable Gothic or "horror" novels. By 1808, as is shown by early letters, his cousin, Harriet Grove, had become a centre for his affections. From a private school—Syon House Academy at Brentford, Middlesex, where he acquired the rudiments of science—in 1804 he went on to Eton. The brutalities, misery, and sense of isolation he endured there must have affected him. But among the more beneficial results of his schooldays were the fluency he derived from an early training in Latin composition and that robustness of intellect which the classical tradition stimulates in those whose ability is equal to its demands. Before he left school, he published *Zastrozzi: a Romance* (1810), an early attempt at a Gothic romance in the style of Mrs. Radcliffe (*q.v.*). *Original Poetry by Victor and Cazire*, written with his favourite sister, Elizabeth, and published in the same year, revealed a fluency of versifying in metres then fashionable.

At Eton an important direction had been given to Shelley's mind by his elderly friend, Dr. James Lind (1736–1817), a fellow of the Royal Society living at Windsor, who introduced him to Plato and notably to the *Symposium*. In October 1810 he entered University College, Oxford. Though his Greek does not seem, at this stage, to have been as good as his Latin, he read other Platonic dialogues in translation and was eloquently enthusiastic. With this enthusiasm went an eager interest in chemistry, physics, and astronomy. For Shelley, however, as for Lucretius—one of his favourite authors—the "sciences" were a background to the theme of "Man in the Universe," on which he enjoyed discoursing with his fellow student, Thomas Jefferson Hogg (*q.v.*). In November 1810 appeared, supposedly edited by "John Fitz-Victor," *Posthumous Fragments of Margaret Nicholson*, a mock-revolutionary glorification of the mad washerwoman who had tried to stab George III.

At heart, however, Shelley was always repelled by the violence attending revolutionary action. And so, when he was drawn to William Godwin (*q.v.*) and the 18th-century French philosophers, it was not as a revolutionary but as a reformer. For some years he was to be primarily a writer of inquiring prose on politics and theology. But in England at that period, with its panic fear of revolutionary infection from France, even questioning could be dangerous. In disgust at churchmen who professed Christian charity but regarded those reformers who sought to apply its principles to the alleviation of human misery as heretics, Shelley came to question religion itself, and in March 1811 he circulated to the bishops and heads of colleges a pamphlet entitled *The Necessity of Atheism.* Though it was an inquiry into, rather than a denial of, religion, it provided the Oxford authorities with a reason for expelling him, with Hogg, his friend and supporter.

BY COURTESY OF THE NATIONAL PORTRAIT GALLERY, LONDON

SHELLEY, PORTRAIT BY AMELIA CURRAN, 1819

Harriet Westbrook.—Shelley's father demanded an immediate recantation of the pamphlet and a complete break with Hogg. Refusing both, Shelley, banished from home, settled in London. The views that had already cost him his life at Oxford and at Field Place next lost him the affection of his cousin Harriet. Some clue to the emotional vacuum troubling this part of his adolescence may be found in the epigraph (in Latin) attached to some verses written before he left Oxford, which may be translated: "I was not yet in love and was in love with loving; I was looking for something to love, loving to love." This epigraph was to be used again for *Alastor; or the Spirit of Solitude* (published 1816, preface dated Dec. 14, 1815). It heralds a quest which was to run throughout his life: the quest for a Woman-Symbol, expressive of Intellectual Beauty. Harriet Westbrook, the daughter of a retired hotelkeeper and a school friend of his sisters at Clapham, whom he had met before he left Oxford, became the symbol he needed: in his eyes she was Love and Beauty, demanding to be rescued by Shelley-the-Libertarian from family prejudice, and platonically molded on the lines of the *Symposium.* Five months after leaving Oxford he eloped with her to Edinburgh, where they were married on Aug. 28, 1811; they were remarried in England in March 1814.

His father's anger at the marriage arose more from consideration of Shelley's position and prospects than from any fault in Harriet. Her social inequality has been exaggerated: she was well-bred, charming, and adaptable. All too easily she fell into Shelley's belief that incompatibilities could be smoothed away by study and by the share which a 19-year-old husband and his 16-year-old wife might reasonably expect to have in the Shelleyan agape—that love-power which was to radiate from their household for the reconstitution of mankind. Harriet threw herself enthusiastically into Shelley's schemes, from land reclamation in Wales to the emancipation of Ireland. Forward-looking idealism failed to consider the poverty that might come from this new affront to Shelley's father, combined with Shelley's extravagant generosity to the needy and importunate; nor could present happiness foresee the time when the claims of babies and the learning of Latin would pull the young wife in opposite directions. In their inexperience, Shelley and Harriet failed to recognize the threat to their marriage latent in the presence, almost invariably, of a third person in their household: first Hogg, whose brotherly love for them both soon turned to a more than brotherly love for Harriet; then Miss Hitchener, the Sussex schoolmistress who, from the status of "soul sister" in an immense correspondence, was to become, domestically, the hated "Brown Demon," pensioned off by Shelley; and, finally, the possessive Eliza, Harriet's elder sister.

But the years 1811–12, during which the Shelleys moved energetically about Scotland, England, Wales, and Ireland, were happy. They have their memorial in *Queen Mab*, finished in 1812, and in the delightful dedicatory verses addressed with it to Harriet. This is Shelley's first major poem. Mab, queen of the fairies, takes aloft the soul of the sleeping heroine, Ianthe, and explains to her the significance of what is happening on earth. The voice is the voice of Shelley explaining Godwinism to Harriet, and the exposition marks an important transitional stage in Shelley's devel-

opment as a poet. For a time, in his reaction against Oxford, Shelley had turned from the humanizing influence of Plato to the Godwinian belief in Necessity, a grim power that swayed the forces of destiny, making only a limited allowance for the power of the human will and bringing about, by the process of "mutability," cycles of good and evil in human affairs. As the talk in the poem proceeds, the Lucretian love which "recreates all things" passes into Platonism; and this gradually overlays the Godwinism, till in the end "Necessity" has become identified with the concept of the "World Spirit" (W. B. Yeats's "Spiritus Mundi"). This was to be, throughout Shelley's poetry, a substitute for God. The nine cantos of *Queen Mab* are written in unrhymed Southeyan lines of unequal length that pass into blank verse. The copious notes published with it reveal an immense reading in four languages and a remarkable gift of exposition. Like the poem, they mark the passing of Shelley the potential crank into Shelley the man of genius.

Queen Mab; a Philosophical Poem was privately printed in the summer of 1813. This summer brought new friends and influences. At Bracknell, Berkshire, where the Shelleys took a house in July, Shelley became a member of the intellectual circle of Harriet de Boinville and her daughter Cornelia Turner, with whom he studied Italian. It was at this time too that he began to be intimate with the learned and witty Thomas Love Peacock (*q.v.*). Peacock's conversation gave him some of the stimulus that he might in other circumstances have derived from contemporaries at Oxford. One of its lasting results was his throwing himself into serious study of the Greek writers in their own language. Platonism, from this time onward, became the main fabric of his thought. This by no means excluded other threads. Into the general pattern of his reading and thought went Dante, Calderón, Goethe, and ancient and modern writers too numerous to mention; much scientific literature was included. Shelley was a scholar, humanist, and, always, a man of his time.

Mary Godwin.—Personal poems relating to the accomplished Cornelia Turner, and a reference to his "contemplation of female excellence," may be clues to the comparisons that Shelley was making in 1814. As a Woman-Symbol of Intellectual Beauty, Harriet was no longer adequate. Peacock, the Boinvilles, and others that understood his world of intermingled ideas and feeling could understand this, though they naturally sympathized (as we must) with the affectionate wife who had done no wrong. When, in March 1814, Shelley met Mary Godwin (*see* SHELLEY, MARY WOLLSTONECRAFT), the daughter of Godwin and Mary Wollstonecraft (*see* GODWIN, MARY WOLLSTONECRAFT), she became the new Woman-Symbol, a role for which her parentage alone was a qualification in Shelley's eyes. Godwin's behaviour was typical. When on July 28 Shelley and Mary eloped to the continent (going first to Paris and then to Switzerland), accompanied by Claire Clairmont, Mary's step-sister, to make up the inevitable, recurrent, domestic triangle, Godwin, the philosopher who had preached against marriage, banished Shelley from his friendship. He never ceased, however, to importune him for money. Claire too was to be in many ways a burden to Shelley for the rest of his life.

In September, after financial difficulties in Switzerland, the travelers returned to London and settled at Bishopgate. In January 1815 Shelley's grandfather died, and his father, who succeeded to the baronetcy, decided to allow him £1,000 a year in respect of his expectations. As Harriet and their son, Charles Bysshe (b. 1814), had to be provided for, as well as Mary and the Godwins, this was little, and in a year of anxiety and maladjustment debts loomed large. After leaving England with Mary, Shelley had asked Harriet to join the agape-party as a sister, a Platonic arrangement that seemed to him very reasonable but one that she, not unnaturally, declined. *Alastor*, published, with other poems, in 1816, is a self-projection. It tells of a poet torn by good and evil passions symbolized by pursuing daemons.

In May 1816 Shelley, Mary, and Claire went to Geneva, and, while Claire made the most of the liaison into which she had enticed Byron (*q.v.*), Shelley and Mary enjoyed a Platonic idyll in the scenes consecrated by the lovers in Rousseau's *La Nouvelle Héloïse*. "Mont Blanc" and the "Hymn to Intellectual Beauty" are manifestoes of Shelley's poetic Platonism inspired by those days and scenes; they also inspired Mary's novel *Frankenstein* and the third canto of Byron's *Childe Harold*. Tragedy followed the Shelleys' return to England in the autumn. In November Fanny Imlay, another daughter of Mary Wollstonecraft, committed suicide at Bath, and in December Harriet Shelley was found drowned in the Serpentine, in Hyde Park, London. Harriet's suicide, if suicide it was, is surrounded by mystery. Most biographers have accepted her landlady's evidence that she "appeared to be pregnant," but there is no medical corroboration of this. Other than hearsay there is no evidence either to impugn Harriet's conduct or to suggest, beyond the fact of their separation, any responsibility on Shelley's part. Later disasters always made both Shelley and Mary view her death as the beginning of a chain of encompassing destiny. The events of 1816 ended with Shelley's marriage to Mary (Dec. 30) and the birth to Claire (in January 1817) of Byron's daughter, Allegra.

During 1817 Shelley and Mary, with Claire and Allegra, lived at Marlow, whence they paid many visits to London. It was, in many ways, a happy year. A son, called William, had been born in January 1816; a daughter, Clara, was born in September 1817. Leigh Hunt and his family were constant visitors, and Shelley enjoyed the friendly rivalry of John Keats (*q.v.*), who was then writing *Endymion*, while Shelley was busy with the poem that eventually became *The Revolt of Islam*. There were learned talks and walks with Hogg and Peacock, Claire's singing, and much boating on the Thames—boats, for Shelley, were a symbol, recurrent in his poetry, of the mind's escape from sorrow and evil to a world of goodness, love, and beauty. *The Revolt of Islam* (first printed in 1817 as *Laon and Cythna; or, the Revolution of a Golden City: a Vision of the 19th Century*, and revised under its new title in 1818) is a 12-canto poem in the Spenserian stanza (*q.v.*). It is a transcendental epic of the bloodless revolution and the paradise on earth which can be brought about by the Shelleyan-Platonic agape-love of Laon and Cythna, a hero and heroine who, by a strange adaptation of the brother-brother concept of idealized friendship in Plato's *Symposium*, are also brother and sister. Tediousness is diluted by some fine eloquence, and the dedication to Mary is a delightful example of Shelley's performance in his graceful, lighter vein.

Then came a new disaster. The lord chancellor finally, after a prolonged case, decreed that Shelley was unfit to have custody of Ianthe and Charles, his children by Harriet. Shelley's health was giving way, and he was beginning to fear that Clara and William might also be snatched from him. In March 1818 he and Mary left England forever. Claire and Allegra went with them.

Italy.—The Shelley who fled to Italy with Mary in 1818 was no longer the young reformer who had pursued liberty and justice round England, Wales, and Ireland with Harriet. Though his notebooks are full of statistics and memoranda on contemporary problems, by 1818 the reformer had become merged in the poet, and poetry and allegory, rather than political pamphlets, were to be the vehicle for his social and religious ideas. In Italy he found fulfillment. The sea and sky, the lakes, rivers, and mountains delighted the poet of Nature, while the treasures of antiquity and the Renaissance inspired and exalted the humanist. The Italian years had their sorrows. All three of the children he had brought from England died of disease: Clara in September 1818; William on June 7, 1819; Allegra on April 19, 1822. There were periods of tension with Mary, and there was chronic anxiety about the troubles of Godwin and Claire. Despite these sorrows, and despite constant travel, Shelley maintained a life of unremitting study, composition, and correspondence. Much of the Shelley of 1818–19 went into the great philosophical verse-drama *Prometheus Unbound, a Lyrical Drama in four acts*, published, with other poems, in 1820. Its hero is a libertarian and reformer, persecuted by the ruler of the universe for his attempts to improve the lot of humanity. Not till he has forsworn revenge can he unseat the tyrant. Then, following the union of Prometheus (Mind) with Asia (Love and Beauty), there ensues a reign of peace and happiness on earth. The same gospel of "Spirit, Patience, Gentleness" is the Shelleyan prescription for the workingmen of Manchester

in "The Masque of Anarchy," inspired by the Peterloo Massacre (*q.v.*) of August 1819—though the gentleness of the policy urged is in strong contrast to the ferocity of Shelley's language against purblind refusal to see the need for it. *The Cenci: a Tragedy, in five acts*, written in the summer of 1819, was a triumph of sheer ability, a pastiche written by a man ignorant of the theatre, under the conviction, shared by Keats and others, that a neo-Elizabethan play could be produced by sedulous copying of Elizabethan verse, situations, and effects. Many who, like Mary Shelley, find difficulty in understanding the philosophical content of Shelley's more essential works have followed her in rating this work higher than he did.

The writing of Act I of *Prometheus Unbound* had been followed by the death of Clara Shelley at Este. Acts II and III are full of delight in the Roman spring that preceded the death of William. October 1819 found Shelley and Mary at Florence awaiting the birth of another child. Shelley was full of grief for the revolution he believed to be impending in England and for the loss of his children; he was angry too about the fierce reviews of *The Revolt of Islam*. It seemed that an evil destiny gripped all he had ever loved: were the offspring both of his mind and his body always to perish? But as he walked in the Cascine he watched the windswept leaves, and they became for him a symbol. The symbol grew into the "Ode to the West Wind," Shelley's prayer that his ideas for mankind may be driven over the universe, quickening a new birth of thought and action as the wind quickens the leaves. There could be no better example than this great ode of the power that Coleridge called "the mystery of genius in the fine arts"—the power "to make Nature thought and thought Nature." Though intellectually divorced from institutional religion, Shelley had the power of faith of the naturally devout. Faith brought him renewed strength, out of which came a fourth act to Prometheus, a hymn of joy that corresponds to the choral finale Beethoven was at this time meditating for his Ninth Symphony.

Here are Earth's great lines about the mission of Man:

> All things confess his strength. Through the cold mass
> Of marble and of colour his dreams pass;
> Bright threads whence mothers weave the robes their children wear . . .

Shelley goes on to what seems almost a prevision of the airplane and the atomic age:

> The lightning is his slave; heaven's utmost deep
> Gives up her stars, and like a flock of sheep
> They pass before his eye, are numbered, and roll on!
> The tempest is his steed, he strides the air;
> And the abyss shouts from her depth laid bare,
> Heaven, hast thou secrets? Man unveils me; I have none.

Nor did Shelley omit to set down some thoughts for more immediate dissemination. In *A Philosophical View of Reform* he urges the rulers that the way to avoid revolution and anarchy is not through bloody repression but through gradual change. As *A Philosophical View* was first published in 1920 it cannot be said that its ideas influenced his countrymen; but since it was in this way that the reforms of the 19th century were achieved in England it must be remembered that Shelley was the precursor, if not the inspirer, of constitutional reform.

A Defence of Poetry (written, 1821; published, 1840) is his most effective prose work. In it he gives a warning of what can happen when "science" progresses more rapidly than the arts and humanities, so that man acquires a mechanical power which he lacks the wisdom to use. Mary Shelley's *Frankenstein* had already given a similar warning. Other thoughts of Shelley's about uses for contemporary science appeared in his scheme in 1819, that *annus mirabilis* of his poetry, for starting a service of the newly invented steamboats between Leghorn and Marseilles.

Epipsychidion, a long poem in rhymed couplets, composed in January 1821 and published anonymously later in the same year, is the culmination of Shelley's quest for the Woman-Symbol. It derives from the situation in a Pisan convent of Emilia Viviani, a young noblewoman whom he believed to be the victim of parental tyranny. He imagines a space voyage with his Platonic love to an enchanted isle, radiant, like the Cave of Asia and Prometheus, with truth, beauty, and freedom. By the spring of 1821 he had turned against both Emilia and his poem. "The error," he wrote, "consists in seeing in a mortal image the likeness of what is perhaps eternal." So much for the ideal that had cost so much to the idealist and to several who loved him.

But close upon this disillusionment came news that roused Shelley to unsurpassed thought and feeling: Keats had died in Rome in February. Quite wrongly, Shelley believed him to have been killed by the stupidity and malice of reviewers. *Adonaïs: an Elegy on the Death of John Keats* (1821) is less a lament for a friend than the elegy of one bard for another. Into Spenserian stanzas adapted in the Greek pastoral style, he poured the accumulated power of philosophic imagery culled from Plato, Dante, and Calderón. The darkness of ignorance has put out a light of poetry: so, at any rate, it seems to us, on our side of the veil, but, asks the poem, how can we judge the reality of the world of eternity by our own transitory world of semblances? Among the poem's images is that of winter passing into spring, by which thoughts of Keats's immortality are joined to Shelley's mood of strength-through-sorrow in October 1819, when the might of the west wind became the symbol both of his own poetry and of the poetic power itself. It is, indeed, in language seeming to borrow the strength of the wind that the elegy on Keats thunders past the miserable, ephemeral reviewer—"Thou noteless blot on a remembered name"—bearing Shelley toward the vision wherein

> The soul of Adonaïs, like a star,
> Beacons from the abode where the eternal are.

No "error" now. Shelley has found "the likeness of what is eternal" symbolized not in a living woman, but in a dead poet.

The year 1821 was concluded triumphantly, with the composition of *Hellas, A Lyrical Drama* (published 1822). This is a paean of joy in which liberated Greece stands for the whole heritage of Western civilization. What new turnings of thought his final masterpiece, "The Triumph of Life," might have had can only be guessed. It was unfinished when, on July 8, 1822, in his little boat, the "Don Juan," he was drowned in a storm off the Tuscan coast. His body was cremated on the shore in the presence of his friends Leigh Hunt, Byron, and E. J. Trelawny (*q.v.*), the Cornish adventurer.

Conclusion.—Shelley believed it to be his special function and ability

> . . . to apprehend minute and remote distinctions of feeling, whether relative to external nature or to the living beings which surround us, and to communicate the conceptions which result from considering either the moral or the material universe as a whole.

This function and ability are shown throughout his major poetry. There must always be many who are unable to apprehend feelings as subtle as Shelley's and conceptions so broad and deep. Many will be attracted to Shelley's poetry by a limited understanding of it. Much that has been written about it is based on misjudgment of the complexities of his life, rather than on true judgment of the complexities of his thought. Shelley's lyrics are much anthologized. Personal, minor poetry for the most part, they have a hard philosophical core which is often unperceived. Their more obvious beauties demand no very high level of understanding, and for this reason they have been disproportionately praised while being anomalously described as "fragile." This critical cliché, combined with Matthew Arnold's biographical one—"the ineffectual angel"—has done much to popularize a distorted impression of Shelley and his poetry. By the mid-20th century, scholarship, in conjunction with the evidence of his manuscripts, had established Shelley as a thinker in advance of his time. His translations, from both classical and modern languages, his power of Byronic wit (in "Peter Bell the Third," for example), and his Aristophanic satire (as in *Swellfoot the Tyrant*) still await the evaluation that is their due. His music and language have influenced many poets. Only Yeats, perhaps, among later poets writing in English, has approached his metaphysic.

Shelley saw himself and contemporary writers as ". . . the companions and forerunners of some unimagined change in our social condition or the opinions which cement it." Among those to whom his ideas have been "companions and forerunners" have been reformers from Robert Owen and William Morris to George Orwell and movements ranging from the Chartists and Fabians to the Ger-

man revolutions of 1848, the beginnings of the Risorgimento, and the anti-Fascist campaigns of Gaetano Salvemini and Lauro De Bosis.

BIBLIOGRAPHY.—*Editions, Correspondence, etc.*: The most nearly complete edition of Shelley's *Works* is the Julian edition, by R. Ingpen and W. E. Peck, 10 vol. (1926–30; reprinted 1965). Of many collected editions of the *Poetical Works* the standard is that by T. Hutchinson (1904; rev. by B. P. Kurtz, 1934). Earlier editions still of value are that of Mary Shelley, with preface and important critical and biographical notes, 4 vol. (1839); and the 2-vol. edition by C. D. Locock, with notes and introduction by A. Clutton-Brock (1911). *See* also C. H. Taylor, *The Early Collected Editions of Shelley's Poems* (1959). Unpublished poems are contained in *The Esdaile Notebook*, ed. by K. N. Cameron (1964); *see* also *The Esdaile Poems of Percy Bysshe Shelley*, critical edition by N. Rogers (1966). *Shelley's "The Triumph of Life,"* by D. H. Reiman (1965), is valuable both as a textual and as a critical study.

Shelley's Prose (excluding letters and the romances) is edited by D. L. Clark (1954). *A Philosophical View of Reform* and *A Defence of Poetry* are included in *Political Tracts of Wordsworth, Coleridge, and Shelley*, ed. by R. J. White (1953). The edition of the *Letters* by F. L. Jones (2 vol., 1964) supersedes that by R. Ingpen (2 vol., 1909; rev. 1914). For Mary Shelley's valuable journal and letters, *see* the bibliography to the article on her.

Life.—Early lives by T. J. Hogg (2 vol., 1858), E. J. Trelawny (*Recollections of the Last Days of Shelley and Byron*, 1858; revised as *Records of Shelley, Byron, and the Author*, 1878), and T. L. Peacock (*Memoirs of Shelley*, in *Fraser's Magazine*, 1858–62; ed., with Shelley's letters to Peacock, by H. F. B. Brett-Smith, 1909), were collected and ed. by Humbert Wolfe, for Everyman's Library (1933). Later biographies include that by E. Dowden, 2 vol. (1886; rev. and condensed, 1896); N. I. White, *Shelley*, 2 vol. (1940; "corrected," 1947; abbreviated as *Portrait of Shelley*, 1945); E. Blunden, *Shelley: a Life Story* (1946); A. M. D. Hughes, *The Nascent Mind of Shelley* (1947). Documents in the Carl H. Pforzheimer Library are chronologically presented in *Shelley and His Circle, 1773–1822*, ed. by K. N. Cameron, 8 vol. (1961–).

Criticism.—See Mary Shelley's critical notes to the edition mentioned above, included in *Keats and Shelley: Complete Poetical Works* (1932); H. N. Brailsford, *Shelley, Godwin and Their Circle* (1913); C. Baker, *Shelley's Major Poetry* (1948); J. A. Notopoulos, *The Platonism of Shelley* (1949); K. N. Cameron, *The Young Shelley: Genesis of a Radical* (1950); N. Rogers, *Shelley at Work* (1956); D. King-Hele, *Shelley: His Thought and Life* (1960). *See* also *The Keats-Shelley Memorial Bulletin*, ed. by D. Hewlett (1950), and *The Keats-Shelley Journal*, ed. by M. Steele (1952–), with annual bibliographies. (N. W. Ro.)

SHELL MOUNDS (KITCHEN MIDDENS; Dan. *kjökkenmöddinger*) are prehistoric refuse heaps or mounds, found distributed throughout the world, that consist chiefly of the shells of edible mollusks intermingled with fragments of animal bones, implements of stone, bone, and horn, and other evidence of human occupancy.

Midden living first developed after the retreat of the glaciers and the disappearance of large Pleistocene fauna hunted by prehistoric man (*see* PLEISTOCENE EPOCH). Primitive peoples adopting these hunting-collecting economies became more settled; thus the oldest pottery of northern Europe, eastern North America, and Central America occurs in shell mounds (*see* ARCHAEOLOGY: *The Materials of Archaeology*).

Many shell mounds have been examined, notably on the eastern coast of Denmark (*see* SCANDINAVIAN ARCHAEOLOGY). Such accumulations indicate seasonal or year-round waterside sites. These were at first thought to be raised beaches, but cursory examination readily indicated their origin in human activities. Further investigation showed these shell mounds to belong to the late Mesolithic Ertebølle culture (*c.* 4000–2500 B.C.). These mounds contained the remains of quadrupeds, birds, and fishes apparently used as food by prehistoric human inhabitants. Among the bones were those of the wild bull or aurochs (*q.v.*), beaver, seal, and great auk, all now extinct or rare in this region. Moreover, shell mounds may contain full-sized remains of the common oyster, which at present cannot live in the brackish waters of the Baltic except near its entrance, the inference being that the shores where oysters flourished were open to the salt sea at the time the mounds were made. Thus also the shells of edible cockle, mussel, and periwinkle that abound in some kitchen middens are of full ocean size, whereas those now living in the adjoining waters are a third of full size for want of salinity. This extension of the North Sea is called the "Littorina Sea." In some places the debris is 10 to 20 ft. deep.

The more recent of these mounds give evidence that their builders had long-range contacts with emerging civilizations of western Asia (yielding remains of cultivated plants, domesticated animals, polished stone implements, and pottery), but incomplete acceptance of the agricultural orientation of the "Neolithic Revolution" (V. G. Childe). The kitchen middens of Denmark apparently were not mere summer quarters: ancient fishermen seem to have stayed in the neighbourhood for much if not all of the year; the mounds contain bones of wild animals that often give reliable clues to fix the time of year when they were killed. Remains of a winter visitor, the wild swan (*Cygnus musicus*), that leaves the Danish coast in March and returns in November, are found in abundance. Additional evidence is found among mammalian remains in the light of such periodic events as the shedding of stag's antlers and the birth and growth of young. Flint implements found include flakes, axes, awls, slingstones or net weights, and rude lance heads. One fragment of a polished axe had been worked into a scraper. Small pieces of coarse pottery are also encountered, typical vessels having pointed bases.

Middens of the British Isles (Oban, Scot., and others), of France (Tardenoisian: Téviec; *see* BRITTANY: *Archaeology*), of Italy, Spain, and Portugal (Asturian, Mugem), and of North Africa (Capsian: Ibero-Maurusian: Mechta, Afalou) also cluster along the Mesolithic-Neolithic transition line. In remote regions like southern Africa (Wilton C) and northern Japan (Jômon; *see* JAPAN: *History*), where Neolithic cultures endured longer, midden accumulations continued until the coming of iron; and in the Pacific islands they accumulated until recently.

In the Americas middens are represented by radiocarbon dates of 5000–2000 B.C. from Panama and eastern North America (*see* GEOCHRONOLOGY). Middens of Brazil, Argentina, northern Chile, Peru, and California probably antedate 2000 B.C. As in the Old World, midden living persisted outside the high civilizations, continuing in California, Florida, West Indies, and Patagonia until European conquest.

See also WORSAAE, JENS JACOB ASMUSSEN.

See "The Kitchen Middens," in G. Bibby, *The Testimony of the Spade* (1956); M. R. Matteson, "Reconstruction of Prehistoric Environments Through the Analysis of Molluscan Collections from Shell Middens," *American Antiquity*, 26:117–20 (July 1960). (JH. R. C.)

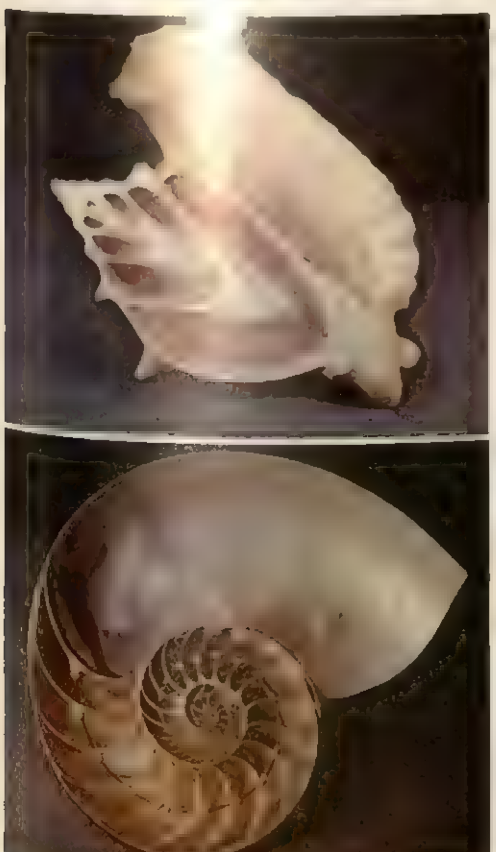
SHELLS AND SHELL COLLECTING. Shells of mollusks, because of their bright colours, rich variety of shapes and designs, and abundance along sea shores, have probably from the earliest times attracted the attention of man. He used them for ornaments, tools and coin, and dined on the mollusks that formed them. Aristotle and Pliny wrote extensively about them. In the ruins of ancient Pompeii and in a crypt in a Mayan pyramid in Yucatán shells were found that may well be the remains of ancient collections.

Shell collecting as we understand it today, however, on the same plane as coin, stamp and china collecting, is of quite recent origin. Shell collecting reached its apex in England during the late 18th and early and middle 19th centuries. This was the period of the burgeoning Pacific and China trade; new islands were being discovered and the shells found there were first imported as curiosities, later as specimens for the collectors among the newly rich merchant princes. This period reached its highest point in the 1850s and '60s when shell auctions became a common occurrence and relatively high prices were paid for particularly rare and perfect specimens. There followed a period of relative decline, probably because many formerly rare shells suddenly became common as their haunts were discovered and exploited. The shells were then dumped on the market and the prices suffered a sudden decline. An example is *Voluta junonia* of Florida which fetched $50 or $60 for choice specimens when that state was still relatively unexplored and difficult of access. Then when Florida became one of the nation's winter playgrounds and *V. junonia* was collected in large quantities, the price fell rapidly until superb specimens brought no more than $5. This drop happened in hundreds of other cases too.

Nevertheless, shells are so strikingly interesting and beautiful in themselves that the hobby was rapidly revived. The U.S. troops stationed on lonely islands in the Pacific after the fighting

Shells of unusual shape [illegible] Spindle shell, *Tibia fusus* (Philippines); (2) *X*[illegible] *pallidula* (Japan); (3) *Murex tenuispina* (Ry[illegible]ands); (4) *Brechites radix* (Torres strait); (5) [illegible] *bednalli* (Australia); (6) *Trophon triangulatus* [illegible]rnia); (7) *Malleus albus* (Japan); (8) *Arca to*[illegible] (China); (9) Bubble shell, *Hydatina albocincta* [illegible] (10) *Latiaxis deburghiae* (Japan); (11) *Terebr*[illegible]*ata* (Japan); (12) *Murex pele* (Hawaii); (13) *An*[illegible]*pis hirasei* (Japan)

Florida shells: (1) Sunrise tellin, *Tellina radiata;* (2) Thorny oyster, *Spondylus americanus;* (3) Lace murex, *Murex florifer;* (4) common sea snail, *Janthina janthina;* (5) Long-spined star-shell, *Astraea longispina;* (6) Fargo's worm shell, *Vermicularia fargoi*



Left: Cross sections of the common giant conch, *Strombus gigas* (above) and the chambered nautilus, *Nautilus pompilius*

Below: The ear shell or abalone, *Haliotis,* found on the California coast. Exterior of shell is shown at top left; polished exterior after outer coating has been removed is shown at right. The inside of the shell, bottom left, shows the mother-of-pearl lining. Holes in the shell are used for breathing by the abalone

PHOTOGRAPHS, (TOP LEFT, BOTTOM LEFT, BOTTOM RIGHT) RUTHERFORD PLATT, (OTHERS) SHERMAN—PHOTO RESEARCHERS FROM PUBLIX

SHELLS AND SHELL COLLECTING

Snails: (1) *Umbonium giganteum* (Japan); (2) *Epitonium scalare* (China); (3) *Ancilla hilgendorfi* (China); (4) *Pyramidella acus*; (5) *Trochus maculatus* (Oceania); (6) *Cancellaria nodulifera* (Japan); (7) *Phasianella australis* (Australia); (8) Sundial shell, *Architectonica maxima* (China); (9) *Babylonia japonica* (Japan); (10) *Voluta fulgetrum* (Australia); (11) *Murex erythrostomus* (Gulf of California); (12) *Oliva irisans* (Japan)

Chitons, mollusks of the class *Amphineura* with shells [illegible]sting of eight transverse plates. The somewhat flat, elongate sh[illegible]nown in the specimens in the photograph is typical. Chitons are fo[illegible]n shallow littoral waters throughout the world

Volutes (genus *Voluta*): (1) *V. piperita* (Solomon Islands); (2) *V. pulchra* (Australia); (3) *V. grayi* (Australia); (4) *V. ponsonbyi* (South Africa); (5) *V. rossiniana* (New Caledonia); (6) *V. thatcheri* (New Caledonia); (7) *V. sowerbyi* (Australia); (8) *V. hamillei* (Japan); (9) *V. imperialis* (Sulu sea); (10) *V. keatsiana* (Australia); (11) *V. ellioti* (Australia); (12) *V. arausiaca* (Ceylon); (13) *V. costata* (Mauritius); (14) *V. delicata* (Japan); (15) *V. damoni* (Australia); (16) *V. hirasei* (Japan); (17) *V. daviesi* (Japan); (18) *V. prevostiana* (Japan)

Cones (genus *Conus*), a sample of the 400–500 known species. Common inhabitants of tropical coral reefs, many feed on living worms, and a few species capture living fish. They are among the shells most prized by collectors

PHOTOGRAPHS, RUTHERFORD PLATT

Rare and colourful [illegible] shells: (1) *Bathybembix argenteonitens* (Japan); (2) *Tric*[illegible]*ae* (East Indies); (3) *Turritella terebra* (Philippines); (4) [illegible] *pulcherrima* (Admiralty islands); (5) *Cantharidus opalu*[illegible] [illegible]ealand); (6) *Cypraea aurantium* (Fiji); (7) *Oliva tigrina* [illegible] Pacific); (8) *Chlamys senatorius nobilis* (Philippines); (9) [illegible]*heliotropium* (New Zealand). *P. pulcherrima* is a land snail; [illegible]e marine shells

Pacific marine shells: (1) *Calliostoma cunninghami;* (2) *Cymatium rubecula;* (3) *Turbo undulatus;* (4) *Neritina communis;* (5) Elephant's tooth, *Dentalium elephantinum;* (6) *Conus lividus;* (7) *Strombus auris-dianae aratrum;* (8) *Lambis digitata;* (9) *Turbo petholatus;* (10) Golden cowry, *Cypraea aurantium;* (11) Triton's trumpet, *Charonia tritonis;* (12) *Purpura persica;* (13) *Oliva sericea miniacea;* (14) *Trochus acutangulus*

Fine specimen shells from the F. A. Constable collection: (1) *Conus victoriae* (Australia); (2) *Guildfordia triumphans* (Hong Kong); (3) *Conus adamsoni* (Australia); (4) *Conus geographicus mappa* (Pacific islands); (5) *Bursa pustulosa* (Pacific ocean); (6) *Voluta lyraeformis* (east Africa); (7) *Mactra violacea* (China); (8) *Corculum cardissa* (Philippines); (9) *Fusinus ustulatus* (West Indies); (10) *Cypraecassis testiculus* (West Indies); (11) *Voluta grayi* (Tasmania)

Collection from New Caledonia, except where noted: (1) Chambered nautilus, *Nautilus scrobiculatus* (Indo-Pacific); (2) *Trochus niloticus;* (3) Scorpion shell, *Lambis rugosa;* (4) *Amusium pleuronectes;* (5) *Mitra episcopalis;* (6) *Turbo petholatus;* (7) *Cellana testudinaria;* (8) *Cypraea talpa;* (9) *Cypraea mappa;* (10) *Cypraea mauritiana* (Moluccas); (11) *Charonia tritonis*

PHOTOGRAPHS, RUTHERFORD PLATT

SHELLS AND SHELL COLLECTING

Philippine forest snails. Many of the more than 1,000 varieties are found on only a single mountain or on a single small island. Occasionally the soft parts of some species are eaten

Cuban tree snails belonging to the genus *Liguus*. *L. p*[illegible]*anus*, which is shown in two views at top centre, is a rarity in that som[illegible]lls coil to the right and others to the left

Rare Florida marine shells: (1) *Chlamys mildredae*; (2) *Chlamys imbricatus*; (3) *Murex cabritii*; (4) *Tellina magna*; (5) *Terebra taurina*; (6) Golden panama, *Oliva sayana*; (7) Lion's paw, *Lyropecten nodosus*; (8) *Conus regius*; (9) *Conus sozoni*; (10) Carrier shell, *Xenophora conchyliophora*; (11) *Scaphella junonia*

New England marine mollusks: (1) *Thais lapillus*; (2) *Littorina obtusata*; (3) periwinkles, *Littorina littorea*; (4) Jingle shell, *Anomia simplex*; (5) Moon shell, *Lunatia heros*; (6) limpets, *Acmaea testudinalis*; (7) Blue mussels, *Mytilus edulis*; (8) *Modiolus modiolus*; (9) Quahog clam, *Mercenaria mercenaria*

PHOTOGRAPHS, RUTHERFORD PLATT

had moved on learned to gather shells to while away the time. Stay-at-homes became interested in shells and soon the hobby began to rival stamp and coin collecting in popularity. Shells, because of their beauty and exotic points of origin, will long cause large numbers of persons to be interested in them.

The shell itself is the exoskeleton (external skeleton) of a soft bodied animal that secretes it to serve as protection as well as support for its various organs. The animal is bound securely to the shell and can no more survive separation from it than we can from our bony framework. After a mollusk has died other marine animals may become occasional inhabitants of the shells. Such an animal is the hermit crab, which makes use of the shell as a temporary abode. Most marine mollusks are able to shut themselves off completely from their watery world when they find themselves in danger, by closing their aperture by a horny or calcareous door, or operculum. The careful collector makes it a point to preserve the door with the shell in his collection.

The rarest, the most valuable, and in many senses the most beautiful shells are those that are found in the sea, but large numbers of species live only on land or in fresh water. The aristocrats of the shell world are the various members of the families Cypraeidae (cowrie shells), Conidae (cone shells), Volutidae (volute shells), and Muricidae (rock shells). One of the rarest and most desirable single species is *Conus gloria maris*, the glory of the sea. This is a handsome cone shell, of which only about 30 exceptionally beautiful specimens have ever been known. They always command a very good price on the market—$1,250 in 1957. It is also the only shell known to have been stolen from a large museum.

The question might be asked: why are some shells rare, others relatively common, and still others exceedingly abundant? The answer is intimately bound up with the habits of mollusks. Many species can tolerate only the conditions of life that are found in the intertidal zone on sandy or rocky shores. Such shells are most easily collected, since one need only wait for a low tide to begin gathering them. In the intertidal zone some very handsome shells are found, *e.g.*, the glistening, richly coloured Olividae (olive shells), the Naticidae (moon shells), and Strombidae (strombs) as well as Terebridae (augur shells) and many bivalves (clams) that burrow in water-soaked sand. On the rocks and under them are found Littorinidae (periwinkles), Turbo (turban shells) and limpets, as well as the Chitonidae (pill bug or coat of mail shells) that have the curious habit of curling up like the land crustacean (pill bug) for which they are popularly named.

Other shells live only in the shallow water below the extreme low tide line. These shells are washed up on the beach after storms. Since luck and some degree of effort—dredging or straining—is involved in capturing them, they tend to be scarcer and more costly than the intertidal species.

The really rare shells are species that can live only in deep water. There live the rarest and costliest volutes, cones, cowries and rock shells. They can be obtained only by expensive and difficult deep-sea dredging operations from specially designed ships. Such shells would be quite unavailable to the average collector if he had to depend only upon these sources for his specimens. Luckily, however, they are frequently obtained as a side line in shrimp or fish dredging. Many fishermen, especially in Japan, have discovered that selling shells can supplement their regular incomes. They keep the shells on board instead of discarding them with the muck that their dredges bring up and sell them to dealers or collectors. Some of our greatest rarities, such as *Busycon coarctatum* (the turnip shell) are obtained in this manner.

Another source of deep-water shells are fish stomachs. Many ground-feeding fish species such as haddock and cod are frequently caught with their stomachs full of nicely cleaned and faultless shell specimens. A few rare shells such as the costly *Cypraea fultoni* (Fulton's cowrie shell) are known only from this sort of "habitat," *ex pisce* as the scientist puts it. SCUBA and skin divers bring up many rare specimens from depths up to 30 ft. Many shells that hide regularly in coral crevices can be collected in no other manner.

Although most mollusks live in the sea, many colourful species are air-breathing land dwellers. These shells are found everywhere that their few simple requirements of shade, moisture and food (usually plant or fungi) are satisfied. The largest majority are tiny (some only a few millimetres in size) and quite plain looking. They excite the interest of only the dedicated scientist or the enthusiastic amateur collector who makes a specialty of the shells of special regions. Some of these shells, however, are brightly coloured, large and quite showy. Best known are the *Liguus* (tree snail) of southern Florida and Cuba, *Papuina* of Papua, *Helicostyla* of the Philippines and the very large *Achatina* of Africa. These shells are not so popular among most enthusiasts as are marine shells because few collectors are willing to undergo the discomforts and dangers of collecting them. The tree snail of Florida shares its habitat with diamondback rattlers and cottonmouths, as well as voracious mosquitoes; *Papuina* and *Helicostyla* are found on trees in hot malarial regions haunted by savage headhunters; the ground-dwelling *Achatina* of Africa frequently live in the very heart of the continent.

The shells that live in fresh water are even less popular with collectors, chiefly because none of them are colourful, and only very few have any claim to beauty. The river mussels of the midwestern United States have a beautiful internal pearllike covering, or nacre, and are sometimes handsomely ornamented with tubercles and wavy ridges. They are often sold in bulk to button manufacturers. Few if any command a respectable price per specimen. Their slight claims to beauty are further impaired by the erosion of the shell caused by acids in solution. Fresh-water snail shells are of considerable interest to the student of tropical diseases, since some snails serve as unwilling hosts of severe human and animal diseases.

One of the pleasantest aspects of shell collecting is that the shells need little service to be stored. They do not require the careful attention and handling necessary in a butterfly, beetle or herbarium collection. They are not as heavy as minerals. They are not subject to insect or fungus attack. Once the shell is reasonably well cleaned and provided with a complete label, it can remain unaltered for years in the drawer or cabinet to which it has been consigned.

Most collectors tend to specialize, since they realize very soon that their chance of obtaining even a fair proportion of the about 100,000 shell species known is very slight. Most collectors confine their interest to the aristocrats mentioned earlier. Others limit themselves to localities, collecting only the shells of Florida or California or even more immediate locales. Such specialists frequently contribute valuable scientific data in the form of exact locality records and observations on life habits of many mollusks. Another type of collector gathers only specimens of the type genus of shells, that is the single species of a shell that characterizes a particular genus. Such a collection is a valuable lesson in taxonomy and evolution and gives a perspective view of the entire field of conchology, the study of shells.

There are also specialists in land shells. Land shells have the virtue of being less bulky than marine shells and lend themselves to easier storing. They are also less expensive than marine shells, and complete family or generic collections can be made at low cost. The *Liguus* enthusiasts are a class apart, and in their zeal for completeness of their collections are sometimes ready to pay respectable prices for particularly desired specimens.

Shell collecting also can form the base for many interesting studies. All stamps and coins of the same type are machine made and therefore identical; however, like all living organisms, no two shells are exactly alike. Many an amateur has become interested in the variations of a single species and has often accumulated so much data that he could prove that what were considered two distinct and valid species are actually only the extreme variations of a single species. This aspect adds an element of excitement to shell collecting that is not found in many other collecting hobbies. In fact, many of the best-known scientists in the field of conchology began as amateur shell collectors.

A word might be said about the equipment of the shell collector. All that most collectors use are a basket or a small quantity of bags and small boxes; some sort of tool to scrape away

the sand or leaves; a knife to detach rock clingers from their perches; a hammer and chisel to remove rock borers from their nests; a sieve to strain out specimens from shallow water; a notebook to record locality and ecological data and a stout back for much stooping. In addition there should also be a library of books on shells, limited only by financial means and availability of the classics of the shell world. These books, together with recourse to museum collections, are the only sources of the proper Latin names for shell species.

See also references under "Shells and Shell Collecting" in the Index.

BIBLIOGRAPHY.—R. Tucker Abbott, *American Seashells* (1954); American Malacological Union, *How to Collect Shells* (a symposium) (1955); A. H. Cooke, *Cambridge Natural History*, vol. iii (Molluscs) (1895); Kathleen Yerger Johnstone, *Sea Treasure, A Guide to Shell Collecting* (1957); Rutherford Platt, "Shells Take You Over World Horizons," *Nat. Geogr. Mag.*, 96:33–84 (July, 1949); Johannes Thiele, *Handbuch der systematischen Weichtierkunde*, vol. 1–2 (1931–35); George W. Tryon and Henry A. Pilsbry, *Manual of Conchology*, series 1, vol. 1–17, series 2, vol. 1–27 (1879–1927); Walter Freeman Webb, *Handbook for Shell Collectors*, 4th ed. (1936). (M. K. J.)

SHELL SHOCK, a nonspecific term applied to World War I combatants suffering from the condition that was generally called combat fatigue in World War II. *See* COMBAT FATIGUE.

SHELTON, THOMAS (fl. 1612–1620), English translator of *Don Quixote,* seems to have been employed by Theophilus Howard, 2nd Baron Howard de Walden and afterward earl of Suffolk, to whom he dedicated *The History of the Valorous and Wittie Knight-Errant Don Quixote of the Mancha* (1612). In his dedication he states that he did the work "in the space of fourty dayes" at the request of "a very deer friend." Shelton did not use the original edition of Cervantes (1605) but that published at Brussels in 1607. On the appearance of the Brussels reprint of the second part of *Don Quixote* in 1616 he translated that into English also, and published it in 1620 with a revised version of the first part.

Shelton's version is the happiest English translation of the great Spanish classic, combining robust vigour with grace and ease. It is possible that the translator was the Thomas Shelton who wrote the sonnet prefixed to *A Restitution of Decayed Intelligence* by Richard Verstegan (1605).

Shelton's *Don Quixote* was edited by J. F. Fitzmaurice-Kelly, four volumes (1896), A. W. Pollard (1900), and F. J. H. Darton (1923).

See also Edwin B. Knowles, *Four Articles on Don Quixote in England* (1941). (V. DE S. P.)

SHELTON: *see* ANSONIA.

SHEM was the eldest of Noah's three sons (Gen. v, 32; x, 1). In the Old Testament genealogical tables tracing the origins and connections of the various peoples, the Hebrews, Aramaeans and Arabs are counted among the descendants of Shem, and from the name is derived the adjective Semitic. Certain other peoples (as the Elamites) are referred to also as his descendants, on a geographical or political rather than linguistic basis. Roughly speaking, the nations supposed to share Shem as their ancestor lie geographically between the descendants of his brothers Japheth (*q.v.*), on the north, and those of Ham (*q.v.*) on the south.

SHEMA (Heb. "hear"), the initial word of the scriptural verse "Hear, O Israel: the Lord our God is one Lord" (or ". . . the Lord our God, the Lord is One"; Deut. 6, 4), the watchword of Jewish faith. It is chanted at various points of the Jewish liturgy and recited mornings and evenings as a part of a text that includes Deut. 6:4–9 (defined as "the acceptance of the yoke of the kingdom of God"), 11:13–21 ("the acceptance of the yoke of the commandments"), and Num. 15:37–41 ("the remembrance of the redemption from Egypt"). It is inscribed on certain ritual objects. Following the example of the Palestinian sage and martyr Rabbi Akiba (2nd century A.D.), the Shema is uttered by the Jewish martyrs through the ages as the final profession of faith in the one God of mankind and of love for him. This theme was variously elaborated by medieval liturgic poets. *See* further LITURGY, JEWISH.

See I. Elbogen, *Der jüdische Gottesdienst in seiner geschichtlichen Entwicklung* (1931). (N. N. G.)

SHENANDOAH NATIONAL PARK, in Virginia, U.S., was established in 1935 to preserve 193,646 ac. (78,369 ha.) in the Blue Ridge Mountains. The most easterly of the Appalachian ridges, the Blue Ridge is geologically ancient, and has been eroded to its present elevation through countless thousands of years. (*See* APPALACHIAN MOUNTAINS.) In general, mountain profiles are rounded, but rocky outcrops and precipitous ledges appear in a few places, such as on the highest point, Hawksbill Mountain, 4,049 ft. (1,234 m.) above sea level, on Old Rag Mountain and at Franklin Cliffs.

The park is noted for its scenery, which affords some of the widest views in the eastern states. Skyline Drive runs the length of the park, winding among the summits, dipping through highland valleys and gaps for about 105 mi. (170 km.), from the town of Front Royal on the north to Jarman Gap on the south. Except for a few grassy meadows along the crest, the park is heavily forested with hardwood trees, including black gum, yellow, gray and black birches, linden, tulip tree, and several species of oak. Among the conifers are eastern hemlock, Virginia, white, Table mountain, and pitch pines, red spruce, and, on the highest peaks, Fraser fir. Wild flowers typical of eastern woodlands are abundant. In spring, from April to June, there is a succession of blooms—redbud, flowering dogwood, azalea, and mountain laurel—and in autumn a display of brilliant foliage. White-tailed deer, bobcat, red and gray fox, and gray and flying squirrels inhabit the park, and the greatest number of bird species occurs in spring, when such migrants as rose-breasted grosbeak, brown thrasher, towhee, scarlet tanager, and numerous warblers arrive to make their summer homes.

This park is joined to Great Smoky Mountains National Park (*q.v.*) by the 470-mi. (756 km.) Blue Ridge Parkway. (Dx. B.)

SHENANDOAH VALLEY, in Virginia, U.S., extends southwest from the vicinity of Harpers Ferry on the Potomac River and lies between the Blue Ridge Mountains and the Alleghenies. It is a part of the Great Valley of Virginia. In a narrow sense the Shenandoah Valley embraces only nine counties drained by the Shenandoah River: Berkeley and Jefferson in West Virginia, and Frederick, Clarke, Shenandoah, Warren, Rockingham, Page, and Augusta in Virginia.

In popular usage, however, the Shenandoah Valley extends to the James River and includes Rockbridge County: it is approximately 95 mi. (153 km.) long and averages 25 mi. (40 km.) in width. From the floor of the valley near Harrisonburg rises Massanutten Mountain, which stretches to the northwest for 50 mi.; it divides the two forks of the Shenandoah River. Among the historic passes through the Blue Ridge are Swift Run Gap and Rockfish Gap.

The valley was once inhabited by Indians of the Tuscarora and Shawnee tribes whose Indian road ran the length of the valley. The road became the famous Valley Pike, a main artery in the westward movement, and is now a U.S. highway. In 1707 Louis Michelle explored the lower valley, and in 1716 Gov. Alexander Spotswood led an expedition over the Blue Ridge to the Shenandoah River. Settlement of the valley by white men began about 1730; at this time buffalo and elk abounded. Most of the early settlers, German Lutherans and Scotch-Irish Presbyterians, came from Pennsylvania, though some English came across the Blue Ridge. The various national groups usually settled in separate communities. In antebellum days the Shenandoah Valley was a region of family-operated farms and few slaves. During the Civil War "Stonewall" Jackson won immortality by his "Valley campaign" (*see* JACKSON, THOMAS JONATHAN). Near the end of the conflict the valley was devastated by Federal forces.

The backbone of the valley's economy is agriculture, with emphasis upon purebred cattle, sheep, poultry, and dairy farming. Harrisonburg is a turkey capital and Winchester is noted for its apple orchards. Industry has enjoyed a remarkable growth with carpets, chemicals, cutting instruments, electric appliances, and rubber goods being particularly important. The native limestone is suitable for building. The labour force tends to commute from rural areas instead of moving to mill towns.

Colleges in the valley include Madison in Harrisonburg, Mary

Baldwin in Staunton, and Virginia Military Institute and Washington and Lee University in Lexington. There are numerous military preparatory schools in the area and many summer camps.

Large numbers of tourists are attracted by the Skyline Drive and Blue Ridge Parkway (both on the crest of the Blue Ridge), the Shenandoah National Park (*q.v.*), the Natural Bridge, near Lexington, and many limestone caverns. (G. M. BE.)

SHENANDOAH VALLEY CAMPAIGNS. In the American Civil War (*q.v.*) the Shenandoah valley was often the scene of military operations, some of them of considerable importance. This valley lies between the Blue Ridge and the Allegheny mountains. The river from which it takes its name rises below Lexington, Va., and flows 155 mi. to Harpers Ferry where it enters the Potomac river. It is part of the great valley that extends on up into Maryland and Pennsylvania providing a well-sheltered route for invading the North. On the eastern side of the valley a series of gaps opens the way through the Blue Ridge mountains to the Piedmont plateau. The low Massanutten mountain rises abruptly from the floor of the valley for about 50 mi. between Strasburg and Harrisonburg, affording excellent opportunities for military maneuvering. The inhabitants of the valley were mostly southern in their loyalties, making it an important recruiting ground as well as a rich source of cattle and grain.

The valley was more useful to the Confederates than to the Federals. Movement up and down, and in and out on the eastern side, was made easy by a network of good roads connecting all the towns and running through the several gaps into eastern Virginia. In the upper part of the valley two railroads served Richmond and at the northern end the Baltimore and Ohio railway, connecting Washington, D.C., with the west, ran along the line of the Potomac river. Whenever a Confederate army moved north it drew nearer to Washington, but when a Union army moved up the valley it was pulled away from Richmond. When a southern army crossed the Potomac it cut across the Baltimore and Ohio railroad and was 60 mi. N. of Washington. Hence the presence of a Confederate army in the northern part of the Shenandoah valley was often considered a sufficient menace to justify calling back troops from campaigns elsewhere to ensure the security of the capital. The southerners used the advantages offered by the valley so effectively that it often became for the North the "valley of humiliation" until late in the war when the Union forces took undisputed control.

The initial campaign of the war in the east turned on the use of forces posted in the lower valley. In July 1861, when the Federal invasion of Virginia began, this Shenandoah army united with the forces of *Gen. P. G. T. Beauregard* on the field of Bull Run (*q.v.*) in time to meet the attacks of Gen. Irvin McDowell and turn the tide in favour of the southerners. (To aid the reader the names of Confederates are put in italics.)

In the spring of 1862 when Gen. George B. McClellan was advancing up the Yorktown peninsula toward Richmond, instead of calling *Gen. T. J. "Stonewall" Jackson* from the Shenandoah to the defense of Richmond, *Lee* sent *Gen. Richard S. Ewell* with reinforcements so he could threaten Washington. Operations in the valley opened with *Jackson* retiring from Winchester on March 12, and Gen. N. P. Banks with 20,000 Union troops taking possession. Banks then sent Gen. James Shields on to Strasburg and a week later *Jackson* withdrew his small force to Mt. Jackson. When Banks recalled Shields, *Jackson* followed him, but Shields turned on him at Kernstown and defeated him with a force that outnumbered his two to one. Banks then decided to drive *Jackson* from the valley, but *Jackson* took refuge in the passes of the Blue Ridge where he could await reinforcements and a favourable opportunity to fall upon the Federals. In May, leaving *Ewell* to watch Banks, *Jackson* proceeded secretly by rail through Rockfish Gap with the remainder of his command to join *Gen. Edward Johnson's* brigade, then beset by Gen. Robert H. Milroy west of Staunton, and together they claimed their first victory on May 8. Meanwhile Banks's army near Strasburg had largely been sent elsewhere, and *Jackson* moved quickly through Luray valley on the east side of the Massanutten mountain. He surprised part of Banks's army at Front Royal on May 23 and followed the retreating Banks through Winchester, driving him across the Potomac. At this turn of events, Washington authorities rescinded the order for McDowell's corps at Fredericksburg to join McClellan. McDowell was ordered to send Shields to converge with Gen. John C. Frémont's troops in the southwestern part of the valley and intercept *Jackson's* army as it returned. *Jackson* had luck, a good road and *Ewell's* division to slow up Frémont. The result was that the Confederates were able to defeat Frémont on June 8, at Cross Keys, and were also able to cross the valley to intercept and defeat Shields at Port Republic the next day. *Jackson* then went on to aid *Lee* and the main army before Richmond, and Lincoln recalled the divisions of Frémont, Banks and McDowell, leaving the valley at peace for a time.

When *Lee* entered Maryland in the campaign that was to reach its climax at Antietam (*q.v.*) in Sept. 1862, he expected the Federals to leave Harpers Ferry to defend Washington. When they failed to do so, he planned while moving to Hagerstown to take Harpers Ferry and reopen communications with the Shenandoah valley. *Jackson* marched with three divisions to invest the town from the rear, another division took up a position on Maryland Heights overlooking Harpers Ferry from north of the Potomac, while a fifth climbed Loudon Heights across the Shenandoah from the city. Surrounded and subject to a triple fire the Federals surrendered and *Jackson* rejoined *Lee*.

The concentration of Federal forces in northern Virginia in May of 1864 in preparation for Gen. U. S. Grant's advance upon Richmond from the north involved a fresh series of operations in the

BY COURTESY OF THE LIBRARY OF CONGRESS

BATTLE OF WINCHESTER, VA., MARCH 23, 1862; A PENCIL DRAWING BY A. R. WARD WHICH APPEARED IN "HARPERS WEEKLY," APRIL 12, 1862

valley. The Confederates defeated a containing force there under Gen. Franz Sigel. Sigel was replaced by Gen. David Hunter who was to co-operate with Grant's army of the Potomac. Grant sent Gen. Philip H. Sheridan with the cavalry to cut the valley's communications with the Confederate capital, but *Gen. J. E. B. Stuart* stopped him at Yellow Tavern, losing his life in the encounter. *Lee* then sent *Gen. Jubal A. Early* to the valley once again to remove the pressure on Richmond by threatening Washington. *Early* was as famous in his way as *Jackson*, the other great Confederate leader in the valley campaigns. He was an aggressive fighter, an exponent of the quick march and the audacious action later to be associated with the blitzkrieg. Grant detached Sheridan to join Hunter at Charlottesville, but *Lee* sent *Gen. Wade Hampton's* cavalry to intercept him and compel him to return and leave Hunter to his fate. To divert northern troops from *Lee's* front, *Early* drove Hunter out of Virginia by the Kanawha river route and moved down the valley unopposed. He crossed the Potomac, reached Hagerstown and Frederick in Maryland and on July 11 led 8,000 men past Silver Spring, Md., and into sight of the U.S. capital. Though he frightened official Washington he made no attempt to take the city.

This raid added insult to injury. Grant therefore sent Sheridan to clear the valley of Confederates by intercepting *Early* on his return from Maryland, or, failing that, to drive him up the valley and to destroy all the supplies he could not use himself. Sheridan made Harpers Ferry his headquarters and on Aug. 10 moved out in strength, only to find that *Early* was also reinforced. In a series of encounters beginning on Sept. 19 near Winchester, and shortly after at Fisher's Hill, he drove *Early* back to Mt. Jackson and finally defeated him, but not without one more reminder of the "valley of humiliation." After the victory at Fisher's Hill, Sheridan retired to Winchester apparently feeling that *Early* would trouble him no more. But *Early* followed him to prevent his sending aid to Grant at Petersburg. Two of the three corps of Sheridan's army were routed in a well-planned and well-executed attack at Cedar creek but *Early* failed to follow up and complete his victory. Sheridan was absent from the first part of the fighting but returned to the scene in time to rally his forces, attack and drive back the Confederates, recovering all the ground lost and recapturing his abandoned guns and baggage. *Early* then retreated up the valley to Waynesboro where, on March 2, 1865, he suffered the defeat that ended Confederate resistance in the valley. Sheridan meanwhile had destroyed supplies and communications in the valley in a manner reminiscent of Gen. W. T. Sherman's march through Georgia and the Carolinas.

See J. Davis, *The Shenandoah* (1945); D. S. Freeman, *Lee's Lieutenants*, 3 vol. (1942–44). (C. W. Te.)

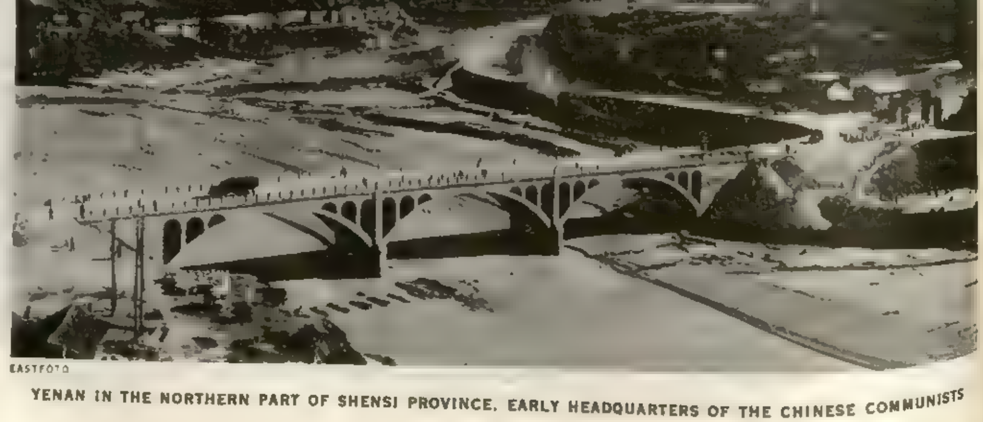
SHENSI (Shen-hsi Sheng) is a northwest province in the loesslands of China between Kansu and Shansi provinces. It was one of the earliest seats of Chinese civilization, the site of a historic capital centred in the vicinity of Sian (*q.v.*), the present provincial seat. Area 75,598 sq.mi. (195,800 sq.km.); pop. (1957 est.) 18,130,000.

Topographic barriers and the fertile productivity of the well-watered Wei Ho (river) Plain supported its command over strategic gateways to central Asia in the northwest, Inner Mongolia in the north, Szechwan across the Tsinling Shan in the south, and the Yellow River Plain in the east. Shaped in a slender wedge 300 mi. (480 km.) long east and west, the Wei Ho Plain begins with a width of about 1 mi. at Pao-chi in the west and widens near Sian to 30–40 mi., continuing thus to the fortress city of T'ung-kuan at the Yellow River elbow. Northward the land rises through a region of deeply gullied badlands to plateau remnants 3,000–6,500 ft. (900–2,000 m.) high near the Ordos Desert and eastern Kansu. Halfway between the Wei Ho and the Ordos Desert lies the town of Yenan, the Chinese Communist capital from 1935 to 1949. In the east, the land overlooks the deep gorge of the Yellow River which drops turbulently there and is only about 330 ft. (100 m.) wide through much of its course until Lung-men is reached. From there to T'ung-kuan the river broadens to 1–2 mi., then turns and leaves the province to flow eastward. Southward of the Wei Ho, the northern front of the Tsinling Shan rises abruptly from the plain, mantled with the same yellow loess on its northern slopes as that covering northern Shensi. Southern Shensi comprises the south slopes of the Tsinling Shan and the north slopes of the Ta-pa Shan and represents an entirely different landscape. Between them in the Han Shui Valley lies the Han-chung Basin, about 50 mi. long by 6 mi. wide.

North and south Shensi have greatly differing climates. The Tsinling Shan blocks the cold winter winds from moving south, while it greatly reduces the moisture in summer winds moving north. Thus in winter the Han-chung Basin is mild and damp, while northern Shensi is cold and dry. Precipitation reaches a yearly average of 30 in. (760 mm.) south of the Tsinling Shan where the annual average temperature is 15° C (60° F), similar to that of Szechwan. By contrast, the Wei Ho plain gets under 20 in. (510 mm.), while in the north this is reduced to as little as 4 in. The January average at Sian falls to 0° C (31° F), while at Yu-lin in the extreme north it falls to −7° C (17° F). July averages at these two cities are 27° C (82° F) and 23° C (75° F) respectively. The autumn rainfall concentration is combined with great fluctuations in annual amounts and seasonal occurrences, presenting great farming hazards. Canal and well irrigation is practised where possible.

The Wei Ho Plain, the economic heart of the province, is especially noted for winter wheat and summer cotton production. In the north, spring wheat and millet are significant crops. Beans, kaoliang, corn, and tobacco also are grown. In the Han-chung Basin, paddy rice in summer is followed by winter wheat. Fruits grown include pears, apples, apricots, and grapes. Farm sidelines include the raising of sheep, horses, donkeys, mules, goats, and pigs. Forest remnants exist in the heights of the Tsinling Shan, but north Shensi is rather barren and brown. Sand encroachment from the Ordos Desert in the north is countered with planted shelter belts.

Shensi has important coal deposits which are largely located in the drainages of the Ching and Lo rivers. Petroleum is being



EASTFOTO

YENAN IN THE NORTHERN PART OF SHENSI PROVINCE, EARLY HEADQUARTERS OF THE CHINESE COMMUNISTS

exploited in the Yen-ch'uan and Yen-ch'ang districts. Oil shale occurs in the Han-chung Basin. Alluvial gold is panned in the streambeds of the upper Chialing Chiang and the south-slope rivers of the Tsinling Shan. Salt is obtained from lakes in the Wei Ho Plain. Manganese and graphite also are mined in the province.

Cotton manufacturing is an important industry at the chief cities of Sian, Hsien-yang, Pao-chi, and Han-chung. Sian is the major industrial, cultural, and political centre. The Lunghai Railroad from Honan follows the Wei Ho Plain westward into Kansu. From Pao-chi on this line, a railroad crosses the Tsinling Shan to Szechwan. Highways radiate from Sian to Yenan and Yu-lin in the north, Kansu in the west, and Honan in the southeast. A highway also crosses the Tsinling Shan from Pao-chi.

Much industrial progress has taken place since the Communists consolidated their power and took control. In the 1960s output of electric power, coal, and crude oil was reported to be much higher than before 1949. Many other new industries were introduced, including textile plants, flour mills, and brick factories.

(H. J. Ws.)

SHENSTONE, WILLIAM (1714–1763), English poet, essayist, landscape gardener, and arbiter of taste, who was a key figure in the movement away from Augustan poetry and aesthetics (*see* ENGLISH LITERATURE: *The 18th Century*). Born on Nov. 18, 1714, at the Leasowes, Halesowen, Shropshire, he was educated at the village school, whose "dame," Sarah Lloyd, he celebrated in "The School-Mistress," and at Solihull. In 1732 he entered Pembroke College, Oxford. In 1735 he inherited the Leasowes, and from 1745 devoted his income of £300 a year to making it what he described as a "*ferme ornée*" (*i.e.*, an ornamented, but still useful, farm), blending classical elegance with early Romantic sensibility. His ideas on landscape gardening (a term he was the first to use) are explained in "Unconnected Thoughts on Gardening": "landscape should contain variety enough to form a picture upon canvas . . . Objects should . . . be less calculated to strike the eye than the judgment or . . . imagination."

BY COURTESY OF THE NATIONAL PORTRAIT GALLERY, LONDON

WILLIAM SHENSTONE, DETAIL OF A PAINTING BY E. ALCOCK

"The School-Mistress," Shenstone's most famous poem, one of the earliest 18th-century imitations of Spenser, uses the Spenserian stanza with delicate humour. Included in his *Poems on Various Occasions* . . . (1737), it was expanded in 1742 and 1748, when it appeared in Robert Dodsley's *Collection* . . . (vol. i, 2nd ed.). Of the later volumes Shenstone became, in effect, editor. "A Pastoral Ballad," with its fresh, lyrical measure, appeared in vol. iv (1749), and many other of his verses in vol. v (1758). He was frequently consulted, and his influence is revealed in his correspondence—notably that with Bishop Thomas Percy (*q.v.*), proving that Shenstone's collaboration in *Reliques of Ancient English Poetry* (1765) amounted to co-editorship. He helped to revive the ballad (*q.v.*), and established easy lyrical movement and rural simplicity as criteria for poetry, foreshadowing many of the preoccupations of Romanticism (*q.v.*). He edited a poetic miscellany to illustrate his theories, but died (Feb. 11, 1763, at the Leasowes) before he had submitted it to his friend John Baskerville for printing. It remained in manuscript until 1952. His *Essays on Men and Manners* (1764) is notable for its development of the maxim, or *pensée*, and, as Dodsley said, shows "the acuteness of his understanding and his profound knowledge of the human heart."

BIBLIOGRAPHY.—*The Works in Verse and Prose* . . . , ed. by R. Dodsley, 2 vol. (1764), vol. iii (1769) adds *Letters to Particular Friends. The Letters*, ed. by M. Williams, and by D. Mallam (both 1939); *Shenstone's Miscellany, 1759–1763*, ed. by I. A. Gordon (1952); *Men and Manners*, ed. by H. Ellis (1927). Johnson's *Life of Shenstone* (1781) fairly assesses his poetry, but ridicules him as a person; R. Graves, *Recollections of . . . . Shenstone* (1788), presents a fairer assessment. *See* also M. Williams, *William Shenstone and his Friends* (1933), *William Shenstone* (1935); A. R. Humphreys, *William Shenstone* (1937).

(I. A. G.)

SHEN-YANG: *see* MUKDEN.

SHEOL, in the Old Testament, one of several names used to designate the abode of the dead; it is the commonest of these names, being used 66 times. The exact meaning of the term is not known, though it appears to have some connection with necromancy; that is, revealing the future through communication with the dead. *See* also HELL: *Judaism*.

SHEPHERD, JOHN (*c.* 1520–?), English composer of sacred music. In 1542 he succeeded Thomas Appleby as instructor of the choristers at Magdalen College, Oxford. He appears to have resigned in 1543, to be replaced by Thomas Preston, but in 1545 he was reappointed, this time remaining, apparently, for two years. In 1552 he was made a Gentleman of the Chapel Royal and continued so under Queen Mary, being described in several manuscripts as "of the queen's chapel." In 1554, however, he was again in Oxford, supplicating for the degree of doctor of music and described as a student of music for 20 years. In June 1555 he was in trouble with the authorities of Magdalen College over two incidents that reflect unpleasantly on his character; in 1557 he was once more at the Chapel Royal. The date of his death is unknown. A substantial quantity of Shepherd's music for both the Latin and the English rites remains. It embraces a variety of liturgical forms including five complete Masses, antiphons, responds, hymns, psalm-motets, troped lessons, processionals, English services, anthems, and keyboard pieces. The quality of his work suggests a reason for the tolerance with which his employers seem to have regarded his restless instability. One of the manuscripts containing his setting of *Haec dies* bears the annotation "a good songe, an excellent good songe" and "the best songe in England."

(Jo. D. B.)

SHEPPARD, JOHN (JACK) (1702–1724), English criminal, became famous for his daring escapes from Newgate Prison. Many tracts and plays were written about him, and *A Narrative of all the Robberies, Escapes . . . of John Sheppard* (1724) is attributed to Daniel Defoe. Sheppard was made the hero of much romance, notably in Harrison Ainsworth's novel, *Jack Sheppard* (1839).

He had a short career. He was born at Stepney, near London, on March 4, 1702, the son and grandson of carpenters, was brought up in the Bishopsgate Workhouse, and in April 1717 was apprenticed as a carpenter by one of his father's employers. At a neighbouring Drury Lane tavern, in 1723, young Sheppard met Elizabeth Lyon, known as "Edgeworth Bess," a woman with whom he lived. To gratify her tastes, he committed many of his crimes. At the end of 1723 he was arrested as a runaway apprentice, and thenceforward he said "I fell to robbing almost everyone that stood in my way." Joseph Blake, known as "Blueskin," was a frequent accomplice.

In the first six months of 1724 he twice escaped from prison, and toward the end of that period he was responsible for an almost daily robbery in and around London. Eventually, however, his independent activity provoked the bitter enmity of Jonathan Wild, controller of a large criminal organization, who procured his arrest on July 23.

Sheppard was tried at Old Bailey on Aug. 13 and condemned to death. Largely thanks to Edgeworth Bess he managed to escape from the condemned cell on Aug. 31, and was soon back in his old haunts. Rearrested on Sept. 10, he was imprisoned in the strongest part of Newgate, where he was handcuffed and chained to the floor. By a combination of skill and strength he freed himself and, after climbing through the chimney to the room above, forced five doors and lowered himself from the prison roof into an adjoining house. After a few days concealment he reappeared in the Drury Lane quarter, where he was recaptured, hopelessly drunk, in a

Clare Market tavern on Oct. 31. He was again taken to Newgate, and this time his cell was watched night and day. He was hanged at Tyburn on Nov. 16, 1724, before a crowd estimated at 200,000.

See H. W. Bleackley, *Jack Sheppard*, ed. by S. M. Ellis (1933); C. Hibbert, *The Road to Tyburn* (1957).

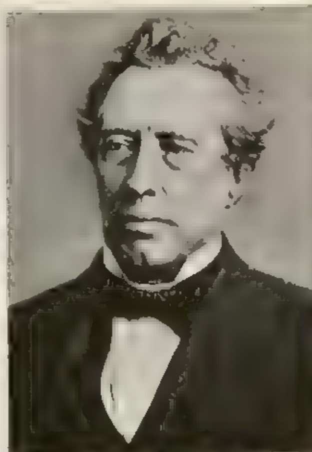
SHEPPEY, an island off the Kentish coast of England, the largest of several low islands separated from the mainland by the ramifying creeks about the mouth of the Medway. The Kingsferry road and rail bridge over the Swale is the only direct link with the mainland. Pop. (1961) 27,211. Area 35 sq.mi. (91 sq.km.). Sheppey is 10½ mi. in extreme length from east to west, and 5 mi. wide at its broadest point.

It is low lying with a small elevation near the north coast which presents slight cliffs toward the shallow sea. There are occasional incursions from the sea but protective measures have been taken to prevent further erosion and flooding. Except for the higher ground, the island is treeless but very fertile, producing grain and vegetables; its name, meaning the "island of sheep," is still appropriate. Leysdown and Minster are rapidly developing as holiday camping resorts. On the west are the ancient borough and port of Queenborough and the naval establishment of Sheerness (*qq.v.*).

SHEPSTONE, SIR THEOPHILUS (1817–1893), British South African statesman, was largely responsible for the peaceful development of Natal (*q.v.*). He was born at Westbury, near Bristol, Eng., on Jan. 8, 1817, the son of the Rev. William Shepstone, a Wesleyan minister. The family emigrated in 1820 to Cape Colony, where Shepstone was educated at his father's mission station. He acquired great proficiency in the Xhosa (Kaffir) languages and insight into Xhosa culture, both of which later helped to make him an outstanding administrator. He served as headquarters interpreter on the staff of the governor of Natal, Sir Benjamin D'Urban (*q.v.*), in the sixth Kaffir War (1834–35), and at the end of the campaign remained on the frontier as clerk to the agent for the native tribes. In 1838 he was one of the force sent to occupy Port Natal at Durban, and when the force was recalled in 1839, Shepstone was appointed British resident among the Fingo and other tribes in Kaffraria.

After Natal was placed under separate government in 1845, Shepstone was made agent for the native tribes. In 1848 he became captain general of the native levies, and in 1855 judicial assessor in native causes. When Natal became a colony, he was secretary for native affairs and a member of the executive and legislative councils from 1856 to 1877. The main line of his policy was to maintain tribal customs as far as was consistent with humane principles, and not to attempt to force civilization on the tribes. When the Zulu king Dingane was murdered in 1840 Shepstone resettled the tribal remnants scattered by the king's reign of terror on new locations under the successor king Panda (Mpande). So well was a tribal structure of sorts achieved that the only subsequent major risings in Natal were in 1873 and 1906. His policy was based on security of tribal tenure and the custom based on that tenure, both subject to his own authority as supreme chief. Also, apart from the crisis of 1873, he kept Natal tribes from being involved in frontier wars.

On two occasions he moved into Zululand (*q.v.*) on political errands. In 1861 he persuaded Panda formally to recognize his son Cetshwayo (*q.v.*) as successor, and in 1873 Shepstone attended the proclamation of Cetshwayo as king and made him swear to rule without illegal violence. During Shepstone's absence from Africa in 1879, Cetshwayo rejected an ultimatum to stop his forays over the border in pursuit of fugitives, was himself captured and exiled, but at his partial restoration in 1883, Shepstone was specially commissioned to install him. In 1874 Shepstone was consulted in London on native policy by the secretary of state for colonies, and in 1876 Sir Theophilus attended the conference there on South African affairs to discuss how to achieve a federal approach to native problems. He returned with a commission empowering him to confer with the Transvaal (*q.v.*) executive on the question of the federation of South African states and, should he deem it necessary, to annex the country, subject to the confirmation of the British government. He went to Pretoria in January 1877 with a force of 25 mounted policemen, and on April 12 issued a proclamation announcing the establishment of British authority over the Transvaal though he did not have majority approval of the *volksraad*. Outwitted and obstructed in the Transvaal and ill supported both by the Cape government and the Colonial Office, Shepstone remained in Pretoria as administrator until January 1879, when he was replaced and summoned home to London to advise the Colonial Office on South African affairs. On his return to Natal in 1880, he retired. During his retirement he was the victim of controversy, not the least was because he opposed the grant of responsible government to Natal out of fear that native policy would become a political football between two parties. He died in Pietermaritzburg on June 23, 1893.



RADIO TIMES HULTON PICTURE LIBRARY

SHEPSTONE

Shepstone was knighted in 1876, and the town of Port Shepstone, Natal, was named after him. The natives called him "father" and "Somsteu," in reference to his hunting abilities.

His younger brother JOHN WESLEY SHEPSTONE (1827–1916) was in the civil service in Natal, acting as secretary for native affairs and judicial assessor (1861, 1874, 1876–83).

Of Sir Theophilus' six sons, Theophilus (1843–1907) was advocate of the supreme court of Natal and a member of the legislative council; Arthur J. (1852–1912) served in various native expeditions, as assistant commissioner in Zululand (1887–89, 1889–90), in the South African War (1899–1902), and also as magistrate at various times from 1901 through 1907.

See C. J. Uys, *In the Era of Shepstone* (1933); E. H. Brookes, *The History of Native Policy in South Africa from 1830*, 2nd ed. (1927); D. R. Morris, *The Washing of the Spears* (1965). (W. A. ML.; X.)

SHER ALI KHAN (1825–1879), amir (king) of Afghanistan from 1863 to 1879, was born at Kabul in 1825. When his father, Dost Mohammed Khan (*q.v.*), died in 1863 he left behind him more than a dozen sons, of whom five aspired to the throne, though the amir had nominated Sher Ali, his third surviving son, as his heir. Sher Ali promptly proclaimed himself amir, but from the outset he was confronted by innumerable difficulties with his two elder brothers, Muhammad Azam and Muhammad Afdal, and with his nephew, Abd al-Rahman, son of Afdal. In 1866 Abd al-Rahman seized Kabul and in the next year drove Sher Ali from Kandahar. Afdal having died in 1867, Abd al-Rahman shifted his support to Azam, but Sher Ali and his eldest son Yakub defeated them in 1868.

In the early part of his reign Sher Ali was on excellent terms with the British, particularly with Lord Lawrence, the viceroy of India. In March 1869 he accepted an invitation from Lord Mayo, Lord Lawrence's successor, and met the new viceroy at Ambala. Sher Ali imprisoned Yakub when he rebelled at the naming of his younger brother, Abdullah Jan, as successor. Mayo refused to recognize him, thus arousing Sher Ali's resentment. Similarly British arbitration of a border dispute which gave a good part of the most fertile lands in Seistan to Persia in 1873 further alienated the amir. In the same year British failure to assure Sher Ali of support against Russian territorial advances turned the amir to Russia, who assured him they would not support Abd al-Rahman. In 1876 Lord Lytton succeeded Lord Northbrook as viceroy of India. An advocate of the "forward policy," Lytton moved British troops closer to Afghanistan and sought to have a British mission received by Sher Ali. The amir refused the request, in 1878 received a Russian mission, and turned back a British mission at the border. Abdullah Jan died. Thus the British government drifted into the Second Afghan War (1878–81). On Nov. 21, 1878, the three great passes leading into Afghanistan were entered by British armies, which took Kabul. Sher Ali, hoping to get help from the Russians, left Yakub Khan in Kabul

as regent and went to Mazar-i-Sharif, hoping to get help from the Russians. The Russians advised him to make his peace with the British. The amir, already worn out by physical disease and mental worry, died at Mazar on Feb. 21, 1879. *See* also AFGHANISTAN. (MD. A.)

SHERATON, THOMAS (1751–1806), English furniture designer, was born at Stockton-on-Tees, County Durham, the son of a working cabinetmaker. He appears to have been apprenticed to a cabinetmaker (an obituary notice in *The Gentleman's Magazine* [London, 1806] stated that Sheraton "worked for many years" at that trade), but he was ever a strange blend of mechanic, inventor, artist, mystic, and religious controversialist. Indeed, it is as a writer on theological subjects that we first hear of him. In 1782 he published at Stockton *A Scriptural Illustration of the Doctrine of Regeneration*, to which was added *A Letter on the Subject of Baptism*, describing himself on the title page as a "mechanic, one who never had the advantage of collegiate or academical education." He settled in London probably about 1790, and his trade card gave his address as Wardour Street, Soho, where he "teaches perspective, architecture and ornaments, makes designs for cabinetmakers, sells all kinds of drawing-books." Nothing is said of him as a maker of furniture, and all he tells us in his works is that "having possessed a strong attachment for carving in my youth, I was necessarily inclined to make attempts in this art . . . and was employed in the County occasionally in it."

Supporting himself "mainly by his exertions as an author" Sheraton published in four parts between 1791 and 1794 *The Cabinet-Maker and Upholsterer's Drawing Book*, by far his most important undertaking. The first part, which throws a curious sidelight on the author's mentality, is devoted to verbose dissertations on perspective, architecture, and geometry, which strike the reader as a naïve attempt to display the author's learning. But the notes on the plates are much fuller than in any other comparable publication and reveal sound technical knowledge. They leave no doubt that Sheraton's primary aim was "to exhibit the present taste in furniture"; as he put it himself, "I have made it my business to apply to the best workmen in different shops, to obtain their assistance in the explanation of such pieces as they have

BY COURTESY OF (LEFT) THE VICTORIA AND ALBERT MUSEUM; PHOTOGRAPH BY C. R. CANNINGS, (ABOVE) A. C. COOPER LTD.

(LEFT) "A NEW DESIGN OF A BOOKCASE & WRITING DRAWER" FROM THE CABINET-MAKER AND UPHOLSTERER'S DRAWING BOOK BY THOMAS SHERATON, 3RD ED., 1802. (RIGHT) 18TH-CENTURY SATINWOOD WORKTABLE MADE FROM A DESIGN IN THE SAME BOOK: IN A PRIVATE COLLECTION

been most acquainted with." In one or two instances he even informs the reader where specialized pieces of furniture may be obtained; a worktable is "taken from one executed by Mr. McLean in Marylebone Street . . . who finishes these small articles in the neatest manner." It is on the designs in the second part of the *Drawing Book* that Sheraton's reputation (in a great measure misleading) is certainly based. Throughout they are admirable in draftsmanship, rarely at fault in form or proportion, while the distribution and propriety of the ornament may be explained in part by the author's "strong attachment and inclination for carving."

In 1803 Sheraton brought out his *Cabinet Dictionary* (with plates), containing *an Explanation of all Terms Used in the Cabinet, Chair and Upholstery Branches With Dictionary for Varnishing, Polishing and Gilding*. It does not fulfill the promise of the title: the selection of terms is arbitrary and eclectic, suggesting that the author's marked tendency to eccentricity was on the increase. It is, however, sober and informative compared with Sheraton's final project, the *Cabinet-Maker, Upholsterer and General Artists' Encyclopaedia*, of which only one volume covering A to C appeared in 1805, the year before his death. The text warrants the inference that he suffered at least from mental instability, while the garish colour plates may be said almost to travesty the prevailing Regency style, so marked is the eccentricity of some of the designs.

Adam Black, a Scottish publisher and politician who was employed by Sheraton in connection with the encyclopaedia, refers to the latter's "painfully humble circumstances" and describes his house as "half shop, half dwelling-house"—firsthand evidence which should alone have sufficed to refute the legend that Sheraton was a fashionable cabinetmaker. Even now, however, pieces are often described as "by Sheraton" in catalogues and press reports of sales. There are several examples which are either virtually identical with, or closely resemble, a Sheraton design, but this affords no presumption that he was the maker. Sheraton's books, like those of the other great cabinetmakers of the second half of the 18th century, were intended not for the general reader, but for the practical use of the trade, which, no doubt, copied their designs extensively; indeed, hundreds of cabinetmakers and joiners were among the subscribers to the *Drawing Book*. It was in fact a trade catalogue, the plates of which were reproduced throughout the country with varying degrees of accuracy. It is reasonable to suppose that he himself obtained orders by the publication of his books and employed other cabinetmakers to manufacture the work; it seems certain that he himself never possessed anything more than a small shop. Of his own manufacture only one piece is known with certainty—a glass-fronted bookcase, stamped "T. S." inside one of the drawers.

Though Sheraton doubtless borrowed what served his turn, there is good reason to suppose that the bulk of the plates in his publications were engraved from his own designs. The version of the neoclassical style represented in Hepplewhite's *Guide*, which appeared only three years earlier than the *Drawing Book*, conveys a general impression distinctly different from that associated with Sheraton. The style in the earlier book, simple, elegant, and notably free from extravagance and eccentricity, as interpreted by Sheraton becomes more refined and feminine in character; the eccentricity so conspicuous in Sheraton's later works is as yet scarcely perceptible. The designs bear the mark of a single controlling intelligence, eclectic and even now and again ready to reproduce existing models, but gifted with an innate sense of proportion and style. Much of what he borrowed he adapted and improved upon. The term "Sheraton" has been recklessly bestowed upon vast quantities of late 18th-century painted and inlaid satinwood furniture, but properly understood and used in a generic sense it is an appropriate recognition of the most powerful source of inspiration behind the furniture of the century's closing years. Sheraton writes slightingly of Hepplewhite's *Guide* in his preface, and it is significant that in the 3rd edition of the *Guide* (1794) the designs for chairs are drastically amended, some of those with square backs then newly introduced being scarcely distinguishable from the chairs illustrated by Sheraton.

That Sheraton can have been personally popular is incredible. His books make it evident that his character was tart, angular, and self-assertive, and that he was little disposed to be generous toward the work of predecessors or rivals. Such an attitude toward the world would suffice to explain his lack of substantial success. He appears to have preached occasionally to the end, and even in his furniture books he sometimes falls into improving remarks of religious character. He died in Soho on Oct. 22, 1806.

See R. Edwards and M. Jourdain, *Georgian Cabinet-makers*, 3rd ed. (1955); R. Fastnedge, *Sheraton Furniture* (1962). (R. ED.)

SHERBORNE, a market town and urban district in the West Dorset parliamentary division of Dorset, Eng., 19 mi. (31 km.) NNW of Dorchester by road. Pop. (1961) 6,053. Area 2.6 sq.mi. It lies near the border of Somerset on the southern slope of a hill overlooking the River Yeo, in a fertile, well-wooded district. In 1951 traces of Romano-British occupation were found in the southern part of the town, dating from the 2nd–3rd century A.D. Other sites include a large Roman settlement, 10 ac. in extent, overlying occupation of the British (Belgic) period. It is first mentioned in 705 as the place where St. Aldhelm fixed his bishopstool for the new diocese of Western Wessex. In 1075 the see of Sherborne was transferred to Old Sarum and the cathedral of Sherborne became the abbey church of St. Mary the Virgin. Roger de Caen, bishop of Old Sarum from 1107 to 1139, replaced the Saxon church by a Norman building, parts of which remain today. Pre-Norman work appears in the western wall, and there are an Early English chapel and some Decorated windows. The church, however, was almost entirely reconstructed in the Perpendicular period; the colour decorations in the choir and the stone-vaulted roof with fan tracery being particularly fine. The completion of this work in 1490 gave existence to the Pack Monday Fair, which is still held annually on the first Monday after Oct. 10. Of the original building of the old castle, also built by Bishop Roger, only the gatehouse, portions of the curtain wall, the chapel, the keep, and certain domestic buildings remain. In 1139 the castle was seized by Stephen and remained with the crown until 1355, when it was returned to the bishops and retained by them until 1599, when it was given by Elizabeth I to Sir Walter Ralegh who built the present castle, now used as a museum. During the Great Rebellion the old castle was twice besieged, Cromwell himself being present in 1645 when it fell to Gen. Thomas Fairfax after a defense lasting 16 days. The almshouse of St. John Baptist and St. John the Evangelist was refounded in 1437 and retains a Perpendicular chapel, hall, and other portions. The town is an educational centre. Sherborne Boys' School now occupies the site of the school for which Edward VI granted a charter in 1550. There are also Sherborne Girls' School and Lord Digby's School for Girls, founded in 1743. Silk weaving was introduced by Huguenots in 1740 and the town also specializes in glove making and agriculture. It has a milk factory, light engineering and printing works, and a brewery.

SHERBROOKE, ROBERT LOWE, VISCOUNT (1811–1892), British statesman, leader of the opposition to the Reform Bills of 1866 and 1867, chancellor of the exchequer from 1868 to 1873, was born on Dec. 4, 1811, at Bingham, Nottinghamshire, the son of an Anglican minister. He was educated at Winchester and at University College, Oxford, where he distinguished himself by taking first place in classics and a second in mathematics, besides taking a leading part in the Union debates. He was fellow of Magdalen College (1835–36) until he married Georgiana Orred. Lowe became renowned as the outstanding tutor of his day, a savage examiner, and an energetic controversialist. Called to the bar in 1842, he emigrated in the same year to New South Wales, where he was soon prominent in the courts and in the Legislative Council as a supporter of efficient but representative government. Lowe made his mark in the political world by his clever speeches, particularly on finance and education, and besides obtaining a large legal practice, he was one of the principal writers for the *Atlas*, a weekly newspaper. In 1850 he returned to England in order to enter political life there, "with a tolerable fortune and a detestation of democracy," and became a familiar figure in society, his previous university reputation and connections, his colonial experience, and his caustic wit standing him in good stead. He began writing leaders (editorials) for *The Times* in 1851, and until J. T. Delane left the editorship in 1877 Lowe's influence remained considerable.

Lowe was Liberal member of Parliament for Kidderminster, Worcestershire (1852–59), and afterward represented Calne, Wiltshire (1859–68), and the University of London (1868–80). He was joint secretary of the Board of Control for India (December 1852–February 1855) under Lord Aberdeen, and vice-president of the Board of Trade (August 1855–March 1858) and vice-president of the Committee of Council for Education (June 1859–April 1864) under Lord Palmerston. In these minor offices he helped to establish competitive entry to the Indian civil service; his Joint-Stock Company Acts of 1856 and 1857 accepted the principle of limited liability; and he introduced the Revised Code (1862) of "payment by results" in elementary education (*see* EDUCATION, HISTORY OF), opposition to which caused his resignation in 1864. He rather scandalized his old university friends by the stress he laid on physical sciences as opposed to classical studies.

Lowe proved an effective speaker and a capable administrator; but his brusque manner and sarcastic contempt for both vested interests and eager reformers gave much offense. An albino, and almost blind, he ascribed his faults to "physical rather than moral deficiencies." However, there can be no doubt of his intellectual arrogance, confidently and lucidly expressed, nor of the rigour with which he applied his harshly Benthamite views, stressing the sole criteria of merit for the individual and of utility and economy for the actions of the state. Fortified by experience in both England and Australia, and on his visit to the United States in 1856, Lowe regarded democracy as the enemy of good government, which he considered should be based upon a limited, informed electorate. He was satisfied with Palmerston's benign conservatism, tempered by Gladstonian efficiency. In 1866, by a series of shrewd and brilliant speeches, he caused the defeat of Lord John Russell's Reform Bill and brought down the Liberal ministry. Few men have more completely (or unexpectedly) dominated the House of Commons.

As leader of the Adullamites (*q.v.*), the group of antireform Liberals, Lowe's influence attained its zenith; yet his very success stimulated the demand for parliamentary reform and thus inadvertently helped to produce Benjamin Disraeli's more drastic measure, the Second Reform Act (1867). Despite Lowe's mutiny of 1866, Gladstone made him chancellor of the exchequer in 1868, admiring his talents and trusting his opinions on finance; and Lowe gave the reforms of Gladstone's most notable administration his full support. Maintaining Gladstone's tradition at the Treasury, Lowe was at first a success; but after 1871 his old lack of tact became a liability. He was transferred to the Home Office in 1873, but his active political life came to an end with the ministry's defeat in the following year. When the Liberals returned to power in 1880, Lowe, already almost forgotten, was created Viscount Sherbrooke. His first wife having died in 1884, Lowe married Caroline Sneyd in 1885. He had no children by either marriage. He died at Sherbrooke Lodge, Warlingham, Surrey, on July 27, 1892, and the title became extinct.

See Asa Briggs, *Victorian People* (1954). (A. F. T.; X.)

SHERBROOKE, a city and seat of Sherbrooke County, Quebec, Can., is about 100 mi. (161 km.) E of Montreal and 30 mi. (48 km.) N of the U.S. boundary at Derby Line, Vt. It was named after Sir John Coape Sherbrooke (1764–1830) who from 1816 to 1818 was governor-general of Canada. Its situation at the confluence of the Magog and St. Francis rivers first brought it importance as a trading centre where the French bought furs from the Indians. Later, along with the surrounding regions (known as the Eastern Townships), it was further developed by United Empire Loyalists and their descendants as a water-powered grist mill centre for the grain from their agricultural holdings. It was incorporated as a town in 1852 and as a city in 1875. The population in 1966 was 74,229, more than 85% being of French origin. The intersection there of the principal east-west and north-south road and rail routes of the Eastern Townships, plus a large labour force and ready markets, have contributed toward the emergence of Sherbrooke as an important industrial, commercial, and cultural centre.

It is the seat of a Roman Catholic bishopric and of the Université de Sherbrooke, a French-language Roman Catholic university chartered in 1954. Hydroelectric power developed from neighbouring rivers helped in attracting over 100 industries to the city. The principal of these are engaged in the manufacture of textiles, transportation, quarrying, and mining equipment, machinery, oil burners, gloves, dairy products, leather and rubber goods, boots and shoes, hosiery and paper. A television station broadcasts mainly in the French language. Tourism and summer

cabin settlements are important in the surrounding hilly areas and lake districts and favour the city through increased trade and commerce in the summer months. (W. F. Ss.)

SHERIDAN, PHILIP HENRY (1831–1888), most successful Union cavalry leader of the American Civil War, was born March 6, 1831. Sheridan left his birthplace in doubt, at different times listing Albany, N.Y., Boston, Mass., and Somerset, O. No supporting evidence can be found for any one of the three. He entered West Point from his boyhood home of Somerset in 1848, and, while there, his habit of fighting cost him a suspension that delayed his graduation until 1853. Later this aggressive spirit, under better control, marked Sheridan as a military leader. Nine years of duty, mostly at frontier posts in the Southwest and the Northwest, brought him into the Civil War as a quartermaster officer in Southwest Missouri.


THE GRANGER COLLECTION
GENERAL SHERIDAN, PHOTOGRAPHED BY MATHEW BRADY

His desire to get into combat was satisfied in the spring of 1862 when he was appointed colonel of the 2nd Michigan Cavalry. On July 1 at Booneville, Miss., Sheridan skilfully split his outnumbered command to rout a large Confederate force, and Gen. W. S. Rosecrans declared that he was worth his weight in gold. He was made a brigadier general. At Perryville, Ky., (Oct. 8) he led the 11th Division, Army of the Ohio, as it held its position against repeated attacks. At the battle of Stones River (Dec. 31, 1862, to Jan. 3, 1863) he was made a major general of volunteers for his unyielding defense of the Union right centre. Later the Army's move south to drive the Confederates from Tennessee brought defeat to Rosecrans at Chickamauga, Ga. (Sept. 19–20). Though forced to withdraw, Sheridan managed to regroup his division but he was too late to aid Gen. G. H. Thomas, which exposed Sheridan to some criticism after the war. In the assault on Missionary Ridge below Chattanooga (Nov. 25, 1863), Sheridan brought his fighting in the west to a brilliant close, and Grant called him east in the spring of 1864 to head the cavalry of the Army of the Potomac.

Sheridan at once lifted the morale of the troops by re-outfitting the command, curtailing picket and escort duty to give the command more independence of action, and dismounting his men whenever possible to exploit the firepower of their new repeating weapons. Following action in the Wilderness (May 5–6), Sheridan led a raid toward Richmond that destroyed supplies and rolling stock, and resulted in the death of the South's great cavalry leader, "Jeb" Stuart.

On Aug. 4 Sheridan took command of the Army of the Shenandoah, charged with clearing that rich valley of Confederates and destroying their means of subsistence. A six-week delay for preparations taxed Northern patience, but in the battle of Opequon (Sept. 19) Sheridan drove Gen. Jubal Early out of Winchester. Further victories at Fisher's Hill (Sept. 22) and Cedar Creek (Oct. 19) quickly made him a major general. It was at Cedar Creek that Sheridan arrived after his famous ride from Winchester to stabilize his front and reverse what might have been a defeat. This battle also brought him the thanks of Congress. Meanwhile, on Grant's orders, Sheridan had destroyed the capacity of the Shenandoah Valley to support operations by the South. Near the end of March 1865 he rejoined his cavalry before Petersburg. With the 5th Corps Infantry added to his command, he circled south and west of the city to cut Lee's rail communications. On March 31 at Dinwiddie Court House and next day at Five Forks, Sheridan broke into the Confederate right and rear, forcing Lee to retire westward from his Richmond–Petersburg lines. But Sheridan, despite his success, felt that General Warren had lagged in pressing the attack of his 5th Corps and so removed him from command. The order brought Sheridan under considerable fire both at the time and later. To Grant, however, he remained the ideal officer, one who had shown that driving ahead at all cost was the only way to win the war. Sheridan continued his pressure against Lee's southern flank and at the end helped close off his escape near Appomattox. By now Sheridan's driving force had made him a dynamic and popular general. Soldiers responded enthusiastically to his banner-waving gallops along the lines. His reputation as a combat leader was secure.

The war over, Sheridan was ordered to the Gulf, where his presence along the Texas border hastened the fall of Maximilian, the French puppet emperor in Mexico. But later, as military commander of Louisiana and Texas in 1867, Sheridan's failure to adjust to the complex requirements of civilian authority led to his removal. He spent the next years until 1883 in western command. He planned and conducted a successful Indian campaign during the winter of 1868–69, after which he was promoted to the rank of lieutenant general. Abroad in 1870–71, Sheridan remained for two months as an observer at German headquarters near the close of the Franco-German War. Back at his Chicago headquarters, he married Irene Rucker, daughter of a fellow officer, in 1875. He became general in chief of the Army Nov. 1, 1883, and on June 1, 1888, received his commission as full general. He spent his last months writing his *Personal Memoirs* (2 vol., 1888), and died on Aug. 5 of that year in Nonquitt, Mass.

See Richard O'Connor, *Sheridan the Inevitable* (1953).

(E. J. N.; X.)

SHERIDAN, RICHARD BRINSLEY BUTLER (1751–1816), Irish-born dramatist remembered for his brilliant comedies of manners, but also important in his own period as a Whig politician, as adviser to the prince of Wales (later George IV), and as a political orator, was the third son of Thomas and Frances Sheridan. The date of his birth is not known, but he was baptized (mistakenly as "Thos. Brinsley") at St. Mary's Church, Dublin, on Nov. 4, 1751. His family later moved to London, and he never returned to Ireland. He was educated (1762–*c.* 1767–68) at Harrow, and in 1770 moved with his family to Bath.

While there he corresponded with Nathaniel Halhed, a schoolfriend then at Oxford. Together they wrote *Jupiter*, a farce, which after revision, Sheridan renamed *Ixion*. They also collaborated in a metrical version (1771) of the epistles of Aristaenetus. Sheridan's poems "The Ridotto of Bath" and "Clio's Protest" also date from this period.

In Bath, Sheridan met Elizabeth Ann Linley (1754–92), whose fine soprano voice had, from an early age, delighted audiences at the concerts and festivals conducted by her father, Thomas Linley (*q.v.*). She thought herself persecuted by a Welsh squire, Thomas Mathews of Llandaff, and decided to take refuge in a French nunnery. In March 1772, Sheridan accompanied her to Lille, but returned to fight duels with Mathews, in London on May 4 and at Bath on July 1, 1772. Meanwhile, Elizabeth had returned home with her father, and Sheridan was ordered by his father to Waltham Abbey, Essex, to pursue his studies. He was entered at the Middle Temple on April 6, 1773, but broke with his father and a legal career on April 13, when he married Elizabeth at Marylebone Church, London.

Dramatic Career.—After his marriage, Sheridan turned to the theatre for a livelihood, and his comedy, *The Rivals*, was produced at Covent Garden Theatre on Jan. 17, 1775. It ran for an hour longer than was usual; and, because of the offensive nature of the character of Sir Lucius O'Trigger, and also because this part was poorly acted, it was hardly a success. Drastically revised for its second performance on Jan. 28, and with a new actor as Sir Lucius, it won immediate applause. The situations and characters were not entirely new, but Sheridan gave them freshness by the buoyancy of his wit and his richly humorous presentation. The scenes between Sir Anthony Absolute and his son are particularly striking, and the whole play reveals Sheridan's remarkable sense of theatrical effect. Characteristic too, is the genial mockery of affectation: the absurdities of Lydia Languish and Acres are unforgettable. Even the "malapropisms" that slow down the play give a

proper sense of caricature to the character of Mrs. Malaprop.

Some of the play's success was due to Lawrence Clinch's acting as Sir Lucius. Sheridan showed his gratitude by writing for Clinch's benefit night (May 2, 1775) the amusing little farce, *St. Patrick's Day: or, The Scheming Lieutenant*. Another example of his ability to invent an interesting plot from well-worn materials is *The Duenna* (Covent Garden, Nov. 21, 1775). The characters, except for the delightful Isaac Mendoza, are undeveloped, but the clever intrigue and charming lyrics, and the music by Thomas Linley and his son, gave this ballad opera great popularity. Its 75 performances even exceeded the record number of 62 credited to Gay's *The Beggar's Opera* (1728), and it is still revived.

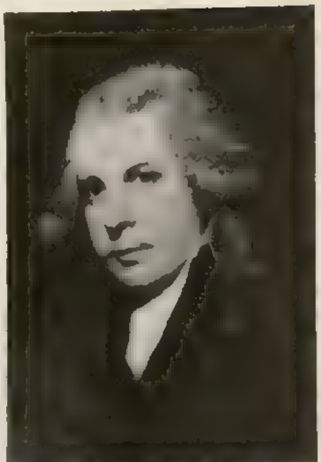
Thus, in less than a year, Sheridan had brought himself to the forefront of contemporary dramatists. David Garrick (*q.v.*), looking for someone to succeed him as manager and proprietor of Drury Lane Theatre, saw in him a young man with energy, shrewdness, and a real sense of theatre. A successful physician, Dr. James Ford, agreed with this estimate and decided to increase the amount of money he had thought of investing in the playhouse. In 1776, Sheridan and Linley became partners with Ford in a half-share of Drury Lane Theatre, and they bought the other half from Willoughby Lacy, Garrick's partner, two years later. Sheridan had only £1,300 in the concern, and the transaction was only possible because Garrick and Ford backed him with mortgages in the belief that Drury Lane would flourish under his direction.

In fact, Sheridan's interest in his theatre soon began to seem rather fitful, but he was responsible for the renewed appreciation of Restoration comedy that followed the revival at Drury Lane of the plays of William Congreve (*q.v.*). In February 1777, he also brought out his version of Vanbrugh's *The Relapse* (1696) as *A Trip to Scarborough*. This again showed his talent in revision. He gave the rambling plot a neater shape, and removed much indelicacy from the dialogue, but the result was disappointing, probably because the play had lost much of its gusto.

What Sheridan learned from the Restoration dramatists can be seen in *The School for Scandal* (Drury Lane, May 8, 1777), which earned him the title of "the modern Congreve." Although resembling Congreve in that its satirical wit is so brilliant and so general that it does not always distinguish character, it contains two subtle portraits in Joseph Surface and Lady Teazle. There were several Restoration models (*e.g.*, Mrs. Pinchwife in Wycherley's *The Country Wife*, and Miss Hoyden in Vanbrugh's *The Relapse*) for the portrayal of a country girl, amazed and delighted by the sexual freedom of high society, but Sheridan softened his Lady Teazle to suit the more refined taste of his day. The part combined innocence and sophistication and was incomparably acted by Frances Abington (*q.v.*). The other parts were written with equal care to suit the members of his company, and the whole work is a triumph of intelligence and imaginative calculation. Further, its spirited ridicule of affectation and pretentiousness makes it the best comedy of manners in English.

The auction and screen scenes in *The School for Scandal* demonstrate Sheridan's flair for stage effect, shown again in his delightful satire on stage conventions, *The Critic*, which, ever since its first performance (Oct. 30, 1779), has been thought much funnier than its model, *The Rehearsal* (1671) by the 2nd duke of Buckingham (*q.v.*). Sheridan himself considered the first act his finest piece of writing. Its character sketches are cleverly drawn and, although Puff is little more than a type, Sir Fretful Plagiary is not only a caricature of the dramatist Richard Cumberland (*q.v.*), but an epitome of the vanity of authors in every age.

Political Career.—Sheridan continued to adapt plays and to improvise spectacular shows at Drury Lane, but, as a succession of acting managers took over the burden of direction, his time was increasingly given to politics, and his only full-length later play was the worthless but popular patriotic melodrama, *Pizarro* (Drury Lane, May 24, 1799), based on August von Kotzebue's *Die Spanier in Peru* (1796). Sheridan had become member of Parliament for Stafford in September 1780, and was under-secretary for foreign affairs in Rockingham's ministry (1782) and secretary to the treasury in the duke of Portland's coalition ministry (1783). He



BY COURTESY OF THE LORD PLUNKET

RICHARD BRINSLEY SHERIDAN, DETAIL OF A FULL-LENGTH PORTRAIT BY SIR JOSHUA REYNOLDS, 1788–89

took office again, in the "ministry of all talents" (1806), as treasurer of the navy (1806–07), and became a privy councilor. The rest of his 32 years in Parliament were spent in opposition. His critical acumen and command over language had full scope in oratory, and were seen at their best in his speeches as manager of the impeachment of Warren Hastings (*q.v.*). Although recognized as one of the most persuasive orators of his time, however, he never acquired the political influence he and others thought he deserved, because he was regarded as an unreliable intriguer. There is some support for this view in his behaviour during the Regency crisis (1788–89), when he acted as the prince of Wales' adviser, and in the party squabbles (1791–93) with Edmund Burke (*q.v.*).

He was one of the few members courageous enough openly to defend those who suffered for their support of the French Revolution. Indeed, Sheridan liked taking an individual stand, and although he supported Charles James Fox (*q.v.*), he came out on the side of government when he condemned the Nore mutineers (1797). Much to Fox's disgust, he also gave Addington's administration (1801–04) some support.

In November 1806, Sheridan succeeded Fox as member for Westminster (although not, as he had hoped, as leader of the Whigs), but he lost the seat in May 1807. The prince of Wales then returned him as member for the pocket borough of Ilchester, but his dependence on the prince's favour rankled with Sheridan, because they differed in their attitude to Catholic emancipation. Sheridan, who was determined to support emancipation, stood for Stafford again in 1812, but could not pay those who had previously supported him as much as they expected, and as a result was defeated.

Last Years.—His financial difficulties were largely brought about by his own extravagance and procrastination, as well as by the destruction by fire in February 1809 of Drury Lane Theatre (which had been rebuilt in 1791–94). When he lost his parliamentary seat, and his income from the theatre, he became a prey to his many creditors. His last years were unhappy because of these difficulties, and because of worry over his circulatory complaints, and the cancer that afflicted his second wife, Hester Jane Ogle, daughter of the dean of Winchester. He had married her in April 1795, three years after Elizabeth's death. Sheridan himself died in London on July 7, 1816, pestered by bailiffs to the end. Even in decline, he made a strong impression on Byron, who wrote a *Monody on the Death of the Right Honourable Richard Brinsley Sheridan* (1816), to be spoken at Drury Lane Theatre.

Conclusion.—Sheridan's genius both as dramatist and politician lay in humorous criticism, and ability to size up situations and relate them effectively. These gifts were often exercised in the House of Commons on other men's speeches, and at Drury Lane in the revision of other men's plays. They are seen at their best in *The School for Scandal*, where he shaped, from two mediocre draft-plays of his own, a plot and dialogue of unusual brilliance. In person he was often drunken, moody, and indiscreet, but he possessed great charm and powers of persuasion. As a wit he delivered his sallies against the follies of society with a polish that makes him the natural link between Congreve and Wilde.

THOMAS SHERIDAN (1687–1738), grandfather of the dramatist, was born at Cavan, Ire., in 1687. He is chiefly known as a companion and confidant of Jonathan Swift, whom he offended by fulfilling an old promise to tell Swift if he ever saw signs of avarice in him. He died in poverty at Rathfarnham, Ire., Oct. 10, 1738.

THOMAS SHERIDAN (1719–1788), son of the above, was noted for his *Pronouncing Dictionary* and for his advocacy of a scheme of public education, in which elocution was to play an important part. He was born in Dublin in 1719 and died at Margate on Aug. 14, 1788.

FRANCES SHERIDAN (1724–1766), wife of the above and mother of the dramatist, gained some fame as a playwright. She died on Sept. 26, 1766, at Blois.

BIBLIOGRAPHY.—*Editions and Bibliography: The Plays and Poems of R. B. Sheridan*, ed. by R. Crompton Rhodes, 3 vol. (1928) is the standard text. It should be supplemented by R. L. Purdy's excellent edition of *The Rivals, a Comedy, as it was first Acted at the Theatre Royal in Covent Garden* (1935), and by G. H. Nettleton's texts of *The Rivals, The School for Scandal*, and *The Critic*, in *British Dramatists from Dryden to Sheridan*, ed. by Nettleton and A. E. Case (1939). No full bibliography of Sheridan's works has been published: valuable bibliographical material is to be found in J. P. Anderson's lists in L. C. Sanders' *Life of R. B. Sheridan* (1890); in *The Major Dramas of R. B. Sheridan*, ed. by G. H. Nettleton (1906); in the works by Sichel (*see* below) and Rhodes (*see* above); and in I. A. Williams, *Seven XVIIIth-Century Bibliographies* (1924). On bibliographical points *see* also F. W. Bateson, "Notes on the Text of Two Sheridan Plays," in *Review of English Studies*, vol. xvi (1940); G. W. Williams, "A New Source of Evidence for Sheridan's Authorship of *The Camp* and *The Wonders of Derbyshire*," in *Studies in Philology*, vol. xlvii (1950); W. van Lennep, "The Chetwynd Manuscript of *The School for Scandal*," in *Theatre Notebook*, vol. vi (1951); C. Price, "Hymen and Hirco: a Vision," in the *Times Literary Supplement* (July 11, 1958), and "The Text of the First Performance of *The Duenna*," in *Papers of the Bibliographical Society of America*, vol. liii (1959).

Biography and Criticism: Thomas Moore's *Memoirs of . . . Sheridan*, 5th ed., 2 vol. (1827) is indispensable. William Smyth's *Memoir of Mr. Sheridan* (1840) is a brief but vivid account from personal acquaintance. W. Fraser Rae, *Sheridan*, 2 vol. (1896) and W. Sichel, *Sheridan*, 2 vol. (1909) are full accounts, printing much new material, but are not altogether reliable. R. C. Rhodes, *Harlequin Sheridan* (1933), provides a scholarly review of evidence about disputed passages in Sheridan's life. L. Gibbs, *Sheridan* (1947), gives a useful summary of accepted material. J. Dulck, *Les Comédies de R. B. Sheridan* (1962), is a close examination of Sheridan's achievement from the French point of view. (C. J. L. P.)

SHERIFF (HIGH SHERIFF), a senior executive officer in an English county or smaller area, who performs a variety of ancient administrative and judicial duties. Officers of this name also exist in Scotland, Northern Ireland, and the United States.

England.—The office of sheriff is of pre-Conquest origin. In the 10th century the Anglo-Saxon shire was controlled by the ealdorman, who presided in its court jointly with the bishop and the king's reeve, the sheriff. By the 11th century the ealdorman had become responsible for a whole province rather than a single shire; and the separation of the ecclesiastical from the secular courts under William I left the sheriff supreme in the county and as president of its court. He convened and led the military forces of the shire, executed all writs, and for the first century after the Conquest judged both criminal and civil cases. But the increasing use, from the time of Henry II, of itinerant justices from the *curia regis* ("king's court") and the reservation to the *curia* and its developing branches of all serious criminal and some civil cases, severely restricted his jurisdiction. Henceforth his duty was to investigate allegations of crime from within his shire, to conduct a preliminary examination of the accused, try lesser offenses, and detain those accused of major crimes for the judge's assize. Twice a year he made his "tourn" in the hundred (*q.v.*) courts. He was also responsible for all dues owed to the Exchequer, to which the sheriff alone had audience and direct obligation for his shire. In the later part of the 12th century this duty was also limited by the supervisory powers of itinerant justices and eventually he ceased to attend the Exchequer.

In the late 11th and the 12th centuries the sheriffs wielded great power. There was always danger of the office becoming hereditary. In 1170 Henry II held an inquest on the conduct of his sheriffs and many were dismissed. A statute of 1300 ordering sheriffs to be elected by the county was superseded in 1315 by one arranging for their appointment by various officers of state. In 1340 the office became annual only. The new offices of coroner (first mentioned in 1194), of local constable (first mentioned in 1242), and of justice of the peace (first known, in the 12th century, as *custodes pacis*, "keepers of the peace") all took work and duties from the sheriffs. The office was not included in the Tudor reorganization of local government, and since that time has been largely ceremonial. English law was consolidated in the Sheriffs Act, 1877; local variations disappeared in a unified set of duties. The sheriff attends on the judges at assizes and election petitions and is responsible for the execution of writs; he prepares the panel of jurors for assizes, is liable for the safe custody of prisoners, and acts as returning officer at parliamentary elections. Until the Murder (Abolition of Death Penalty) Act, 1965, he was responsible for the execution of sentences of death. The sheriff still holds his county court for a few limited purposes, but this court is quite distinct from the modern County Courts established by statute in 1934. His administrative duties have been reduced also by his subordinacy in the county to the Lord Lieutenant, the principal executive officer. Sheriffs are usually nominated by government for the counties, and by corporations for cities and towns which formerly gained the right by royal charter to have sheriffs of their own.

Scotland.—The sheriff, as a ministerial and judicial officer, appears in records from the 12th century, and, despite the later introduction of the office of justice of the peace, the sheriff has retained considerable and greater powers. Sheriffs and sheriff substitutes for the 15 sheriffdoms are appointed and removed by the crown on recommendation or order of the secretary of state for Scotland. Grounds for removal are personal unfitness. A sheriff must be an advocate or sheriff substitute of five years' standing; a sheriff substitute, who is a judicial officer salaried by the crown, must be an advocate or solicitor of similar standing. The appointment, tenure of office, and judicial duties of the sheriff and substitute are fixed by the Sheriff Courts (Scotland) Acts, 1876, 1907, and 1913. The original jurisdiction of the Sheriff Court is largely exercised by sheriff substitutes, and in certain civil matters appeal lies to the sheriff, and through him to the Court of Session. The civil jurisdiction of the Sheriff Court extends to nearly all actions, and the Small Debt Court is also presided over by the sheriff or sheriff substitute. Only grave crimes are excepted from the criminal jurisdiction of the Sheriff Court, these being tried by the High Court. Honorary sheriff substitutes are appointees of sheriffs; the sheriff of Chancery is appointed by the Crown.

Northern Ireland.—The sheriff and salaried under-sheriff for a county are appointed by the governor, and their duties, similar to those in England, are set out in the Sheriffs (Ireland) Act, 1920.

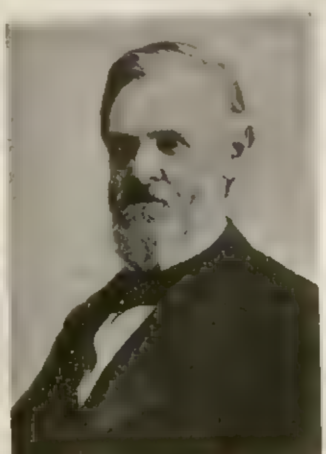
United States.—The sheriff ordinarily is an elected public officer in his county, the chief executive officer and an officer of the court, with a term usually from two to four years. The deputy is appointed by the sheriff and is delegated duties. The sheriff and deputy are peace officers, and thus have the power of police officers in the enforcement of the criminal law. They may assume the functions of the local police department, but carry personal responsibility. The city marshal plays a similar role in his city or municipality, but is generally regarded as chief of police. As a means of preserving public order the sheriff is empowered to call out the *posse comitatus* (*q.v.*). The sheriff's judicial duties principally concern the service and return of process, and the execution of writs, particularly judgments by sale or distress. Though usually a salaried official, he is sometimes paid from fees, which, in populous counties, can be lucrative.

See W. S. Holdsworth, *A History of English Law*, vol. 1, 7th ed. (1956); P. W. Wager (ed.), *County Government Across the Nation* (1950). (D. W. M. W.)

SHERIFFMUIR, a battlefield near the western end of the Ochil Hills, in the parish of Dunblane, Perth County, Scot., and 2½ mi. (4 km.) E by N of the town. It was the site of a drawn battle (Nov. 13, 1715) between the Jacobites, about 12,000 strong, under John Erskine, earl of Mar, and 4,000 Hanoverians, under John Campbell, 2nd duke of Argyll. About 500 men were lost on each side. Both armies claimed the victory, but Mar felt it prudent to withdraw, and the encounter, occurring one day before the defeat of the Jacobites at Preston in England, was fatal to the Jacobite cause. The "Gathering Stone of the Clans," upon which

the Jacobites are supposed to have sharpened their dirks before the battle, can still be seen. (G. S. P.)

SHERMAN, JOHN (1823–1900), U.S. statesman, financial administrator, and author of major legislation concerning the currency and regulation of commerce. He was born at Lancaster, O., on May 10, 1823, a younger brother of William Tecumseh Sherman (*q.v.*), who was to win fame as a general in the American Civil War. The son of a judge of the state supreme court, he studied law with an uncle and a brother at Mansfield, O., and was admitted to the bar in 1844. Law was his major concern for ten years, although he also took an interest in politics and was a delegate to two Whig national conventions. After the repeal of the Missouri Compromise by the Kansas-Nebraska Act of 1854, he joined the Anti-Nebraska movement against the extension of slavery and was elected to Congress in the autumn. He won early distinction in the House of Representatives by drafting the majority report of a committee sent to investigate troubles in Kansas in 1856 (*see* KANSAS: *History*). In 1859 he was the Republican choice for the speakership but fell short of a majority in the badly divided House. Elected to the Senate in 1861, he served in that body until 1897, except for the years 1877–81 when he was President Hayes's secretary of the treasury.



THE GRANGER COLLECTION

SHERMAN

In both House and Senate his major interest was in financial legislation, and after 1867 he was chairman of the Senate finance committee. He was associated with the enactment of important war and postwar fiscal measures. For a time he favoured payment of government bonds in Civil War greenback currency, a concession to western inflationist sentiment, but in 1875 he was largely responsible for the Specie Resumption Act that provided for redeeming them in coin at face value. In 1879, as secretary of the treasury, he carried out the policy successfully. He opposed free coinage of silver, but accepted the limited coinage authorized by the Bland-Allison Act of 1878.

During the Harrison administration the Anti-Trust Act of 1890 and the Silver Purchase Act of the same year bore his name, but both represented compromises which had only his qualified approval. (*See* UNITED STATES [OF AMERICA]: *History*.)

Sherman's name was presented to three Republican national conventions—1880, 1884, and 1888—but the presidential nomination eluded him. His middle-of-the-road course on monetary policies suited neither the inflationist West nor the conservative East; his sober, colourless personality attracted no devoted following; and his candidacy lacked whole-hearted support in his state.

During the last years of his senatorial career he was chairman of the Senate committee on foreign affairs. Upon the accession of President McKinley he accepted the position of secretary of state, thus creating a senatorial vacancy for Marcus A. Hanna, McKinley's campaign manager. Difficulties with Spain over Cuba made his new duties too exacting for his strength, and in April 1898 he resigned. He died at Washington, D.C., on Oct. 22, 1900.

BIBLIOGRAPHY.—T. E. Burton, *John Sherman* (1906); W. S. Kerr, *John Sherman, His Life and Public Services* (2 vol., 1908); *John Sherman's Recollection of Forty Years in the House, Senate and Cabinet*, 2 vol. (1895); Rachel Sherman Thorndike, ed., *The Sherman Letters* (correspondence between John Sherman and William Tecumseh Sherman from 1837 to 1891; 1894). (E. H. Ro.)

SHERMAN, ROGER (1721–1793), American political leader of the Revolutionary era and signer of the Declaration of Independence. He was born at Newton, Mass., April 19, 1721, and grew up in Stoughton, Mass. In 1743 he moved to New Milford, Conn., where he combined shoemaking with farming and studying until he was appointed to the lucrative post of county surveyor. After being admitted to the bar in 1754 he practised law, engaged in trade, and wrote an almanac. Moving to New Haven, Conn., in 1761, Sherman established a general store and took an active part in the affairs of his church, town, and colony. He was treasurer of Yale College, 1765–76; a member of the governor's council, 1766–85; and mayor of New Haven from 1784 to 1793. Though Sherman's stand on questions ranging from paper money to religion was consistently conservative, during the dispute with England he was among the first to realize the impossibility of reconciling the British and American points of view.

As a delegate to the Continental Congress, Sherman signed the Non-Importation Articles of Association in 1774 and the Declaration of Independence in 1776. He helped draft the Articles of Confederation and approved the peace treaty with England. Perhaps his greatest service was performed during the Constitutional Convention of 1787 when he helped to bring about a compromise (often called the Connecticut Compromise) between the large and small states and thus prevented a deadlock. (*See* CONSTITUTIONAL CONVENTION [U.S.].) In Congress under the new Constitution, first as a representative (1789–91) and then as a senator (1791–93), he supported Alexander Hamilton's program for assuming state debts, establishing a national bank, and enacting a protective tariff. Sherman's rugged homespun qualities served as a symbol of patriotic dedication and puritan simplicity to his own and later generations. He died in New Haven on July 23, 1793.

See Roger S. Boardman, *Roger Sherman, Signer and Statesman* (1938); Lewis H. Boutell, *Life of Roger Sherman* (1896). (H. I. BE.)

SHERMAN, WILLIAM TECUMSEH (1820–1891), one of the ablest Union generals in the American Civil War, leader of the famous march through Georgia in 1864, was born on Feb. 8, 1820, at Lancaster, O. When his father, a judge of the supreme court of Ohio, died in 1829, leaving a widow with a family of 11 children, Thomas Ewing, a prominent Ohio politician and close friend of the father, took William into his home. When the boy came of age he entered West Point and graduated near the head of his class in 1840. His first field service was in Florida against the Seminole Indians. Thereafter, amid the usual changes of station and detached duty, he studied law. When the war with Mexico began in 1846 he asked for field duty and was sent to join an expedition going to California by sea. He became executive officer in the administration of local government until California became a part of the United States in 1848. In 1850 he married Ellen Ewing, daughter of his adoptive father, who was then serving as the first secretary of the interior in the president's cabinet.

Three years later Sherman resigned from the army and, attracted by the possibilities he had observed in California, went to San Francisco to manage a branch of a St. Louis bank. The panic of 1857 interrupted a promising career in business. He then engaged in business in New York for a brief time and practised law in Leavenworth, Kan., before becoming superintendent of a newly established Louisiana military academy, which opened its doors on Jan 1, 1860. When the southern states seceded Sherman resigned his post and returned to St. Louis. His devotion to the Union was strong but he was greatly distressed at what he considered an unnecessary conflict between the states.

He offered his services to the U.S. army and on May 4, 1861 was appointed colonel of a new infantry regiment. He was soon assigned to command a brigade in Brig. Gen. Irvin McDowell's army and served with it in the disastrous first battle of Bull Run. Promoted to the rank of brigadier general of volunteers, he was ordered to Kentucky to serve as second-in-command to Gen. Robert Anderson, who had commanded Fort Sumter when the first shots of the war were fired. In October Sherman succeeded to the command of the department of the Cumberland but soon lost it when he reported to his superiors that he needed 200,000 men for the Kentucky campaign. At this point in his career his eccentric behaviour and suspicion of news reporters led some newspapers to describe him as insane. He was nevertheless placed at the head of a division of new recruits and accompanied Grant's army to Pittsburg Landing. At the nearby battle of

Shiloh his raw troops received the first shock of the Confederate assault. His conduct won him promotion to the rank of major general.

Sherman took part in the succeeding attack on Corinth, Miss. At the close of 1862 he led the Mississippi column in the first Vicksburg campaign. Though he suffered defeat at Chickasaw Bluffs, his capture of Fort Hindman, near Arkansas Post, served to restore his reputation. In Grant's final Vicksburg campaign Sherman commanded the right of the line.

After the capture of Vicksburg on July 4, 1863, Sherman was sent to oppose Gen. *Joseph E. Johnston* near Jackson, Miss. (For convenience in identification, Confederate leaders are italicized.) His star, along with Grant's, was now in the ascendant. Their careers would thereafter be closely linked as they worked together to bring about the end of the conflict. When Grant was placed in supreme command of the west, Sherman succeeded to the command of the Army of the Tennessee and in that capacity he took part with Grant in the battle of Chattanooga in November. In March 1864, when Grant became general-in-chief of the Union armies, Sherman was made commander of the military division of the Mississippi, including the Army of the Tennessee, then under Gen. J. B. McPherson, the Army of the Cumberland under Gen. G. H. Thomas, and the Army of the Ohio under Gen. J. M. Schofield. After detaching some troops for garrisons and minor operations in a theatre of war over 500 mi. wide, he assembled near Chattanooga his three armies, aggregating 100,000 men, and began in May the invasion of Georgia. *Johnston* retreated slowly and strategically ahead of him. Outside Atlanta, *Johnston* was replaced by *Hood*, who gave battle but was forced back into the city and compelled on Sept. 1 to evacuate and leave the field to Sherman. The war was not going well in the east and this much-needed victory restored northern morale and helped ensure the reelection of Lincoln in November.

CULVER PICTURES, INC

SHERMAN

Hood in early October began a vigorous movement designed to carry the war back into Tennessee. He moved across Alabama to northern Mississippi but Sherman refused to be diverted. Union superiority in manpower was beginning to tell. Sherman could leave Thomas and Schofield behind to deal with *Hood* while he went ahead with his celebrated "March to the Sea" from Atlanta to Savannah with 60,000 men. This was total war with a vengeance. Cutting loose from his supply base, he set out for Savannah, living off the country, destroying railroads and supplies, reducing the war-making potential of the Confederacy and bringing the war home to the people themselves.

He reached Savannah in time for Christmas. In Jan. 1865, once more abandoning his base, Sherman marched northward toward Virginia where Grant and *Lee* were approaching a final showdown. *Johnston*, recalled to raise an army to stop him, could not muster a sufficient force, and Sherman was in Raleigh, the capital of North Carolina, by April 13. During this march Columbia, the capital of South Carolina, had burned. Though Sherman denied responsibility for it, Carolinians charged him with it. His army had destroyed much property, had burned some homes and had done considerable plundering along the way, making their commander the symbol to southerners of the ruthless invader. The war was now drawing to a close. Thomas and Schofield had utterly destroyed *Hood's* army and Schofield had joined Sherman in North Carolina. *Lee's* position in Virginia was increasingly desperate, and Sherman with 90,000 men was steadily pushing *Johnston* ahead of him as he moved toward *Lee* from the south. A few days after *Lee* surrendered *Johnston* also gave up the struggle. Sherman gave him even more generous terms of surrender than Grant had given *Lee* but was compelled to recall those parts dealing with the political settlement to follow the end of the conflict.

Sherman remained a soldier to the end, though his view of warfare was succinctly put in his oft-quoted assertion that "War is hell." A century later, in the nuclear age, it is tragically amusing to read that Sherman believed that another war could not last four years because the new weapons of destruction, such as semiautomatic rifles and quick-firing artillery, would leave nobody to fight that long. Unlike Grant, Sherman refused to become involved in politics, saying he would not run if nominated and would not serve if elected. At the end of the war he went first to St. Louis and continued his interest in the growing west by working with the army to pacify Indians and build the transcontinental railroad. When Grant became a full general in 1866 Sherman moved up to the rank of lieutenant general. When Grant became president in 1869 he made his old comrade-in-arms commanding general of the army, a post he held until Nov. 1883. One of his lasting achievements was the establishment of a centre for military training at Fort Leavenworth, Kan.

Early in the year 1884 Sherman retired from active duty, living for a time in St. Louis and then in New York city. He lived until Feb. 14, 1891, a well-known and popular, if sometimes controversial, figure to the end. He published two volumes of memoirs in 1875.

See AMERICAN CIVIL WAR; *see* also references under "Sherman, William Tecumseh" in the Index.

BIBLIOGRAPHY.—Lloyd Lewis, *Sherman, Fighting Prophet* (1932); B. H. Liddell Hart, *Sherman: Soldier, Realist, American* (1929); E. S. Miers, *The General Who Marched to Hell* (1951); R. G. Athearn, *William Tecumseh Sherman and the Settlement of the West* (1956).

(C. W. TE.)

SHERRINGTON, SIR CHARLES SCOTT (1857–1952), English neurologist noted for his studies of reflex action, in 1932 received (with E. D. Adrian, *q.v.*) the Nobel Prize for Medicine for discoveries regarding the function of neurons. He was born in London on Nov. 27, 1857. He was educated at Gonville and Caius College, Cambridge, receiving his B.A. (1883) with a first-class honours in natural science; and at St. Thomas Hospital Medical School, London, (M.B., Cambridge, 1885). After conducting research on epidemic cholera in Spain and Italy, he was a fellow of Caius College and lecturer at St. Thomas Hospital (1887–93). In 1891 Sherrington became Brown professor of Pathology, London, and superintendent of the Brown Animal Sanatory Institution; in 1895 Holt professor of physiology, Liverpool; and was Waynflete professor of physiology, Oxford, 1913–35, retiring in 1936. He also held the Fullerian chair of physiology at the Royal Institution, 1914–17. Made a fellow of the Royal Society in 1893, he served as president, 1920–25. He was knighted in 1922.

While at Caius College, he investigated the anatomy of nerve paths between the brain and the spinal cord. Later he turned to the physiology of the spinal cord after isolation from the brain and proved the presence of sense organs in skeletal muscles, thus ending the ancient controversy about "muscle sense." Next, he showed how complex was spinal reflex action in the scratch reflex. He suggested that the lability of reflex nerve centres came from physicochemical states in the points of contact between nerve cells, which points of adjunction he called synapses. His Silliman lectures at Yale in 1904, published as *The Integrative Action of the Nervous System* (1906; reprinted 1947), became a standard work and established his worldwide reputation, marking out new fields for physiological research. (*See* REFLEX: *Development of Reflex Theory*; FEELING, PSYCHOLOGY OF: *Theory of Feeling: Sense Organs Involved.*)

Sir Charles died at Eastbourne, Sussex, on March 4, 1952.

(E. G. T. L.; X.)

's HERTOGENBOSCH, capital of North Brabant Province, Neth., lies at the confluence of the Dommel and Aa rivers, which unite to form the Dieze, and 29 mi. (47 km.) SSE of Utrecht. Pop. (1960) 71,597. Founded by Henry I, duke of Brabant, the town was granted city rights in 1185 and remained an important fortress until 1874. St. John's Cathedral, with a well-preserved

interior, is one of the finest in the country, and the Baroque town hall has a carillon of 30 bells. Opposite the town hall stands the statue of the painter Hieronymus Bosch (*q.v.*), who was born there. The covered cattle market (1933) is the largest in Western Europe. An international vocalist competition and a summer course for vocalists are held annually. The town is a rail and road junction, with links to all parts of the Netherlands and to Belgium and Germany.

SHERWOOD, ROBERT EMMET (1896–1955). U.S. author whose plays reflect the gradual shift of the intellectual position of his generation—from irresponsible detachment to responsible involvement in human problems. He was born in New Rochelle, N.Y., April 4, 1896. He was an indifferent student at Milton Academy and Harvard, failing the freshman rhetoric course while performing well and happily on the *Lampoon*, the humour magazine, and Hasty Pudding, the college musical comedy show. He left before graduation to enlist in 1917 in the Canadian Black Watch Battalion, served in France, was gassed, and was discharged in 1919. He was drama editor of *Vanity Fair* (1919–20) and with his colleagues there, Dorothy Parker and Robert Benchley, found his way to the Algonquin Round Table, the centre of a New York literary coterie that included Harold Ross, Alexander Woollcott, Heywood Broun, Franklin P. Adams, Frank Sullivan, Marc Connelly, George S. Kaufman, and Donald Ogden Stewart. Sherwood was associate editor (1920–24) and editor (1924–28) of the humour magazine *Life*. His first play, the urbanely witty historical *The Road to Rome* (1927), criticizes war as pointless. The heroes of *The Petrified Forest* (1935) and *Idiot's Delight* (Pulitzer Prize, 1936) begin as detached cynics but realize their own bankruptcy and sacrifice themselves for their fellowmen. In *Abe Lincoln in Illinois* (Pulitzer Prize, 1939) and *There Shall Be No Night* (Pulitzer Prize, 1941), in which his pacifist heroes decide to fight, Sherwood shows that only by losing his life for other men can a man make his own life significant. In 1938 Sherwood formed with Maxwell Anderson, Sidney Howard, Elmer Rice, and S. N. Behrman the Playwrights' Company, which became a major producing company.


© KARSH, OTTAWA
SHERWOOD

The Lincoln play led to Sherwood's introduction to Eleanor Roosevelt and ultimately to Pres. Franklin Delano Roosevelt, for whom he served as speechwriter and adviser. Sherwood's speechwriting did much to make ghostwriting for public figures an accepted practice. Between service as special assistant to the secretary of war (1940) and secretary of the Navy (1945), Sherwood served as director of the overseas branch of the Office of War Information (1941–44). From his wartime association with Roosevelt came much of the material for *Roosevelt and Hopkins: an Intimate History*, for which he received his fourth Pulitzer Prize in 1949. Except for his Academy-award-winning film *The Best Years of Our Lives* (1946), Sherwood's theatrical work after World War II was negligible. He married Mary Brandon in 1922, and, after divorcing her in 1934, married Madeline Hurlock in 1935. Sherwood died in New York, Nov. 14, 1955.

See John Mason Brown, *The Worlds of Robert E. Sherwood* (1965). (Js. T. N.; X.)

SHERWOOD FOREST, one of the ancient English forests, in Nottinghamshire. It extended from Nottingham northward to Worksop, being over 20 mi. (32 km.) long by 5 to 9 mi. (8 to 14 km.) broad. Though subject to forest law, it is not mentioned in Domesday Book because it was not liable for tax. Sherwood, a crown forest from the time of Henry I, is traditionally noted as the retreat of the outlaw Robin Hood. It once contained vast numbers of deer. Today about 12,500 ac. (5,059 ha.) are under national ownership, administered by the Forestry Commission as productive plantations, mostly of pines; many veteran oaks remain and there is some birch and beech, and much bracken. The forest is on the Bunter sandstone and much of the ground is too poor for agriculture, but in the 18th century great parks, with extensive woodlands, were laid out to form the "Dukeries." Deep-lying coal seams have been extensively developed and colliery settlements built. Mansfield (*q.v.*) is the main town. (H. L. En.)

SHESHONK (SHISHAK, SHOSHENQ), pharaoh of Egypt, in the 22nd dynasty, from *c.* 945 to 924 B.C. In the Old Testament he is called Shishak. He came from a line of princes or sheikhs of Libyan tribal descent, whose title was "great chief of the Meshwesh" and who appear to have settled in Heracleopolis, in middle Egypt, though Manetho says the family came from Bubastis in the eastern delta. Sheshonk appears to have ascended the throne without a struggle, and he married his son Osorkon to a daughter of Psusennes II, the last king of the foregoing dynasty.

The Bubastite portal at Karnak (*see* THEBES) bears a relief and inscriptions celebrating this king's victories in Palestine. This was the occasion (*c.* 930 B.C.) when "Shishak king of Egypt came up against Jerusalem" (I Kings 14:25–26) in support of Jeroboam, the pretender, who challenged the right of Solomon's son Rehoboam to succeed to the Israelite throne. The biblical account speaks of the looting of the palace and Temple, though the name Jerusalem has not survived in the Egyptian record. A fragment bearing Sheshonk's name has, however, been found at Megiddo.

There were three other pharaohs of the same name later in the dynasty. *See* also EGYPT: *History*. (M. S. Dr.)

SHETLAND, or ZETLAND, as it is spelled officially, is a group of about 100 islands, islets, and skerries (fewer than 20 of them inhabited) lying about 130 mi. (209 km.) to the north of the Scottish mainland, constituting the most northerly county of Scotland. The main island, Mainland, is much the largest, being, though narrow, about 50 mi. (80 km.) long. The total area of land in the county is 550.5 sq.mi. (1,426 sq.km.). Lerwick, the county town, lies about a third of the way up Mainland on the east side.

Physical Features.—The scenery of Shetland is wild and in places very fine. Its most distinct characteristics are perhaps the voes or sea lochs which indent all the coasts, often enclosed by steep hills and giving from the higher ground fine views over sea and coast. Though subject to fierce storms in fine weather (especially in summer when the light lasts all night) sea and land take on intense and sometimes rather unearthly colourings. There are few trees. The winds are so continuous and strong that even vegetables require shelter. As the main form of agriculture is crofting, each croft having a few acres of arable land and the right to keep so many sheep on the *scattald* or common grazings, the lower ground bears the typical aspect of crofting districts. The croft houses are usually low cottages and the arable ground lies around them giving way up the hillsides to rock and heather. There are streams and some freshwater lochs but no rivers.

The common seal is abundant in the Shetland waters, especially between Lerwick and the north isles. Brilliant red and purple sea urchins occur along the shores, due to the presence of the Atlantic drift.

The geological character of the islands resembles that of northern Scotland. Old Red Sandstone, red grits, sandstones, and marls and conglomerate occur in a narrow belt on the east side of Mainland. The remainder of the island is occupied by metamorphic schists and gneisses with which are associated dikes and masses of intrusive igneous rock.

Mainland is almost cut in half just north of Brae. The northern portion, Northmaven, is sparsely inhabited. It contains Ronas Hill, the highest point in the islands (1,475 ft. [450 m.]). The western parishes, stretching out to Walls and Sandness, are also thinly populated. Tingwall, the site of the ancient "Ting" or Norse parliament, lying directly north of Lerwick, is the most fertile part of the island and contains one or two sizable farms. South of Lerwick in Dunrossness there is some good land, and the districts are more thickly populated.

To the north of Mainland are Yell, the next in size, and Unst, the most northerly and populous of the islands. Offshore and 1 mi. NW of the entrance to Burra Firth is Muckle Flugga Lighthouse, the most northerly point of the British Isles. Ponies are bred on Fetlar, northeast of Mainland, and Whalsay, lying farther south, is inhabited by fishermen.

Off the western seaboard are six inhabited islands (including Muckle Roe, now joined to Mainland by a bridge). Among these Papa Stour has fine caves. Foula lies in the Atlantic, 16 mi. (26 km.) from the west mainland and is rocky and exposed. Fair Isle, 20 mi. (32 km.) S of Sumburgh Head, the southernmost point of Mainland, belongs to the National Trust for Scotland and has two lighthouses, and an ornithological observatory under the Fair Isle Trust. Noss has been a nature reserve since 1955.

History.—Zetland or Shetland is a modernized version of the old Norse name *Hjaltland*. The original inhabitants were probably a primitive people who lived in rough, sunken dwellings. Various stone circles survive as well as brochs (circular stone towers) of Pictish origin. In the 7th or 8th century the islanders began to be converted to Christianity by missionaries of the Celtic Church from Ireland or the west of Scotland, although the work was not completed until later. In the 8th and 9th centuries the islands were invaded by the Norsemen who ruled them until the 15th century. The Norsemen left a heavy imprint on the people and the local place-names, most of which are Norse in origin. The language, Norn, survived down to the 18th century, and various sea birds and parts of boats are generally known by their Norse names. In 1469 the islands, like Orkney in 1468, were given as a pledge for the dowry of Margaret, princess of Denmark, who married (1469) James III of Scotland. The pledge was never redeemed, and both groups were annexed to the Scottish crown in 1472. Large tracts of land were granted to Scottish nobles, and from the 16th century began a steady influx of Scots which, since many of them came as oppressive lairds, was by no means always popular with the native Shetlanders. For several centuries the islands have been famous for wool and knitwear, and fish. The origins of the "Fair Isle" patterns are disputed. They are probably native, although it has been suggested that they came from the crew of a Spanish Armada galleon which was wrecked on the Fair Isle. Development of herring fishing, which at one time supported stations in almost every voe and packed Bressay Sound and Baltasound with ships, was largely due to the Dutch in the 18th century; but the Shetlanders have always been great seamen. Many local customs, mostly of Norse origin, survive, such as Up Helly A', the winter festival (*see* LERWICK). The islands, however, on the whole have stood outside the mainstream of Scottish history; the kilt has never been a native dress nor Gaelic a native language. In fact, they have no share in the Celtic traditions of the Highlands. There are many dwellings and burial places dating from Pictish or Norse days, particularly Jarlshof at Sumburgh and the broch on Mousa. Shetland is the scene of *The Pirate* by Sir Walter Scott.

Stimulated by a Viking conference attended by representatives of the major Norse countries and held at Lerwick in 1955, archaeological investigations were begun in the south of the island. In 1957 considerable remains of a Celtic church were discovered. This church is believed to be one of a chain of such churches associated with St. Ninian and, in fact, to have been the mother church of the island.

Some months later a remarkable treasure of bronze ornaments was discovered beneath this church including 12 bronze brooches, six bowls, and a hanging lamp, as well as two unusual carved objects which may also have been used as brooches. These objects are believed to have been placed near the "founder's tomb" of the church. They form a particularly fine and well-preserved collection of Celtic work decorated with semiprecious stones.

Population and Administration.—The total population of the county was 17,812 in 1961, a decrease of 1,540 since 1951. There were 269 people on Bressay (335 in 1951), 127 (161) on Fetlar, 1,148 (1,101) on Unst, 764 (859) on Whalsay, and 1,155 (1,483) on Yell.

Lerwick (*q.v.*), a small burgh, is the capital and main port. Pop. (1961) 5,908 (5,538 in 1951). The only other town is Scalloway, the ancient capital, which lies almost opposite Lerwick on the west coast. It is a considerable fishing centre. It has the ruins of a castle built by Patrick Stewart, one of the most notorious of the Scottish landlords.

Shetland unites with Orkney to return a member to parliament. The island is divided into six districts. It forms a sheriffdom with Orkney and Caithness, and there is a resident sheriff-substitute at Lerwick. The system of udal tenure is still in existence on the islands.

Agriculture and Industries.—The soil of Shetland, except in certain districts, is poor and the climate inimical to many kinds of agriculture. The system of land tenure is mostly crofting and

J. ALLAN CASH FROM RAPHO GUILLUMETTE

(Above) Seagulls following fishing boats at Scalloway, Shetland (Right) Shetland ponies on the island of Unst. (Left) Eroded cliffs at Eshaness, Shetland. (Below) Typical small farm on Weisdale voe, Shetland

the main source of the crofters' income is wool. The Shetland breed of sheep produces a light fleece of very fine wool which is knitted by the crofters' wives and daughters into gloves, pullovers, socks, berets, etc., in the patterns known as "Shetland" or "Fair Isle." Many of the crofts do not yield a sufficient income for a family, however, and Shetland men often seek supplementary means of livelihood, usually at sea by fishing or in the navy, merchant service, or Antarctic whaling expeditions. Efforts were made to break out more land from the hill and increase the number of cattle, the native breed of which gives fair supplies of milk and, though light, the beef is of good quality. However, Aberdeen Angus and crossbred cattle have to some extent taken their place. Shetland ponies which once did all the work on the crofts are found in considerable numbers, especially on Unst. But demand for both croft and farm work and for the mines has much diminished, and the main market is now for childrens' ponies.

Lerwick is an important herring port; boats arrive there from the east coast of Scotland and England to join the local boats for the summer fishing. At Scalloway, too, much herring is landed early in the season. There is a fish-meal plant on Bressay and a quick-freezing factory at Lerwick, as well as curing facilities. Most of the herring stations in the countryside and the outlying islands have, however, been closed. and the number of boats fishing from Lerwick has decreased. All the year round white fish of all kinds, principally haddock and whiting, are caught by seine nets. The main fishing centres are Whalsay, Burra Isle, Lerwick, and Scalloway. For lack of adequate piers at Whalsay and Burra Isle most of the boats land at Lerwick or Scalloway. White fishing is also carried on elsewhere, notably from Out Skerries, Cullivoe, and Voe. Many crofters keep small boats from which they catch some haddock, whiting, or piltock for their own use. Fishing boats of other nations frequently call at Shetland harbours, the Norwegian dogfish and shark boats being a common sight. For many years the Swedes voyaged every summer to Baltasound to fish ling, but this decreased. Whales are occasionally seen around the Shetland coast, but whaling locally has ceased and the caaing whale is no longer driven ashore as it once was for food and lamp oil.

Apart from agriculture, wool, knitting, and fishing, the main native source of livelihood is a small amount of weaving. The stone in Unst is also quarried for talc. From time to time, efforts were made to start the working of peat and iron ore, but without success. Unemployment in the 20th century has been high.

Communications.—The north isles and Whalsay can be reached by regular ferry services and by steamer from Lerwick. The remainder of the islands are served by mail boats. There is an airport at Sumburgh from which there are daily services to Orkney, Inverness, and Aberdeen, and there are regular sailings between Lerwick and Orkney and Aberdeen.

BIBLIOGRAPHY.—A. C. O'Dell, *Zetland*, Land of Britain series no. 5 (1940); A. T. Cluness, *The Shetland Isles* (1951); J. Peterson, *Shetland: a Photographer's Notebook* (1948); P. A. Jamieson, *Letters on Shetland* (1949); U. Venables, *Tempestuous Eden: The Story of a Naturalist in the Shetlands* (1952); L. S. V. and U. M. Venables, *Birds and Mammals of Shetland* (1955). (J. GD.)

SHEVKET, MAHMUD (MAHMUD SHEVKET PASHA) (1858–1913), grand vizier of Ottoman Turkey in 1913, was born in Baghdad in 1858, the son of a junior administrator (*mutisarrif*). After attending a primary school in Baghdad he went to the Cadet School in Constantinople and graduated as a staff captain in 1882. He then served on the general staff of the Ministry of War and was posted to Crete for a short time before returning to Constantinople to teach at the Cadet School. In 1886 he joined the commission which was sent to Germany to supervise the manufacture of war materials for Turkey. On his return he was appointed director of the Inspection and Control Department, with the rank of general.

After the Young Turk Revolution of 1908 Mahmud Shevket became commander of the 3rd Army which marched on Istanbul to suppress the counterrevolution of April 13, 1909. He supervised the suppression of the uprising and forced Sultan Abdul-Hamid II to abdicate. He then became inspector of the 1st, 2nd, and 3rd armies and minister of war. On Jan. 23, 1913, during the first Balkan War, some members of the Committee of Union and Progress marched on the Sublime Porte, killing the war minister, Nazim Pasha, and forcing Kiamil Pasha to resign. Mahmud Shevket Pasha was then made grand vizier. He preferred soldiering to politics but served his country well as grand vizier during his tenure of office. He was assassinated in Istanbul on June 11, 1913. He wrote several books on mathematics and military science, and a history of the Ottoman Army. *See* also TURKEY: *History*. (M. P. P.)

SHEWBREAD, in the Old Testament, an archaic English form of SHOWBREAD, to indicate that it is bread on display; also known as BREAD OF THE PRESENCE, for it is presented and shown in the Temple, in the presence of God.

The bread consisted of 12 loaves, arranged on a table in two rows of 6 each (Lev. 24:5 f.). This arrangement of the loaves was an important aspect of the presentation; I Chron. 9:32 and 23:29, and Neh. 10:33 literally speak of "the bread of the arrangement." The table, which stood at the west end of the "holy place" of the Temple, next to the Holy of Holies, was also important. In Ex. 25:23–30 is a description of its construction, and it is called "the table of the bread of the Presence" (Num. 4:7). The bread was changed every sabbath and the priests ate that which had been on display; though in emergency it was given to David and his men (I Sam. 21:6).

The 12 loaves stood for the 12 tribes of Israel. They were a symbolic acknowledgment that God was the resource for Israel's life and nourishment; at the same time they served as Israel's act of thanksgiving to God for this. Many aspects of the Christian Eucharist show influence of Israel's showbread. (J. C. RY.)

SHIBARGHAN, a town and capital of the minor province of Shibarghan in northern Afghanistan, lies 80 mi. (129 km.) W of Mazar-i-Sharif. Pop. (1962 est.) 50,264. It was formerly the capital of a small independent Uzbek khanate, which was allotted to the Afghans by the Anglo-Russian boundary agreement of 1873. It has a citadel but is not otherwise fortified, and is surrounded by gardens. It is on the main road from Kabul to Herat via Mazar-i-Sharif.

SHIBARGHAN PROVINCE (pop. 1962 est. 395,560) stretches northward over the desert plains of Afghan Turkistan to the Amu Darya (Oxus) and the Soviet frontier, and is bounded east by Mazar-i-Sharif and west by Maimana. In the south and southwest it extends to the provinces of Herat and Kabul and includes part of the northern foothills of the Band-i-Turkistan range. The province is drained by the Ab-i-Safed River and its chief crops are cotton, maize (corn), and melons. Carpet weaving, sheep grazing, and the preparation of karakul lambskins are the main occupations; the province produces the highest quality karakul. Oil has been found at Sar-i-Pul (south) and around Aq Chah (north). (J. P. C. N. H.)

SHIDEHARA KIJŪRŌ, BARON (1872–1951), Japanese diplomat and statesman who opposed the growth of militarism in his country, was born in Osaka on Aug. 11, 1872. A graduate of the Law College of the Tokyo Imperial University in 1895, he entered government service, serving briefly in the Agriculture and Commerce Department. Then he entered the diplomatic service serving as consul at Chemulpo, a Korean Japanese treaty port (1899), at London (1899), Antwerp, and other cities. He was secretary at the Head Office in 1911, Counsellor to the embassy at Washington (1912) and at London (1914), and minister to the Netherlands (1914–15). In 1915 he was vice-foreign minister and ambassador to the United States in 1919. In that capacity he argued in vain against U.S. immigration laws that discriminated against the Japanese. He was created a baron in 1920 and was a member of the House of Peers in 1925. As the chief Japanese delegate to the Washington Conference (1921–22) he accepted agreements favouring China, and as Japan's foreign minister from 1924 to 1927, and again from 1929 to 1931 he became known as an advocate of conciliation toward China and of peaceful policies generally. Forced from office by the militarists in 1931, he did not again play a significant role until October 1945 when, at the age of 73, he was accepted by the military occupation authorities as prime minister during the demilitarization period (until May 1946). He was then elected as a conservative to the lower house of the Diet where he served as speaker until his death on March 10, 1951. The

paradox of his liberalism abroad and conservatism at home may in part be explained by his long association with Mitsubishi financial interests. His wife, Iwasaki Masako, was the daughter of the founder of that *zaibatsu* combine. His book *Gaikō Gojūnen* (1951) describes his public career. (F. H. Co.; X.)

SHIEL, LOCH, a glacially excavated ribbon lake in the northwest Highlands of Scotland draining to a bay on the southern shore of Loch Moidart (a sea-loch) by the 31-mi.-long (50 km.) River Shiel. The loch is 17 mi. (27 km.) long, from 200 yd. to 1 mi. (1.6 km.) wide, and only 11½ ft. (3.5 m.) above sea level, with a maximum depth of 420 ft. (128 m.) (mean depth 81½ ft [24.8 m.]). The loch acts as a boundary between Moidart in Inverness and Sunart and Ardgour in Argyll. The lower 5 mi. (8 km.) runs east-west among gentle hills rising to about 600 ft. (183 m.); but Glen Finnan and the upper loch run fairly straight northeast-southwest among fine-conical glaciated hills rising to about 3,000 ft. The area around the head of the loch and the mouth of Glen Finnan has roads running east to Loch Eil and Fort William and west to Loch Ailort, Morar, Mallaig, and thence to Skye. A monument commemorates the raising of Prince Charles Edward's standard in 1745. On St. Finnan's isle are an ancient chapel and the traditional burial place of the Macdonalds. (A. T. A. L.)

SHIFNAL, a market town (formerly known as Idsall) of Shropshire, Eng., lies 17 mi. (27 km.) ESE of Shrewsbury by road. Pop. (1961) 3,896. The 12th-century church of St. Andrew is cruciform, combining examples of every period from late Norman to late Perpendicular. The town was destroyed by fire in 1591 and now consists chiefly of half-timbered and brick houses of the 17th and 18th centuries. Bishop Thomas Percy (*q.v.*) reputedly found the manuscript upon which he based his *Reliques of Ancient Poetry* (1765) in an old house in the marketplace. Trade is mainly agricultural, with some engineering. There are fat and store stock sales every Monday, but the November fairs have been discontinued. (M. C. Hl.)

SHIGA, prefecture (*ken*) northeast of Kyōto in South Honshu, Japan. Area 1,551 sq.mi. (4,017 sq.km.). Pop. (1965) 853,370. Biwa-ko (*q.v.*), the largest lake in Japan, and the surrounding mountainous district form Biwa-ko recreational park and occupy about 25% of the total area of Shiga. The percentage of paddy fields is the highest in Japan. Ōmi cows are famous as beef cattle. The textile industry at Ōtsu (*q.v.*), Hikone, and Nagahama owes its development in part to the chemical qualities of the lake water. Ōtsu, the prefectural capital, is also an important tourist centre. The tomb of Bashō (1644–94), greatest *haiku* poet, is at Shiga. From Ōtsu boats provide tours to the scenic "Eight Views" and other points on Biwa-ko. (R. B. H.)

SHIGA NAOYA (1883–), one of the deans of modern Japanese fiction, whose writings have served as model and inspiration for many younger authors, was born on Feb. 20, 1883, at the village of Ishinomaki in Miyagi prefecture but in 1885 was taken by his parents to Tokyo to live in his grandfather's home. From about 1899 he was drawn to Christianity, but later abandoned it, apparently because he could not accept such doctrines as that of original sin. Literature and art next claimed his allegiance. In 1906 he graduated from the Peers' School and entered the Department of English Literature at Tokyo Imperial University, leaving after two years. As a young writer Shiga developed an objective style, perceptively delineating the most sensitive reactions of his somewhat delicate characters. Spurts of literary activity, which earned him a reputation as a fine short-story writer, were separated by long periods of inactivity and he was unable to rely on writing for his living. In 1910 he joined Mushanokōji Saneatsu and other friends of his Peers' School days in founding the journal *Shirakaba* ("White Birch"), which emphasized individualism and Tolstoyan humanitarianism.

Much of Shiga's fiction is concerned with difficult family relationships, the result no doubt of his own problems with his family, especially with his father. Both the novelette *Wakai* ("Reconciliation"; 1917) and the long novel *An'ya kōro* or *Road Through the Dark Night* (written between 1925 and 1937) describe the hero's search for peace of mind in the face of family involvements. In *Wakai,* the first-person hero distresses his father by marrying the family maid. In *An'ya kōro,* the hero is the son of his apparent father's father; when he marries, his wife in turn contracts a liaison with a cousin, from which a child is born. Shiga's concern with the psychological involvements of his first-person heroes places some of his stories in the category of *shi-shōsetsu* ("I" fiction, private fiction).

Synopses of Shiga's work, with a biographical note, are found in *Introduction to Contemporary Japanese Literature,* part ii (1936–55) compiled by the Kokusai Bunka Shinkōkai (1959); English translations of his short stories are included in *Modern Japanese Literature,* ed. by Donald Keene (1957), *Modern Japanese Stories,* ed. by Ivan Morris (1962), and *The Heart Is Alone,* ed. by Richard N. McKinnon (1957). (J. K. Y.)

SHIH HUANG TI (259–210 B.C.), "First Sovereign Emperor," is the title assumed in 221 B.C. by King Cheng of the Chinese state of Ch'in (in northwest China) when he completed Ch'in's conquest of the other independent states into which China was then divided, thereby creating the first unified Chinese empire. It is from the state and empire of Ch'in that the name China probably derived.

The first emperor was born in 259 B.C., allegedly the natural son of a rich merchant and Ch'in statesman, Lü Pu-wei, though this story is possibly a slanderous invention. He ruled as king of Ch'in from 246 to 221 and thereafter as first emperor of all China until his death in 210. Major innovations instituted in and after 221 included: (1) abolition of the formerly independent Chinese states with their landed aristocracy and replacement of them by an administrative system of 36 (later 42) *chün* or provinces, subdivided into lesser *hsien* or prefectures and all governed by centrally appointed, nonhereditary, salaried officials; (2) simplification and standardization of the Chinese script according to the written characters current in Ch'in; (3) extension of the Ch'in system of laws, weights, and measures throughout the empire; (4) completion of the Great Wall as a barrier between China and its nomadic neighbours to the north and northwest; and (5) the "burning of the books," a governmental proscription of literature ordered in 213 to suppress Confucianism and other dissident schools of thought in favour of Legalism and to make the Ch'in version of past history the accepted one. These innovations, though formally ordered by the first emperor, were largely inspired by his prime minister, Li Ssŭ (*q.v.*).

The first emperor was a mighty conqueror, Machiavellian schemer, bold innovator, and at the same time a superstitious megalomaniac whose fear of death caused him to be searching for the Taoist elixir of immortality when he died in 210. His harsh reign was followed within a year by rebellion, leading to the collapse of the Ch'in dynasty and its replacement by that of Han (206 B.C.–A.D. 220). Though the Confucian historians reviled him as a ruthless tyrant, the first emperor's reign was crucially important for bringing China's feudal age to an end and inaugurating norms of empire that persisted until the founding of the Chinese republic in 1912. *See* also CHINA: *History.*

BIBLIOGRAPHY.—Derk Bodde, *China's First Unifier* (1938), *Statesman, Patriot, and General in Ancient China* (1940); E. Chavannes, *Les Mémoires historiques de Se-ma Ts'ien,* 5 vol. (1895–1905).

(D. Be.)

SHIHKIACHWANG (SHIH-CHIA-CHUANG, SHIH-MEN), an industrial, rail junction-terminal city in west Hopeh Province, China. Pop. (1958 est.) 623,000. In the 1920s it was only a small village near the old city of Chenting close to the Peking-Hankow railway. It lay at the junction of a branch line built to T'ai-yüan, Shansi, in the 1900s. The village became a terminal point and grew into a city of 217,000 by 1935. A rail link built to Te-chou, Shantung, during World War II increased Shihkiachwang's importance. The Chinese Communists, by the 1960s, had developed the city into a modern regional industrial, transport, and trade centre, with an airport, rail yards, cotton mills, and many new agricultural processing plants. (J. E. Sr.)

SHI'ISM, one of the two major branches of the religion of Islam (*q.v.*) distinguished from the majority Sunnism. One of the factions which disputed power over the early Muslim community and its vast conquests supported the claims of 'Ali (*q.v.*),

who was nominated fourth caliph, and then of his descendants. From a political faction (Shi'at 'Ali, "Party of 'Ali") this gradually developed into a religious movement, Shi'ism, which deeply influenced all Sunni Islam and also produced a number of important sects to which, especially, the term Shi'a is applied. In the mid-20th century, perhaps one-tenth of all Muslims were Shi'ites; *i.e.*, about 40,000,000. Shi'ism is the majority faith in the Iranian plateau and Iraq and perhaps Yemen, and is found in Syria, Lebanon, east Arabia, northern India (especially around Lucknow), in the Deccan and Bombay, Pakistan, and elsewhere.

'Ali was a first cousin and son-in-law of Mohammed, father of Mohammed's two grandsons, Hasan and Husain, by the Prophet's daughter Fatima. 'Ali was noted for his devoted piety and his valour in war. Raised to the caliphate (*q.v.*; 656) with the support of the murderers of the third caliph (another son-in-law of Mohammed), he never received the allegiance of all the Muslims but had to wage increasingly unsuccessful civil wars. When he was murdered (661), his chief opponent, Mu'awiya, was generally acknowledged caliph; for some time 'Ali was officially cursed from the pulpits of Islam. But many Muslims, especially at Kufa in Iraq, 'Ali's headquarters, hoped for an 'Alid restoration. On Mu'awiya's death they invited Husain, 'Ali's son, to become caliph. But they failed to support him at the crisis and he and his little band were cut down (680) near Kufa at Karbala, now a pilgrimage spot. (*See* HASAN AND HUSAIN.) The Kufans thereupon bewailed penitently the death of Mohammed's grandson and swore vengeance against the triumphant Islamic government. Repeatedly they supported insurrections by members of 'Ali's family, but without success.

The 'Alid cause soon gained support from other groups which opposed the *status quo*—for instance, from the aristocratic Muslim families of Medina, who eventually established a tradition of 'Alid rule in the holy cities; from pious men protesting against a too worldly interpretation of Islam; and from non-Arab Muslims, especially in Iraq, who demanded an equality refused them by the ruling Arabs. The 'Alids never won power (though it was initial Shi'ite help which set up the Abbasid dynasty [750], descended from another cousin of Mohammed). Yet 'Ali was rehabilitated as a major hero of Sunni Islam, and his descendants by Fatima have received a privileged status as *sayyids* and *sharifs*.

But the Shi'ites were not satisfied with this. Some, called Zaidis (Zaydis), whose principles were worked out in the 9th century, demanded, sword in hand, that the ruler must be whichever descendant of Hasan or Husain proved qualified, at a given time, by his knowledge and his practical ability; otherwise, they differed little from the Sunnis. They set up several small states along the Caspian and in Yemen. Other Shi'ites, called Imamis, asserted a more exalted religious role for the 'Alid claimants. They insisted that, in power or not, a given descendant of 'Ali was the divinely appointed imam, and sole authority in his time on all matters of faith and law. The more speculative among them, called Ghulat ("extremists"), sometimes paid the imams practically divine honours. The more moderate came in time to claim at least that a supernatural "Mohammedan light," embodied in them, gave them superhuman knowledge and power, and that their sufferings were means of divine grace to their devotees. Love of the imams and of their persecuted cause became equally obligatory with belief in God's oneness and in Mohammed's mission. Every year during the month Muharram, when Husain was killed, the Shi'ites mourned his death (sometimes with bloody self-lacerations) and condemned the guilty Sunnis. Under bigoted Sunni rule, they felt they might have to protect themselves from persecution by dissimulating their faith (*taqiyya*); but in the end the imam, as mahdi, the "well-guided," would deliver the faithful and punish their enemies. (*See* IMAM; MAHDI.)

Several sects acknowledged as imams the line of Mohammed ibn al-Hanafiyya, a son of 'Ali (not by Fatima), but these died out in the 9th century. Most Shi'ite sects have acknowledged one of two lines stemming from Husain's great-grandson Ja'far al-Sadiq (d. *c.* 765). One is that of the Isma'ilis, who developed a unique religious system and for a time established a powerful Fatimid caliphate, and are represented by the modern Khojas and Bohras, merchant communities of India and east Africa; from them split the Druze of Syria. (*See* DRUZE; ISMA'ILISM.)

The majority of modern Shi'ites acknowledge a younger line down to a 12th imam, Mohammed al-Muntazar, supposed to have gone into hiding in 878 and expected to return as mahdi before the Last Judgment to establish justice in the earth. Among these are the Nusairis (Nusayris) or 'Alawites of northern Syria, with a secret faith of complex origin; the 'Ali-Ilahis or Ahl-e Haqq, scattered peasants and herdsmen of Kurdistan, Turkey, and Iran; and the order of Bektashi dervishes in Turkey and Albania. But most who acknowledge the 12 imams belong to the Ithna 'Ashariya or "Twelver" sect (in Syria and Lebanon, Mutawali).

Despite occasional Shi'ite rulers, the Shi'a remained almost everywhere a minority faith till at the start of the 16th century the Persian Safavid dynasty made it the sole legal faith of their empire, embracing the Turks of Azerbaijan, the Persians of Iran, and the Arabs of Iraq proper; these have since been all overwhelmingly Twelver Shi'ites and have given that sect a vigorous modern life. Usulis, claiming the primacy of legal principle, have debated with Akhbaris, attached to established texts; with the speculative Shaykhis; and especially with the Babis and Baha'is of the 19th century. At the Shi'ite pilgrimage places, especially Najaf near Kufa in Iraq, scholars qualified to make independent judicial decisions, mujtahids, regarded as representatives of the hidden 12th imam, expound Islamic law according to the school of Ja'far, which differs in some particulars from each of the Sunni schools. They also have a political role as popular authorities, especially in Iran.

See also references under "Shi'ism" in the Index.

BIBLIOGRAPHY.—Dwight M. Donaldson, *The Shi'ite Religion* (1933), on Twelver Shi'ism, uncritical; Edward G. Browne, *Literary History of Persia*, vol. iv (1924); R. Strothmann, *Staatsrecht der Zaiditen* (1912); René Dussaud, *Histoire et religion des Nosairis* (1900); Marshall Hodgson, "How Did the Early Shi'a Become Sectarian," *Journal of the American Oriental Society*, 75:1–13 (1955).

(M. G. S. H.)

SHIKOKU, the smallest of the four main islands of Japan, is separated from Honshu Island by the Inland Sea. Area 6,857 sq.mi. Pop. (1960; Kagawa, Ehime, Tokushima, and Kōchi prefectures) 4,121,423. Most of the population is concentrated in urban areas along the coast, and Takamatsu in the northeast is the major city. *See* JAPAN: *Geographic Regions*.

SHILLONG, the capital of Assam, India, and the headquarters of the United Khasi and Jaintia Hills District, is situated on the Shillong Plateau, about 60 mi. (97 km.) S of Gauhati, at an altitude of 4,987 ft. (1,520 m.) above sea level. Pop. (1961 Town Group, including the cantonment) 102,398. Shillong first came into prominence in 1864 when it was made headquarters of the Khasi and Jaintia Hills District in place of Cherrapunji. Ten years later it became the seat of government of the newly formed province of Assam. In 1954 the administrative headquarters of the North East Frontier Agency were established at Shillong. The town contains, besides the military cantonment, the Pasteur Institute and Research Laboratory, a dairy farm, five colleges connected with Gauhati University, and two large hospitals. Barapani hydroelectric power station, one of the largest in Assam, lies about 12 mi. (19 km.) N of the town. Shillong experiences frequent earth tremors and was devastated by the great earthquake of June 12, 1897. (M. BA.)

SHILLUK, a Nilotic nation, numbering 120,000 in the 1960s, settled along the west bank of the Nile, between Lake No and latitude 12° N in the south of the Republic of the Sudan (*q.v.*). They are sedentary negroid agriculturalists with strong pastoral interests (30,000 cattle, many sheep and goats). After long contacts with northern Sudan, and incorporation into the Ottoman Empire, they were under Anglo-Egyptian administration from 1898 until Sudan independence in 1956. Formal education, once entirely in the hands of Roman Catholic and Protestant missions, began to receive state aid in 1926. The Shilluk are headed by a divine king (*reth*) chosen from the sons of previous kings; his physical and ritual well-being is held to ensure the prosperity of the whole land. The large royal clan traces descent from the first king and culture hero, Nyikang (Nyikango). The

kingdom (capital, Fashoda) has two main provinces that are symbolically united at the king's installation, and are subdivided into settlements under local chiefs. *See* also NILOTES.

BIBLIOGRAPHY.—E. E. Evans-Pritchard, *The Divine Kingship of the Shilluk of the Nilotic Sudan* (1948), *Essays in Social Anthropology* (1963); W. Hofmayr, *Die Schilluk. Geschichte, Religion und Lebens eines Niloten Stammes* (1925); D. Westermann, *The Shilluk People, Their Language and Folklore* (1912); reports of the Jonglei Investigation Team, Sudan Government. (Go. L.)

SHILOH, a town in Canaan where the tabernacle and ark were installed after the Israelite conquest. The site at Saylun (Seilun, in Jordan), 25 mi. N of Jerusalem, was partially excavated by the Danish archaeologists H. Kjaer and Aage Schmidt during 1926–32; they proved that the Philistines had destroyed Shiloh (*cf.* I Sam. 4) and that it remained for centuries in ruin (Jer. 7). For about a century and a half (1200–1050 B.C.) it was the chief centre of the Israelite tribal confederation, under a series of high priests, from Eleazar and Phinehas to Eli. (W. F. A.)

SHILOH, BATTLE OF. This, the second great battle in the American Civil War, also called the battle of Pittsburg Landing, was fought on April 6 and 7, 1862, between the Union forces under Gen. U. S. Grant and Gen. D. C. Buell and the Confederates under *Gen. A. S. Johnston* and *P. G. T. Beauregard.* (For clarity, the names of those fighting on the Confederate side are in *italics.*)

In February 1862 Grant had taken Ft. Henry on the Tennessee River and Ft. Donelson on the Cumberland. The Confederates had acknowledged the importance of these forts by abandoning their strong position at Columbus, Ken., and evacuating Nashville on the upper Cumberland. Grant then sought to extend his advantage by an amphibious move up the Tennessee to attack the line of the Memphis and Charleston Railroad, which followed the line of the upper river. High water frustrated these early efforts at raids from the river as a base. He then disposed his five divisions in camps around Pittsburg Landing on the Tennessee River near Corinth, Miss. There *Johnston,* commanding Confederate forces in the west, and *Beauregard* were collecting a force aimed at recovering some of their recent losses. Buell's army was marching somewhat leisurely across the country from Nashville to join Grant in an attack upon Corinth. Since they were planning for an offensive the Union troops had not fortified their camps. *Johnston,* to their surprise, seized the initiative and decided to attack Grant before Buell could arrive. Accordingly he led his army from Corinth against Pittsburg Landing early on Sunday morning, April 6.

This battle was fought by inexperienced troops on both sides. Gen. W. T. Sherman was holding the most exposed position at Shiloh Church about 2 mi. W of the Landing with recently recruited raw troops. The other Union divisions were scattered in several camps out of sight of each other. Sherman's men received the first assault and the two advanced divisions were swiftly driven in on the others, who had only a little more time to prepare themselves. Confederate leaders in the woods were unable to control and maneuver their untrained and excited men. But the Confederates continued to push each isolated Union division fighting hard toward the Landing. Thus the day passed in confused and savage scuffles between raw troops.

By late afternoon, Grant, who had come up from his headquarters at Savannah 9 mi. down the river, had rallied his troops. With one brigade of Buell's leading division, which had arrived on the previous evening, he formed a defense line in a naturally strong position barely 600 yd. from the Landing. Earlier in the afternoon *Johnston* had been killed, a sore and irreparable loss to the Confederate cause. At sunset *Beauregard,* as was customary, suspended the attack. During the night Buell brought up 25,000 troops and he and Grant took the offensive early the next day. *Beauregard* thereupon decided to extricate his hard-pressed army and to retire fighting toward Corinth. A strong rear-guard action under *Gen. Braxton Bragg* repulsed the attacks of Grant and Buell at Shiloh Church for six hours and the Union forces succeeded in doing little more than reoccupying the camp they had lost the day before while the Confederates returned to Corinth. It was a Confederate failure but not a Union victory and, each side being weakened by about 10,000 men, neither made any movements for the next three weeks.

See also AMERICAN CIVIL WAR.

BY COURTESY OF THE PRESTON PLAYER PRINT COLLECTION, KNOX COLLEGE, GALESBURG, ILL.

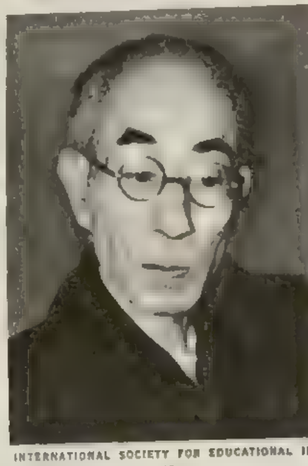
THE GUNBOATS "TYLOR" AND "LEXINGTON" SUPPORTING THE TROOPS AT PITTSBURG LANDING DURING THE BATTLE OF SHILOH. SKETCHED BY A. F. MATHEWS

BIBLIOGRAPHY.—S. M. Howard, *The Illustrated Comprehensive History of the Great Battle of Shiloh* (1921); Bruce Catton, *Grant Moves South* (1960), C. C. Buel and R. U. Johnson (eds.), *Battles and Leaders of the Civil War* (1884–87). (C. W. TE.)

SHIMANE, a prefecture (*ken*) in southwest Honshu, west Japan. Area, including the island group of Oki-Guntō, 2,558 sq.mi. (6,625 sq.km.). Pop. (1965) 821,620. It is located on the Sea of Japan, distant from the central part of the country. Because of this location and its unfavourable natural conditions, Shimane's culture, notable in ancient times, and its industry have become retarded. Forests cover most of the land, and forestry and stock breeding are well developed. However, agriculture supports most of the population. Paddy fields are scattered throughout the prefecture. Matsue (*q.v.*), the prefectural capital, is known to the west through the writings of Lafcadio Hearn. The ancient shrine at Izumo is justly famous. (R. B. H.)

SHIMAZAKI TŌSON (1872–1943), Japanese poet and novelist, whose fiction does much to illuminate the clash of old and new forces in a Japan feverishly modernizing itself, and the unsettling effects of this collision on the intellectual. He was born in Nagano Prefecture on March 25, 1872, and educated at a Tokyo mission school where he was also baptized, although Chris-

tianity did not lastingly affect either his life or his thought. In the early 1890s he joined the short-lived "Romantic" movement of young poets and writers, which he later described in his novel *Haru* ("The Spring"; 1908). Most of his nature and love lyrics, still widely appreciated because of their authentic freshness, had already appeared by the turn of the century, as had a few essays and short stories.



INTERNATIONAL SOCIETY FOR EDUCATIONAL INFORMATION, TOKYO, INC.
SHIMAZAKI

The first of his major novels, *Hakai* ("The Broken Commandment"; 1906), the story of a young schoolteacher's vain fight against village obscurantism, has been called representative of the "Naturalist School," then the vogue in Japan, although it definitely reveals the influence more of Rousseau than of Zola. *Ie* ("The Family"; 1910) depicts his and his wife's relatives drifting from their rural home into the large city but, while escaping the rigour of the patriarchal system, acquiring no new certainties. *Shinsei* ("The New Life"; 1918) narrates in a manner that carries the confessional principle to embarrassing excesses the unsavory affair of a writer with his niece. Later essays and stories, including exquisite tales for children, reveal an increasing maturity of style.

After remarrying in 1928, Shimazaki began research for the novel that was to be his greatest work as well as one of the masterpieces of modern Japanese literature, *Yoake Mae* ("Before the Dawn"; 1935). This is a story of the struggle for the imperial restoration in the 1860s as mirrored in a rural community, the tragic hero of the novel, faithfully modeled after the writer's own father, dying an embittered death, convinced that the cause of pure patriotism had been betrayed by the glib modernizers of post-Restoration Japan. Of a final novel, *Tōhō no Mon* ("Gate to the East"), only the first few chapters appeared in a monthly journal during the period just before his death, on Aug. 22, 1943. The fragment seems to invoke the Buddhist wisdom of medieval Japan as a way out of the impasse of the present. None of Shimazaki's work has been published in English. (J. Rf.)

SHIMOGA, a town in Mysore, India, and the administrative headquarters of the district of the same name, lies on the left bank of the Tunga River, about 150 mi. (241 km.) NW of Bangalore by rail. Pop. (1961) 63,764. Educational institutions include an engineering college and a degree-granting college, both affiliated to Mysore University. The town has rice and oil mills, and cotton spinning and pressing factories; handloom cloth production is a cottage or family unit type industry. It is also a trading centre, reexporting areca nuts, rice, coffee, and pepper.

Shimoga District has an area of 4,066 sq.mi. (10,531 sq.km.). Pop. (1961) 1,017,368. In the east it is drained by the twin rivers Tunga and Bhadra which unite at Kudli and join the Kistna (Krishna) River as the Tungabhadra. In the west the district is drained by the Sharavati River, which, before it enters the western coastal plain, forms a magnificent cataract, the Gersoppa Falls. The watershed of the Tungabhadra and Sharavati river basins lies on the western edge of the Ghats. The western half of the district, which is mountainous and forested, is known as *malnad* or hill country. Coffee is grown on the hill slopes and the valley bottoms are given over to the cultivation of rice and sugarcane. The eastern half of the district is open, undulating country and there dry crops such as cotton, millet, and oilseeds are grown. Natural resources include iron ore, limestone, and manganese. Bhadravati (pop. [1961] 24,495), on the Bhadra River, is an industrial centre with an iron and steel plant and a paper mill. At Gersoppa there is a hydroelectric works.

In the 16th century Shimoga was under the Keladi chiefs. In 1763 it was annexed by Haidar Ali, being finally brought under British administration in 1830. (G. K. Gh.)

SHIMONOSEKI, largest city of Yamaguchi Prefecture in extreme western Honshu, Japan. Pop. (1965) 254,380. Its strategic location on the Straits of Shimonoseki brought it to transportation and commercial prominence. The city was formerly called Akamagaseki or Bakan. Modern development began in 1905 with the opening of railroad ferry service with Moji in Kyushu; more recent links include a 2.3 mi. (3.7 km.) railroad tunnel (1942) and a vehicular-pedestrian tunnel (1958) under the straits. Shimonoseki became a heavy industrial centre after 1942 and in 1940 its port was formally amalgamated with Moji (*see* Kita-Kyushu) and Kokura ports into the single port of Kammon.

The Straits of Shimonoseki (Shimonoseki-kaikyō), renamed Kammon Straits, are the narrow western passageway between Japan's Inland Sea and east Asian waters. Although about 2,200 ft. (670 m.) wide at the narrowest point and partially blocked by Hiko Island at the western end, the straits are 40–65 ft. (12–20 m.) deep and are easily navigable. (J. D. Ee.)

SHINGLES: *see* Skin, Diseases of.

SHINNECOCK, a North American Indian tribe whose aboriginal home was the eastern end of what is now Long Island, N.Y. In the 17th century they were part of the Montauk Confederacy, along with such fellow Algonkian tribes (*q.v.*) as the Manhasset, Massapequa, Montauk proper, Patchogue, and Rockaway. During this period the confederacy was attacked and dominated from the north, first by the Pequot and later by the Narraganset, and suffered serious additional population losses through epidemic disease. Most of the survivors organized to form the Brotherton tribe, moving in 1788 to lands near Marshall, N.Y., donated by the Oneida (*q.v.*). Descendants of those members of the Montauk Confederacy who remained on Long Island are recognized by New York State as the Shinnecock tribe. In the 1960s, having lost almost all their earlier customs and language through acculturation (*q.v.*) and marriage with Negroes, about 150 people officially called Shinnecock lived on their reservation near Southampton, N.Y.

Bibliography.—M. R. Harrington, *Ancient Village Site of the Shinnecock Indians* (1924); L. G. Carr and C. Westez, "Surviving Folk Tales and Herbal Lore Among the Shinnecock Indians of Long Island," *Journal of American Folk-lore*, vol. 58 (1945); L. M. Hunter, *Shinnecock Indians* (1950); F. W. Hodge (ed.), *Handbook of American Indians North of Mexico*, 2 vol. (1959).

SHINRAN (called Shinran Shōnin, "St. Shinran") (1173–1262), founder of Jōdo Shinshū (True Pure Land School, Shin Buddhism), the representative school of devotional Buddhism that arose in the Kamakura period and remains one of the chief Japanese Buddhist sects, was born in Kyōto, the son of Arinori Hino, a member of the illustrious Fujiwara family which ruled Japan during the Heian period. Orphaned at an early age, Shinran entered a monastery on Mt. Hiei in Kyōto and began intense religious training in the Tendai tradition. After 20 years, realizing his inability to purge himself of ignorance and human passions, he became a disciple of Hōnen (1133–1212), founder of the Jōdo, or Pure Land School. Hōnen taught that because Amida Buddha's vow of salvation was universal and unconditional, it was not through good works but through the exclusive recitation of Amida's name (*Namu Amida Butsu*) that man could attain to Buddhahood. The sudden popularity of the new doctrine aroused the jealousy of the established schools, and Hōnen and some of his disciples were banished. Shinran continued his ministry in distant provinces, marrying and thus breaking the age-old rule of celibacy for the priesthood. In 1224 he completed his main work, the *Kyō-gyō-shin-shō*, in which he systematized his unique teaching, Jōdo Shinshū. He not only accepted Hōnen's doctrine but also sought to refine it by setting forth the significance of faith—which was, according to Shinran, itself a gift from Amida Buddha—as the true cause of birth in the Pure Land after death. In clarifying Hōnen's concept of the *Nembutsu* (reciting the Holy Name), he stated that the recitation was the spontaneous expression of one's innermost joy and gratitude and not the means to accumulate merit.

Shinran developed the Pure Land School to its utmost simplicity in order to reach all levels of people. Shin Buddhism became one of the most popular schools of Buddhism in Japan. In accordance with the progressive and liberal attitude of its founder, it has involved itself in extensive educational and social welfare programs.

Since the late 19th century it has gone beyond its native boundaries and established itself abroad, especially in the Americas, where numerous congregations are to be found.

English translations are available of Shinran's *Kyō-gyō-shin-shō* ("Teaching, Practice, Faith and Attainment"), by Kōshō Yamamoto, with notes (1958), and, separately, of his *Shō-shin-ge* ("Gāthā of True Faith"), by Daien Fugen, also with notes (1961). His *Wasan* (more than 350 verses in praise of Buddha), translated by Yamamoto, are included in the *Shinshū Seiten* (1955). The *Tannishō*, compiled by one of his disciples but including many important sayings of Shinran, has been translated into English and other European languages. Shinran's *Private Letters* also have been translated by Yamamoto (1956).

See also BUDDHISM: *Regional Variations in Buddhism: Japan.*

BIBLIOGRAPHY.—Shōdō Tsunoda *et al.* (eds.), *Buddhism and Jōdo Shinshū* (1955); Gendō Nakai, *Shinran and His Religion of Pure Faith* (1937); Daisetz T. Suzuki, *Miscellany on the Shin Teaching of Buddhism* (1949), *Mysticism: Christian and Buddhist* (1957); Kōshō Yamamoto, *An Introduction to Shin Buddhism* (1963); S. Inaba and I. Funabashi, *Jōdo Shinshū*, Eng. trans. by I. Fukuhara and W. Flygare (1961). (Ry. F.)

SHINTŌ, which literally means the way or teaching of the gods (*kami*), is the loosely organized, indigenous religious cult of Japan. This designation arose in order to distinguish the traditional religion from Buddhism when the latter was introduced to Japan in the 6th century A.D. While Shintō has no founder, no official sacred scriptures, and no dogma, it has preserved its ethos throughout the ages. In the modern period it has been used as a tool of ethnocentric nationalism and chauvinistic militarism, and it is often thought of in these terms by outsiders.

Primitive Shintō.—Even before the historical period the inhabitants of Japan had some kind of religion, but the origins of the ancient cult that gradually developed and was later called Shintō are unknown. The ancient Japanese people were an admixture of various groups who had migrated to the Japanese islands from the Asian continent, and their religion betrayed northern and southern Asian influences. It took several centuries for these ethnic groups and peoples to become assimilated under the leadership of the so-called Yamato clan, about the 4th century A.D. The early Japanese myths, which are often unrelated and contradictory, indicate that they were in part the legacy of the dominant Yamato clan and in part the sacred traditions of other clans and peoples that were subjugated by the Yamato.

The early Japanese by and large did not distinguish sharply between the celestial and earthly domains. They had only a vague notion about the life to come, and the veneration of ancestors that characterized religious life later was hardly known. Their religion was a simple polytheistic nature worship, emphasizing gratitude to the beneficent forces of nature, while also to some degree appeasing the malevolent forces. These forces were indiscriminately called *kami*, usually translated as "gods" or "deities" but meaning "above," "superior," or "divine," and signifying anything that was the object of reverence and respect. Accordingly, all the heavenly and earthly forces, great men both living and dead, and many animate and inanimate beings such as plants, rocks, birds, beasts, and fishes as well as earthquakes, thunder, water, sun, and moon were *kami*. The early myths spoke of 800 myriads of *kami*, usually divided into heavenly *kami* (those who resided in the heavenly abode) and earthly *kami*. The Japanese did not regard spirit as superior to matter; rather, they believed that the two coexisted as equals. *Kami* were thus regarded as inseparable from *shintai* (*kami-* or god-body) or *mitama-shiro* (visible representations of *kami*). The religious rites of the early Japanese were mostly fertility cults. Ceremonial purification was emphasized. Three kinds of persons performed religious rites: (1) the heads of families or clans, considered priests *de facto*; (2) shamans (*see* SHAMANISM) who performed divination, sorcery, and lustration; and (3), in at least some clans, hereditary lines of priests and shamans. When the Yamato clan overpowered the other clans its hereditary priest-king (Tennō or "emperor") acquired supreme prestige. Significantly, the function of the throne was to take charge of both religious rites (*matsuri*) and political administration (*matsuri-goto*), which were regarded as inseparable.

With the ascendancy of the Yamato clan its myths began to provide foundations for political theories. (*See* JAPANESE MYTHOLOGY.) Their prominent feature was the divine origin of the Yamato people in the sun-goddess, whose grandson, Ninigi, was said to have been the grandfather of the first emperor, Jimmu Tennō, the earthly ancestor of the imperial family. The sun-goddess was later enshrined at the Grand Shrine of Ise, which became both the tutelary shrine of the imperial family and the central shrine of the whole nation (*see* ISE). Eventually the myth of the solar ancestry of the imperial clan was widely accepted among the various clans, including many of Chinese and Korean descent. The sun-goddess was worshiped side by side with other clan deities and nature deities, all of them comprising a rich Shintō pantheon. While in the history of Japan the political authority of the throne was often precarious, it never lost its bond with the masses because of Shintō, which was rooted both in the religious tradition of the imperial clan and in the sacred memories of other clans.

Foreign Influences.—With the introduction of Chinese civilization in the 5th century A.D., the religious and cultural situation in Japan became complex. Confucianism not only provided Japan with systematic theories of social and political institutions but also with ethical norms for individuals, ranked in a hierarchical society. With Confucianism came Taoism and the Yin-Yang philosophy, and they too contributed philosophical concepts. (*See* CONFUCIANISM; TAOISM.) Gradually, Confucian moral virtues began to be attributed to Shintō deities, and imperial edicts stressed such Confucian virtues as uprightness, sincerity, and honesty as guiding moral principles. The concepts of filial piety and the veneration of ancestral spirits were soon adopted by the Japanese.

Buddhism (*q.v.*), introduced to Japan in the 6th century A.D., also had far-reaching influence, becoming in the 7th century for all practical purposes the state religion. Shintō, however, retained its prestige and influence among the masses, and a government department of Shintō was established to manage and control the festivals and rites at the imperial household as well as at the tutelary and ancestral shrines of powerful clans. The cause of Shintō was greatly enhanced by the compilation in the 8th century of the *Kojiki* and the *Nihongi* (*qq.v.*), in which the ancient oral traditions were assembled and recorded for the first time. Since Chinese script was used, the myths and legends were inevitably influenced by Chinese thought; nevertheless, these records are an important storehouse of early Japanese myths and they are often considered as quasi-sacred scriptures of Shintō, although Shintō has no concept of scripture comparable to the Bible or the Koran.

Coexistence of Buddhism and Shintō.—Buddhism began to overshadow Shintō during the Nara period (8th century A.D.), especially in the capital city of Nara, but it had to come to terms with the deeply rooted Shintō beliefs and practices of the people. Even the construction of the Tōdaiji, the national cathedral of Buddhism at Nara, required the blessing of the Shintō deities, communicated to the throne by means of oracles; and emperors and empresses who considered themselves humble slaves of Buddha could not neglect the Shintō rites altogether. Buddhism and Shintō did not attempt to exterminate each other. What emerged was a pattern of coexistence that is often referred to with some exaggeration as the amalgamation of Shintō and Buddhism.

This pattern developed gradually. For example, wealthy land-owning Buddhist monastic institutions found their tenants deriving solidarity from Shintō deities and Shintō cults; by allowing Shintō shrines to remain, the Buddhist institutions were protected by Shintō taboos. Eventually, Shintō shrines found their way into the sacred premises of Buddhist temples and monasteries, and in turn Buddhist chapels were built near Shintō shrines. During the Nara period Buddhism became so powerful in the court that Shintō had nothing to lose by allying itself with Buddhism. Even the popular Shintō deity Hachiman came to be known as *Daijizaiten-bosatsu* (Bodhisattva).

When the capital was moved from Nara to Heian-Kyō (the present Kyōto) toward the end of the 8th century, ostensibly to be freed from the political pressure of Buddhist institutions, the pattern of coexistence persisted. As early as 794, Buddhist sacred

scriptures were recited at some Shintō shrines. Two new Buddhist schools, Tendai and Shingon, inaugurated in the 9th century not only welcomed an alliance with Shintō but also attempted to provide theoretical justification for coexistence. The Shingon Buddhist theory was called Ryōbu ("Two Aspects") Shintō, the Tendai theory Ichi-jitsu ("One Reality") Shintō. Both interpreted Shintō deities as manifestations or incarnations of the Buddha, and each attempted to incorporate Shintō into its framework of Buddhism. Shintō priests were dominated by the Buddhist ecclesiastics and had to be content to play only a minor role even in Shintō rites. Shintō continued to enjoy favour among the masses in the outlying districts, and the folk elements of the Shintō tradition were represented in the Shugen-dō (order of Mountain Priests), but during the middle of the Heian period the Shugen-dō also allied itself with the Tendai and Shingon Buddhist schools.

Shintō Reaction Against Buddhism.—During the Kamakura period, which succeeded the Heian, the feudal regime (*bakufu*) ruled by warrior-statesmen (*shōgun*) was established, and courtly elegance gave way to the austere culture of warriors. Moreover, after centuries of peace and tranquility, Japan was threatened by the Mongols. But this period, with its social and political disruption, was an age of spiritual awakening reflected in Shintō by an attempt to emancipate itself from the domination of the Tendai and Shingon schools. The spokesmen for Shintō were hereditary priests of the Watarai family who served at the Outer Shrine of Ise; hence the movement was called Ise Shintō. Buddhism was too deeply rooted to be rejected altogether, and recognizing this, the spokesmen of the Ise Shintō movement attempted to reverse the Buddhist claim that Shintō deities were incarnations of the Buddha; they taught instead that Buddhas and Bodhisattvas were manifestations of the great *kami* nature of Shintō. By the middle of the Kamakura period the five-volume apologetic work, *Shintō Gobushō* ("Shintō Pentateuch"), was completed.

The decline of the Kamakura feudal regime was followed by a period of imperial rule which greatly encouraged the Ise Shintō movement. The historian Kitabatake Chikafusa (*q.v.*) wrote in his *Junnō Shōtōki* (1369), "Great Yamato is a divine nation. It is only our land whose foundations were first laid by the divine ancestor. It alone has been transmitted by the sun-goddess to a long line of her descendants." The short-lived imperial rule was followed by the Ashikaga feudal regime, during which, with some notable exceptions, Shintō was again at a low ebb. A monotheistic view developed, formulated by Ichijō Kanera (d. 1481) and by Yoshida Kanetomo (d. 1511), who, deeply influenced by Taoist metaphysics, initiated Yoshida or Yui-itsu ("One and Only") Shintō, teaching that one unique *kami*-nature is the underlying substance of all Shintō deities and Buddhas. The influence of Shintō, however, was negligible during the Ashikaga period and the "dark age" that followed it.

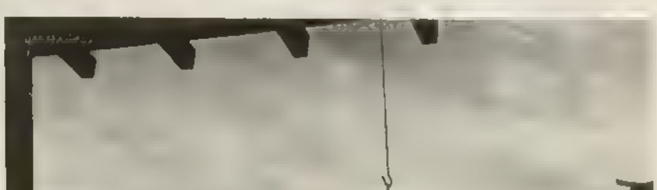
Confucian Shintō.—Japanese unification was completed with the establishment of the Tokugawa feudal regime early in the 17th century. While the Tokugawa rulers gave financial support and certain prerogatives to both Buddhist and Shintō clergy, ecclesiastics were strictly governed by the "commissioner of temples and shrines" appointed by the regime, and the guiding ideology of the regime was the Chu Hsi school (Shushi-gaku) of Neo-Confucianism (*see* CHU HSI; CONFUCIANISM). Curiously enough it was the Confucian scholars of this period who allied themselves with the cause of Shintō and eventually brought about the Shintō revival. (*See* further JAPANESE PHILOSOPHY: *Philosophy Before the Modern Period: Search for a New Universal Principle.*)

Shintō leaders welcomed an alliance with Neo-Confucianism, which provided the cosmology that was lacking in traditional Shintō. Deguchi Nobuyoshi (d. 1690) attempted to interpret Shintō according to the Confucian classic *I Ching* ("The Book of Changes"). Kikawa Koretaru (d. 1694) identified the "Supreme Ultimate" of Chu Hsi with a Shintō deity, Kuni-tokotachi, whose name now meant the Cosmic Lord and the Reason of Heaven. Meanwhile, under the leadership of Yamazaki Ansai (*q.v.*), a new movement called Suika Shintō emerged. This school identified the Supreme Ultimate of Neo-Confucianism with two Shintō deities, Kuni-tokotachi and Ameno-minaka-nushi, who in turn were regarded as essentially one reality which is both the substance of the universe and the source of morality; and the sun-goddess is their incarnation. Yamazaki taught the importance of devotional prayer and the virtue of honesty and advocated an extreme respect for the emperor. Similar emphasis on respect for the emperor was taught by the Mito school of Neo-Confucianism, laying the foundation for the loyalist movement later.

Shintō Revival.—The revival of Shintō studies developed gradually in the 18th century along with the study of the Japanese classics. Stimulated by philological research in ancient Japanese poetry, scholars of the Koku-gaku ("Japanese studies") such as Kamo no Mabuchi (1697–1769) devoted themselves to the study of the *Manyōshū*, a collection of ancient poems, and *Norito*, a collection of old Shintō liturgies. Rejecting both the Buddhist- and Confucian-centred interpretations of Shintō, Kamo tried to restore its pre-Buddhist and pre-Confucian meaning. Unconsciously espousing the Taoist ideal, he stressed pure simplicity, in accordance with the order of heaven and earth, as the best morality.

Kamo's disciple Motoori Norinaga (1730–1801) rejected this Taoist-oriented interpretation and insisted that Shintō was based on the revelation of *Takami-musubi-no-kami* ("August Producing Deity") transmitted by the sun-goddess. Motoori is credited with the most systematic interpretation of the concept of *kami* as "anything whatsoever which was outside the ordinary, which possessed superior power or which was awe-inspiring." In 1798 he completed his voluminous commentary on the *Kojiki*, which has remained the authoritative interpretation of the theoretical aspects of Shintō.

Another noted Shintō scholar, Hirata Atsutane (1776–1843), a Confucian scholar before he was



BY COURTESY OF CONSULATE GENERAL OF JAPAN, N.Y.

(LEFT) LANTERNS AT KASUGA SHRINE; (ABOVE) ITSUKU-SHIMA SHRINE

Both Itsuku-shima and Kasuga shrines are ancient, dating back to the 9th and 8th centuries respectively. The island shrine of Itsuku-shima, in Hiroshima Bay, has been rebuilt several times, always according to the original plans. The 3,000 bronze lanterns of Kasuga Shrine, at Nara, were placed by worshipers as offerings to the deities.

influenced by Motoori's writings, tried to systematize Shintō "theologically" by holding *Ame-no-minaka-nushi* ("Heavenly Central Lord") as the creator god, viewed above the sun-goddess. He also developed an eschatological concept within the Shintō framework. His *Hongyō-gaihen* ("Supplemental Compilation of Shintō," 1806) is divided into two parts, one based on writings of the Jesuit Matteo Ricci (*q.v.*), who wrote in Chinese, the other on selections from a manual of Christian ethics written in 1614 by another Jesuit missionary in China, Didace de Pantoja. Hirata's thought was to have a great influence on the development of modern Shintō.

Toward the end of the Tokugawa period, the folk elements of Shintō began to break through its traditional framework. The Konotabi movement, inaugurated by an illiterate peasant woman, Kinō (d. 1826); the Tenrikyō movement, founded by another peasant woman, Miki (d. 1887); and the Kurozumi-kyō movement, initiated by an obscure Shintō priest, Kurozumi Munetada (d. 1849), were but a few examples of this dynamic religious awakening among the poor and oppressed. (*See* below, *Sect Shintō.*) They were destined to play important roles in subsequent periods. Meanwhile, these messianic movements among the peasantry were looked down upon as impure Shintō by the orthodox hierarchy and scholars.

Shintō as the National Faith.—In 1867 the modern age began in Japan. The Meiji regime that followed the Tokugawa shogunate had two diametrically opposed aims—the revival of the ancient Japanese pattern of *saisei-itchi* (unity of religion and the state) and the modernization of Japan. In 1868 the department of Shintō (*Jingi-kan* or *Shingi-kan*) was established and immediately issued a separation edict intended to abolish the age-old pattern of Shintō and Buddhist coexistence. Buddhist priests who had been connected with Shintō shrines were given the choice of returning to lay life or being re-ordained as Shintō priests. The traditional custom of burying Shintō priests according to Buddhist funeral rites was forbidden, and a new Shintō burial office was introduced. In 1869 the department of Shintō was made independent of the cabinet and placed above the grand council of state. In 1871 Shintō was proclaimed the national religion and Shintō shrines were decreed to be the place of worship for all subjects of the emperor. All Shintō priests were to be appointed by the government. To counteract movements to restore the feudal regime, the government began to advocate emperor worship. But the measures taken to uphold the supremacy of Shintō at the expense of Buddhism failed: Buddhism was too deeply imbedded in the fabric of Japanese life.

The department of Shinto was replaced during 1872–77 by the ministry of religion and education (*Kyōbu-shō*), which was given jurisdiction over both Shintō and Buddhism. In 1873 the edict banning Christianity was lifted. Recognizing the impossibility of suppressing the growing tide of spontaneous folk religious movements, the government in 1882 decided to classify them as Sect Shintō, to be differentiated from Shrine Shintō, which had about 200,000 large and small shrines throughout the nation. The constitution promulgated in 1889 included a clause guaranteeing freedom of religious belief, but the government continued to favour Shintō, not so much as a religion but as a national cult. In 1900 shrines and religions were placed in separate bureaus, and in 1913 the religious bureau was transferred to the ministry of education

In a real sense, modern Japan was caught between its two objectives. One of these, to reestablish the ancient system of *saisei-itchi*, drove Japan to assert the centrality of Shintō as the national religion. The second, modernization, however, drove it to pay lip service to freedom of religious belief. Toward the end of the 19th century an uneasy compromise was worked out, based on the theory that Shintō was not a religion but a suprareligious national cult, and as such could be superimposed on the nation. The emperor cult was almost arbitrarily devised and a course in *shūshin* ("moral teaching") was made the basis of compulsory education. Beginning with the Chinese-Japanese War (1894–95), Japan followed an expansionist policy, and from that time until World War II Shintō was manipulated by the militarists and jingoistic nationalists as the spiritual weapon for mobilizing the nation to guard the prosperity of the throne.

Shintō After World War II.—State Shintō was disestablished immediately after World War II. Orders from the supreme commander for the Allied powers to the Japanese government stated explicitly that:

> The sponsorship, support, perpetuation, control, and dissemination of Shintō by the Japanese nation, prefectural, and local governments, or by public officials, subordinates, and employees acting in their official capacity are prohibited and will cease immediately; all financial support from public funds and all official affiliation with Shintō and Shintō shrines are prohibited and will cease immediately.

Equally drastic was the abolition of *shūshin* from the school curriculum and the emperor's public statement:

> The ties between Us and Our people have always stood upon mutual trust and affection. They do not depend upon mere legends and myths. They are not predicated upon the false conception that the emperor is divine and that the Japanese people are superior to other races and are fated to rule the world.

Although the emperor cult was thus abolished, the time-honoured tradition of imperial family Shintō continues. Four shrines have been set aside for the imperial household rites, the most important among them being the *Kashiko-dokoro*, dedicated to the sun-goddess, a branch of the Grand Shrine of Ise.

Organization and Worship.—With disestablishment, Shintō shrines lost their financial subsidies from public funds. Of about 110,000 shrines governed before the war by the home ministry, over two-thirds came to belong to the Association of Shintō Shrines (*Jinja Honchō*); the rest are either independent or belong to small local associations.

Many Shintō priests are so by heredity, but anyone with proper training can become a priest. The main training centre is Kokugakuin University in Tokyo. From the Meiji period until the end of World War II, Shintō priests were government officials, graded into several ranks. After the war they had to be supported by income from their shrines and other offerings from the laity. Each shrine has *ujiko* (parishioners), whose representatives act as a governing board for shrine affairs. In addition, *sūkeisha* (worshipers who are not regular parishioners) are welcome to participate in the ceremonies at any time. Many shrines have started various kinds of social welfare work, and Shintō wedding rites are popular.

There are numerous types and kinds of shrines, from the Grand Shrine of Ise, with its 14 subsidiary shrines, to small obscure roadside oratories unattended by a priest. Before disestablishment, shrines were classified as national, special national, governmental, prefectural, and district or village shrines, besides about 63,000 shrines "without rank." Most shrines are dedicated to deities, but some are dedicated to historical figures such as Emperor Meiji and General Nogi Maresuke. Most of the special national shrines enshrine the spirits of persons who made conspicuous contributions to the nation. Genealogies of deities enshrined in rural Shintō shrines reveal varied backgrounds.

The shrines assume various forms, but most are built in scenic surroundings. The two main units are an inner sanctuary (*honden*) and an oratory (*haiden*). According to Shintō, deities are present in the *shintai* (*kami-* or god-body) or in the *mitama-shiro* (visible representation of *kami*) which is kept in the inner sanctuary. A shrine dedicated to mountain or forest deities, however, has no need of an inner sanctuary. Usually, only the priests and their attendants are permitted to approach the inner sanctuary, where they recite prayers. Larger shrines have additional buildings, such as a hall of reciting prayers (*norito-den*), a hall of offerings (*hei-den*). and a hall of liturgical dance (*kagura-den*).

The entrance to the shrine is marked by a *torii*, a simple gate marking off the sacred compound. Worshipers wash their hands and rinse their mouths at an ablution basin (*te-mizu-ya*) and then approach the oratory, bowing reverently to the inner sanctuary and clapping their hands. A small offering to the deities is usually made. On special occasions the priest may be asked to perform a simple rite of purification (*harai*) in which a branch of the sacred tree (*Cleyera japonica*, an evergreen of the tea family) is waved three times before the worshiper, after which he makes offerings and secures charms at the shrine office.

Parishioners (*ujiko*) belong hereditarily to certain shrines, in the sense that the deity of a given shrine is the tutelary *kami* for all members of the *ujiko*. But priests have no pastoral relation

with *ujikos*, their main duty being to serve the deities and offer prayers. Shintō religious activities are not confined to the shrines. The traditional Japanese family performs a simple daily rite before the family *kami*-shelf (*kami-dana*), dedicated to the family's tutelary deity.

Shintō prayers (*norito*) are based on the ancient belief that the spoken word has a spiritual potency; therefore the prayers must be recited reverently. Prayers usually include, in elegant classical language, words of praise for the deities, lists of offerings, and petitions. The norms of Shintō rituals and prayers are found in the *Engi Shiki* ("Ritual Notes"), a 50-volume work compiled in the 10th century.

Festivals.—Before 1945 most national holidays were based on Shintō festivals: New Year's day, when at the imperial court and elsewhere people worshiped in the four directions; Empire Foundation day (Feb. 11), the day when the first emperor, Jimmu, is supposed to have been enthroned; the Spring Season Imperial Spirit festival (spring equinox); the Anniversary of the Death of Jimmu (April 3); the Autumn Season Imperial Spirit festival (autumn equinox); the Festival of the Presentation of First Rice to the Deities at the Grand Shrine of Ise (Oct. 17); and the Autumn Thanksgiving festival (Nov. 23). After 1945 some of these holidays lost their religious meaning, but they remained holidays. For example, Nov. 3, previously set aside to commemorate the great achievement of Emperor Meiji, became the National Day of Arts and Culture, while Nov. 23, the traditional day of the Shintō festival at the sanctuary of the imperial household, became the Japanese Labour Day.

Doctrine.—Throughout its history Shintō has been conspicuously indifferent to any theoretical systematization of its beliefs. Though there exist movements to formulate a coherent theological system, none has been widely accepted. Such central concepts as *kami*, *musubi* (creating and harmonizing power), *makoto* (truthfulness), and *harai* (purification) are integrated into the total Shintō way of life and worship. Probably the genius of Shintō lies in its refusal to formulate a *summa theologica*. An ancient Japanese poem expresses its nebulous but genuine religious sentiment: "Unknown to me who resideth here; tears flow from a sense of unworthiness and gratitude."

Sect Shintō.—Throughout its long history, Shintō has been supported by the folk piety of the masses, uninterested in complicated concepts and theories of religion. The common people of Japan developed arts and crafts to a high degree, and they had a vivid poetical sense and appreciation of the mystery of life and the universe. Living close to nature, they invoked deities of all sorts, depending on their physical and spiritual needs. Divination, spirit possession, protection from misfortune and disease, and magical formulas for other benefits played important roles in popular belief and practice, and they infiltrated official Shintō and Buddhism as well. In times of social unrest or natural calamity this folk piety was often intensified and supported various messianic movements, which frequently became established and eventually assimilated into Shintō or Buddhism. The 19th century, a period of social upheaval and transition, witnessed the emergence of many such messianic movements. Between 1882, when these popular cults were separated from the suprareligious national cult and classified as Sect (Kyōha) Shintō, and 1908, 13 Sect Shintō denominations came to be recognized as "churches" (*kyōha* or *kyōkai*), dependent, like Buddhist sects and Christian denominations, on private initiative for their propagation, organization, and financial support. Not all of the 13, however, were Shintō in origin or in ethos.

Usually the Sect Shintō denominations are classed as follows: (1) Pure Shintō sects based primarily on certain Shintō beliefs and practices and emphasizing loyalty to the throne and gratitude to ancestors; (2) Confucian sects, which blend Confucian moral doctrines and certain Shintō beliefs; (3) Mountain sects, which hold that their deities reside in certain sacred mountains; (4) Purification sects, which stress mental and physical purification from evil and contamination; and (5) Utopian or Faith-Healing sects. The denominations have little in common. Some of them worship certain traditional Shintō deities and add others of their own; the Konkō-kyō sect, however, has no Shintō deities, and the god of Tenri-kyō is not included in the Shintō pantheon. Most active among the 13 denominations is the Tenri-kyō, which boasts of more than 500 churches overseas, in addition to its well-organized ecclesiastical institution in Japan.

Most adherents of Sect Shintō sought worldly benefits, such as cures for sickness, protection from disasters and misfortunes, wealth and success in life, which traditional Shintō as well as Buddhism failed to offer. Gradually, however, the new denominations tended to become institutionalized and less concerned with the immediate needs of their adherents. The people began to look for new prophets and shamans, creating a number of small splinter groups within the denominational framework. These splinter groups were emancipated from their parent bodies after World War II and initiated the postwar boom of "New Religious Cults" (*Shinkō Shukyō*). An extremely precarious relationship exists between the Sect Shintō denominations, now organized as the Federation of Sect Shintō, and the New Religious Cults. *See* also JAPAN: *History* and *The People: Religion*.

BIBLIOGRAPHY.—W. G. Aston, *Shintō (the Way of the Gods)* (1905); D. C. Holtom, *The National Faith of Japan* (1938), *Modern Japan and Shintō Nationalism*, rev. ed. (1947); Genchi Katō, *A Study of Shintō, the Religion of the Japanese Nation* (1926); Genchi Katō *et al.*, *A Bibliography of Shintō in Western Languages from the Oldest Times till 1952* (1953). (J. M. KA.)

SHINYANGA REGION, TANGANYIKA, established in May 1963 following a reorganization of administrative units in Tanganyika (Tanzania), comprises Shinyanga and Maswa districts, which previously belonged to the former Lake Region, and Kahama District, which previously belonged to the former Western Region.

The region is bounded north by West Lake, Mwanza, and Mara regions, east by Arusha Region, south by Singida and Tabora regions, and west by Kigoma Region. Area 19,600 sq.mi. (50,764 sq.km.). Pop. (1963 est.) 749,000. The region lies on the central plateau of Tanganyika across the ill-defined watersheds between the Lake Victoria Basin to the north, the Malagarasi River system to the southwest, and the Wembere Depression and Lake Eyasi in the eastern rift to the east. The plateau is largely composed of granites with hills of banded ironstone of Nyanzian Age. Maswa District extends eastward to include part of the Serengeti National Park. Mean annual rainfall ranges from less than 20 in. (500 mm.) in Maswa District to more than 40 in. (1,000 mm) north of Kahama, the rain falling mainly during the period from November to April. Soils include old lacustrine soils, sandy clay loams on slopes, black clays in valley bottoms, and some with a lateritic horizon.

Regional headquarters are at Shinyanga, and the population is most dense in Shinyanga District, which is famous for the Williamson Diamond Mine. Shinyanga and Maswa districts, which form part of Sukumuland, are characterized by an intensive agricultural economy based on cattle and the cultivation of cotton. The Nyamwezi and Sumbwa tribes predominate in Kahama District, in the western part of which the population is relatively sparse and the tsetse fly is prevalent. Maize, sorghum, and millet are among the staple food crops. (J. M. KE.)

SHIP, the vehicle by means of which man conveys himself and his goods by water; more precisely, the larger and more seaworthy of such vehicles, the smaller and simpler being boats. To merit classification as a ship or boat it is necessary for the vessel to support weight, not merely because of the buoyancy of the material of which it is made, but because of its displacement of water. (Structures depending only on their inherent buoyancy are rafts, or, in their simplest form, floats.)

After an initial discussion of the various kinds of shipbuilding and the limits of this discussion, the article is divided into the following sections:

For the practice and theory of modern shipbuilding, *see* NAVAL ARCHITECTURE; MARINE ENGINEERING. *See* also SHIPPING. For ships in naval history, *see e.g.*, AIRCRAFT CARRIER; DESTROYER; articles on battles and wars, and sections on history in articles on countries. *See* also separate articles on types of ships, such as CANOE; GALLEY; etc.

All kinds of ships, from the most primitive to the most modern and elaborate, can be classified under six heads: (1) rafts (floating logs or bundles of reeds, etc., either singly or connected to form a platform); (2) dugouts (hollowed trees); (3) canoes of bark or skin with an internal framework; (4) canoes or boats formed from planks stitched together; (5) vessels with planking nailed together and with a framework inserted; (6) vessels of which the framework is first set up and the planking (or plating) attached afterward. These classes sometimes shade into one another.

Rafts were no doubt used in all parts of the world at a certain stage of local culture. As seagoing craft in modern or comparatively recent times they are best known from the catamarans of India and the balsas of South America. The Ecuadorian balsa was a large sailing raft fitted with something like the modern centreboard. Its seaworthiness was sufficiently proved in 1947, when a party of Norwegian scientists, wishing to test a theory with regard to the peopling of the Pacific islands, built a similar raft, the "Kon-Tiki," at Callao, Peru, and succeeded after a voyage of three and one-half months in reaching the islands east of Tahiti. Dugouts are still characteristic of East Indian and Polynesian waters, bark canoes are (or were) chiefly North American, skin boats belong to the arctic, and the practice of building with sewn planks can still be observed in the East Indies. The two last classes correspond to the clinker building and carvel building of modern European small craft.

The following survey deals only with what may be called European ships, though this has to include the whole of the Mediterranean and, in later times, the shipping of North America. Eastern, Far Eastern (especially Chinese) and Polynesian craft form subjects of their own. (*See* BOAT: *Existing Boat Types*.)

The more or less standardized European full-rigged sailing ship was produced in the 15th century by the combination in a single vessel of characteristics derived from two almost distinct lines of descent, one from the Mediterranean and the other from northern and western Europe. Roughly speaking, construction came from the south and rig from the north. After this combination—from about 1460 to the last days of deep-water sailing ships—the European ship, in spite of a constant process of modification and improvement, remained unchanged in essentials.

It must be remembered that the southern line was very much older than the northern, or at least that it had reached a stage of comparative maturity much earlier. At the beginning of the Christian era the Mediterranean ship had a history of about 3,000 years and its builders had been producing more and more elaborate vessels for many centuries, while their fellow craftsmen in the north had only just evolved the method of building with keel, stem, sternpost, and planks and in so doing passed the line separating the mere canoe from the potential ship.

I. EARLY CRAFT

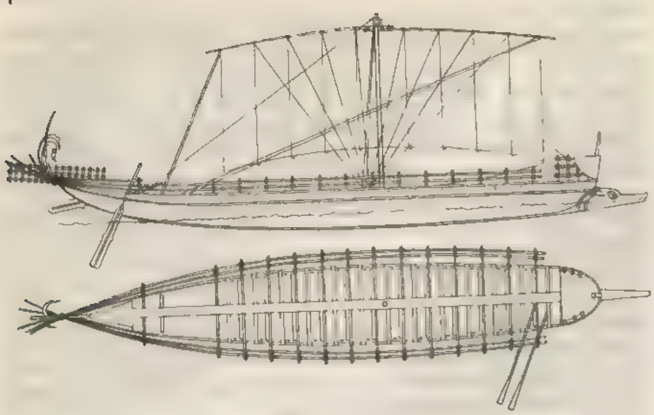
1. Origins.—The earliest knowledge of boats and ships comes from Egypt, where as early as 4000 B.C. boats were already far advanced from the primitive form from which they must have been derived, in this case probably a bundle of reeds. It has been claimed that "naval architecture is an Egyptian art and that the main lines of the history of ship-building for the whole world were laid down in Egypt towards the end of the 4th millennium B.C." (Elliot Smith, *Ships as Evidence* [1917]). There is, however, at least one extremely weighty objection to this sweeping claim—the fact that the Egyptian method of construction was not transmitted to the neighbours and successors of that kingdom. Egyptian vessels were essentially built-up dugouts (to use a contradiction in terms); they had no keel, stem, or sternpost and no internal framing, but consisted simply of a heavily built skin formed of many pieces of timber doweled or dovetailed together. With regard to matters of equipment such as mast and sail or steering gear, the claim is more easily substantiated.

Unfortunately, comparatively little is known of the ships of the Cretans, who were the dominant sea power in the eastern Mediterranean about 1500 B.C., or of the Phoenicians, who took their place; but that little suggests that both nations had begun to differentiate between the fighting vessel and the merchantman and between the rowing galley and the sailing ship. The few representations of Phoenician vessels (about 700 B.C.) do, however, illustrate two striking developments: the arrangement of oars in two banks at different levels and the fitting of a ram bow in galleys intended for fighting. It is possible that the ram was an Egyptian invention, and something of the sort is shown in action as early as 1200 B.C., but it was in Phoenician, Greek, and Roman galleys that its importance was most clearly emphasized.

With the ships and galleys of the Greeks firmer ground is reached, though there is still much that remains a subject of controversy. One thing seems certain—Greek vessels were built on a system entirely different from that of Egypt, having keel, stem, sternpost, and internal framing with the planks attached edge to edge on the sixth method mentioned above, practically the carvel building of modern times. In all probability the true Greeks, who were originally a race of nomadic shepherds, learned this from their predecessors in the Aegean, but how or where these first evolved so advanced a method of ship construction is still unknown.

2. Galleys.—The galley, an oar-propelled fighting vessel, is also included under the definition at the head of this article. At the time of the siege of Troy the normal large Greek rowing vessel seems to have been a 50-oared boat with a single row of 25 oars on each side, but suggestions could already be seen of still larger vessels, which must have had their oars arranged in some way which allowed more men to work in a given length, and it was this development which produced the bireme and began the process which led to the many-banked galleys of about 300 B.C.

Since galleys were always narrow, shallow craft, there was a very definite obstacle to increasing their length to any great extent because of the weakness of such a vessel against "hogging" or "sagging" strains tending to break it in half. Thus, the number



FROM BJORN LANDSTROM, "THE SHIP," ILLUSTRATION COPYRIGHT 1961 BY BJORN LANDSTROM; REPRODUCED BY PERMISSION OF DOUBLEDAY & COMPANY, INC.

FIG. 1.—ANCIENT GREEK BIREME, A GALLEY WITH TWO BANKS OF OARS

of oars in a single simple row could not be increased very much and it became necessary to find some other method of getting greater power. It would seem the obvious course to put two or more men on each oar, but it was not until much later that the Mediterranean peoples did this. What was done was to arrange the oars in two staggered rows, the uppermost high enough to clear the heads of the lower rowers and their greater length allowing the upper rowers to sit far enough inboard for their legs to be clear of the ends of the lower oars. Another method was to have two oars quite close together at the same level with their two rowers sitting side by side, the foremost oar of a pair being the shorter. This plan was common in medieval times and was apparently employed to some extent in classical or postclassical times also, but it was from the first kind of bireme that the classical trireme and the later many-banked galleys were derived.

The Phoenician vessels of about 700 B.C. mentioned above were biremes of this type, and the trireme followed a little before 500. Its design is still to some extent a subject of dispute, but the view generally accepted is that its third set of rowers sat roughly above the lowest and at about the same level as the second, and that their oars worked on the outer edge of a long outrigger running almost the full length of the hull. This outrigger, once introduced, remained an almost invariable feature of the Mediterranean galley to the very last. Probably the trireme was invented in Greece or in one of the Greek colonies, and it was the Greeks of the last few centuries B.C. who produced the many-banked galleys which have still to be satisfactorily explained.

Galleys carried a mast and sail but were primarily rowing vessels, whereas the merchantman of Greece and Rome was a sailing ship pure and simple. It was much deeper and no doubt also much wider than the galley, without a ram and with stem and sternpost curving down to a comparatively short keel. Both ends rose well above the general level of the sides, but this was more pronounced at the stern, where the chief accommodation was situated. By Roman times, and probably before, the hull was what would now be called carvel-built, but one noticeable feature was that the ends of the deck beams usually projected through the side planking, as they had done in ancient Egypt.

The sail was what is called a square sail, set on a yard slung horizontally from the mast and carried more or less across the ship's fore-and-aft line. In early Egyptian vessels the mast had been like a narrow inverted V, but this had been replaced by a single stick long before classical times. At first there was only one mast and one sail, but by the beginning of the Christian era Mediterranean merchantmen had a second much smaller sail on a mast projecting over the bows, and could also set a triangular topsail or topsails above the maih yard. Some representations show the foresail a good deal larger and its mast more nearly upright, but the other arrangement was probably the more usual. The steering gear was inherited from Egypt, at least in principle, and consisted of a large paddle-shaped rudder on each quarter, held to the ship's side, but free to rotate on its axis and worked by a thwartship tiller. This side rudder, which was found in galleys as well, had obviously been evolved from a mere steering oar, free to move in any direction, and that step had been taken in Egypt before 1500 B.C.

II. MEDITERRANEAN AND NORTHERN VESSELS

1. Mediterranean Craft.—It will be convenient to take the story of Mediterranean sailing ships down to about A.D. 1300 before turning to parallel developments in the north. Actually, there was surprisingly little change in hulls, but there was a complete transformation in rig by the substitution of the lateen for the square sail. The lateen belongs to the family of fore-and-aft sails, carried so as to receive the wind on either side but to keep the same edge forward, whereas the square sail does just the reverse. The lateen is triangular in shape and is set on a long yard coming down nearly to the deck forward and rising well above the masthead at the other end. Evidence which came to light in 1955 showed that this sail was in use in the eastern Mediterranean at least as early as the 2nd century A.D. Where it was first used is uncertain, though the probabilities seem to point to Egypt or the Persian Gulf; in any case it is clear that the tide of Arab and Muslim conquest was mainly responsible for its spread. Even in modern times the area in which the lateen is the typical sail of local craft is much the same as that over which Muslim control once extended or threatened to extend.

The rig of a medieval lateener comprised two masts and two sails, the larger forward, and this rig was carried on a carvel-built hull with flush planking attached to a framework of keel, stem, sternpost and ribs. The deck beams still often projected through the side planking and the steering gear was still a semi-permanent oar-shaped rudder on either quarter. Except for an increase in the accommodation aft and a consequent widening and raising of the stern, the hull was similar to that of a Roman ship.

2. Northern Craft.—Northern vessels of the 13th century had developed on different lines from their southern rivals and had, as has been said, a much shorter history, but in their own way they were no less efficient, while in one respect—that of steering gear—they were superior. All evidence indicates that they had descended from simple dugouts, and many such craft have been unearthed in various countries; but it has not been possible to say with confidence that any of these are actually older than the first of the more advanced types that have been discovered, while early at-

FROM BJORN LANDSTROM, "THE SHIP," ILLUSTRATION COPYRIGHT 1961 BY BJORN LANDSTROM; REPRODUCED BY PERMISSION OF DOUBLEDAY & COMPANY, INC.

FIG. 2.—LATEEN-RIGGED SHIP, FIRST USED ABOUT A.D. 100 IN THE EASTERN MEDITERRANEAN AREA

tempts at depicting northern vessels are too crude to show much.

The earliest-known specimen of an actual northern boat is believed to date from about 300 B.C. and was found in Als, Den., in 1921. Its dugout ancestry is clearly shown by the way in which its two ends are carved from solid blocks, but between them it is formed of five planks overlapping in clinker-built fashion and sewed together. These planks, which must have been worked with an adze, have projecting cleats left on the inside at intervals of rather more than three feet and bent timbers are tied to holes in these cleats and kept in shape by passing through slots at the ends of the thwarts and in another series of transverse timbers beneath them. This boat is about 45 ft. (14 m.) long.

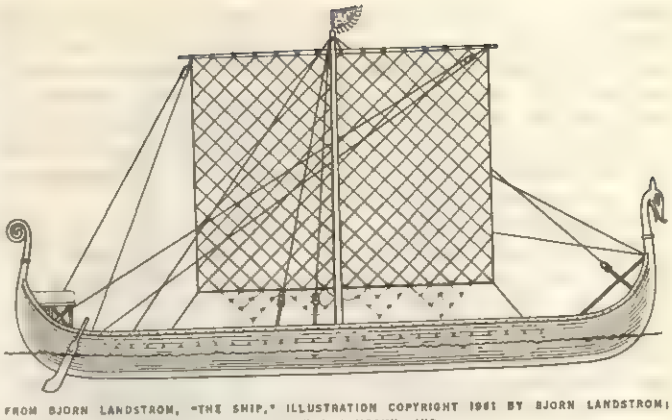
The remains of two vessels discovered at North Ferriby in Yorkshire in 1937 show a different form of development from the primitive dugout, or perhaps from the raft. In their case there is a flat bottom formed from three planks curving upward at the end with the side planks curving round to meet it. The planks are much heavier than those of the Als boat and are fitted edge to edge with a sort of rudimentary tongue-and-groove joint; they also are sewed together and have internal battens covering the seams. The bottom planks have internal cleats, but the short, straight transverse timbers run through holes in these instead of being lashed to them; how the side planking was supported is uncertain. The date of this find is believed, on geological evidence, to lie somewhere between 200 B.C. and A.D. 100.

As far as can be judged, this type of construction was a dead end, perhaps merely local; certainly later developments in northern Europe followed more on the lines of the Nydam boat, found near Flensburg in Schleswig in 1863. In this 75-ft. (23-m.) vessel, dating from about A.D. 250 or a little later, one of the chief problems of the primitive shipbuilder, the treatment of the ends, was solved by the use of an almost modern-looking stem and sternpost, though there is still no true keel but only a centre plank rather heavier than the others. The planks, too, are nailed together in modern clinker-built fashion, but the ribs—cut to shape instead of bent—are still lashed to cleats on their inside. This vessel shows another feature which remained characteristic of northern ships for almost another 1,000 years, the single side rudder on the starboard (steer-board) quarter.

The remains, or rather the traces, of a similar, rather larger boat dating from the early part of the 7th century were found at Sutton Hoo in Suffolk in 1939. Many objects of great interest were found in the excavation, but the actual wood of the boat had entirely disappeared, though changes in the sand where it had lain made it possible to determine the vessel's shape and even details of its construction, which was almost the same as that of the Nydam boat.

Both these last two boats were purely rowing craft, but those unearthed at Gokstad and Oseberg in Norway in 1880 and 1903 were definitely sailing ships, though still using oars as well. They date from about A.D. 900 and 800, respectively, and are built on much the same system as that from Nydam, clinker-built with ribs lashed to cleats and double ended, with the stern almost exactly similar to the bow. Each carried a single square sail as did the larger Viking ships. The Vikings' fighting ships, or longships, for short voyages in comparatively sheltered waters, were shallow, narrow in the beam, and pointed at both ends, and were thus maneuverable with oars in creeks and bays. Probably the largest had but ten oars to a side. It seems likely that a ship's complement would contain twice as many men for fighting as for rowing; therefore, a 20-oared vessel would carry 60 men. The warriors' round and painted shields hung outside along the bulwarks, and the vessel was steered by an oar at the starboard side. Prow and stern rose high, the former being carved, most often as a snake's or dragon's head.

The Viking *hafskip*, for the carriage of families and goods over open sea in summer, sometimes was more than 70 ft. (21 m.) long and up to 20 ft. (6 m.) wide. Relying more on its large square sail (suspended from a 40-ft. mast) than on its oars, it had fewer rowers than the longship. It was steered by a rudder on the starboard quarter; had two boats, one on board, the other in tow; and could be covered with awnings. On the voyages of colonization



FROM BJORN LANDSTROM, "THE SHIP," ILLUSTRATION COPYRIGHT 1961 BY BJORN LANDSTROM; REPRODUCED BY PERMISSION OF DOUBLEDAY & COMPANY, INC.

FIG. 3.—LONG SHIP OF THE TYPE USED BY WILLIAM THE CONQUEROR, AS SHOWN IN THE BAYEUX TAPESTRY

such a ship would carry wives, children and livestock, stores, and at least 30 men. A 19th-century replica of the Gokstad ship made a successful Atlantic crossing in 27 days.

This double-ended type of ship remained in use for several centuries, but the method of lashing the ribs to cleats was superseded by nailing, and the keel plank was replaced by a true keel. This stage had been reached by the boats found in 1933–34 at Ohra near Danzig and believed to date from about 1000; the abandonment of internal cleats was perhaps due to the use of saws instead of adzes for shaping the planks.

The ships used by William I in 1066 (shown in the well-known Bayeux tapestry) had probably reached the same state of development. Some of them have their masts supported by shrouds, and this suggests that by the time the tapestry was made—probably not earlier than 1150—northern ships were able to sail with the wind at least abeam. Roman ships had used shrouds centuries before, but this is their first appearance in the north; after this they are almost invariably shown in ships on seals and elsewhere.

From such evidence it can be said that there was little change during the 12th century apart from the fitting of light "castles" at either end of the larger vessels; but at the end of that period came one of the great steps in the history of sailing ships, the introduction of the stern rudder. When or where this first appeared is uncertain, but the date cannot have been far from 1200 and the place was probably the Netherlands or thereabouts. The step was important, not only for the increased efficiency it gave, but also for the fact that it was followed—almost necessarily—by a differentiation between bow and stern and the transformation of the double-ended northern ship into something far more like its contemporaries in the Mediterranean. When about 1300 the nations of southern Europe recognized the superiority of the new type and adopted both the stern rudder and the square sail, a standard European ship was not far off, though there was still the distinction between the clinker-built northerner and the carvel-built southerner. *See* also RUDDER.

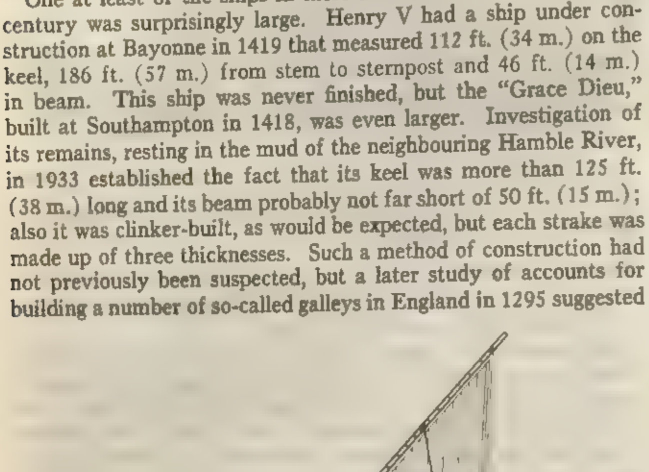
III. MEDIEVAL SHIPS

With the invention of gunpowder and the first use of guns on board ship soon after 1350 there began a process which caused the sailing man-of-war to become more and more distinct from the merchantman. Otherwise the 14th century brought little change, but developments in the 15th more than made up for this. In 1400 ships still had one mast and one sail; before 1450 some of them had three masts and three sails; by the end of the century the largest had four masts and eight sails. Meanwhile, carvel building had spread from the Mediterranean to the north and the ships of all Europe had become similar in both hull and rig, though there were still national or even local variations of type.

In view of the great importance of this short period it must be treated in some detail. The change from one mast to three was very rapid, but was not accomplished in a single step; there was an intermediate two-masted stage, at least in some instances. In inventories of the ships of the Royal Navy in 1410–12 may be

found one ship, and only one, with "1 mast magn" and "1 mast parv"—one big mast and one small. Whether this second mast was before or abaft the mainmast is still only a matter of opinion, but the ship in which it was carried was apparently southern in origin and thus more likely, because of its lateen-rigged ancestry, to have had its smaller mast aft rather than forward. In any case, the two-masted stage was so brief that by 1435 or thereabouts the third mast had come to stay.

One at least of the ships in these inventories of the early 15th century was surprisingly large. Henry V had a ship under construction at Bayonne in 1419 that measured 112 ft. (34 m.) on the keel, 186 ft. (57 m.) from stem to sternpost and 46 ft. (14 m.) in beam. This ship was never finished, but the "Grace Dieu," built at Southampton in 1418, was even larger. Investigation of its remains, resting in the mud of the neighbouring Hamble River, in 1933 established the fact that its keel was more than 125 ft. (38 m.) long and its beam probably not far short of 50 ft. (15 m.); also it was clinker-built, as would be expected, but each strake was made up of three thicknesses. Such a method of construction had not previously been suspected, but a later study of accounts for building a number of so-called galleys in England in 1295 suggested



FROM BJORN LANDSTROM, "THE SHIP," ILLUSTRATION COPYRIGHT 1961 BY BJORN LANDSTROM; REPRODUCED BY PERMISSION OF DOUBLEDAY & COMPANY, INC.

FIG. 4.—THE CARAVEL "NINA," ONE OF THE SHIPS USED ON COLUMBUS' VOYAGE TO AMERICA

that these were of two-layer clinker work and the same may well have been the case in many of the larger vessels of the 14th and 15th centuries.

It is probable that the "Grace Dieu" had three masts, though neither the foremast nor the mizzen would have been of much size or at all elaborately rigged. The foresail would undoubtedly have been a square sail, but the mizzen is more likely to have been a lateen; certainly it became the rule within the next few years for the third mast (and afterward the fourth) to carry a sail of that kind. This was one of the two essentially southern features of the newly developed full-rigged ship; the other, carvel building with flush-fitting planks attached to frames previously erected, reached the north a little later. It may be that it had been used on the west coast of France as long ago as the time of Julius Caesar, who fought a naval battle on the Loire in 56 B.C. and described the ships of his opponents, the Veneti, in some detail without suggesting that their construction was in any way radically different from that of Mediterranean ships. In any case, Brittany seems now to have been the centre from which the new method of building began to spread, but it was soon almost universal for large ships, whether men-of-war or merchantmen.

Toward the end of the 15th century ships of moderate size had three masts and five sails: main, fore, mizzen, main topsail, and the spritsail under the bowsprit; these were, for instance, the sails carried by Columbus' "Santa Maria" in 1492. Considering that bowsprits had been present for more than 200 years, the spritsail was surprisingly late in appearing, as it did, at about the same time as the main topsail in the third quarter of the 15th century.

The original function of the bowsprit is somewhat doubtful; it has usually been supposed that it was to give a better lead for the bowlines, which controlled the windward edge of the mainsail, and it is certain that it was afterward used for a similar purpose, but there is much to be said for a suggestion that the first bowsprits were used in connection with the anchors.

Besides the sails just mentioned, the largest and most elaborately rigged ships of about 1500 would carry a fore-topsail, a main topgallant sail above the topsail, and perhaps a lateen topsail above one or both of the two lateen mizzens, the aftermost of which was usually called the bonaventure. They were still armed for the most part with a great number of small guns disposed in the towering structures which had grown from the two castles of the 13th century, but the time had almost come for larger guns to be carried and for them to be mounted between decks and to fire through openings in the main hull, as they did until the last days of sailing men-of-war and even later. The invention of these gun ports is traditionally ascribed to a French shipbuilder in 1501, and the date, at least, is approximately correct.

1. Caravel and Galleon.—Another change came about now in the shape of the stern. Starting from the pointed shape produced by giving the double-ended ship a straight sternpost, to carry the rudder, the stern had gradually had its upper part widened more and more by means of a transverse timber, the transom, on which the aftercastle was based, and as a result the lines of the stern had become fuller and fuller, until the shape of its lower part was not far from being a quarter of a sphere. In the new form, which probably originated in the south and in a comparatively small type of vessel, the caravel, the attempt to make the planking conform to the required shape was abandoned and the stern from a little below the water line to the transom was given a flat finish, while the lines below this were made much finer.

The true caravel was a lateen-rigged vessel with three masts, the largest forward. Being a southerner it was carvel-built and it is more than likely that the two words were connected. In any event, they soon became hopelessly confused. The "Santa Maria" is often said to have been a caravel, but this is incorrect, though its two companions were caravels, at least in the shape of their hulls.

Then came the galleon, in which the general principles of the design of sailing men-of-war were finally established. As the name implies, it had something of the galley about it, though it was still purely a sailing ship. The Mediterranean galley had retained the ram of its classical predecessor, but had raised it from the water line and lengthened it into a long beak, and this feature with a square-ended forecastle rising abaft it was incorporated in the galleon in place of the triangular overhanging forecastle of the previous carrack type. At the same time, the hull was made longer, so that the keel became about three times the beam instead of two and a half times or less, while the number of heavy guns between decks was increased until they ran in one or two tiers for the full length of the broadside. This last change was well shown when the famous "Henry Grace à Dieu" of 1514 was rebuilt in 1536, though in shape it remained a carrack rather than a galleon, as did most other large northern ships for another 40 years or more.

2. Notable Ships.—Competition among various countries in the first half of the 16th century produced a number of outstanding ships such as the Scottish "Great Michael," the English "Henry Grace à Dieu" (or "Great Harry"), the Portuguese "São João" and others, but none of these was as large as the "Grace Dieu" of

1418 and after their time there came a reaction. The biggest ship of Elizabeth I's reign, the "Triumph" of 1561, was only 100 ft. (30 m.) on the keel and 40 ft. (12 m.) in beam, while a Portuguese vessel captured in 1592 and then believed to be the largest ship in the world was about 25 ft. shorter and 3 ft. narrower than the "Grace Dieu" of nearly 200 years before.

Some idea of the size of merchantmen at this time can be obtained from the list of the fleet which met the Spanish Armada in 1588. The largest of the privately owned vessels was of 400 tons, the smallest 20. Of these some of the larger were really private men-of-war built as such, while the smaller were mere coasters of no fighting value. The average tonnage of the 30 ships equipped by the city of London was 150, which would correspond to a keel length of 60 ft. (18 m.) at the most.

By this time, however, the opening of the sea route to India had led to the building of large state-owned merchantmen in Portugal, the ship just mentioned being one of these, and when the English and Dutch began at the opening of the 17th century to force their way into the same waters, they too began to use larger merchantmen: indeed, the Netherlands East Indiamen were larger than any man-of-war.

IV. 17TH AND 18TH CENTURIES

The first half of the 17th century saw the culmination of the galleon type in the "Prince Royal" of 1610, the "Sovereign of the Seas" of 1637, and its French contemporary, the "Couronne." The two English ships were three-deckers and the "Sovereign" was the first ship to carry what was for nearly a century the standard armament of a first rate, 100 guns. The "Couronne," though of almost the same size as the "Sovereign," was a two-decker of 72 guns, an early example of the French tendency to build ships very large for their armaments. The same period also saw certain noticeable changes in sail plan. The fourth mast (the bonaventure or after mizzen) disappeared, the lateen mizzen topsail was replaced by a square sail, the spritsail topmast appeared, standing upright on the end of the bowsprit, and royals above the topgallants became possible, though probably seldom set.

1. The Frigate.—Then came the frigate, not the well-known type of Nelson's day but its 17th-century namesake. Originally the *fregata* was a small member of the galley family used in the Mediterranean chiefly as a dispatch boat. The name was then applied by the Spaniards to the small fast vessels used to bring treasure from America and from them passed to the ships of the semipiratical privateers of Dunkirk, from which the first English frigates were copied. The exact essentials of such a frigate are by no means clear, but they certainly included an increase in length (or a decrease in beam) and a reduction in top hamper. Soon, however, the name ceased to mean much more than a ship of modern design, since even three-deckers were occasionally called frigates, at least in England.

The chief reason for the growth of the frigate was no doubt the fact that the second half of the 17th century was a time of frequent naval fighting between well-matched opponents and of many hard-fought actions between enormous fleets. Experience soon showed that ships armed almost solely on the broadside should fight in line ahead in a prearranged order, and this in its turn led to the conclusion that the ships in such a line must be of a certain minimum strength, 40 guns or more. Such was the origin of the term ship of the line or line-of-battle ship—abbreviated later to battleship.

2. Improvements in Rigging and Equipment.—Just before the appearance of the 17th-century frigate, sometime about 1630, English shipbuilders had made a change in the shape of the stern by rounding off the sudden turn between the side planking and the flat stern of the galleon type sufficiently to admit of working the planking round to the sternpost and transom while maintaining the fine lines below water. Although the flat stern continued to occur at intervals almost to the last days of sailing men-of-war, especially in the smaller vessels, the new form of stern was more or less a distinguishing feature of English ships for at least 50 years; it was not imitated by other countries until the end of the 17th century or the beginning of the 18th.

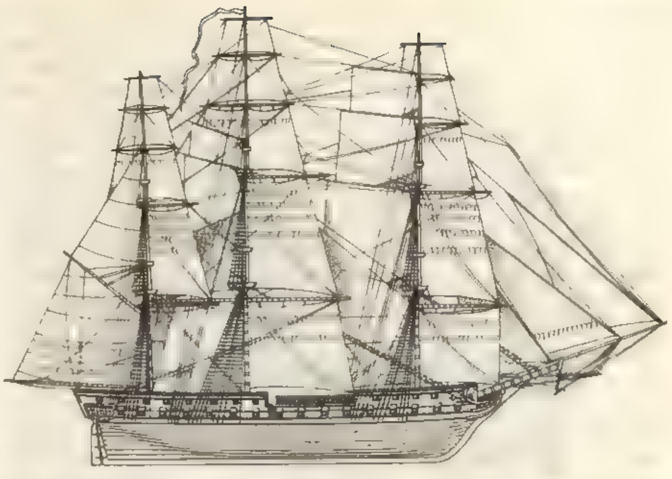


FIG. 5.—U.S. FRIGATE "PRESIDENT," LAUNCHED IN 1800

By then the French Navy had become important, not only because of its great expansion under Jean Baptiste Colbert, but also because of the quality of its ships. This was not merely a matter of increased dimensions; another reason was the greater attention paid in France to the scientific side of naval architecture. Because of these two factors the French, and at a later date the Spaniards, built ships more efficient, class for class, than any produced in England. The first half of the 18th century was, indeed, a period of comparative stagnation in English shipbuilding, at least as far as men-of-war were concerned. It was the period of "establishments" laying down standard dimensions for ships of each class and thus cramping the initiative of the various builders. Each successive establishment certainly allowed some increase in size, but this was never enough to overtake what was being done on the continent.

The chief interest of the period lies in its developments in rig and equipment. Staysails had been in use for 50 years or more, and there were already such sails on the three stays leading to the bowsprit, but in 1700 or thereabouts a new sail of the same kind—the jib—was introduced, carried between the fore-topmast head and the jib boom, which was a small spar fitted as a prolongation of the bowsprit. Obviously this sail demanded the removal of the spritsail topmast, but an attempt was made for at least another 25 years to carry both together; then the spritsail topmast vanished and its sail was shifted to beneath the jib boom. The bobstay, to hold down the bowsprit, seems to have come just before the jib. Why it should have taken so long to produce so apparently essential a piece of rigging is hard to say, but no trace of a bobstay earlier than 1690 has yet been found. Bowsprit shrouds came a little later. At the other end of the ship the part of the lateen mizzen before the mast was done away with and, though the biggest ships kept the whole yard for a long time, it was gradually replaced by a gaff such as was carried in the 19th century; by 1800 the long mizzen yard was a thing of the past.

The introduction of the steering wheel was a less conspicuous but extremely important change. From the middle of the 16th century to the end of the 17th the tiller had been worked by means of the whipstaff, a vertical lever acting on its inboard end and passing through a pivot in the deck above, a device which must have been much less efficient than the wheel with its ropes leading to the end of the tiller. In England the wheel was first used about 1705; there is a suggestion that something of the sort may have been tried in France somewhat earlier, but other countries seem to have waited a few years before adopting it.

By 1750 the smallest ships considered fit for the line of battle carried 64 or 60 guns, though there were still smaller two-deckers of 50 and even 40 guns. After them came the 24-gun ships, which were really small two-deckers with only four ports on the lower deck, and it was these which developed into the frigates of Nelson's time, where the lower deck, though still called in England the gun deck, carried no guns at all. The first of these new-type frigates were built in England in 1756 and in France a few years

FROM BJORN LANDSTROM, "THE SHIP," ILLUSTRATION COPYRIGHT 1961 BY BJORN LANDSTROM; REPRODUCED BY PERMISSION OF DOUBLEDAY & COMPANY, INC.

FIG. 6.—ADMIRAL NELSON'S FLAGSHIP, THE BRITISH SHIP OF THE LINE "VICTORY"

sooner; they carried 28, 32 or 36 guns, but soon grew, until the largest of them actually mounted as many as 56 guns, though classed, because of the peculiarities of official rating, as 44s. One of these heavy frigates, the USS "Constitution" of 1797, has been more or less restored to its appearance in the War of 1812; while in England the 100-gun "Victory," built in 1765, is preserved in drydock as it was at Trafalgar in 1805.

At the time of its launching the "Victory" was the largest British ship afloat, but even so it was hardly larger than the "Grace Dieu" of 1418 and no English (or British) ship in the meantime had been as large. It was 186 ft. (57 m.) on the gun deck and 52 ft. (16 m.) in beam, whereas its Spanish contemporary, the "Santissima Trinidad," measured 204 ft. by 53 ft. and the French "Commerce de Marseille" of 1790 was 208 ft. by 54 ft. Before the end of sailing men-of-war the beam of the largest ships had risen to as much as 60 ft.; but the length never went beyond the 210 ft. of the USS "Pennsylvania" of 1837, and about 205 ft. was usually the maximum. The reason for this was the tendency of long wooden ships with guns all along the broadside to "hog" or drop at the ends. Merchantmen did not present quite the same difficulty and in their case with improved methods of construction it proved possible to make the length much greater.

V. THE 19TH CENTURY

In merchantmen the increase of length, both actually and in relation to the beam, was associated with the rise and development of the clipper ship in the middle of the 19th century. At the time of the Napoleonic Wars the largest and finest merchantmen afloat were those in the service of the East India Company, heavily built vessels not unlike men-of-war and actually used as such on occasion. The largest of them, the 1,200-ton class, measured about 165 ft. (50 m.) by 42 ft. (13 m.). With the ending of the company's monopoly of trade with the East in 1833, the various owners who had supplied it with ships began trading on their own account and were at once driven by mutual competition to employ ships of greater efficiency as speedy carriers of cargo. The first of these were comparatively small, but by 1842 they were back at 1,200 tons with the "Prince of Wales" measuring 179 ft. by 39 ft., much longer and at the same time narrower than the older type.

1. Clipper Ships.—By then the clipper was coming to the fore. It is a matter of controversy which was the first clipper ship and what were the essential characteristics dividing it from its predecessors, but on the whole the U.S. "Ann McKim" of 1833 seems to have the best claim to the title. In any case, it was a matter of gradual evolution rather than sudden change, since this ship was little more than a slightly enlarged version of the existing Baltimore clipper type, well-known as successful privateers and slavers. Other claimants are the British "Glentanar" of 1842, developed from the schooner "Scottish Maid" of 1839, and the U.S. "Rainbow" of 1845. In the case of the "Scottish Maid" and its Aberdeen-built successors, it is possible that the most obviously revolutionary feature of the design, the long overhanging bow, was primarily a means of cheating the existing rule for measuring tonnage and that the increased speed it gave was to some extent a by-product. However this may be, clippers, in the strictest sense of that ill-defined and very elastic term, were being built in ever-increasing numbers on both sides of the Atlantic by the end of the 1840s.

They were chiefly employed in the tea trade from China, in carrying passengers from the eastern part of the U.S. to San Francisco after the discovery of gold in California in 1848, in similar voyages from Great Britain to Australia a few years later, and finally in the Australian wool trade. On the whole, the U.S. clippers were larger and more powerful ships than the British and therefore capable of greater maximum speeds, but in the matter of all-round performance there was little to choose between the two types. The American "Lightning" of 1854 is believed to have made the best day's run ever recorded by a sailing ship, 436 mi., but the palm for consistently fast passages in all conditions should probably be given to two of the last British tea clippers, the "Thermopylae" and "Cutty Sark" of 1868–69.

Apart from their finer lines as compared with ships of an older

FROM BJORN LANDSTROM, "THE SHIP," ILLUSTRATION COPYRIGHT 1961 BY BJORN LANDSTROM; REPRODUCED BY PERMISSION OF DOUBLEDAY & COMPANY, INC.

FIG. 7.—BRITISH CLIPPER SHIP "CUTTY SARK"

type and the relatively greater spread of canvas which they carried, the clippers owed much of their speed to a further increase in length as compared with beam. The process had, as has been said, begun with the first successors of the old East Indiamen, but was taken much further in the clippers. The "Rainbow" had a length about 4 times its beam, as had been the rule toward the beginning of the century, but in the "Lightning" the ratio reached 5.6. British builders went in for ships still longer and narrower; the usual standard, from the "Stornoway," built three years before the "Lightning," to the "Cutty Sark," was a length of about six times the beam, and this was sometimes exceeded, as in the "Lord of the Isles" of 1853 with 6.4 and the "Queensberry" of 1856 with the extreme figure of 7.2. This increase of length was to a great extent made possible by the advent of steam, which allowed these comparatively unhandy ships to be towed in and out of harbour.

Meanwhile, there had been many improvements in construction and rig. Even with new methods of framing introduced soon after the beginning of the 19th century it was still difficult to build wooden ships beyond a certain length. The use of iron did away with this difficulty and the "Lord of the Isles" was an iron ship, but this was found to be bad for a tea cargo and a compromise was effected by means of the composite method of building with wooden planking on iron frames. In the matter of equipment the great step was the introduction of wire in place of hemp for standing rigging in the early 1850s, soon followed by the invention of the double-topsail rig, which, though perhaps no more efficient, was much easier to handle than the large single sails previously carried.

2. Last of the Sailing Ships.—By the time these changes took place, the sailing man-of-war was rapidly becoming more and more obsolete. There had been paddle steamers in the British fleet from 1822 and some of the last of these had been ships of considerable importance, but it was not until the adoption of the screw in the 1840s that it became possible to combine steam propulsion with the complete broadside armament of the larger classes. When once this stage had been reached, ship after ship was given engines of steadily increasing power, till by 1850 no more purely sailing men-of-war were being built. Masts and sails were retained for a long time, but gradually the two systems of propulsion exchanged roles, and the man-of-war, instead of being a sailing ship with an auxiliary engine, became a steamship with auxiliary sail.

In the case of merchantmen the distinction between steamships and sailing ships was made much sooner and more definitely. It is true that unsuccessful attempts to give deepwater sailing ships auxiliary steam power were made before 1840 and that steamers, especially those employed on long voyages, continued to have masts and sails, though they seldom had anything like full ship rig; but on the whole the sailing merchantman was one thing and the steamship another almost from the first.

In one trade after another, particularly those where perishable cargoes had to be carried, steamers began to take the place of sailing ships. The opening of the Suez Canal in 1869 was a heavy blow to the sailing ship, since, besides shortening the route to the East, it reduced the distance between coaling stations and so allowed steamers to devote more space to cargo. After this the surviving China clippers turned for the most part to the Australian wool trade and it was in this that such ships as the "Thermopylae" and "Cutty Sark" made their reputations. Then came grain from Australia and nitrate from the west coast of South America, till at length the Panama Canal, opened in 1914, gave steamers the advantage even there.

For nearly 250 years the standard rig of men-of-war and of the larger ocean-going merchantmen had been that of the full-rigged ship with three masts all carrying square sails. When the square mizzen topsail was first introduced, the long yard of the lateen mizzen beneath it had made it impossible to carry a square course or lower sail; the yard spreading the foot of the topsail, the "crojack" in English, was in fact called in French the *vergue sèche* or "barren yard," because it set no sail of its own. The square mizzen course did not appear until about 50 years later, in the middle of the 19th century.

Meanwhile, many smaller vessels, in particular the collier barks of the North Sea, had carried no square sails on their mizzenmasts. It is probable that originally the term bark marked a distinction in size and in form of hull rather than in rig; but when large ships began once more to do without square sails on the mizzenmast, they were called barks or barques, the French form of the word being for some reason usually preferred. This became the fashion soon after the clipper ship era, when it was necessary for sailing ships to reduce their crews, and thus their running costs, to meet the competition of steam.

At the same time the need for greater cargo-carrying capacity produced ships both larger and fuller bodied, though the increased length and power did a good deal to compensate for the loss of the finer lines of the clippers. Composite building had only a short life, and iron (and afterward steel) ships became the rule; while with the increase of length it was found desirable to add a fourth mast, the result being the four-masted ship or four-masted barque, according to the rig of the aftermost or jigger mast. Four-masted square-rigged vessels had been almost unknown since about 1630, but not entirely so; the French privateer "L'Invention" of 1801 had been a four-masted ship in the strictest sense of the term, while the famous "Great Republic" of 1853, the largest of the U.S. clippers, 325 ft. (99 m.) long, had been a four-masted barque. On the whole, the barque rig proved the more satisfactory and at the beginning of the 20th century the four-masted barque could be considered the standard type, though there were some noteworthy exceptions.

For the most part this final development of the sailing merchantman was in the hands of British, French, and German builders. In the U.S., which had taken so conspicuous a place in the days of the clippers, the building of square-rigged vessels declined rapidly after 1860; but as some compensation for this there came a boom in large fore-and-aft rigged vessels, the original two-masted schooner of the 18th century becoming by degrees three-masted, four-masted, and even six- and seven-masted, the climax being reached in the "Thomas W. Lawson" of 1902. At the same time, many square-rigged vessels bought from foreign owners were rerigged as many-masted schooners.

French shipowners had the benefit of a system of bounties which are said to have been high enough to allow sending a ship round the world in ballast, without cargo, and still making a profit on the voyage, the result being that some very large ships were to be found under the French flag toward the end of the sailing-ship era.

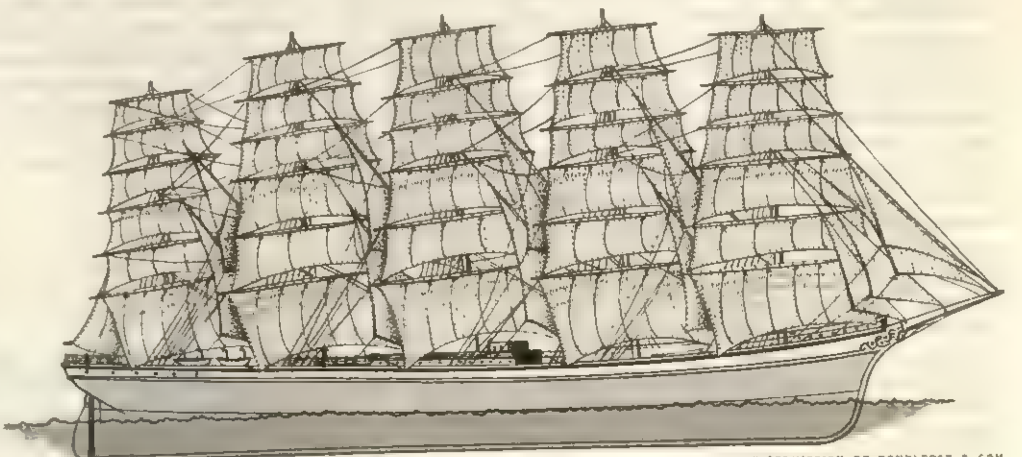


FIG. 8.—"PREUSSEN," ONLY FIVE-MASTED, FULL-RIGGED SHIP EVER BUILT

The five-masted barque "France" of 1890, with a length of 361 ft. (110 m.), was built in Scotland, but the next ship of the name, the largest sailing ship ever built, was launched at Bordeaux in 1911. It, too, was a five-masted barque and was 418 ft. (127 m.) long with a gross tonnage of 5,633. This ship was wrecked in 1922; its predecessor had been lost at sea in 1901.

In the same way the first of the German five-masters, the "Maria Rickmers," was built in Scotland in 1890 and was of about the same size as the first "France"; but the "Potosi" (1895), "Preussen" (1902) and "R. C. Rickmers" (1906) were all built in Germany and each was larger than the one before, the last being 410 ft. long and measuring 5,548 tons. All save the "Preussen" were rigged as barques.

World War I caused the loss of many sailing ships and it was no longer financially practicable to replace them. The Germans built a few curious five-masted vessels with auxiliary motors, rigged as schooners with square topsails on the first and third masts; while the Danish East Asiatic Company had a large auxiliary five-masted barque, the "København," built in Scotland in 1921, but this was intended primarily for a training ship. It, too, was unfortunate, being lost at sea in 1929. Meanwhile, the survivors of the days of sail were lost or laid up, until in 1929 the last British square-rigged sailing ship left in service was also lost; this was the "Garthpool," built as the "Juteopolis" in 1891. A few, mainly German in origin, passed into Finnish ownership and except for a break caused by World War II continued to bring grain from Australia as late as 1949. In 1950 the last two of these were for sale. The only large square-rigged sailing ships still in service in the 1960s were those used by several countries as training vessels for their naval cadets. (Rr. C. An.)

VI. INTRODUCTION OF STEAM AND IRON

A revolution in the history of the ship may be said to have occurred with the changes from sails to steam engines for propulsion and from wood to iron for construction. These proceeded together, but at first slowly. There was still much ignorance of the principle of flotation by displacement, and it was urged by the unenlightened that iron would not float and was therefore unsuitable for ship construction. Even among those who realized the fallacy of this argument it was asserted that an iron ship would be far more easily damaged in the event of her touching the ground than a wooden one, while there existed the real difficulties of preserving the bottom from the action of the sea and fouling by weeds and barnacles, and of compensating the compass for the errors produced by local attraction. With regard to the strength of the ship, experience showed that iron construction was better able to withstand rough usage than was wood, and examples, such as the "Garry Owen" in 1834, were not lacking of iron and wooden ships being stranded together by the same gale and under similar circumstances and the iron ship getting off little the worse, while the wooden ship became a total wreck. Another remarkable instance of the endurance of iron ships was that of the "Great Britain," which, in 1846, ran ashore in Dundrum Bay in Ireland and settled on two detached rocks; it remained aground for 11 months, was subsequently got off, and afterward did good service. In due course a suitable composition was discovered for painting the underwater surface of iron ships, while trials carried out in the "Rainbow" at Deptford and the "Ironsides" at Liverpool went far toward providing the solution for correcting the compass.

One of the earliest iron craft on record was a boat apparently intended for passenger service, built on the banks of the Foss River in Yorkshire in 1777. In 1787 a canal lighter was constructed with a shell of iron plates, and for many years iron and wood were used in conjunction for the construction of what were known as composite ships.

For warships iron was at first objected to because it was thought that the enemy's shot would cause more serious damage to them than it would to a wooden ship, but this again was proved to be a fallacy. The first iron steamer was the "Aaron Manby," built at Horsley in 1821 and assembled in London.

1. The First Steamboats.—One of the earliest proposals for a steam-driven boat was made in 1690 by Denis Papin. In 1707 he built a paddle boat and tried it on the Fulda, but the paddle wheels were turned by man power, not by steam. In 1736 Jonathan Hulls, of Gloucestershire, patented a steam tugboat, but it was never tried. The first experimental steamboat to be built, in France, by Comte J. B. d'Auxiron in 1774, foundered before it could be tried. In 1775, however, Jacques C. Périer contrived, for the first time in history, to move a small boat by steam power on the Seine at Paris. After unsuccessful experiments with a 43-ft. steamboat and palmipede or "duck-foot" paddles, on the Doubs in 1778, the Marquis Claude de Jouffroy d'Abbans built his "Pyroscaphe" of 182-tons displacement, fitted with a double-ratchet mechanism to produce continuous rotation of the paddle wheels. In 1783 this vessel, the first really successful steamboat, mounted the Saône River near Lyons. In 1787 James Rumsey drove a boat on the Potomac four miles an hour by means of a power pump. About the same time John Fitch produced his oar-driven steamboats.

A more practicable device was to be the paddle wheel. The "Charlotte Dundas," constructed by William Symington in Scotland in 1802, was one of the earliest of these vessels. It proved its utility for towing work on the Forth and Clyde Canal. Robert Fulton, having witnessed the success of this craft, in 1807 constructed the "Clermont" on the Hudson River in the U.S. The engines for the vessel were made by Boulton and Watt in England. It proved popular as a passenger boat between New York and Albany. The first steamer to make a regular sea voyage was the "Phoenix" which, in 1809, steamed from Hoboken, N.J., to Philadelphia, Pa. In 1812 Henry Bell built his steamer "Comet," which carried passengers between Glasgow, Greenock and Helensburgh. It was 43 ft. long, 11 ft. broad and 5½ ft. deep and was driven by a one-cylinder engine. The success of these early steamers soon produced others. In 1815 the "Margery" of 38 tons, constructed on the Clyde, was brought through the Forth and Clyde Canal and then down the east coast to the Thames. In 1816 this same vessel, renamed the "Elise," was the first steamer to cross the channel to France.

In Great Britain the steamship was first used only as a passenger carrier and tug along rivers and canals, but gradually increased competition forced the shipowners to seek other spheres and in doing so they found that steamers could perform short coasting voyages. The steamship then became a passenger carrier along the coast to the various holiday resorts, its novelty being the principal attraction. The owners soon realized that the steamship was capable of much more than this and vessels were built to cross the North Sea, the Straits of Dover, and the Irish Sea. They were modeled on the sailing ships that they replaced and all carried auxiliary sail either on masts or on their funnels, as the low-pressure machinery was wasteful and incapable of giving any considerable speed. The various methods of propulsion which had been experimented with in the 18th and early 19th century, including a rudimentary screw propeller and jet propulsion, gave way to the side paddle wheel, which remained in favour for many years.

2. Atlantic Crossing.—It was not long before steamship owners aspired to cross the Atlantic by steam. The sailing packet "Savannah," which had been designed to run on the service between New York and Le Havre, was given auxiliary steam machinery and crossed the Atlantic from Savannah, Ga., in May and June 1819. It only used its engines for about 85 hours of the voyage, but

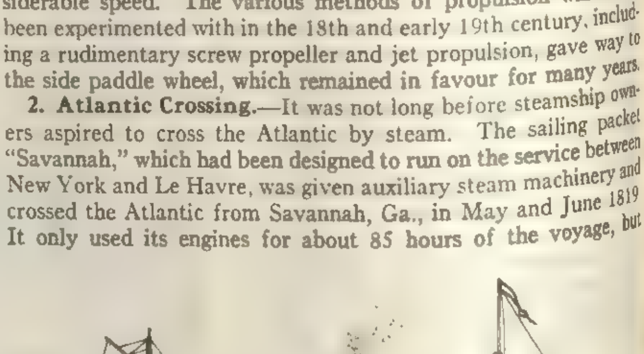


FIG. 9.—ROBERT FULTON'S "CLERMONT"

arrived off Ireland with its coal consumed. Afterward the "Savannah" visited the Baltic, where it aroused great interest. Although this is generally recorded as the first steam Atlantic crossing, the vessel returned to the U.S. under sail alone. The first to use steam power on the Atlantic in a westerly direction was the "Rising Star" of 428 tons, built at Rotherhithe in 1821 for Thomas Cochrane, earl of Dundonald. It had internal paddle wheels, and in 1821–22 crossed from Gravesend to Valparaiso, Chile. In 1824–25 the "Caroline," first steamer in the French Navy, crossed from Brest, France, to Cayenne, French Guiana.

The steamer "Curaçao," built at Dover in 1825, a wooden paddler of 438 tons, was purchased by the Dutch government as a man-of-war but was employed on the mail service to the Dutch West Indian colonies. It left Rotterdam on its first passage to the West Indies in April 1827 and took 28 days to do the voyage, after which it made the regular sailing each year, until it was required as a warship during the troubles in Belgium in 1830, after which it never returned to the mail service.

The first transatlantic steamer of the Royal Navy was the "Rhadamanthus" of 813 tons, which crossed from Plymouth to Barbados in 1832. The next steamer to perform the feat was the "Royal William," whose performance was particularly remarkable because of the fact that it was built in Quebec with the idea of running from that port to Halifax. Samuel Cunard, who afterward founded the great Atlantic company, was one of its owners. Trade depression and an epidemic of cholera spoiled its chances in the trade for which it was designed and it did no better as a tug. Its owners, being forced to consider its sale and thinking that they would get a better price for it in Europe than in Canada, sent it across the Atlantic in 1833. It took 25 days to do the passage and burned 324 tons of coal. Afterward it was sold for £10,000 and later served in the Portuguese Navy as a transport and in the Spanish Navy as a warship until it was finally condemned in 1840.

3. Regular Atlantic Passages.—These early steamship passages across the Atlantic were more or less haphazard, but it soon became the object of the owners to provide a regular service. Pending the construction of suitable tonnage the 703-ton steamer "Sirius," which had been built for the Irish Sea service, was chartered in 1838 by the British and American Steam Navigation Company. It was considered a big steamship in its day and was one of the first steamers to be fitted with a surface condenser, patented by Samuel Hall in 1834, instead of using salt water in its boilers; this represented one of the milestones of steam engineering at sea. It sailed from London to New York by way of Cork with 40 passengers and although it was grossly overloaded to modern ideas it made the passage in safety. Within a few hours of its arrival in New York a very much bigger and finer steamer, the "Great Western," which had been constructed with the idea of continuing the Great Western Railway across the Atlantic, arrived after a crossing of 15 days from Bristol. It had a tonnage of 1,320 and was regarded as the finest steamship of its day.

After this several other Atlantic liners were built but all the services were irregular and maintained by a heterogeneous collection of ships, suitable and unsuitable. It was when Samuel Cunard founded his transatlantic line in 1840 that a new policy in shipping produced a revolution in shipbuilding—the construction of sister ships. He started operations with four transatlantic ships and one small feeding steamer, the "Unicorn," in Canadian waters.

The sister ships "Britannia," "Acadia," "Columbia" and "Caledonia" were wooden steamers built on the Clyde, their tonnage according to the rule then in use being about 1,150 on dimensions 207 ft. by 34 ft. 2 in. by 22 ft. 2 in. depth of hold and their two-cylinder side-lever paddle engines of 740 i.h.p. (indicated horsepower) being sufficient for an average speed of nine knots in favourable circumstances. Charles Dickens crossed on the "Britannia" in 1842. The coal supply of these ships was the chief anxiety of their designers, and their passenger accommodation was not equal to the sailing packets which they rivaled, but the regularity of their passages compensated for the fact that they were frequently beaten by the sailing ships in a fair wind. They were barque-rigged and had a considerable area of canvas which was set whenever circumstances were favourable.

4. Long-Distance Steamers.—In the meantime, although the North Atlantic crossing had attracted popular attention to the exclusion of most other services, great progress was being made in the long-distance routes. When the General Steam Navigation Company was founded in 1824 its promoters had the intention of running steamship services all over the world as material improved, and had every confidence that this would come about.

In 1825 the steamship "Enterprise" had proved that it was possible for a steamer to reach India, although at the same time it proved that it was not a commercial proposition. It had been suggested in 1822 to establish a company for the purpose of maintaining a steam service to India and a naval officer was sent out to arouse popular enthusiasm. Within two years 80,000 rupees had been raised in Bengal by public subscription, to which the government of India added 20,000 rupees and announced that the whole sum would be given as a prize to the first steamer that could contrive two round voyages between Great Britain and India before 1826, the stipulated time for each passage being 70 days.

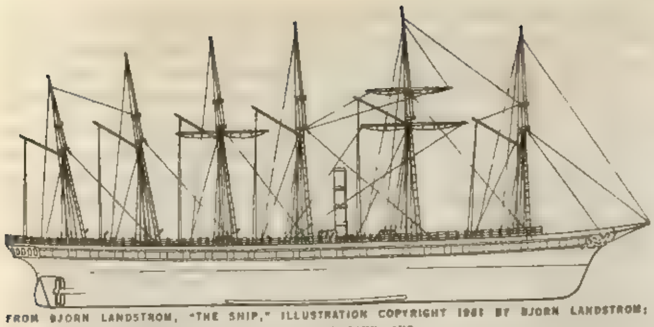
This prize caused a syndicate to purchase the paddle steamer "Enterprise" of 470 tons when under construction on the Thames and to fit it out with a fore-and-aft rig to compete for the prize. Its dimensions were 122 ft. by 27 ft. and it had an engine of 120 nominal horsepower which was designed for nine knots speed but which could be relied upon for six or seven only. The "Enterprise" sailed from Falmouth on Aug. 16, 1825, and reached Calcutta 113 days out, including 10 days spent coaling at St. Thomas and the Cape of Good Hope. The government of India awarded it half the promised prize and then purchased it as a warship.

The improvement and finally the real practicability of the overland route was brought about by the Peninsular and Oriental Steam Navigation Company, which started in 1834 to maintain a mail service between Great Britain and the Spanish and Portuguese ports in competition with the sailing mail packets maintained by the government. It was then the Peninsular service, but in 1840 the service was extended to Alexandria to connect with the East India Company's steamers and the company became the Peninsular and Oriental Steam Navigation Company. It was not until 1854 that the East India Company abandoned its end of the service, although its irregularities and the poor steamships employed on it were the cause of constant complaint.

Steamships of a type practically identical with those of the Cunard Line, but of rather greater tonnage varying from 1,700 to 1,900, were built in 1841 for the West Indian mail service for which the Royal Mail Steam Packet Company had obtained a royal charter. At that period the British West Indian colonies were of much greater importance than in more modern times, and the service was heavily subsidized in order to overcome the great difficulties of coal supply on the route. This service was afterward extended to the Brazilian coast and the Rio de la Plata, causing a steady increase in the size of the ships, although they were long built on the same principle.

5. The Screw Propeller.—Private shipowners who were untrammeled by the conditions of their subsidies were anxious to improve on the paddle wheel and in 1836 both Francis Pettit Smith, an Englishman, and Capt. John Ericsson, a Swede, patented practical screw propellers although the principle was not new. The year 1839 saw a ship built to each of these principles, the "Archimedes" to Smith's patent and the "Robert F. Stockton" to Ericsson's. Neither was a large vessel, but they both proved the superior economy and power of the screw and led to more important ships being built.

The most important of these steamers was the "Great Britain," which was laid down in dry dock at Bristol in 1839 and floated out in 1843. It was noteworthy not only because of its dimensions, which were 322 ft. by 51 ft. by 32 ft. 6 in. depth of hold, which gave it a tonnage of 3,270 according to the old burden measurement, but also because it was constructed of iron in spite of the great prejudice of the navy against that material. It was designed by I. K. Brunel, whose intention was originally to make



FROM BJORN LANDSTROM, "THE SHIP," ILLUSTRATION COPYRIGHT 1961 BY BJORN LANDSTROM; REPRODUCED BY PERMISSION OF DOUBLEDAY & COMPANY, INC.

FIG. 10.—"GREAT BRITAIN," FIRST PROPELLER-DRIVEN SHIP TO CROSS THE ATLANTIC

it a paddle steamer. But he was so greatly impressed by the performance of the "Archimedes" on a cruise round the British coast that he altered the "Great Britain" and fitted screw machinery, reinforced by a big sail area on six masts.

Its machinery consisted of a simple engine with four cylinders each 88 in. in diameter, with a stroke of 72 and with indicated horsepower of about 1,500 at 15 lb. per square inch pressure, a speed of 11 knots being obtained on trial. Its stranding on the Irish coast in 1846, when it had to withstand the gales of a whole winter in an exposed position, finally convinced shipowners that iron construction was both strong and practical.

6. Atlantic Development.—In 1847 the U.S. brought out its first transatlantic steamers, the "Hermann" and "Washington," run by the Ocean Line between New York and Bremen by way of Southampton in return for a heavy subsidy. They were ships of 1,750 tons each with paddle engines of 1,100 i.h.p. giving them a speed of 11 knots, their hulls being on the lines of the less extreme Atlantic sailing packets. They maintained their service with fair success until the reversal of the United States subsidy policy in 1857, when they were sent to end their days in the Pacific. In 1850 a second U.S. transatlantic service between New York and Le Havre was started with the "Humboldt" and "Franklin."

The straight stem of these two ships made them differ greatly in appearance from their British rivals and was an American conception adopted by E. K. Collins when he started the Collins Line (U.S. Mail Steamship Company) in 1850. The first fleet consisted of the wooden paddle steamers "Atlantic," "Arctic," "Baltic" and "Pacific," practically sister ships of 2,860 tons, with side-lever engines of 2,000 i.h.p. They were a great improvement on the existing material and although they were so well built and extravagantly fitted that it was necessary to obtain an additional subsidy they were the most noteworthy ships on the Atlantic in their day. In 1856 they were joined by the "Adriatic," of 3,670 tons gross with engines of about 3,600 i.h.p. designed for a speed of 13½ knots, which marked the high-water mark of U.S. paddle-steamer construction on the Atlantic. When the subsidy on which the company relied was suddenly withdrawn after two disasters the service collapsed and although Cornelius Vanderbilt constructed somewhat similar ships to maintain U.S. interest on the Atlantic they were soon withdrawn and for more than 30 years U.S. steamship development was practically confined to the rivers and coasts of the country and to the Pacific trade.

While these big paddle steamers were being built the screw propeller was being introduced into European Atlantic companies, beginning seriously with the foundation of the Inman Line in 1850. This company, which was intended to improve the emigrant service, began with iron screw steamers—the first of less than 2,000 tons—barque-rigged and still maintaining a full spread of canvas. Its example was soon followed. The North German Lloyd (Norddeutscher Lloyd) and Hamburg-American (Hamburg-Amerika Linie) companies in Germany, and the Compagnie Générale Transatlantique in France, saw the opportunities of the great continental fields of emigration with up-to-date steam tonnage and were soon encroaching on the third-class traffic which was the last stronghold of the sailing packet.

The Cunard Line was endeavouring to carry on in its traditional way, in spite of the fact that by then the naval conditions were considerably relaxed, but it was gradually being forced into line. For its mail ships it remained faithful to the paddle until the early 1860s, the "Scotia" of 3,871 tons which was launched in 1861 being the last and finest of the type. It had, however, changed from wood to iron with the "Persia" of 1855. The first Cunard screw steamer was the "China" of 2,529 tons, built in 1862, which in spite of its smaller size showed its advantages in competitive service trials.

7. "Himalaya."—In 1853 the Peninsular and Oriental Line, although not destined to abandon the paddle for several years afterward, built the iron screw steamer "Himalaya," which was the largest vessel of its type in the world, having a gross tonnage of 3,438 on dimensions 340 ft. by 46.2 ft. by 34.9 ft. depth of hold. Its trunk engines gave it a speed of 13.9 knots on trial, and yet were sufficiently economical to permit it to stow enough fuel to undertake long voyages under steam which showed a profit. After one or two voyages on its owners' service it was taken up as a transport for the Crimean War and so impressed the admiralty that they bought it and employed it as a naval trooper until the 1890s. It was then converted into a coal hulk for use at Portland, where it was destroyed in World War II.

8. "Great Eastern."—The "Great Eastern" was one of the most discussed steamships ever built, and the most historic failure. It was originally conceived by I. K. Brunel on the success of the "Great Britain," and in 1851 a company was floated for the purpose of building it and trading to the East. At that period steam navigation to the East and Australia was greatly handicapped by the lack of coaling facilities, and the "Great Eastern," was specially designed to ply between England and either Calcutta or Colombo, where smaller steamers and sailing vessels could pick up its cargo and passengers and distribute them to various destinations. Its dimensions of 692 ft. on the upper deck by 82.5-ft. beam and 30-ft. draft gave it a gross tonnage of 18,914, and it became necessary to take particular precautions that its hull should have the requisite strength. It was, therefore, given not only a double bottom but a tubular upper deck and was one of the strongest ships ever built. The hull and the paddle engines were built by Scott Russell and Company on the Thames, while James Watt and Company of Birmingham built the screw engines, for Brunel had decided to provide alternate methods of propulsion, the greatest fault in the original design. Scott Russell of the building firm designed the details of its hull and gave it the wave-line principle in which he believed. The paddle engines had 3,411 i.h.p., while the screw engines which drove a four-bladed propeller had 4,886 i.h.p. Altogether 6,500 sq.yd. of canvas were set on its six masts and it was fitted with ample bunker accommodation for a long voyage, in addition to large holds and passenger accommodation. Fearing for the narrowness of the river, Brunel insisted that it should be launched broadside on and was so anxious to avoid the huge mass taking charge that he checked it too soon on the ways and it stuck fast for three months. This delay, and the subsequent work of launching it, drove the original company into liquidation and it was purchased for use on the North Atlantic, a service for which it was most unsuitable. The result was that it was a most expensive failure except for the work that it did in laying the Atlantic cable.

9. Steam Colliers.—By this time the screw steamer was invading most of the trades of the world, including the coastal. The U.S. was building up a fine coasting fleet, differing very materially from European ideas of design. Most of the European nations were following suit, while steam colliers were even invading the coal trade between the northeast coast of England and the Thames, which was regarded as the stronghold of the sailing ship. The first was the 273-ton "Q.E.D.," which was really an auxiliary, schooner rigged, with the smoke from its low-powered screw machinery carried up through the mizzenmast. This vessel made its appearance in 1844. Eight years later, however, the first real iron steam collier was put into service, the "John Bowes." This steamer had dimensions 151 ft. 9 in. by 26 ft. 3 in. and was one of the first ships to be fitted with tanks for water ballast. Rebuilt out of all recognition, it was later transferred to the Spanish coasting trade as the "Valentin Fierro."

An important result of the success of the "John Bowes" was that it caused the introduction of the steam tramp, a cargo vessel which was open to charter on any trade instead of running to a definite schedule. Before that time all steamers were built for definite services and chartering for bulk cargoes was unusual, this business being left almost entirely to the sailing ship.

10. The Compound-Expansion Engine.—Increased economy was still the aim of the engineer, and this led to the introduction of the compound-expansion engine, in which the steam was used in a second cylinder at a lower pressure after it had done its work in the first. It was an invention of James P. Allaire, U.S. engineer, in the year 1824, but at that time it failed because of the low pressure used. As steam pressure increased John Elder, the head of what later became the Fairfield Shipbuilding and Engineering Company on the Clyde, brought out his compound engine, which employed the steam in two stages and which did much to overcome the disadvantages of the steamship on long routes on which coaling stations were rare.

The first vessel to be so fitted was the "Brandon" of 1854, a screw steamer designed for the trade between London and Limerick and fitted with a vertical engine having the cranks diametrically opposite to one another. Its coal consumption on trial was returned as 3¼ lb. per indicated horsepower per hour, as compared with the 4 to 4½ lb. which was the utmost economy to which the simple engine could aspire. Although built for a short-distance service, it was employed as a transport in the Crimean War and its success caused the compound engine to be generally adopted, while the steam pressure steadily increased. The first Atlantic steamship to be fitted with compound-expansion engines was the "Holland" of the National Line in 1869.

As pressures still increased it became possible to add a third stage to the engine and triple-expansion machinery came into being. In France this system was sponsored by Benjamin Normand of the famous firm of Le Havre shipbuilders, who took out a patent in 1871 and installed his first set two years later. In England a patent was taken out by A. C. Kirk, a colleague of John Elder, who tried a triple-expansion engine and machinery first in 1874 in the 2,083-ton steamer "Propontis."

11. Water-Tube Boilers.—The original installation of the "Propontis" was fed by water-tube boilers, in which the water passed in tubes through the flame instead of the flame passing in tubes through the water as in ordinary mercantile practice. These boilers, which were later generally adopted in all navies and in many merchant ships, had already been introduced in France but had given much trouble. Those installed in the "Propontis" were no more satisfactory than those of French men-of-war which had been so fitted in the 1840s, and it is generally described as a failure, but when new boilers were installed with reduced pressure it continued to work satisfactorily for many years.

In 1881 the "Aberdeen" of 3,616 tons, designed to run on the Australian trade in which economy was more obviously necessary than in any other, was fitted with Kirk's triple-expansion engines. On trial it reduced its coal consumption to 1.28 lb. per indicated horsepower per hour and was a most satisfactory vessel on service, although this abnormally small consumption was naturally increased under working conditions. Its success caused a large number of steamers of all kinds, which had originally carried compound engines, to be tripled by the addition of a third cylinder during the 1880s and 1890s.

At the end of the 19th century steam expansion was taken one stage further, to obtain the best economy with steam pressures above about 180 lb. per square inch, and quadruple-expansion engines were built for marine use. Such engines were fitted in 1904 on the twin-screw "Caronia" of 19,687 tons, which was the last Cunard liner to be fitted with reciprocating engines. Its sister-ship, the "Carmania" (19,524 tons) of 1905, marked the transition of the Cunard Line to turbine propulsion.

12. Twin Screws.—In 1862 the first full-powered twin-screw steamer, apart from certain experiments in the earliest days of steam, was built on the Thames. It was the 400-ton "Flora" and although its builders, the Dudgeons, made a specialty of twin-screw steamers and built the far larger "Ruahine" in 1865 for the Panama, New Zealand and Australian Royal Mail Company, the idea did not attract the liner companies until the "Notting Hill" of 1881 proved the advantages of the system on the North Atlantic. Even so the single screw was not abandoned at once and as late as 1896 first-class liners were built with one shaft only, in spite of the uneasiness caused by several serious accidents.

13. Turbine Ships.—The experimental "Turbinia," built at Wallsend on the Tyne in 1894 and given turbine engines invented by Sir Charles Parsons, revived one of the earliest principles of generating power and proved that it was capable of being used at sea. Its dimensions were 100 ft. by 9-ft. beam by 3-ft. draft, its displacement being 44½ tons. The original machinery installation consisted of three steam turbines, totaling 2,000 s.h.p. (shaft horsepower) and each driving a shaft carrying three screws in order to overcome the disadvantages of high propeller speed. During the Diamond Jubilee naval review of 1897 it was taken down to Cowes secretly and suddenly dashed out among the assembled ships at what was then the astounding speed of 34½ knots. Naturally it caused a great sensation and the admiralty built two turbine destroyers which unluckily came to grief, one by stranding and the other because of its light construction. There was a tendency to blame the machinery, which had nothing to do with either mishap, and it was some time before Parsons could persuade commercial shipowners to take an interest in his invention.

The "King Edward" of 1901 was the first merchant steamer to be given turbine machinery, a Clyde passenger steamer which ran for a considerable time after material alterations. It was followed by the cross-channel packet "The Queen" of 1903, which proved itself greatly superior in speed and far more economical than the reciprocating-engined steamers which had preceded it on the service between Dover and Calais. The next important ships to be fitted with turbines were the Allan liners "Virginian" and "Victorian" of 1904, ships of about 10,750 tons gross each with a trial speed of nearly 20 knots. With all these ships there was a certain amount of trouble in the early days, but the turbine made progress and proved its reliability in the Cunard liner "Carmania," whose success resulted in the turbine's being adopted for the Cunard steamers "Mauretania" and "Lusitania" which were the biggest and fastest liners in the world when they were built in 1906. The "Mauretania" held the blue riband of the Atlantic for 22 years.

The 22-ft. steam launch "Charmian" was the first vessel in which intermediate gearing between the turbine and the propeller was tried, the experiments taking place on the Tyne as early as 1897 and being practically contemporary with similar experiments carried out on the continent. It had long been realized that the efficient speed of the turbine was far too great for the propeller, which was the reason why multiple screws were fitted to the shafts of the early passenger steamers. In 1909 these experiments had produced sufficient promise to warrant the cargo steamer "Vespasian" having its old triple-expansion engines taken out and turbine machinery with single-reduction mechanical gearing substituted. The great increase in speed and economy which was immediately obtained drew attention to the possibilities of the turbine for cargo as well as fast passenger vessels, and from this it became evident that the gearing, whether it was single or double reduction, was of the greatest advantage even at the maximum speed so that the direct coupled turbine came to be regarded as obsolete at sea.

14. Introduction of Steel.—Improvements in the hulls of steamers were introduced steadily while machinery was being improved, principally with the idea of increasing the strength and carrying capacity and reducing the weight of hull necessary. Iron had shown itself to be superior to wood in these respects, and the next step forward was the introduction of steel for shipbuilding. The first steel-hulled vessel to cross the Atlantic was the paddle steamer "Banshee" of 325 tons, built at Liverpool in 1862 for use as a blockade runner in the American Civil War. In 1876 a small steel paddler was built for river service in Burma and in the following year the Royal Navy built two fast dispatch vessels, the "Iris" and "Mercury." The first sizable merchant ship to be so built was the "Rotomahana," a ship of 1,777 tons built by William Denny on the Clyde for the Union Steamship Company of New Zealand in 1879. Within a few months it was followed by a

much bigger and more important ship, the Allan liner "Buenos Ayrean" of 4,005 tons, for the Canadian mail service. Although it was recognized that steel offered advantages in every direction except, possibly, durability, its general adoption was checked by the difficulty of obtaining supplies; however, in the early '80s this was overcome and many steel ships were constructed.

The form of the hull was also the subject of numerous experiments in the constant effort to increase the carrying capacity on the same or smaller tonnage, and for this reason several revolutionary designs were brought forward. One of the most striking and permanent of these was the turret deck steamer which was evolved and built by the Doxford shipyard of Sunderland. In this type a curve in the side above the water line resulted in a narrow deck with a broad extreme beam, and for some time it permitted a great economy in harbour and canal-transit fees, since these often were computed on the basis of the cargo capacity of the first deck above the water line.

The "Turret" of 1892 was the first ship of this type and was laid down by the Doxford shipyard on speculation but bought by Peterson Tate and Company for the Canadian trade. Other steamship lines took up the turret deck steamer enthusiastically, particularly the Clan Line, but many of its advantages were negated by the amendment of port and canal regulations, and although these steamers were capable of carrying a large cargo and were excellent sea ships if properly treated, in inexperienced hands they were apt to give trouble and so gradually fell out of favour.

Welded steel-hull construction was tried about 1916, and the first sizable merchant ship to be all welded, without any rivets in its hull, was the M.S. "Fullagar," built at Birkenhead in 1920.

VII. MOTOR SHIPS

The Caspian steamer "Wandal," which was built by the firm of Nobels in 1903, was the first sizable ship to be given an internal-combustion engine, but in this case it was used to generate electricity for the main drive.

Engines in which the charge was exploded by a hot bulb or an electric spark were found suitable for small craft soon after they had become general in automobiles on land, but their size was strictly limited. In the year 1892 Rudolf Diesel took out his patent for an engine in which the charge was exploded by raising its temperature by compression while it was still inside the cylinder, and ships propelled by such units, built up to large sizes and considered by many to be preferable to steam plants, are invariably termed diesel-engined.

The "Wandal" was followed in 1906 by the "Venoge," a motor barge on Lake Geneva, and in 1910 the motor tanker "Vulcanus" of 1,179 tons marked a great improvement in size. One year later the motor ship "Selandia" was commissioned by the East Asiatic Company, Ltd., of Denmark, one of the principal advocates of the diesel engine, and ran successfully until 1942. Its dimensions were 370 ft. by 53.2 ft. by 27.1 ft. depth of hold, its gross tonnage being 4,950. It was a twin-screw ship, each shaft being driven by an eight-cylinder four-stroke-cycle engine with cylinders 20⅞ in. in diameter by 28¾ in. stroke, totaling a brake horsepower of 2,450 at 140 r.p.m. and driving it at a speed of 11 knots. It was rigged as a three-masted schooner and the absence of funnel attracted general attention. Most shipowners continued preferring to run the exhaust from the engines up through funnels, making their motor ships practically indistinguishable from steamers.

After the success of the "Selandia," for although all the early diesel-engined ships had a certain amount of difficulty with their machinery it was a distinct success and economical, a series of ships was built steadily improving in size, efficiency, and economy. World War I practically held up experimental work in that connection as far as merchant ships were concerned, but the diesel improved rapidly in submarines and the Germans brought it to a high pitch of perfection. In Great Britain the results of admiralty experiments were also placed at the disposal of the mercantile engineers at the end of the war. Several submarine diesels were fitted into German cargo ships after the war, their speed being geared down to the efficient speed of the propeller as in the case of the turbine, with satisfactory results.

To begin with, the diesel engine was more or less confined to the cargo vessel, but after World War I it was seen that its adoption in large passenger vessels would mean a considerable saving in space and running costs, although the first costs were considerably higher than in the case of the steamer. After several trials in comparatively small ships, in 1924 the "Aorangi" of 17,491 tons was built at the Fairfield yard on the Clyde for the Union Steamship Company of New Zealand and proved most successful on the service between Vancouver and Australia. It was followed in 1925 by the "Gripsholm" of the Swedish American Line, a ship of 18,134 tons designed for service between Sweden and the United States. There was considerable doubt as to the wisdom of building this ship, for the diesel shows to its best advantage on long runs, but it proved an unqualified success. In the same year a great advance in tonnage was made with the 22,048-ton "Asturias," built at Belfast for the Royal Mail Steam Packet Company, and the diesel-engined passenger liner became firmly established. Harland and Wolff, Ltd., the builders of the "Asturias," specialized in a big double-acting diesel built on the lines of Burmeister and Wain of Copenhagen, one of the earliest diesel builders.

The "Augustus" of 30,418 tons was built for "Italia," Società Anonima di Navigazione, in 1927. In deference to public opinion all these large passenger motor ships were given funnels, the majority of them two, but the builders evolved a typical diesel funnel, short and stout, which does much to obviate the echo which can prove exceedingly troublesome when the exhaust is carried through a funnel of ordinary design, particularly in fog.

1. The Still Engine.—In the Scott-Still engine, first fitted in the Blue Funnel liner (A. Holt and Company) "Dolius" of 5,507 tons in 1924, an effort was made to combine the steam and diesel systems, but although a certain economy was obtained, the system did not become a commercial success.

2. Electric Drive.—Electric propulsion at sea, with power generated by the internal-combustion engine, was first tried by Nobels in the "Wandal" on the Caspian in 1903, a vessel that has already been mentioned as the first motor ship. The German Navy adopted it for use on the submarine salvage vessel "Vulcan" in 1907 and it was also tried in 1913 in a Canadian lake vessel, the "Tynemount," where the current was generated by two diesel engines of 250 b.h.p. (brake horsepower). It was followed by the Swedish "Mjölner," in which the current was generated by two turbines of 450 s.h.p. each. The U.S. Navy, which later brought the system up to its highest pitch of perfection, first tried it in the steam collier "Jupiter" of 20,000 tons displacement, in which the turbine generator was of the Curtiss type and where the results were so successful that the system was adopted in the latest and largest battleships. The "Jupiter" later became the aircraft carrier "Langley." For cruisers and destroyers, however, the United States naval authorities found mechanical gearing better.

After World War I there was a growing tendency to favour electric propulsion for merchant ships, particularly in the U.S., and with the benefit of the experience gained in the U.S. Navy it was possible to install large power plants with perfect success. The most striking case was the Panama-Pacific liner "California," employed on the intercoastal trade between New York and California. It was a ship of 20,325 tons built in 1928, its electric machinery generated by turbines giving it a speed of 18 knots. A similar installation was chosen for the Peninsular and Oriental liner "Viceroy of India," launched on the Clyde in 1928, and the French liner "Normandie" was fitted in 1932 with turboelectric drive of 160,000 h.p.

VIII. STEAMERS FOR SPECIAL PURPOSES

1. Tugboats.—The use of steam power for tugboats was considered from the earliest times. In 1736 Hulls patented his proposal for a steam-propelled tugboat to carry vessels into and out of harbour in a calm. This he described in considerable detail, the vessel was to have a Newcomen steam engine of 30-in. cylinder diameter, fitted with ratchet gear to produce rotative movement of a paddle wheel at the stern. This scheme, so far in advance of its time, met with derision and never had a practical trial.

The first practical tugboat was the "Charlotte Dundas" of 1802

tried on the Forth and Clyde Canal, but condemned because the canal owners feared that the wash from its paddles would cause damage to the canal banks.

The first steam tugboat on the Thames was put into service in 1816; and the first steam vessels in the British Navy, from about 1822 onward, were used as tugboats, to tow the sail warships in and out of harbour. Screw propulsion for tugboats was first tried in the United States about 1850 and soon came into general use. In 1875 triple-screw tugs were in service on British inland waterways.

In the course of World War II a class of ocean tugboat was developed for rescue and salvage work. One notable vessel of this class was the "Turmoil" of 1,800 tons displacement, built in 1944 for the British Admiralty. Its dimensions were 205 ft. by 38.5 ft.; it had diesel engines of 3,200 total b.h.p. and was capable of 15 knots.

For modern ocean liners, the use of tugboats in berthing and docking operations is normally considered indispensable; but it is on record that on Feb. 6, 1953, the "Queen Mary," in exceptional circumstances, was able to berth at New York under its own power and without the help of tugs.

Even after the screw had been adopted for the propulsion of tugboats, the paddle wheel still continued to be used for this purpose, and in 1955 some paddle tugs with diesel-electric drive were built for the British Admiralty.

2. Train Ferries.—The 417-ton steam train ferry "Leviathan," built in 1849 for the North British Railway Company to run across the Firth of Forth from Granton to Burntisland, was the first of a type which later became quite general and was built to large size on several services. Almost at the same time another was built to cross the Tay at Dundee, but both these ferries were later replaced by bridges. In 1878 the Philadelphia and Reading Railroad built its first ferry steamer to run across New York Bay and in the early 1880s the Danes started a train ferry from Korsor to Nyborg. The idea of a similar service to run across the Straits of Dover was put forward in 1870, when the use of double-ended paddle steamers was proposed. Train ferries were introduced in 1936 between Dover and Dunkirk, in connection with the direct London-Paris service. Another train ferry on a large scale in Great Britain is that running from Harwich to Zeebrugge and employed largely for shipping French and Italian fruit to England.

3. Steam Trawlers.—In the early 1870s several attempts were made to introduce steam trawlers in the south of England, but they were not successful. Several years later a shipping slump caused numerous steam tugs to be laid up and the experiment of using some of these for trawling was tried. Most of these were paddlers, the first being the "Messenger" which made its first trip to the fishing grounds in 1877, but the experience gained in it and other converted ships showed that the paddle was unsuitable for trawling. Experiments with screw boats were tried in the early 1880s and were an immediate success, the design developing rapidly and producing some of the finest sea boats afloat. Later, practically every maritime country acquired a big fleet of power trawlers and drifters. The internal-combustion engine made great strides in vessels of this type, and electric machinery was also tried on occasions.

4. Icebreakers.—"Eisbrecher I," launched in 1871, was the first specially designed icebreaker to be built. It was planned on the experience gained a few years previously by a Russian shipowner named Britneff, who reconstructed the bow of the steamer "Pilot" in such a way that it could be driven onto the ice in the hope that it would break it by its own weight. In practice it was too small and light for this purpose, but the idea was appreciated and is embodied in all modern icebreakers, some of which run to a large size. *See* ICEBREAKERS.

5. Tankers.—The increased consumption of oil for various purposes made it necessary to evolve some means of carrying it in bulk instead of in barrels as it had been carried from the earliest days of whaling. The first suggestion was in 1863, when the sailing vessel "Ramsey" had a few tanks built into its hold to carry oil in bulk in addition to stowing barrels in the ordinary way in its 'tween decks. From 1869 to 1872 the sailing vessel "Charles," of 794 tons, was carrying oil in 59 iron tanks which were built into its holds and which completely filled them. The problem of keeping ironwork oiltight was not then fully understood and these tanks were far too weak, with the result that in the working of the ship under sail constant leaks developed and eventually it was burned at sea. Several other ships were converted in similar fashion, mostly for the trade across the Atlantic to Le Havre and Antwerp, but they were all failures.

In 1872 the Red Star Line of Philadelphia started a new system by having the steamship "Vaderland" of 2,748 tons built by Palmer Brothers and Company on the Tyne with the idea of carrying a large cargo of oil in bulk in addition to its passengers. All the authorities were against this scheme on grounds of safety, and it was used for passenger and cargo only, as were also the "Nederland" of 1873 and "Switzerland" of 1874 which were designed on similar lines. This prohibition against shipping oil in passenger vessels still exists.

In 1879 Nobels, who were already noted for their work in building the pioneer motor and electrically driven ship, built a number of tank steamers to carry oil on the Volga River and attained considerable success with them. These steamers attracted some attention at the time, but the idea that oil would be safer in a sailing vessel than a steamer was still generally held and in 1886 the sailing ships "Andromeda" and "Crusader" were fitted with cylindrical tanks on an improved system for the purpose. Their designers attained a considerable measure of success in making their tanks oiltight, but in the same year the firm which later became Sir W. G. Armstrong Whitworth and Company, Ltd., designed and built the first real tank steamer on modern lines.

The "Glückauf" had a dead-weight capacity of 3,000 tons and a speed of 10½ knots, its engines being placed right aft for safety as is still almost invariably done in tankers. Its hull was subdivided into tanks, and arrangements were made for pumping them on a principle which is still used.

After World War I a large number of tank steamers were built, although it is one of the most difficult trades to cater for and to maintain. In spite of numerous experiments oil remained almost invariably a one-way cargo, so that the ship had to be exceedingly economical because of the return journey in ballast. A large cargo is desirable for the ocean route, and 13,000 tons dead weight and more became a usual size, while the speed seldom exceeded 11 knots.

After World War II, however, there has evolved a class of supertanker. One such notable vessel was the "Spyros Niarchos," built at Barrow in 1955 for the Neptune Tanker Corporation. Its dimensions were 757 ft. by 97.2 ft., with a dead-weight capacity of 47,750 tons. The hull was longitudinally framed and nearly all welded. The vessel was propelled by geared steam turbines of 18,000 s.h.p. at a normal service speed of 17 knots. The greatest shipbuilding activity of the 1950s was in tanker construction. By the mid-1960s tankers of more than 150,000 tons dead weight were in service with lengths of up to 1,000 ft. and widths of more than 140 ft. Tankers were constructed not only for transporting oil, but also for carrying liquid chemicals, gases, molasses, and fruit juices. Because of their tremendous length and unequal strain, the design of these tankers involves careful considerations of strength, and all modern types are built on the longitudinal system, with great attention paid to fore-and-aft stresses. (*See also* TANKER.)

6. Packet Steamers.—After the screw superseded the paddle for transoceanic work, the latter survived in the excursion steamers round the British coast, where shallow draft was generally necessary to get alongside the piers, and in the cross-channel services. The growing continental trade and the limitations in draft imposed by the French and Belgian harbours brought the paddle packet steamer to a high pitch of perfection in the last days of the 19th century, ships of considerably more than 2,000 tons gross with a speed of 22 knots and excellent seaworthiness being built in considerable numbers, mostly on the Clyde. No more of these big cross-channel paddlers were built after the turbine had proved its possibilities, but many other steamers propelled in this way, much smaller and rather slower, were built for the various ex-

cursion routes in the summer. Most of these again were excellent sea boats and were able to perform valuable mine-sweeping service during World Wars I and II.

In the United States, however, where the huge inland waterways favour the side wheel, the paddle steamer was brought up to its highest state of development. Notable ships of this type were the "Greater Buffalo" and "Greater Detroit," built in 1923 for service on the Great Lakes, each having a gross tonnage of 7,739 and being driven by three-cylinder compound engines.

7. Standardized Ships.—In the early days of World War I mercantile construction was almost completely suspended in Great Britain, Germany, and France in favour of warships, but it received a great impetus in the United States. When the German submarine campaign caused an acute shortage of tonnage it had to be resumed in an intensified form in Britain, while from 1917 the U.S. launched a shipbuilding campaign unprecedented in the history of the industry. In order to save time and money wherever possible, standardized shipbuilding came into vogue. It had existed within certain establishments for many years, the German yards, particularly, specializing in certain types and turning out large numbers of sister ships for various owners. During the war, however, it was greatly expanded in many countries, and in some cases inland steelworks, usually employed on bridge construction and the like, were utilized for fabricated ships with lines as straight as possible which were only put together and launched in the shipyards.

In the United States wood was largely used, both for full-powered ships and auxiliaries, but there was no time to permit the timber to season and they were built of green timber. Similarly a number of vessels of various types—lighters, tugs, tankers, etc.—were built of ferroconcrete, the rapidity with which they could be turned out compensating for their excessive weight in wartime, although only one or two of them survived more than a few months of peacetime trading.

In criticizing these standard ships built during World War I, there was a general tendency to forget the peculiar circumstances in which they were built and the necessity of getting something that would float into the water as quickly as possible. Their utility in competitive trading in peaceful circumstances was a secondary consideration. The United States Shipping Board, formed to operate the tonnage so built in the U.S., was left with a large number of ships for which no purchasers or managers could be found, though many of them were later converted to diesel or electric power and their running expenses greatly reduced.

In the later years of World War II, quantities of standardized cargo vessels, called Liberty ships, were built in the United States. These were oil-burning ships; a fairly large number of such ships were built by Canada as coal burners. "Standardization and no changes" was an outstanding feature of World War II shipbuilding. The largest number of ships ever built in the world from one design was the Liberty ship. Another outstanding feature was the use of all welding in construction. Welding was not a new process but it was developed and used to a far greater extent than ever before, being applied not only to cargo vessels but also to all classes of ships.

8. The Cabin Liner.—From the early days of regular services, the older and smaller units of the various companies were put on intermediate services at reduced prices, and these ships gradually came to be known as cabin-class ships. Few ships were specially built for the class until 1914, when Canadian Pacific Steamships, Ltd., commenced a long series with the "Metagama" and the "Missanabie," ships with a gross tonnage of about 12,500 and speed of 16 knots, having accommodation for 520 passengers in the cabin class, paying a fare approximating the second-class fare in the bigger ships, and 1,200 in the third class. Afterward the type was taken up enthusiastically by various companies, particularly on the North Atlantic, with such vessels as the "Duchess of Atholl" and her sisters, also owned by the Canadian Pacific Line, having a gross tonnage of roughly 20,000, a sea speed of 17 knots and single reduction-geared turbines supplied by oil-fired high-pressure boilers which gave the remarkable economy of 0.64 lb. of oil fuel per shaft horsepower per hour for all purposes.

9. Fast Cargo Liners.—There was also a noteworthy tendency to build fast cargo liners after World War I, ships with a sea speed of 14, 15 and 16 knots, but carrying cargo only on regular schedule. Both steamers and motor ships were built for these services and found great favour with shippers. Great economy in operation was obtained, with the result that they secured a considerable proportion of the trade which was formerly carried by tramp steamers.

10. High-Pressure Steam.—In steamers this remarkable economy was made possible largely by the employment of high-pressure steam. The Clyde passenger steamer "King George V," built by William Denny and Brothers (Ltd.) in 1926 in conjunction with Yarrow, the boilermakers, and the Parsons Marine Steam Turbine Company, had the first noteworthy installation apart from unsuccessful experiments made about 1880. It was a steamer of 801 tons gross, burning coal under water-tube boilers working at a pressure of 550 lb. to the square inch with the steam temperature raised to a maximum of 750° F. This ship considerably exceeded its contract speed of 20 knots on trial and proved a success in many ways, particularly with regard to economy. But numerous adjustments and alterations had to be made after its launching.

In modified form high-pressure steam resulted in considerable economies, both in fuel consumption and weight. The general tendency was to increase the pressure both for naval and for mercantile work, and 450 lb. became quite usual. While high-pressure steam saw its beginning in 1926, it developed rapidly in use, and in the 1930s very high-pressure and very high-temperature steam installations were made in U.S. Navy destroyers greatly increasing their speed.

These principles were adopted in whole or in part by merchant vessels and became standard practice; *e.g.*, the tanker, "Atlantic Navigator," built in 1951, was equipped with boilers operating at a pressure of 650 lb. per square inch and a temperature of 1,020° F.

11. Gas-Turbine Propulsion.—The earliest recorded description of a gas turbine (in which the hot gases of internal combustion, instead of steam, act on the turbine blades) can be found in the patent obtained by John Barber, of Nuneaton, in 1791. The first vessel to be propelled at sea by a gas turbine was the 110-ft. triple-screw motor gunboat no. 2009 of the British Admiralty, which in 1947 had a gas turbine of 2,500 s.h.p. fitted to the centre shaft. The first ship of the mercantile marine with gas-turbine propulsion was the motor tanker "Auris" of 12,250 tons dead weight, built in 1947 at Newcastle for the Anglo-Saxon Petroleum Co., Ltd. In 1951 one of its four main diesel engines was replaced by a gas-turbine installation of 1,200 b.h.p., constructed by the British Thomson-Houston Co., Ltd.

12. Express Luxury Liners.—The White Star liner "Oceanic" of 1871 was the first of the modern type of express luxury liner, which attracts more attention than any other type, but which can exist economically only on the direct service between Europe and the United States. The White Star Line had then just been founded for the purpose of engaging in the Australian trade, and for that purpose had bought the name and house flag of one of the most famous of the clipper ship companies. Events caused the directors to change their intentions and the first ships were put on service between Liverpool and New York, the "Oceanic" being the pioneer. These ships differed materially from the accepted type of screw steamer, having more than ten beams to their length and being designed on the somewhat revolutionary lines advocated by Sir Edward Harland of Harland and Wolff, the Belfast shipbuilders. The tonnage of the "Oceanic" was 3,707 and while it was fully rigged as a four-masted barque its single-screw engines gave it a sea speed of 14.75 knots. After being generally condemned on theoretical grounds by the experts, it proved itself an unqualified success. Three years later the "Germanic" and "Britannic" were built, each with a gross tonnage of 5,000 and speed of 16 knots.

The Guion Line, which had been engaged principally in the emigrant trade, then entered the race for the Atlantic blue riband with the "Arizona" of 1879, a ship of 5,147 tons with a sea speed of 16.25 knots. An improved edition of the same design was the same company's "Alaska" of 1881, its gross tonnage being 7,142 and its speed 17.75 knots. In the "Oregon," which was

built just before the company got into serious financial difficulties, and which eventually sank in collision with a small schooner, the gross tonnage went up to 7,375 and the speed to 19 knots.

The reply of the Cunard Line was the construction of the "Umbria" and "Etruria" at the same establishment which had turned out the big Guion ships, Elder's Fairfield yard. They were ships of 8,120 tons with a speed of 19.5 knots and they marked the final development of the single-screw express liner on the Atlantic. Although they were specially designed to compete with the "Oregon," that ship had passed to the Cunard Line before they were completed.

The next great improvement was the design of the "City of New York" and "City of Paris," when the twin-screw system was adopted for express ships. They were built for the Inman Line in 1888 but were better known under the flag of the American Line to which they were transferred in 1893. These ships secured the Atlantic blue riband and were the last big Atlantic liners to be built with the old-fashioned clipper stem. When it was desired to transfer them to the American Line, congress permitted them to hoist the U.S. flag only on condition that two ships of at least equal tonnage were built in U.S. yards, which resulted in the "St. Louis" and "St. Paul," of 11,630 tons and 21 knots speed, being built in 1895, the first large Atlantic liners to be built in the United States since the 1860s.

The White Star Line replied to these ships with the "Teutonic" and "Majestic" in 1889, ships which had a great struggle with the "City of New York" and "City of Paris" for the Atlantic record. In 1893 the Cunard Line secured the blue riband without doubt with the "Campania" and "Lucania" of just under 13,000 tons with a sea speed of 22 knots.

In 1897 the German companies, which had built up a big business in comparative obscurity beside the companies racing for the blue riband, came out with the "Kaiser Wilhelm der Grosse" of the North German Lloyd, which lowered the record considerably. It had a gross tonnage of 14,350 and a speed of 23 knots. Their rivals, the Hamburg-American line, built the 14,500-ton "Deutschland" to compete for the record, but on their experience with it they decided to follow the policy of comfort and good cargo capacity on more moderate speed, a policy which was also followed by the White Star Line after the "Teutonic." The North German Lloyd, on the other hand, steadily lowered its own record with the three ships which followed in the same series and to the same general design—the "Kronprinz Wilhelm" of 1901, the "Kaiser Wilhelm II" of 1902 and the "Kronprinzessin Cecilie" of 1906. Each of these ships marked a further advance in size and speed.

Before the last named could be tried at its best the Cunard Line had built the "Lusitania" and "Mauretania" with the financial help of the British government. The dimensions of the latter were 762.2 ft. by 88 ft. by 57.1 ft. depth of hold, giving it a gross tonnage of 31,938, while its quadruple-screw direct-coupled turbines had a designed shaft horsepower of 68,000, intended for a speed of 25 knots. The "Lusitania" was sunk by a German submarine during World War I; the "Mauretania" had the unique distinction of holding the Atlantic blue riband for 22 years and when 20 years old proving itself capable of steaming at 29 knots to the rescue of a disabled cargo ship.

Maintaining their policy of avoiding excessive speed, the Hamburg-American Line answered these Cunarders with the biggest steamers in the world. The first of the series was the "Imperator," later the Cunard "Berengaria." It was launched in 1912, its dimensions being 886.3 ft. by 98.3 ft. by 57.1 ft. depth of hold and its original gross tonnage 51,969. The second of the series was the "Vaterland," which became the "Leviathan" of the United States lines. Its dimensions were increased to 907.6 ft. by 100.3 ft. by 58.2 ft. depth of hold; gross tonnage was 54,282, later 59,957. The third ship was the "Bismarck," later the White Star "Majestic," similar to the "Vaterland" but slightly longer.

Immediately after World War I nearly all the lines had to engage in a shipbuilding program to replace casualties, although shipbuilding prices were at their highest level. The general tendency was to build a moderate-sized ship with good cargo capacity and comfortable passenger accommodation at reasonable rates. The Cunard Line brought out the "Franconia" type of five ships of just over 20,000 tons gross with a speed of 17 knots, while the Hamburg-American Line built four ships of the "Deutschland" type, having a sea speed of 16 knots and a tonnage of slightly more than 21,000. This tendency was broken by the Compagnie Générale Transatlantique, which in 1926 built the "Ile de France," a ship of 43,450 tons gross with direct-acting turbines of 52,000 s.h.p. and a trial speed of 24 knots. The North German Lloyd responded by laying down the "Bremen" and "Europa," ships of about 50,000 tons each, with a speed of 28 knots on service. They had geared turbines, and considerable weight was saved by a boiler pressure of 330 lb. to the square inch. Their design was revolutionary in many features, particularly in the bulb bow under the water line. The "Bremen," launched in 1928, beat all Atlantic records, but was then outstripped by the "Europa," which later became the "Liberté" of the French Line.

The first of the modern 1,000-ft. vessels was the "Normandie" of 83,423 tons, launched at St. Nazaire in 1932 for the Compagnie Générale Transatlantique. It had four screw propellers with turboelectric drive of 160,000 total s.h.p., carried 2,170 passengers and crossed the Atlantic at a mean speed of 31.2 knots. At the outbreak of World War II the "Normandie" was in New York Harbor, but was not utilized by the French for war service. On Dec. 16, 1941, it was expropriated by the U.S. Maritime Commission for a troopship. In the course of conversion, however, fire broke out on Feb. 9, 1942, and the vessel capsized. This occurred the day after the "Normandie" had been turned over to the U.S. Navy by the U.S. Maritime Commission.

The first British vessel to exceed 1,000 ft. was the "Queen Mary" of 81,235 tons, built at Clydebank and launched in 1934. It had four screw propellers 19.6 ft. in diameter, driven by turbines of 160,000 total s.h.p. In 1938 it recaptured the blue riband from the "Normandie" with a mean Atlantic speed of 31.7 knots. Another 1,000-ft. liner was the Cunard White Star "Queen Elizabeth" of 83,673 tons, built at Clydebank in 1938, carrying 2,288 passengers and propelled by turbines of 181,700 total s.h.p. supplied with steam at 450 lb. per square inch pressure.

In 1952 the 53,330-ton liner "United States," built at Newport News, Va., crossed from New York to Southampton at a mean speed of 35.59 knots (nearly 4 knots faster than the "Queen

BY COURTESY OF THE UNITED STATES LINE

SS "UNITED STATES"

Mary") and thereby reclaimed for the U.S. the blue riband of the Atlantic. The "United States" was built to the specifications of the U.S. Navy, to facilitate its rapid conversion into a troop carrier. (E. A.; F. C. Bo.; H. P. St.; E. S. L.)

IX. NEW DESIGN AND DEVELOPMENT

1. Nuclear-Powered Ships.—The late 1950s saw the advent of the nuclear-powered merchant ship. Employing the same basic design of reactor as that used by the United States for nuclear submarines, the U.S.S.R. put into service the nuclear-powered icebreaker "Lenin" in 1959, and in the same year the U.S. passenger and cargo liner "Savannah" was launched.

In Germany a nuclear-powered cargo ship was being built in 1965, but the United Kingdom postponed the construction of a nuclear ship until a reactor could be developed offering some hope of economy in operation comparable to that of a conventionally powered merchant ship. An important development was the increasing use of automatic and remote controls for ships' machinery to allow smaller crews. In Britain, closed-circuit television was introduced in 1963 to assist in the berthing of large ships. The control of main engines and auxiliary machinery began to be centred in control rooms, sometimes with alternative control of the main machinery from the bridge. Mechanization was also introduced into navigation, and the U.S. studied the development of any unmanned ocean-going vessel. Ships continued to grow in size. The most spectacular growth was that of the oil tanker, which in 1963 reached a size of 130,000 tons deadweight on the route between the Persian Gulf and Japan, where depths of water imposed no restriction; later, vessels of 190,000 tons were ordered. Ships of 80,000 tons deadweight carrying dry cargoes in bulk were in service by the mid-1960s. Change in the relative importance of fuel costs and overhead costs made higher speeds for ships economically desirable, and many cargo liners with speeds of more than 20 knots were built. Another feature of cargo liner design was the growing employment of large containers for goods at sea, the disadvantage of waste of hold space through their use being outweighed by greater ease of loading and discharge, and by better protection for the goods. To reduce manual handling of cargo, a design of ship with hatches extending virtually the full width of the ship was developed, the problem of maintaining the strength of the vessel being solved by working additional steel into other parts of it.

2. Hovercraft and Hydrofoil Craft.—Two new types of water craft, the hydrofoil and Hovercraft, came into prominence. The Hovercraft, a British invention, by the mid-1960s reached the stage where orders were being placed for commercial, passenger-carrying craft; models suitable for car ferries were also envisaged. (*See* Ground-Effect Machine.) (I. Br.)

A hydrofoil ship is one that is supported clear of the water during operation by underwater wings called hydrofoils. These wings are of similar shape and function to the wings of an aircraft, except for size. Since water is roughly 775 times as heavy as air, very small hydrofoils will support relatively heavy boats and ships, while very large wings are needed to support aircraft. Another difference is structural strength. Since operating in water imposes very great loads per unit of area on hydrofoils, they are usually built of high-strength steels, or perhaps, in the case of small boats, of high-strength aluminum alloys.

The purpose of raising the hull of boats out of the water and supporting them on hydrofoils is to avoid the resistance of the friction on the hull and the drag of wave-making that occurs when a normal boat proceeds through the water. The result at high speed is to reduce the power required to drive the boat by about half, thus permitting much higher speeds for the same size ship and engine. Another and perhaps more important result is the fact that in a well-designed foil system, the boat does not feel the impact of passing waves in rough water and up to some limited size of a seaway, can fly perfectly level and smooth, thus permitting a very small boat to fly at high speed through a rough sea that would force a conventional boat to slow down.

Hydrofoils are not new. Between 1898 and 1905 Forlanini in Italy built and flew a rudimentary hydrofoil successfully. Considerable interest was stirred among inventors of the time, and even the Wright brothers experimented with hydrofoils before their more noteworthy and successful work on the airplane. In 1918 the world's water speed record was attained by a hydrofoil powered by an aircraft engine and developed by Dr. Alexander Graham Bell and Casey Baldwin, a Canadian inventor. From this period until the 1930s interest lagged, principally because there were no high-power, lightweight engines available to make practical, usable hydrofoils. In the 1930s German inventors were on the verge of a successful seagoing ship when war halted their efforts. After World War II these efforts were renewed and the many commercial hydrofoils that are now in service in western Europe are based on German technology.

In the early 1950s the availability of higher powered gas turbines and a renewed interest in high-speed water travel resulted in the development of hydrofoils in the U.S., Canada, and U.S.S.R. Designers emphasized the military as well as the commercial utility of hydrofoils. Their efforts came to fruition with the completion in the U.S. of the first seagoing patrol craft, the U.S.S. "High Point," a 100-ton vessel designed as a military ship and not as a research tool; the production of an 80-ton commercial prototype, "Dennison;" and the operation on the rivers and waterways of the Soviet Union of passenger ferries carrying hundreds of passengers.

The problem of stabilizing a hydrofoil ship so that it moves without undue rolling or pitching, and the problem of keeping the hull the proper distance from the water have given rise to many different designs of hydrofoils and supporting struts as well as many different control systems. Most successful hydrofoil ships belong to one of two general classes: surface piercing foils and submerged foils. The simplest type, surface piercing, depends for its stabilization on providing about twice as much foil area as is necessary to support the ship, and on operating with the foils piercing the surface of the water so that they are wetted just enough to carry the ship. Then, if more lift is needed because of a shift in weight or the approach of a wave, more area is immersed.

The more complex type of hydrofoil employs fully submerged foils that are wetted at all times. In this type, changes in lifting forces are obtained by changing the angle of the foils to the flow of water until the precise lift is obtained on all foils to just support the weight of the ship. Since a fully submerged foil has no reference to the surface of the water, artificial signals must control its movements. Thus, for the submerged foil type, some mechanical or electrical means must be used to measure the height of the ship above the water and its angles of roll and pitch, and means must be provided to convert these measurements into rotation of the foils to provide the proper response. Thus, the submerged foil ship must be equipped with a system roughly corresponding to an autopilot in an aircraft.

Surface piercing hydrofoils are cheaper, simpler to operate and maintain, and are less susceptible to breakdown; the submerged type, however, is much more seaworthy, provides a smoother ride, and can operate in seas that a surface piercing ship of similar size can not. Thus, most commercial ships and those designed for river or inland waterways have been surface piercing boats, while those designed for unlimited military or open ocean service have employed submerged foils.

Although the largest hydrofoil ships have a displacement of hundreds of tons, there has been considerable controversy as to the maximum size attainable; this is because the so-called cube-square law that has limited the size of aircraft also applies to hydrofoil ships. In accordance with this law, as hydrofoils become larger, their weight increases as the cube of a linear dimension, while the foil area available to support this weight only increases as the square of that dimension. From this, it follows that, at some size, the weight will exceed the capacity of any foils to support it. The general consensus is that this limit is reached somewhere between 500 and 1,500 tons. Therefore, hydrofoil ships are limited to moderate sizes and can not replace aircraft carriers or ocean liners. They are used or planned for high-speed passenger ferries, patrol boats, attack craft, antisubmarine craft,

and other missions where moderate size and high speed are desired.

Hydrofoil ships experience a speed barrier at about 50 knots, which might be likened to the sound barrier in aircraft. At this speed the negative pressure on the top of the hydrofoils reaches the vapour pressure of water and a cavity is formed, causing lift to become erratic; the cavity then collapses, resulting in physical damage to the foils. As a result, a normal hydrofoil cannot operate at more than 50 knots. However, just as aircraft have broken the sound barrier, hydrofoils with sections shaped like an axe (supercavitating foils) instead of a normal airfoil have attained much higher speeds, in excess of 80 knots.

See also references under "Ship" in the Index. (JA. J. S.)

BIBLIOGRAPHY.—*Historical and Sailing Ships:* A. Jal, *Archéologie navale*, 2 vol. (1840), *Glossaire nautique: répertoire polyglotte* (1848–50); G. H. Boehmer, *Prehistoric Naval Architecture of the North of Europe* (1893); C. Torr, *Ancient Ships* (1894); Tjard Schwarz, *Die Entwicklung des Kriegschiffbaues* (1909); A. H. Clark; *The Clipper Ship Era* (New York, 1910; London, 1911); B. Hagedorn, *Die Entwicklung der wichtigsten Schiffstypen* (1914); A. Tenne, *Kriegschiffe der alten Griechen und Römer* (1915); August Köster, *Das antike Seewesen* (1923); L. G. C. Laughton, *Old Ship Figure-Heads and Sterns* (1925); O. T. Howe and F. C. Matthews, *American Clipper Ships, 1833–1858*, 2 vol. (1927); Sir Alan H. Moore, *Last Days of Mast and Sail* (1925), *Sailing Ships of War, 1800–1860* (1926); C. C. Cutler, *Greyhounds of the Sea* (1930); G. S. Laird Clowes, *Sailing Ships* (1930–32); R. C. Leslie, *Old Sea Wings, Ways and Words* (1930); G. la Roërie and J. Vivielle, *Navires et marins, de la rame à l'hélice* (1930), H. Suder, *Vom Einbaum und Floss zum Schiff* (1930); E. K. Chatterton, *Fore and Aft* (1912), *Sailing Ships and Their Story* (1923), *Ships and Ways of Other Days* (1924), *Ship Models* (1923), *Sailing Models, Ancient and Modern* (1934); A. B. Lubbock, *The China Clippers* (Boston, 1914; Glasgow, 1919), *The Colonial Clippers* (1948), *The Blackwall Frigates* (1950), *The Western Ocean Packets* (1925), *The Last of the Windjammers* (1927–29), *The Down Easters* (1929), *The Nitrate Clippers* (1932), *Coolie Ships and Oil Sailers* (1935), *The Arctic Whalers* (1937); J. Hornell, *Water Transport* (1946); R. and R. C. Anderson, *The Sailing Ship* (1926); F. C. Bowen, *The Golden Age of Sail* (1925), *From Carrack to Clipper*, rev. ed. (1948); H. I. Chapelle, *American Sailing Craft* (1936), *History of the American Sailing Navy* (New York, 1949; London, 1950); R. M. Nance, *Sailing-Ship Models* (1949); J. R. Stevens, *An Account of the Construction and Embellishment of Old Time Ships* (1949); L. Casson, *The Ancient Mariners* (1959); R. C. Anderson, *Oared Fighting Ships* (1962).

See also *The Mariner's Mirror*, London (1911–).

Modern and Power Ships: E. J. Reed and E. Simpson, *Modern Ships of War* (1888); H. Fry, *History of North Atlantic Steam Navigation* (1896), *Jane's Fighting Ships* (annually, from 1898); A. J. Maginnis, *The Atlantic Ferry* (1900); A. C. Hardy, *Merchant Ship Types* (1924); H. W. Wilson, *Battleships in Action* (1926); H. P. Spratt, *Merchant Steamers and Motor Ships* (1949), *Transatlantic Paddle Steamers* (1951), *Marine Engineering* (1953), S. G. Sturmey, *British Shipping and World Competition* (1962); J. J. Stilwell, P. W. Nelson, and W. R. Porter, "Hydrofoils at the Crossroads," *Aerospace Engineering* (March 1961). *See also* the British Admiralty *Manuals of Seamanship* and Brassey's *Naval and Shipping Annual*.

SHIPKA PASS, Bulgaria, a pass through the Balkan Mountains, the scene of fierce fighting in the Russo-Turkish War of 1877–78. Situated on the main road from Ruse on the Danube through Stara Zagora to Adrianople (Edirne) in Turkey, this pass was an important point in the Russian plan of operations. It was originally held by a Turkish force of 4,000 men with 12 guns, but the Russian Gen. I. V. Gurko seized it by surprise on July 19, 1877. Gen. Suleiman Pasha, who drove the Russians across the Balkans, attacked Shipka on Aug. 21. The Russian force there, including 7,500 Bulgarian volunteers, held the position against Suleiman's 30,000 Turks. The fighting continued till Sept. 26, when both adversaries entrenched themselves on the pass. After the capitulation of the Turkish forces at Pleven (*q.v.*) in December, however, the Russians began a general advance; and on Jan. 5, 1878, Gen. F. F. Radetski attacked the Turks on Shipka Pass. The fall of Pleven made Turkish resistance at Shipka strategically useless, and on Jan. 9 Gen. Vessil Pasha (who had succeeded Suleiman) surrendered. The Russians had lost 5,500 men on the pass, the Turks 13,000. The manner in which Suleiman had sacrificed his men earned for him the name of the "Shipka butcher."

SHIP LANES: *see* SHIPPING ROUTES.

SHIPLEY, an urban district in the Shipley parliamentary division of the West Riding of Yorkshire, Eng., 3 mi. (5 km.) NNW of Bradford. Pop. (1961) 29,758. Area 3.4 sq.mi. (8.8 sq.km.) Shipley is in Airedale south of the Aire River and on the Leeds-Liverpool Canal.

Mentioned in Domesday Book as Scipeleia, it was developed when Sir Titus Salt started his vast woolen mills (1853) and built the model village of Saltaire with its Italian-style Congregational Church (1859) and the Saltaire Institute (1870).

Engineering is important, but Shipley is also a residential area. The urban district (created 1894) includes Shipley, Windhill, Wrose, part of Frizinghall and Saltaire parishes.

SHIP MONEY, a nonparliamentary tax first levied in medieval times by the English crown for the defense of the country; it is primarily remembered because its enforcement as a general tax by Charles I provoked widespread opposition and led to the famous trial of John Hampden. During the Hundred Years' War the crown had exercised the right of requiring the maritime towns and counties to furnish ships in wartime, and the liability was sometimes commuted for money. Although by the 17th century it was well established that taxes (other than agreed commercial dues) could not be imposed without a specific grant by Parliament, the levying of ship money in time of war had never fallen wholly into abeyance and in 1619 James I aroused no popular opposition by levying £40,000 of ship money on London and £8,550 on other seaport towns.

On Feb. 11, 1628, Charles I issued writs requiring £173,000 for the provision of a fleet to secure the country against possible French invasion and for the protection of commerce. Every county in England was assessed. This was the first occasion when the demand for ship money aroused serious opposition and Charles withdrew the writs. After 1629, however, his determination to rule without Parliament led him to reemploy this financial device.

The first of six writs issued annually between 1634 and 1639 appeared on Oct. 20, 1634, and was directed to the justices of London and other seaports, requiring them to provide a certain number of ships of war, or their equivalent in money, and empowering them to assess the inhabitants for payment according to their substance. The distinctive feature of the writ of 1634 was that it was issued to provide for a possible future emergency rather than against a present danger. The citizens of London immediately claimed exemption under their charter, while other towns demurred to the amount of their assessment; but no resistance on constitutional grounds appears to have been offered, and nearly £80,000 was collected, all of which, as on future occasions, was spent on the fleet.

A second writ of ship money was issued on Aug. 4, 1635, directed on this occasion, as in 1628, to the sheriffs and justices of inland as well as of maritime counties and towns, demanding the sum of nearly £200,000, which was to be obtained by assessment on personal as well as real property, payment to be enforced by distress. This demand excited growing popular discontent, so Charles obtained a written opinion, signed by 10 out of 12 judges consulted, to the effect that in time of national danger, of which the crown was the sole judge, ship money might legally be levied on all parts of the country by writ under the great seal. Had this ruling subsequently gone unchallenged the king's prerogative would have become the sovereign power in the realm. The issue of a third writ of ship money on Oct. 9, 1636, made it evident that the king intended to convert it into a permanent and general form of taxation without parliamentary sanction.

Payment was refused by Viscount Saye and Sele and by John Hampden (*q.v.*). The case against the latter was heard in 1637 before all the common law judges in the exchequer chamber and lasted for six months, Hampden being defended by Oliver St. John and Robert Holborne. Seven of the 12 judges, headed by Sir John (later Baron) Finch, chief justice of the common pleas, gave judgment for the crown, and 5 for Hampden, although 2 found for him only on technical grounds. The narrowness of the king's victory strengthened the opposition to ship money and about 20% of the tax levied in 1638 remained unpaid. Charles issued a writ of ship money for the last time in 1639, but of the £210,400 demanded, only about £50,000 was collected. In 1641, by an act of the Long Parliament, ship money was declared illegal.

BIBLIOGRAPHY.—S. R. Gardiner, *History of England*, vol. iii, vi–viii (1883–84), *Constitutional Documents of the Puritan Revolution*, 3rd ed. (1906); W. S. Holdsworth, *A History of English Law*, vol. vi, 2nd ed. (1937); M. D. Gordon, "The Collection of Ship-Money in the Reign of Charles I," *Transactions of the Royal Historical Society*, 3rd series, vol. iv (1910); D. L. Keir, "The Case of Ship-Money," *Law Quarterly Review*, vol. lii (1936).
(G. Ds.; X.)

SHIPPARD, SIR SIDNEY GODOLPHIN ALEXANDER (1837–1902), British colonial administrator in Bechuanaland, was born in Brussels on May 29, 1837. He went to the Cape Colony in 1868 to practise law. After holding important appointments in Griqualand West as attorney general (1875–77) and as recorder of the high court (1877–78), he became a Cape supreme court judge (1880–85). When Britain formally took over Bechuanaland in September 1885, Shippard became administrator and chief magistrate of British Bechuanaland and resident commissioner of the Bechuanaland protectorate. He was knighted in 1887, and retired in November 1895.

Shippard enthusiastically approved of Cecil Rhodes's plans to extend British influence northward into central Africa and to forestall possible advance there by the Germans or by the Boers. He supported Rhodes on several crucial occasions. In 1887, for instance, Shippard pressed Sir Hercules Robinson (later Baron Rosmead), the high commissioner in South Africa, not to sanction a treaty between Lobengula, ruler of Matabeleland-Mashonaland, and Piet Grobler, who represented the Transvaal, but to send the Rev. J. S. Moffat to negotiate a treaty (February 1888) pledging the chief not to cede territory without Britain's consent. He then helped to persuade Lobengula to grant to Rhodes's agents in October 1888 the Rudd mineral concession, which became the basis of the British South Africa Company, chartered in 1889.

Shippard's complicity in the raid (December 1895) of L. S. Jameson (*q.v.*) was suspected, but never proved. He persuaded two protectorate chiefs in 1894 to let the company administer their territory on the Transvaal border, including Pitsani, the jumping-off point for the raid. He was in Johannesburg during the abortive rising in December, and publicly, though unofficially, asked the Reform committee to lay down its arms. Shippard afterward returned to England and was made a director of the British South Africa Company in 1898. He died in London on March 29, 1902.
(M. F. K.)

SHIPPING. The function of shipping is to transport goods and persons across water, that is, across the world's oceans or along its coastlines or through its inland waterways. The chief work of the shipping industry is to provide, man, and manage the vessels that carry most of the goods entering into world trade; and also to provide and maintain the ancillary services necessary for the receiving of cargoes and for their onward dispatch at ports of destination. As a transport industry it differs from most other industries in that it does not manufacture products to sell but exists to provide a service; and it performs this function almost entirely in an international field—the high seas—for the most part outside the territories of the shipowner's country of domicile. Shipping is an essential element in world commerce and is highly competitive on an international scale. It is also an important factor in the economy of the major trading nations; and, especially since the beginning of the 20th century, the possession of a merchant fleet has become important as a strategic necessity and for national prestige.

The shipping industry had its origin in the Industrial Revolution and its consequences—the invention of the marine steam engine, screw propulsion, and the use of iron, and later steel, for shipbuilding. Although the transition from sail to steam was gradual, the early steam navigation companies were the real founders of the shipping industry. The great expansion of world commerce that followed the Industrial Revolution called for a new system of ship management and shipping services, by which the ship became an instrument of world trade and an essential service to commerce, rather than a private venture on the part of a merchant trader or shipmaster or shipowner-merchant. After the opening of the Suez Canal in 1869, which sounded the death knell of the commercial sailing ship, the expansion of the volume of world seaborne trade became enormous. The development of the mechanically propelled merchant ship, and its continual increase in size, efficiency, and numbers made this growth possible.

Table I demonstrates the expansion of the world fleet of mechanically propelled merchant ships, and the decline of the sailing ship, in the period between 1886 and 1935; Table II illustrates the subsequent development of the world merchant fleet (see *World Merchant Fleets*, below). Though it is not possible to provide comparable statistics of the increase in the volume of world seaborne trade over the same period, particularly for the 19th century, League of Nations statistics estimated that by 1929 the amount of world seaborne trade had reached a level 35% higher than that in 1913. In the depression of the early 1930s the figure dropped at one time almost to the 1913 level, but by 1937 it had fully recovered to the 1929 level. According to United Nations statistics, the volume of seaborne trade reached in 1937 (490,000,000 metric tons) had again been recovered by 1948. Further, by 1962 the volume of seaborne trade had increased to 1,200,000,000 tons.

At the time of the Suez crisis (1956) there was a boom in the ordering of new tonnage. Thereafter, international trade continued to expand, but not at the rate forecast and as a result too many ships became available for too few cargoes, depressing freight rates. The depression of the late 1950s showed little sign of lifting by the mid-1960s.

For the early history of maritime commerce, *see* SHIPPING ROUTES; COMMERCE, HISTORY OF. (The history of various types of vessels is covered under SHIP.)

This article is divided into the following sections:

I. MAIN ELEMENTS OF SEABORNE TRADE

While the volume of seaborne trade has been steadily increasing since the introduction of steam navigation, its nature and direction have been subject to changes. The main elements of seaborne trade are food and raw materials exported by primary producing countries and manufactured goods and capital plant exported by industrial countries in exchange. The shipping industry exists to enable these goods to be interchanged, but the greater part of the cargoes carried by merchant ships consists of basic raw materials (such as coal, oil, mineral ores, and fertilizers) and foodstuffs (particularly grain, sugar, rice, vegetable oils, fruit, meat, and dairy produce).

The growth of the world's population and the expansion of industrial development continue to make additional demands on the shipping industry. The process is best illustrated by taking the development of British trade and shipping. The invention of the steam engine led to a worldwide demand for British coal exports, which increased from about 11,000,000 tons in 1870 to as much as 73,000,000 tons in 1913, not taking into account an additional 21,000,000 tons shipped overseas for use as ships' bunkers. After World War I the growth of motor transport and the development of diesel engines for industry and transport caused a decline in the demand for exports of United Kingdom coal, which amounted to about 36,000,000 tons in 1938, in addition to 10,000,000 tons for bunkers. By the mid-1960s coal exports of all categories, including bunkers, had fallen to barely 6,000,000 tons. Meanwhile, of course, the importation of petroleum products had already assumed large-scale proportions. By the mid-1960s the United Kingdom was importing oil at the rate of more than 60,000,000 tons.

Another long-term development was the increase of the requirements of the steel industry. For many decades British steelworks were able to rely on domestic supplies of native iron ore, but in the first half of the 20th century they had to draw more and more on supplies from sources such as Sweden, France, North Africa, and Newfoundland. Demand for iron ore increased so rapidly after World War II that United Kingdom imports of iron ore rose from 11,500,000 tons in 1954 to about 15,000,000 tons in 1956, some of which had to be drawn from much more distant parts of the world, such as Venezuela, Brazil, and Liberia, so that there was an additional load to be accepted by the shipping industry. By 1964 this figure had reached 18,400,000 tons.

Parallel developments took place in the United States economy. Before World War II the United States was a prominent exporter of petroleum products to the rest of the world, but shortly after the war it became a net importer of oil. At the same time its coal export trade soared to new heights as a result of a deficiency of fuel in Western Europe. The steel industry expanded so rapidly that it, too, could no longer rely on supplies of domestic ore, and a vast expansion of imports from Labrador, South America, and Liberia began.

Technical and economic changes affect demands on the shipping industry, and basic trade routes are laid down according to the pattern of economic geography. It is the geographical position of sources of the principal commodities shipped and that of their main centres of consumption that chiefly shape the pattern of world shipping movements, although these are, of course, influenced by political and physical factors as well.

The chief grain-producing areas of the world, for example, are the United States and Canada, Australia, Argentina, the U.S.S.R., and France. India at one time was an exporter of wheat, but the growing population and rising standard of living of that part of the world have made India and Pakistan substantial importers of grain; much the same situation now applies to the countries under Soviet influence, and the exportable surplus of France as often as not becomes a deficit. There is a steady flow of grain from the three principal exporting areas—North America, Australia, and Argentina—to the continent of Europe and to Japan, as well as to most other countries in the world, but the largest importer of grain is the United Kingdom, as it is of most other foodstuffs. Britain's need for imported foodstuffs is indeed a staple element in seaborne trade, for it must rely on meat from Australia, New Zealand, Argentina, and Uruguay; on fruit from South Africa, Australia, New Zealand, the Mediterranean countries, the West Indies, and Central and North America; on dairy produce from Australia, New Zealand, and Denmark; and on perishables from the near continent, particularly from France, Belgium, and the Netherlands. Sugar comes mainly from Cuba, Mauritius, and Queensland (Australia), which are the main exporters of this commodity to the rest of the world.

Of the sources of energy, oil is steadily surpassing coal in volume, but after World War II coal was a staple export from the United States to Western Europe, mainly to Italy, France, and the Federal Republic of Germany, to Japan, and to South America. The main oil-exporting centres after World War II were the Middle East (Iraq, Iran, Kuwait, Saudi Arabia, Bahrain, and other neighbouring oil fields) and Venezuela and the Caribbean area. The Soviet Union and the United States are prolific producers, but mainly for domestic consumption. Thus oil from the Middle East flows to all parts of the world, the principal consumers being the United Kingdom and Western Europe, with increasing consumption in India, Africa, and the Far East, the United States, South America, and Australia.

Of industrial raw materials, mineral ores became predominant after World War II. The steel industries of the United States, the United Kingdom, the Federal Republic of Germany, and Japan are the greatest consumers of iron ore, which is imported from Labrador, West and North Africa, India, Sweden (as a rule through Norwegian ports), Venezuela, and other parts of South America. Copper is imported by the major industrial countries chiefly from Rhodesia and Chile; tin from Malaya and Bolivia; lead and zinc from Australia. Bauxite, or alumina, the raw material for the production of aluminum, flows mainly from the Caribbean to the United States and Canada, and from West Africa and the south of France to the United Kingdom, Norway, and other European producers. The produced metal is exported by Canada and the United States in ingot form to other manufacturing countries, particularly Great Britain, from which the finished manufactured products may be exported to all parts of the world.

Besides the principal commodities in the seaborne trade mentioned above, many others are handled by the shipping industry in quantities exceeding 1,000,000 tons a year, and they include phosphates, iron and steel scrap, potash, nitrates and other fertilizers, pyrite, sulfur and sulfur compounds, basic slag, cement, rubber, wool, cotton, oilseeds and vegetable oils of various kinds, rice, and coffee. Other commodities, such as tea, jute, timber, tobacco, copra, salt, china clay, and chemicals, are also shipped by sea in large quantities, along with a great variety of shipments of manufactures, machinery, and processed goods.

II. SPECIALIZATION IN SHIPPING

It would be an oversimplification to suppose that international trade is mostly a two-way traffic between industrial manufacturing and primary producing countries, or even an interchange of manufactured goods between one country and another. Much of the world's trade is triangular in character, involving the shipment of raw materials from one country to another, where they are converted or processed for export to a third country. The rate and direction of flow of trade in various commodities, whether raw materials, foodstuffs, or manufactured goods, is subject to continual change, which calls for considerable flexibility in the employment of ships.

The different types of goods carried by sea require, moreover, a wide variety of types of ships, and it is the business of the shipping industry to provide such vessels. The growth of international trade has thus encouraged the development of some specialization in shipping, which has taken various forms, such as the provision of ships designed to carry certain cargoes or to participate in certain trades along well-established routes.

1. The Liner and the Tramp.—The two main branches of the shipping industry are the liner business and tramping. The essential difference between the liner and the tramp is that the former operates regular scheduled services on a predetermined trade route, whereas the tramp is prepared to carry any suitable cargo from any part of the world to any other part of the world. The liner services are the railways of the oceans, running regular services at stated times between scheduled ports, and charging previously advertised tariff rates. The tramps are more like taxicabs or road transport operators, which may ply for hire or will contract to transport goods between any required ports, the terms being arranged by negotiation between the parties on each occasion, or on a contract basis. The tramp system had its origin in the days when it was customary for the master to be owner, or at least part owner, of his own ship. He would load it with a cargo of goods which he hoped to sell in some foreign country, and use the proceeds to buy a cargo which he could sell in another country

or for which he could find a ready market at home.

As more regular trade developed between two overseas countries, such as that between the United States and the United Kingdom, and with the advent of steam navigation, which gave greater assurance of regularity of communications, the steam navigation companies began to operate liner services. The liner services were, and still are, of particular value for passengers and mails, and for small traders not dealing in commodities normally consigned in shiploads. The shipper of commodities in bulk, such as grain or coal, naturally found it more economical to hire ships on the open market, offering his business of transportation to the lowest bidder, or buying and selling his commodity in the market where sea transport costs were for the time being cheapest. Thus the tendency is for tramp ships to be the carriers of homogeneous cargoes of commodities in bulk, while liners are principally carriers of heterogeneous cargoes, made up of hundreds of small consignments of goods of many kinds, as well as mail and passengers.

There thus grew up what is now a worldwide network of liner services between the principal ports of the world. These services link the countries of Europe with the other continents, directly across the Atlantic to the eastern shores of North and South America, through the Panama Canal to the western shores of America, the Pacific Ocean, Australia, and New Zealand; directly to North, West, and South Africa and on to other parts of the world; or through the Suez Canal to East Africa, the Indian Ocean, the East Indies, and the Far East. They link the United States in addition directly, or through the Suez and Panama canals, with Europe, South America, Africa, and the countries of the Indian and Pacific oceans. They link Japan with Africa and America, India with Africa, China with India, and, finally, by transshipment to coastal liner services they provide sea communications between practically all the seaports of the world.

The bulk cargoes carried by tramp ships also tend to flow in volume along regular routes from time to time, and where these coincide with regular liner services the liners will periodically pick up "parcels" of these commodities, particularly grain, but usually only in order to fill up holds that otherwise would be empty.

2. Various Types of Cargoes.—The principal bulk cargoes have varying characteristics (liquid cargoes are in a category apart requiring specially designed ships and must be considered as a separate branch of the shipping industry). The normal dry-cargo tramp ship is adaptable in design, so that it is capable of carrying, as occasion arises, cargoes as different as iron ore, coal, grain, sugar, cotton, or copra. Each of these cargoes has different characteristics. Because of the difference in stowage factors (cubic capacity per ton), for example, a ship large enough to carry 10,000 tons of coal might be loaded down to its maximum draft with less than 5,000 tons of ore, and if it were to carry grain, "shifting boards" would be necessary in the holds to prevent the cargo upsetting the stability of the ship by pouring over to one side when the ship rolled.

Different methods are used for loading and discharging different types of cargo. Cargoes shipped in cases, bales, or bags must be handled by the ship's derricks or by shore cranes, and the same applies to cargoes of commodities like timber, steel, or scrap. Iron ore and coal are discharged by grab, and grain by elevators or suction pipes.

3. Specialized Carriers.—Where any one commodity is carried in large quantities over regular routes, it becomes an economic advantage to build ships specially suited to the cargo concerned, and because of the vast expansion of world trade in bulk commodities the tendency is for specialized bulk-cargo carriers to be built. The specialized ore carrier is a typical example. Large numbers of these have been built since World War II, although they were in operation on the Great Lakes at the end of the 19th century. They rely entirely on shore appliances for cargo handling, not being provided with derricks, although some ore carriers designed for service to ports where shore-handling facilities do not exist are equipped with self-unloading gear. These, of course, are ultra-specialized.

Specialized ships such as ore carriers may still be divided into the two categories of liner and tramp, in that some of them are built by companies owning tramp ships with the object of their being hired out, or chartered, to the shippers or importers of ore, usually on a long-term basis, while others are built and operated by the shippers or consumers of ore themselves, mainly the large steelworks. The latter vessels are liners in the sense that they are designed for and operate on regular routes between predetermined terminals, but they do not operate as common carriers and accordingly are better called industrial carriers. Several other industrial consumers of raw materials in bulk, such as major importers of sugar, newsprint, molasses, gypsum, and chemicals, operate their own fleets of bulk carriers in a similar fashion, as well as hiring ships from tramp shipowners on the open market.

4. Tankers.—The transport of petroleum products by sea is a special branch of the shipping industry which has witnessed a vast expansion in the 20th century, particularly after World War II. Some indication of the rate of expansion is given by comparison of world tanker tonnage in 1939 and 1964. In 1939 the world tanker fleet totaled 11,586,000 tons gross, comprising 16.9% of the total tonnage of all merchant ships. By June 1964 the tanker fleet had reached a total of 50,563,000 tons gross, representing as much as 33% of the total world fleet then in existence. Measured in terms of dead-weight tonnage the world tanker fleet rose from 16,000,000 tons in 1939 to more than 23,000,000 tons in 1946; yet the 1946 figure had been more than tripled by the mid-1960s.

Liquid cargoes such as crude oil or gasoline need ships heavily subdivided into compartments in order to restrict the movement of the free surface derived from the motion of the ship at sea which would otherwise upset the stability of the vessel. No shipborne cargo-handling facilities such as derricks or cranes are required, since the oil is pumped on board through pipelines at the oil terminals and discharged through pipelines by means of the ship's own pumping system at ports of destination. Owing to the inflammable nature of the cargo, special precautions must be taken against the risk of fire. Before World War II the greater part of the world tanker fleet was owned and operated by the major oil companies, operating as industrial carriers, but they depended for their marginal requirements on the chartering of tramp tankers.

After World War II, despite the expansion of their own fleets, the oil companies had to rely to a much greater extent for their ocean transport on tankers operated by tramp companies, generally on a long-term time-charter contract basis. A particular feature of the tanker trade is that it is almost entirely a one-way traffic, and since tankers are not suitable for carrying any other type of cargo without structural alteration, the outward journey must be made in ballast. An exception is made for iron ore, which has a low stowage factor. Sweden, an ore-exporting and oil-importing country, evolved a ship capable of carrying an inward cargo of oil and an outward cargo of ore. As a result of the exploitation of the Labrador ore fields, this combined oil and ore carrier was increasingly built after World War II. In the winter season the ore-loading ports of Labrador are closed by ice, so that these specially designed ships can find alternative employment as oil carriers in the off-season. Other liquid cargoes besides oil products may be carried in considerable bulk, and there are tankers used only for goods such as molasses, caustic soda and other chemicals, wine (and sometimes water), or for semiliquid cargoes such as asphalt and bitumen, and even petroleum gas in liquid form.

5. Coastal and Short-Sea Shipping.—Most of the oceangoing types of ship have their counterpart on a smaller scale in the coastal shipping industry, whose function and extent naturally varies in different parts of the world according to geographical conditions. Coastal shipping (sometimes referred to as cabotage) is another branch of the shipping industry, since its services are generally confined to the coastal waters of a particular country. In countries with continental coast lines, such as the United States, Australia, and India, for example, coastal shipping may not differ operationally from oceangoing shipping, and in the United States the term is extended to include the intercoastal trades between the Pacific and Atlantic coasts via the Panama Canal, and the noncontiguous trades between continental America and outlying

areas such as Hawaii. In Europe, where national coastlines are comparatively short, and the restriction of coastal trades to ships of the national flag is not generally practised, coastal shipping has a more international character.

In the United Kingdom, for example, there can be distinguished the purely coasting trade, plying solely among the islands and along the coast of Great Britain; the home trades, which ply between the United Kingdom and the continent of Europe within the limits of Brest, in France, and the Elbe, in Germany; and the short-sea trades, between the United Kingdom and Scandinavia, the Baltic Sea and the Mediterranean. Some shipping companies, of liners, tramps, or tankers, may operate in all of these trades, while others may specialize in any one.

In the coasting and short-sea trades there are also many specialized vessels such as cross-channel passenger ships, train ferries, ramp-loading or roll-on roll-off motor vehicle transports, ships specially designed for container traffic and the like, as well as colliers and tankers. These tend to be operated by companies specializing in particular routes over which traffic is heavy but localized. As these services often form sea links between or extensions of railway systems, the railway companies often operate them. In fact, the coastal shipping industry may be regarded as complementary to the inland system of transport communications, as well as a link with neighbouring countries; and a most useful characteristic of the coastal ship is its ability to load cargo directly from the oceangoing vessel for transshipment along the coast and distribution to smaller ports whose depth of water or lack of facilities precludes the direct approach of the larger ship.

6. Inland Waterways.—The shipping industry also includes transport services along rivers and inland waterways. Vast quantities of materials are shipped along waterway systems such as those of the Mississippi, the Amazon, the Niger, and the Rhine, sometimes penetrating thousands of miles into the heart of a continent.

See WATER TRANSPORT, INLAND.

7. Ancillary Services.—At the port, where the sea transport and inland transport systems meet, many ancillary services must be provided for ships, such as towage, stevedoring, warehousing, replenishment of stores, repair, and maintenance. At this point industrial practice varies widely between different ports, the various services being provided by independent contractors, by the port authority, or by the shipping companies themselves. In many ports, particularly where large liner companies operate, the companies have their own subsidiary organizations which provide such services, not only for their own ships but also for others. Some of the larger shipping groups operate or control fleets of tugs, loading and discharging facilities and labour, warehouses and refrigerated stores, lighterage and land transport, and even provide for the building, repairing, equipment, and maintenance of ships for themselves and for other shipowners.

Inland waterways and seaports must be kept clear by dredges, and in northern waters which freeze in winter icebreakers must be used to clear shipping lanes; salvage services must be provided for clearing wrecks and other obstructions to navigation and for rescuing ships which have broken down; and cable ships must be provided to maintain telegraphic communications between all parts of the world. All these maritime services, essential to the operation of international seaborne trade, are part of the shipping industry. However, the fishing industry, including whaling, may be regarded as a separate maritime activity unconnected with the commercial operation of merchant ships. *See* FISHERIES; WHALING.

III. SHIP OPERATION

The essential differences between the liner and tramp sections of the shipping industry, namely in the kind of transport service they perform, have been described above. As the carrying performances of liners and tramps differ, so do the operational organization and methods of the companies that run them.

1. Liner Operation.—The growth of liner services resulted from the introduction of steam navigation, which enabled ships to maintain regular services at predictable intervals, mechanically propelled ships being much less dependent than sailing ships on the vagaries of wind and tide. A sailing ship, wherever bound, did not ordinarily sail, either with general merchandise or a bulk cargo, until its holds were full. Sailing dates were therefore uncertain, and arrival dates even more so. The merchant was obliged to ship his goods in speculative anticipation of the state of the market at the time of their arrival at destination, and shippers of small consignments or parcels of cargo were even more uncertain of being able to deliver at a specified time. Fast sailing ships (such as those of the American Black Ball Line) were actually the first to attempt regular liner services across the North Atlantic early in the 19th century, but it was not until the coming of the steamship that such services became firmly established.

The possibility, with liner steamships, of regular sailings from a wider range of loading and discharging ports, with a guarantee that ships would sail whether full or not, and the development of more reliable postal, and, later, telegraphic communications, allowed much more international business to be done, particularly by traders wishing to send small consignments of goods to various destinations at frequent intervals or even only occasionally. Thus liner cargoes are usually composed of a wide variety of general merchandise, of differing values, nature, and quantity, including material that does not customarily move in shipload lots along the trade route concerned. The first modern liner services naturally followed the direction of the greatest demand. The Cunard Line started the first regular steamship liner services from Great Britain to the United States in 1840, and in the same year the Peninsular and Oriental Steam Navigation Company began a liner service from Britain to the Mediterranean area. In 1842 the Royal Mail Steam Packet Company (now Royal Mail Lines) started a regular steam packet service from Britain to the West Indies, with the aid of a government mail subsidy.

The operator of a liner service has special interests in the particular trade route for which he caters. The design and performance of his ships will be directed toward fulfilling in the most economical way possible the requirements of shippers using his services. Some liner routes, such as those between the United Kingdom and Australia and New Zealand, have to cater for a large proportion of refrigerated cargoes such as meat and dairy produce, as well as wool and parcels of grain or base metals, in the direction of Great Britain, and for machinery and motor vehicles and other manufactured goods in the opposite direction. The requirements of liner trades on any particular route range widely in character and fluctuate in quantity. They are apt to be shipped by numerous individual exporters and consigned to no less numerous individual consignees. The collection of such consignments and their assembly into the ship at the loading port, or ports, and their distribution at the ports of discharge obviously require an extensive, permanent organization at each end of the liner route; and the liner company must have staff and offices at the ports that it serves, although in some of the less active ports these functions may be carried out by agents acting for one or more liner companies. Besides this, the services of forwarding agents or travel agents must be employed for soliciting cargo or passengers from inland centres sometimes far removed from the ports of shipment, and it must always be possible to arrange for the transshipment or onward movement of cargo and passengers from terminal ports to eventual destination.

2. The Conference System.—Not only do the liner companies provide regular and frequent services between different parts of the world for the carriage of every conceivable kind of cargo in consignments large and small, but they are prepared to do so at previously advertised fixed charges. They are enabled to do this by means of the liner conference system, which first began to be practised in the United Kingdom-Calcutta trade in 1875. The object of the conference system is to regulate uneconomic competition. The shipping companies of different ownership and nationality that operate services between the same range of ports form a conference agreement to regulate the freight rates that they will charge for each type of goods carried; and in some cases the agreement also allocates a specified number of sailings to each company. Coupled with this agreement there is generally a de-

ferred rebate system, by which regular shippers of goods by conference vessels receive a rebate of a percentage of the tariff freight rate, payable after a period of proved loyalty, provided they use conference vessels exclusively.

The shipping conference system has sometimes come under attack as tending to create monopolies and to restrict competition against the public interest. After investigations, however, the evidence has weighed in favour of the conference system (which in the United States is subject to regulation and approval by the Maritime Administration). It has been felt that no combination of shipping companies can force unreasonable freight rates on shippers, since under those circumstances an outside company will step in to provide a genuine service at reasonable rates; and, on the other hand, shipping companies that provide regular sailings with good ships and maintain staffs and organizations in ports to handle and dispatch cargoes, irrespective of whether trade is good or bad, are entitled to some protection against the casual ship that picks up an occasional cargo at cut rates. It is to the advantage of the shipper that through this system he can rely on a well-managed service, running ships that will carry any desired quantities of his goods at predetermined rates, and that these rates are not usually changed without at least three months' notice.

3. Tramp Operation.—Unlike liners, tramp ships do not provide scheduled sailings between predetermined ports, and they are generally used solely for the carriage of bulk commodities, or homogeneous cargoes, in whole shiploads. There is no fixed tariff of freight rates and no conference system. Every voyage of a tramp ship is the subject of a separate negotiation between the owner of the ship and the owner of the cargo, and its terms are generally arranged through a ship broker in one of the shipping exchanges located in various parts of the world, but mainly in London and New York.

4. Brokerage and Chartering.—Most of the world's tramp ship chartering business is carried out in the Baltic Mercantile and Shipping Exchange, Ltd., in London, commonly known as the Baltic. This exchange had its origins in the 17th century, when merchants and ships' captains were accustomed to meet in coffeehouses (the most famous of which was Lloyd's) to arrange cargoes for ships. The Baltic and the Jerusalem were the two coffeehouses chiefly patronized by merchants and captains until the business was moved to the Baltic tavern in 1810. In 1823 the first rules and regulations of a "Baltic Club" were drawn up, limiting membership to 300 and providing for accommodation and refreshments. At that time tallow was the principal commodity in the Baltic trades, but with the repeal of the Corn Laws in 1846 the buying and selling of grain cargoes became one of the chief activities and called for larger premises. In 1856 larger premises were taken in South Sea House and two years later membership had doubled.

The opening of the Suez Canal and the growth of steam navigation caused a further expansion of trade and in 1903 the Baltic Exchange in its present form was opened in an entirely new building. After World War II air chartering was added to the exchange's activities, and in 1956 a new building was opened as an extension to the 1903 structure, doubling its size.

On the floor of the Baltic Exchange brokers circulate daily, some of them representing the shippers of cargoes such as grain, coal, or ores who require shipping space for their movement, and others representing the owners of tramp ships wanting further employment. When seeking a ship for the carriage of a cargo, the broker naturally looks out for a ship of the right size and specifications, ready at the right time and in the right port, and prepared to carry the cargo at the cheapest possible rate. Conversely, the owner's broker must attempt to anticipate the charterers' requirements by having his ship at the right place at the right time, but he must also try to obtain the highest freight rate possible without running the risk of losing the contract to a competitor. This constant interplay of supply and demand, of ships as well as of cargoes, has the effect of reducing sea transport costs to a minimum, although tramp freight rates fluctuate frequently and widely at times.

These conditions of fluctuating rates also offer opportunities for speculation, as on a stock exchange. The shipper of grain, for example, may decide to charter a ship at what he considers to be a cheap rate even before he has sold the cargo, which he will attempt to dispose of in the best market while it is still afloat. In this case the ship will be chartered for a voyage from, say, the Río de la Plata to the English Channel "for orders," freight rates having been previously arranged to cover a variety of alternative eventual destinations. Again a shipper or an owner may agree to charter a ship for a period of time at a certain rate, in the hope that he will later be able to "relet" it to another owner or shipper at a higher rate, if the market improves. A good deal of "forward" chartering may also occur, in which a contract is entered into for performance at some specified time at a predetermined rate that both owner and hirer hope will prove favourable, according to their view of future market conditions.

5. Methods of Chartering.—There are four principal methods of chartering a tramp ship—on voyage charter, on time charter, on bareboat charter, or on a contract or "lump sum" basis. The voyage charter is the commonest. Under this method a ship is chartered for a one-way voyage between specified ports with a specified cargo at a negotiated rate of freight. The charterer agrees to provide the cargo for loading at the port or ports and the owner to present the ship for loading within an agreed range of dates. As soon as the cargo has been delivered at the port or ports of destination, the ship is free for further employment at the owner's discretion. Sometimes, however, the arrangement is for a series of consecutive voyages, generally for similar cargoes over the same route. The freight rate is expressed in terms of so much per ton of cargo delivered.

On time charter, the charterer undertakes to hire the ship for a stated period of time, or for a specified round trip voyage or occasionally for a stated one-way voyage, the rate of hire being expressed in terms of so much per ton deadweight per month (*see* below for meaning of tonnage measurement). Whereas on a voyage charter the owner bears all the expenses of the voyage (subject to agreement about costs of loading and discharging), on time charter the charterer bears the cost of bunkers and stores consumed.

On bareboat charter, which is less frequently used in ordinary commercial practice, the owner of the ship delivers it up to the charterer for the agreed period without crew, stores, insurance, or any other provision, and the charterer is responsible for running the ship as if it were his own for the period of the contract.

Contracts can also be arranged on a lump-sum basis, when an owner agrees to ship a given quantity of a stated cargo from one port to another for a stated overall sum of money. Sometimes large quantities of cargoes such as coal are arranged for on a contract basis. The shipowner agrees to undertake the shipment of a given quantity over a given period at a fixed price per ton of cargo, but not necessarily in any specified ship, although he will generally use his own ships if they are available, unless he can subsequently charter other ships at a cheaper rate. The question of substituted ships, however, often leads to disputes and the terms of the contract may make special provisions for this eventuality.

6. Legal Aspects of Chartering.—Once the owner's broker and the charterer's broker have agreed on the terms of contract, the ship is "fixed," although by word of mouth alone. The proud motto of the Baltic Exchange is "our word is our bond," and that code of ethics is observed by ship brokers throughout the world, whether they are members of the Baltic or not, for the simple reason that any ship broker who did not abide by it would be unable to transact further business. Legally, however, the final contract is the written "charter party," which for most transactions is accepted as a standard document and is agreed to by all parties normally concerned in a tramp ship "fixture" for a particular trade. The standard form of charter party covers all the main points which experience has shown might later lead to misunderstanding or disputes about the liabilities of each party, while the variable details, such as the particulars of the voyage, cargo, ship, loading and discharging conditions, and ports, rate of freight, etc., can be inserted in accordance with the prior verbal agreement. The

charter party is the document that is subject to scrutiny and interpretation by a court of law in the event of dispute, but in practice most disputes are by agreement submitted to arbitrators (generally independent ship brokers appointed by each participant) for settlement, unless an important legal precedent is involved.

Among the most important clauses in any charter party are generally those which lay down the number of days allowed for loading or unloading, and those which determine who is to bear the expenses involved. As a rule, a certain number of days are allowed for loading, and if the charterer fails to finish loading in the time specified the shipowner is entitled to compensation for demurrage, while the charterer, on the other hand, is entitled to dispatch money if the cargo is loaded in less than the time agreed in the charter party. Similar conditions usually apply in the port of discharge.

Tankers, ore carriers, and other industrial carriers operate either like dry cargo liners or like tramps, except that on scheduled services they do not normally act as common carriers. In other words the tanker or the ore carrier is operated by an oil company or a steel company or an ore exporter exclusively on company business, whether on a regular schedule or not; or it is owned by an independent tramp shipowner and taken on charter by an oil or steel company to fulfill its marginal transport requirements on terms similar to those employed in the dry-cargo tramp shipping market. That is, perhaps, a simplification of the procedure, since in fact the oil company may, for example, find it convenient to charter its own ship to another oil company for a voyage or for a period of time, just as a tramp shipowner might do. Most of the major oil companies (and permanent large shippers of bulk commodities) build and operate for themselves fleets of ocean carriers—and their operation is a matter of company convenience, requiring no contractual obligations or other arrangements except suitable shore staffs and organizations or agencies at the terminals. The great expansion of oil transport requirements after World War II led to a rapid growth of the world tanker fleet, and capital commitments made it more than ever necessary for oil companies to rely on the tankers built and operated by the growing number of independent tank shipowners.

IV. INTERNATIONAL MARITIME LAW

1. The International Conventions.—Ships operate in an international field, the high seas, and are subject not only to the laws of the country in which they are registered and to the laws of the country in whose territorial waters they may be, but also to a series of international conventions, the principles of which are incorporated into the domestic legislation of most maritime countries. There has grown up, particularly in the years since the expansion of steam navigation, a body of international maritime law—"the common law of the sea"—which has been developed through international agreement. For example, nearly all the world's maritime nations have adopted the International Rules for the Prevention of Collision at Sea, which were originally based on British rules formulated in 1862 and made internationally effective after a series of international meetings culminating in a conference at Washington, D.C., in 1889. These rules lay down in great detail how ships must navigate in respect of each other, what lights must be shown and what signals must be given in accordance with circumstances; any infringement of this international code of conduct is accepted in all maritime courts of law as prima facie evidence of liability in case of collision.

Similarly, the internationally accepted requirements for the protection and safety of life at sea, as far as the ship and its equipment are concerned, are embodied in the International Convention for the Safety of Life at Sea. The sinking of the "Titanic" in 1912 gave rise to a general desire to raise the standards of safety of life at sea, and a convention was drawn up in 1914 requiring certain minimum standards for passenger ships; however, this did not become fully operative owing to World War I. A second Safety of Life at Sea Convention was drawn up in 1929; this decided minimum standards for the construction of passenger ships engaged in international voyages and for the provision of lifesaving appliances, and extended its rules for the safety of navigation to all ships on all voyages. The provisions regarding radio equipment were extended to cover cargo ships of more than 1,600 tons gross.

Improved techniques led to the convening of a third Safety of Life at Sea Conference in London in 1948, after which a third International Convention on the Safety of Life at Sea was adopted. This came into force on Nov. 19, 1952, having been ratified one year previously by the following countries in order of acceptance: United Kingdom, New Zealand, the United States, France, the Netherlands, Sweden, Norway, South Africa, Iceland, Portugal, Canada, Pakistan, Denmark, Yugoslavia, Italy, Belgium, Israel, Japan, and the Philippines. The scope of the 1929 convention was again extended to include, in several matters, cargo ships of 500 tons gross and above. Minimum requirements were laid down for the construction of ships, for the provision of watertight bulkheads, for the closing of openings in hulls for lifesaving appliances, for fire appliances, for radio equipment and direction finders, for grain divisions in ships carrying grain cargoes, for the carriage of dangerous goods, and for emergency musters and drills. The 1952 convention also embodied the recommendations of governments associated with the World Meteorological Organization, and provided for the continuance of the International Ice Patrol maintained by the United States Coast Guard in the North Atlantic. This convention was again revised in 1960.

Whereas the Safety of Life at Sea Convention deals with the construction of ships from the safety point of view, particularly in respect of passengers and crews, the International Load Line Convention deals with the strength and seaworthiness of ships in normal operating conditions. This convention emerged from the efforts of Samuel Plimsoll, who succeeded in securing the passage of the British Merchant Shipping Act in 1875. This act provided for the marking of a load line, thereafter popularly known as a Plimsoll mark, on the ship's side, indicating the maximum depth to which a ship could legally be loaded.

In order to protect the competitive position of British ships, the Merchant Shipping Act of 1890 required all foreign ships leaving British ports to comply with the load line regulations. This led to the adoption of the load line rules by most maritime countries, and the International Load Line Convention of 1930 was ratified by 54 nations. The 1930 rules followed the British proposals closely, but provided for the deeper loading of tankers, of ships of special construction, and of vessels carrying deck loads of timber. Provision was also made for the adjustment of load lines to suit seasonal and geographical variations in conditions.

In 1958 the Inter-Governmental Maritime Consultative Organization (IMCO) (*q.v.*), a specialized agency of the United Nations, came into existence. The purpose of this organization, which is advisory in nature, is to promote international cooperation in maritime navigation. By the end of 1960, membership in IMCO had grown to 44 nations. The Maritime Safety Committee of IMCO approved the organizational arrangements for the 1960 Safety of Life at Sea Conference.

2. Commercial Legislation.—International agreements and international law are also concerned with the business dealings between maritime countries. In this connection the International Maritime Committee (Comité International Maritime) and the International Law Association did invaluable work. Delegates to the International Conference on Maritime Law held at Brussels in 1922 recommended to their respective governments the adoption of a set of rules, known as The Hague Rules, which establish the responsibilities, rights, and immunities of carriers under bills of lading (documents acknowledging receipt of cargo for shipment and proving entitlement to the goods). The Hague Rules, which are generally also incorporated into the terms of a charter party, were given widespread legislative sanction by maritime countries, in Britain by the Carriage of Goods by Sea Act, 1924, and in the United States by the Carriage of Goods by Sea Act, 1936, which is also taken in conjunction with the Harter Act, 1893. Legislative sanction is not necessary to secure international agreement or observance in every case.

The York-Antwerp Rules of General Average, for example,

differ in some respects from English law relating to marine insurance and contracts of affreightment. They were formulated by delegates from various maritime countries meeting at York, Eng., in 1864 and at Antwerp in 1877, and were revised at Stockholm in 1924 and at Amsterdam in 1950. They were so widely adopted by the maritime nations that for all practical purposes they have the force of law. *See* MARINE INSURANCE.

V. SHIPOWNERS' AND SEAMEN'S ASSOCIATIONS

Shipowners themselves are generally organized on a national, and sometimes on a regional, basis. The most highly organized associations of this kind grew up in the United Kingdom, where all but a negligible minority of shipowners are members either of the Chamber of Shipping of the United Kingdom, based in London, or the Liverpool Steam Ship Owners' Association. These two associations have sections which look after the special interests of subgroups in the industry, such as tankers, tramps, liners, coasters, etc. In order to speak with one voice in negotiations with the government on matters of highest policy, the two associations created a joint organization known as the General Council of British Shipping. These organizations speak for shipowners as a whole in regard to many industrial matters, such as pilotage and towage agreements and contracts, port organization, dock labour, and national legislation, and to settle countless points at which external matters touch the maritime industries. Associations in other countries often also undertake wage negotiations with seafarers; the British associations, however, do not. Wage negotiations are handled by a separate organization known as the Shipping Federation, which was formed early in the 20th century in order to combat the activities of the newly created National Union of Seamen. Wage negotiations in the British mercantile marine since World War I have been conducted entirely through the medium of the National Maritime Board, which consists of representatives of the Shipping Federation on the one side and representatives of the seafarers' and officers' unions on the other. The work of the National Maritime Board extends, in addition, to the associated subjects of manning scales, hours and conditions of work, and standards of accommodation, etc., as well as to such matters as sickness benefits, pensions, and contracts of employment. Since 1947 it has been possible for British officers and seamen to enter either into individual service contracts with shipping companies, or into a general service agreement with the industry as a whole which guarantees a minimum wage for a minimum period of two years, whether or not the seafarer is in employment, provided that he undertakes to serve in any British ship to which he may be posted. (*See* also LABOUR LAW: *England*.)

On the international level, the national shipowners' associations are members of the International Chamber of Shipping, provided that the national industries concerned are not state controlled. This organization, whose offices and secretariat are housed at the London headquarters of the Chamber of Shipping of the United Kingdom, was formed before World War II, and excluded state shipping organizations of any kind. The ownership of the national fleet, or of a substantial part of it, had to be vested in private individuals or companies. The International Chamber of Shipping was suspended during World War II but in 1946 shipowners representing 14 nations (Australia, Belgium, Canada, Denmark, Finland, Greece, India, the Netherlands, New Zealand, Norway, Spain, Sweden, the United States, and the United Kingdom) met in London to reconstitute it. The assembly wished to avoid the overlap of its functions with those of other organizations and, "bearing in mind the vital need for preserving the principles of private enterprise in a complex industry which must essentially operate in an international sphere," drew up a broad definition of its own purposes as follows:

1. To promote internationally the interests of the shipping industry on all matters of general interest not already dealt with by other specialized organizations within the industry.
2. To coordinate expert advice within the industry in regard to technical questions affecting it and, in particular, with a view to securing improvements in standards of safety and the fullest advantage of technical and scientific progress. In this connection it will be necessary at the outset to consider the procedure which should be followed in the light of postwar circumstances and in relation to government activities in this field.
3. To provide a medium for the exchange of views and information on questions affecting the industry internationally.

To deal with matters of a more routine nature other international shipping organizations exist. Prominent among these is the Baltic and International Maritime Conference, which was started in 1905 as the Baltic and White Sea Conference, in Copenhagen, with the object of putting a stop to the cutthroat competition which then existed between tramp shipowners engaged in the Baltic and White sea trades. The conference was organized primarily to secure minimum freight rates and arrange uniform chartering terms among shipowners of different flags. By 1955 the Baltic and International Maritime Conference consisted of shipowner and ship broker members from 35 different countries representing some 17,000,000 tons gross of tramp shipping. The conference now embraces many kinds of trades in all parts of the world, and one of its principal functions is to inform members of changes in port expenses, in costs of loading and discharging cargoes, and in labour conditions in ports; but perhaps the most valuable part of its work consists of the drawing up of standard forms of charter party for use in a wide variety of tramp shipping trades, and of securing acceptance for them. Most of this work is carried out in cooperation with the Documentary Committee of the Chamber of Shipping of the United Kingdom.

VI. CLASSIFICATION AND TONNAGE MEASUREMENT

1. Classification.—To ensure compliance with the various international safety and load line and other maritime conventions, as well as with the requirements of national legislation concerning the registry and construction of merchant ships, most maritime nations require ships to be built under the supervision of government surveyors or of surveyors belonging to recognized classification societies and in accordance with their standards.

Lloyd's Register.—The leading classification society, operating in almost every country in the world, is Lloyd's Register of Shipping, which began its work long before any national legislation existed for the performance of its purposes. The history of Lloyd's Register of Shipping can be traced back to 1760 (*see* LLOYD'S REGISTER OF SHIPPING). The society was reconstituted in 1834 and again in 1914. The shipping community maintains it voluntarily and its principal work is the supervision of the survey and classification of merchant ships of all nationalities according to rigid standards. Through a worldwide organization of surveyors, initial classifications are made when ships are built, and maintenance surveys of hull, machinery, boilers, refrigerating plant, etc., are carried out periodically or whenever the ship has suffered damage from collision or from any other cause. Lloyd's Register surveyors test and approve, during its manufacture, the steel intended for use in the ship's structure or for its engines and boilers; they survey refrigerating machinery at sea or on land; they supervise the testing of chains, anchors, and pressure vessels, and they are competent authorities for the assignment of freeboard to all classes of ship in accordance with the provisions of merchant shipping acts or the load line regulations.

The society publishes annually a register book which contains in several volumes full details of all merchant ships in the world of more than 100 tons gross; this is kept up-to-date by means of fortnightly supplements. Separate registers are kept of British and American yachts. The society also publishes annual and quarterly statistical summaries of shipping registered or under construction in the world. In 1949 Lloyd's Register of Shipping took into amalgamation the British Corporation for the Survey and Registry of Shipping, a similar organization founded by shipowners in Glasgow in 1890 as a rival classification society.

Lloyd's Register of Shipping operates in most maritime countries, often in cooperation with classification societies established by other nations. These include the American Bureau of Shipping, originally established in 1867, and resuscitated as a result of the large volume of merchant ships built in the United States during World Wars I and II; the Bureau Veritas, which was founded

TABLE I.—*Tonnage of the Vessels of 100 Tons and Upward Prior to World War II, Belonging to Each of the Several Countries of the World*

Countries where owned	1886		1914		1920		1935		
	Steam and motor ships	Sailing vessels	Steam and motor ships	Sailing vessels	Steam and motor ships	Sailing vessels	Steam and motor ships	Sailing vessels	Total
	Gross tons (000)	Net tons (000)	Gross tons (000)	Net tons (000)	Gross tons (000)	Gross tons (000)	Gross tons (000)	Gross tons (000)	Gross tons (000)
Great Britain and Ireland	6,162	3,249	18,892	365	18,111	220	17,298	102	17,400
British Commonwealth	378	1,377	1,632	157	2,032	220	2,986	125	3,111
Total British Empire	6,540	4,626	20,524	522	20,143	440	20,284	227	20,511
United States including Great Lakes	496	1,587	4,330	1,038	14,574	1,475	12,223	629	12,852
Japan	78	32	1,708	...	2,996	...	4,086	...	4,086
Italy	195	705	1,430	238	2,118	124	2,838	46	2,884
France	738	319	1,922	397	2,963	282	2,989	36	3,025
Germany	604	806	5,135	325	419	253	3,693	10	3,703
Norway	140	1,352	1,957	547	1,980	240	3,967	1	3,968
Netherlands	190	229	1,472	25	1,773	20	2,554	5	2,559
Sweden	150	331	1,015	103	996	77	1,541	10	1,551
Spain	362	159	884	15	937	60	1,164	13	1,177
Denmark	143	128	770	50	719	84	1,099	2	1,101
Greece	54	289	821	16	497	33	1,711	...	1,711
Other countries	601	655	3,436	411	3,790	321	5,578	180	5,758
Total	10,291	11,217	45,404	3,686	53,905	3,409	63,727	1,159	64,886

Source: *Lloyd's Register of Shipping.*

in Antwerp in 1828 but moved its headquarters to Paris in 1832; the Norske Veritas, established in Norway in 1894; Germanischer Lloyd, founded in Germany in 1867; and Registro Italiano Navale, originally founded in Italy in 1861. Most of these and other national classification societies work in close conjunction with Lloyd's Register of Shipping.

2. Tonnage Measurement.—Classification societies are largely agreed as to the strength requirements of different types of ship, and the technical, legal, or commercial enactments of the conventions are almost universally accepted; however, complete international agreement on methods of tonnage measurement was slow to develop. The two chief reasons for this were, first the possibility of interpreting the term tonnage in several senses, and secondly the fact that a ship's tonnage is calculated by one of several methods according to the standard of measurement required. In speaking of shipping, the term ton may be used not only to name a unit of weight, in the usually accepted sense, but also to name a unit of volumetric capacity.

Displacement Tonnage and Deadweight Tonnage.—As far as weight is concerned, the ton is the long ton of 2,240 lb. avoirdupois, and this is used to measure the weight of the ship and its contents. According to the law of Archimedes, the weight of a floating vessel and its contents is equal to the weight of water that it displaces. This weight is known as the displacement tonnage, and the term is used most commonly in comparing the size of warships. In a cargo-carrying ship it is obvious that the displacement varies according to the amount of cargo, bunkers, and stores that are in the ship, and as it is important to determine the amount of cargo that can be carried, the ship's displacement must be calculated both when it is in light condition and when it is loaded, to find the difference which represents the weight of cargo that can be carried. The lightweight displacement tonnage is equivalent to the weight of water displaced by the ship's hull, machinery, and equipment, plus the weight of the crew and their effects, when no cargo or bunkers are carried. When the maximum permissible weight of bunkers and cargo is added, the ship has reached its full displacement tonnage. The difference between the full displacement tonnage and the lightweight displacement tonnage is called the deadweight tonnage, and this corresponds to the weight of cargo and bunkers the ship can carry. Deadweight tonnage is the measurement ordinarily used to describe and compare the sizes of dry-cargo ships and tankers.

Many procedures, for example the assessment of harbour dues and canal transit dues, require the use of a different system of tonnage measurement, based on the volumetric ton. The volumetric ton is a measure of the capacity of the enclosed space in a ship, one ton equaling 100 cu.ft. (2.8 cu.m.) of enclosed space. The volume of the ship up to the tonnage deck (which is generally the uppermost continuous deck) excluding certain exempted spaces such as the double-bottom tanks, the steering-gear compartments, the galley, and other spaces not used for the carriage of cargo, is described as the underdeck tonnage, and is expressed in terms of tons of 100 cu.ft. The internal volume of tween-decks and deck erections used for the carriage of cargo is added to the underdeck tonnage to give what is called the gross tonnage, again in terms of tons of 100 cu.ft. In merchant shipping statistics the gross registered tonnage is usually given. It is also given for passenger ships, whose deadweight tonnage is relatively unimportant as a means of comparing them in size. For the purpose of assessing dues payable for port, canal, pilotage, lighthouse, and other services, the freight-earning capacity of the ship is usually accepted as the criterion of measurement, and this is commonly arrived at by deducting from the gross tonnage the space devoted exclusively to such items as machinery, bunkers, crew accommodation, and navigating quarters in order to calculate the net registered tonnage. It is apparent that the final tonnage measurement of a ship according to the volumetric reckoning depends much on the definition of the spaces exempted from computation for various reasons.

The British system of tonnage measurement is the one most used by maritime nations, but in certain areas, notably the Suez and Panama canals, there are different systems of measurement for the assessment of tonnage on which dues are payable, and all ships have to be specially measured for the assessment of their dues when passing through these areas.

For the purpose of assessing freight rates on cargo, yet another form of volumetric tonnage is employed, known as the measurement ton, which is equivalent to 40 cu.ft. of capacity. This term has no connection with the tonnage measurements used for registration and for the assessment of dues. It derives from the fact that the charge made for carrying cargo is normally based either on the weight of the cargo or on the amount of space it occupies. Freight rates cannot all be fairly assessed by weight alone. It is obvious that a ton of feathers would occupy a far greater amount of the ship's hold than a ton of coal, for example. Likewise a heavy item of machinery, although occupying much less space than a ton of coal, would prevent the extra space being used for the carriage of more cargo, for such a deadweight cargo would bring a ship down to its load line limits with much of the cubic capacity of its holds still unfilled. Hence liner freight-rate tariffs are generally expressed in terms of so much per ton weight or measurement, the measurement ton being the equivalent of 40 cu.ft. of capacity, and the rate paid being the higher of the two.

VII. INTERNATIONAL PROBLEMS

Despite the large measure of international agreement and cooperation which has been achieved in the shipping industry, particularly in the 20th century, there is still much competition between merchant fleets. Those of one nation compete among themselves, and international rivalries are also strong. The situation is complicated by the fact that merchant shipping has an economic and strategic as well as a commercial value for most maritime nations. They depend on merchant fleets for the survival of their military power—troops and supplies must be carried by sea, and so must food and industrial raw materials. Many maritime nations too, particularly island nations such as Great Britain and Japan, depend for their economic existence on imports of food and of raw materials for industry, as well as on a flow of exports sufficient to pay for these. All this depends on merchant shipping. Moreover, Great Britain, Japan, Norway, and many

TABLE II.—*Number and Tonnage of Steam and Motor Vessels Registered by Country, 1939–1964*
(100 gross tons and over)

Country	1939 No.	1939 Tons gross	1949 No.	1949 Tons gross	1964 No.	1964 Tons gross
United Kingdom	6,722	17,891,134	6,077	18,093,159	4,538	21,489,948
Other British Commonwealth	2,255	3,110,791	2,585	3,956,581	2,757	6,054,790
Total	8,977	21,001,925	8,662	22,049,740	7,295	27,544,738
Denmark	705	1,174,944	689	1,170,373	901	2,431,020
France	1,231	2,933,933	1,236	3,070,398	1,532	5,116,232
Germany*	2,459	4,482,662	889	300,234	2,504	5,159,186
Greece	607	1,790,666	377	1,329,257	1,290	6,887,624
Italy	1,227	3,424,804	1,013	2,442,659	1,421	5,707,817
Japan	2,337	5,629,845	1,121	1,563,936	5,401	10,813,228
Liberia					1,117	14,549,645
Netherlands	1,523	2,969,578	1,492	2,990,195	1,889	5,110,022
Norway	1,987	4,883,813	2,069	4,916,396	2,732	14,477,112
Panama	159	717,525	535	3,016,227	691	4,269,462
Spain	777	902,251	1,146	1,192,508	1,741	2,047,715
Sweden	1,231	1,577,120	1,278	2,047,664	1,167	4,308,042
United States: Oceangoing	2,345	8,909,892	4,606	25,558,133	3,222	20,351,334
United States: Great Lakes	508	2,451,641	420	2,255,633	315	2,078,915
U.S.S.R.	699	1,305,959	962	2,118,206	1,674	6,957,512
Other Countries	2,991	5,252,874	3,744	6,549,356	5,967	15,190,017
World Total	29,763	69,404,432	30,239	82,570,915	40,859	152,999,621

*West Germany after 1949.
Source: *Lloyd's Register of Shipping.*

means of direct subsidies to liners on specified essential trade routes. In France, Italy, India and Argentina, the state participates directly in the ownership of the larger liner organizations. Ships engaged in international trade, whatever their flag, all have similar expenses. They pay much the same for bunkers, for repairs and maintenance, for cargo handling, and for port dues, except insofar as flag discrimination may be practised. Crews' wages however, do differ among ships of different flags. Shipbuilding costs tend to vary widely from one country to another, but in general shipowners may build where they please. The payment of direct shipbuilding or operating subsidies, therefore, is not as objectionable to unsubsidized shipowners as is the reservation of cargoes to ships of a particular national flag, whether this rule is incorporated in trade agreements or written into legislation. In the United States, for example, after World War II, the shipment of gift cargoes under the Mutual Security Act and the cheap sale of surplus farm products were allowed only on condition that at least half the ships used were of the U.S. flag.

other maritime nations depend to a great extent for their economic well-being on the ability to balance their international trading accounts with the invisible exports represented by the earnings of their shipping industry.

Politics and strategy interfere with the ordinary business activities of the shipping industry in several ways. For internal political reasons, some states decide that the provision of shipping services shall be a state monopoly, but in practice this is difficult to achieve since a state's jurisdiction does not extend beyond territorial waters. Various practices observed from time to time, however, are designed to protect the domestic shipping industry and to discriminate against shipping of other flags. These discriminatory measures may be used to protect the shipping of the national flag, even if the ships are not state owned. Higher port dues may be charged to foreign ships, for example, or national-flag ships may be favoured. In bilateral trade agreements it is sometimes stipulated that a fixed proportion of the cargoes must be carried in ships of the national flag. A common method of assisting the domestic shipping industry is to reserve coastal shipping to ships of the national flag, a policy abandoned by Great Britain with the repeal of the Navigation Acts in 1850. This decision greatly assisted the expansion of the British mercantile marine, but several other nations, notably the United States, adopted the opposite attitude and strictly reserved coastal trade.

Another means of assisting the national shipping industry and of protecting it against economic competition is the payment to it of direct or indirect subsidies from the national exchequer. These subsidies may take many forms, varying from tax exemptions or preferential credit terms to direct subsidy of shipbuilding or of operating costs. In the United States, where shipbuilding costs and manning costs are much higher than in most maritime countries, the Merchant Marine Act of 1936 and its successors attempted to promote the use of U.S.-flag merchant ships by

Although operating costs are much the same for ships of all flags, profits are subject to the rates of taxation in different countries. If all ships paid the same taxes, the margin of profit for each company would depend largely on its efficiency. After World War II, heavy taxes combined with a sharp rise in shipbuilding prices to cause the registration of more and more shipping companies in countries such as Panama and Liberia, where taxation was low. Greek and American shipowners resorted in large numbers to registry under these so-called flags of convenience, which have subsequently varied in number and size with the changing demands of international trade. In 1964 about 20,000,000 gross tons were so registered.

VIII. WORLD MERCHANT FLEETS

Statistics of world merchant shipping tonnage are little more than estimates until the later years of the 19th century, when steamships began to oust sailing ships from predominance in world trade. The statistics given in Table I illustrate the decline of the sailing ship between 1886 and 1935, and the increase in the total tonnage of merchant shipping in the world, which rose from 21,508,000 tons gross in 1886 to 64,886,000 tons gross in 1935. By 1939, before the outbreak of World War II, the world total had reached 69,404,000 tons gross (*see* Table II) and despite the enormous losses suffered during the war (Table III) the world fleet had reached 80,292,000 tons by 1948. The growth continued so that by 1955 the total tonnage exceeded 100,000,000 tons gross for

TABLE III.—*Losses of Merchant Ships, World War II, September 1939–July 31, 1946**
(1,000 gross tons and over)

Flag	1939 No.	1939 Gross	1940 No.	1940 Gross	1941 No.	1941 Gross	1942 No.	1942 Gross	1943 No.	1943 Gross	1944 No.	1944 Gross	1945 No.	1945 Gross	1946† No.	1946† Gross	Total No.	Total Gross
British	86	434,677	460	2,589,984	567	2,800,161	589	3,484,070	253	1,496,835	120	640,558	44	227,391	2	3,799	2,121	[illegible]
French	10	62,258	45	243,705	10	36,629	24	114,524	38	143,006	60	310,971	2	3,711		.	189	[illegible]
German	22	140,738	67	335,847	103	508,964	83	332,733	108	483,018	193	841,878	118	598,325		.	694	[illegible]
Greek	9	42,686	55	237,490	82	294,131	47	211,758	20	87,554	10	37,826	2	8,653		.	225	[illegible]
Italian	2	9,339	44	212,935	129	625,565	104	400,487	154	581,159	51	289,309	7	26,697	..	.	491	[illegible]
Japanese	1	11,930	2	4,274	10	67,226	175	877,278	301	1,419,054	492	2,303,421	194	578,480	.	.	1,175	[illegible]
Netherlands	6	38,845	43	206,367	43	246,979	112	519,909	27	157,089	15	90,534	6	27,282		.	252	[illegible]
Norwegian	16	55,681	100	360,334	89	334,072	152	741,981	67	345,169	36	120,439	24	73,757	2	3,645	486	[illegible]
Panamanian			11	54,255	12	39,104	71	397,987	11	70,558	9	44,389	1	1,349			115	[illegible]
Swedish	18	33,913	52	149,277	25	79,520	36	139,273	20	66,769	10	15,812	1	1,226		.	162	[illegible]
United States	.	.	1	5,883	15	82,245	314	1,937,855	139	955,221	93	652,486	40	279,964	3	18,972	605	[illegible]
Other	22	68,572	85	292,572	75	234,759	125	480,029	52	195,750	26	78,334	8	23,193	2	3,011	395	[illegible]
Total	192	898,639	965	4,692,923	1,160	5,349,355	1,832	9,637,884	1,190	6,001,182	1,115	5,425,957	447	1,850,028	9	29,427	6,910	33,885,395

*From war causes alone. Only those vessels are included whose identity was definitely established. Vessels are charged to the flag under which they operated, regardless of actual ownership. Data, especially for axis losses, are not necessarily complete. Includes vessels on charter to the military services. †Losses caused by striking mines.
Source: U.S. Maritime Commission.

the first time. By 1964 the world merchant fleet totaled 153,000,000 tons gross, one-third of which consisted of tankers.

A significant feature of the development of the world merchant fleet after 1920 was the growth of the number of ships propelled by diesel engines. Table IV illustrates the increase in the proportion of motor ships. A simultaneous development was the replacement of coal by oil as fuel for steamships, a movement which reached its peak just after World War II. By 1964 coal-burning ships still in existence accounted for only 2.6% of the total tonnage, and the number of coal-burning ships being built was negligible.

TABLE IV.—*Fuel Analysis of World Tonnage*
(Tons Gross)

Year	Steamships		Motor ships	Total
	Coal-burning	Oil-burning		
1914 . .	43,859,381	1,310,209	234,287	45,403,877
1929 . .	40,358,396	19,420,895	6,628,102	66,407,393
1939 . .	31,015,069	20,575,676	16,918,687	68,509,432
1949 . .	17,413,643	45,805,553	19,351,719	82,570,915
1954 . .	11,131,288	54,358,576	31,931,662	97,421,526
1964 . .	4,005,292	70,983,467	78,010,862	152,999,621

Source: *Lloyd's Register of Shipping.*

TABLE V.—*World Tonnage of Oil Tankers*
(Tons gross)

Year	Tankers	All ships	% of tankers
1929 . . .	7,071,015	66,407,393	10.6
1939 . . .	11,585,549	68,509,956	16.9
1949 . . .	16,101,720	82,570,915	19.5
1954 . . .	24,624,079	97,421,526	25.3
1964 . . .	50,563,315	152,999,621	33.0

Source: *Lloyd's Register of Shipping.*

After World War II a vast increase in world tonnage took place. This was partly due to the existence of many ships in the United States defense emergency reserve fleet, but chiefly to the rapid increase in the demand for oil tankers (*see* Table V). The United States became a net importer of oil, while the United Kingdom and the continent of Europe also required sea transport for oil. The mid-1950s saw a rapid expansion of world industry, especially of the steel industry in the Federal Republic of Germany, in the United Kingdom, and in Japan. The United States, though its domestic supplies of iron ore were inadequate, likewise expanded its steel industry. These circumstances caused an even greater demand for merchant shipping.

Although the total gross tonnage of the world merchant fleets constantly increased, the actual number of ships remained more or less steady (*see* Table II) because the average new merchant ship was larger, especially after World War II. The average speed of ships also increased, although to what precise extent it is impossible to measure. Large ships tend to have higher speeds than small ones, however, and there can be no doubt that the ton-mile capacity of the world fleet after the end of World War II increased proportionately even more than the total gross tonnage.

1. **British Commonwealth.**—As a result of British war losses and of the U.S. emergency shipbuilding programs of World Wars I and II, the mechanically propelled tonnage registered in the United States on two occasions exceeded that registered under the British flag. Great Britain nevertheless retained the largest active merchant fleet in the world, although its proportion of the world total declined in the 20th century. The British Commonwealth owned 63.6% of the mechanically propelled merchant shipping in the world in 1886, 45.2% in 1914, 37.4% in 1920, and 34.6% in 1927. By 1935 the proportion was 31.8%, and in 1939 it was 30.7%. The tonnage in 1939 was 21,001,925 tons gross, and despite heavy war losses amounting to nearly 12,000,000 tons gross, the 1939 tonnage total was exceeded by 1948. The Commonwealth's proportion of the world total, however, including the U.S. inactive fleet, had been reduced to 27%. By 1955 British tonnage had increased to 23,230,000 tons gross, but its percentage had dropped to 23.2%. In 1964, although total tonnage amounted to 27,544,738 tons gross, the largest commercially active fleet in the world, the percentage of total world tonnage had fallen further.

Most of the British-registered tonnage is owned and operated by the United Kingdom. In 1939, 6,722 ships of 17,891,134 tons gross were registered in the United Kingdom and 2,255 ships of 3,110,791 tons gross in the British Commonwealth outside the United Kingdom, mainly in Australia, Canada, Hong Kong, India, Malaya, New Zealand, South Africa, and the West Indies. Canada acquired a substantial fleet of vessels built during World War I, but after a few years most of the oceangoing ships were re-registered in the United Kingdom, although the beneficial ownership remained in Canada. The general movement toward flags of convenience, referred to above, however, started a tendency from 1956 onward for United Kingdom shipowners to register new tonnage in such commonwealth countries as Bermuda and the Bahamas, so that profits used for the purpose of replacing obsolescent tonnage would not be subject to United Kingdom taxation.

The privately owned United Kingdom merchant fleet at Jan. 1, 1964, excluding fishing vessels, tugs, and other nontrading vessels, as well as ships of less than 500 tons gross in size, exceeded 20,000,000 tons gross. Oil tankers and whaling factory ships accounted for nearly 8,000,000 tons gross of the total gross tonnage. Passenger and cargo liners totaled more than 8,000,000 tons gross and tramps about 3,000,000 tons gross. The remainder consisted of ships engaged in the coastal and short-sea trades. The United Kingdom continued to maintain the largest fleet of purely passenger ships in the world. They are engaged principally in the following trades: (1) Great Britain, France, and Ireland to Canada and the United States; (2) Great Britain to South Africa, Australia, New Zealand, India, and the Far East; (3) Great Britain to Spain, Portugal, and Central and South America; (4) cross voyages, for example Australia to the Far East and New York to Bermuda; (5) cross-channel services between Great Britain and Ireland and the continent of Europe; and (6) pleasure cruising in off-seasons from Great Britain, Australia, the United States, and elsewhere. Specially designed refrigerated cargo liners, often carrying up to 100 passengers, sail between Great Britain and Australia, New Zealand, South America, and South Africa, carrying homeward cargoes consisting mainly of meat and dairy produce. Cargo liners, many of them able to accommodate not more than 12 passengers, operate regular services from Great Britain and the continent of Europe to all parts of the world. They also sail between overseas countries, for example from the United States to India, from Canada to West and South Africa, from the United States to the Far East; and they may go around the world without touching Great Britain. British shipping services make a substantial contribution to the nation's balance of international payments through earnings of foreign currency. An inquiry carried out into the earnings of the British shipping industry for the year 1962 showed that receipts of foreign currency for freight, passenger fares, and charter hire totaled £853,000,000, while disbursements abroad totaled £490,000,000, making a net contribution of £363,000,000 to the nation's balance of payments.

Many of the well-known British shipping companies were founded in the earliest days of steam navigation, and some of them have even longer continuous histories. One of the largest shipping groups in the world is centred on the Peninsular and Oriental Steam Navigation Company, which operates passenger and cargo-liner services between Great Britain and the continent of Europe to India, Pakistan, Ceylon, Malaya, Australia, and the Far East. The group also includes the Orient Line passenger services to Australia, New Zealand, and the Pacific coast of North America and the refrigerated cargo-liner services to Australia and New Zealand operated by ships of the New Zealand Shipping Company and the Federal Steam Navigation Company. It includes the Nourse Line, trading between the West Indies and India; the British India Steam Navigation Company, trading between Britain and India and East Africa, as well as to Australia and in the Indian Ocean; the Hain Steamship Company, engaged in worldwide dry-cargo tramping; the General Steam Navigation Company, with a network of liner services ranging from the Mediterranean to the Baltic Sea; and the Union Steamship Company of New Zealand. In 1956 the Peninsular and Oriental (P. and O.) group broke new ground by ordering a fleet of 25 tankers.

The Cunard Line is famous for its transatlantic liners, but it also operates cargo services to the Mediterranean. The Port

Line, affiliated to the Cunard Line, sends refrigerated liners to Australia and New Zealand. The Brocklebank Line, another Cunard affiliate, runs to the Indian Ocean from Great Britain and the United States. The Union-Castle Line, which operates mail, passenger, and cargo services to South Africa, merged in 1956 with the Clan Line, a company running a worldwide network of cargo-liner services, to form the British Commonwealth Shipping Company. Royal Mail Lines operate passenger and refrigerated cargo-liner services to South America, and cargo-liner services to the Caribbean and the Pacific coast of North America, while their subsidiary, the Pacific Steam Navigation Company, operates passenger and cargo services to the west coast of South America. Another large organization is the Furness Withy group, operating passenger and cargo-liner services in many parts of the world. Most of the companies referred to above are public companies, but there are also some private companies or partnerships in which there is no public participation by means of shareholdings. Some of these, for example Alfred Holt and Company (the Blue Funnel Line) and Ellerman Lines, operate substantial fleets. The P. and O. group and several other shipping companies have substantial interests in independent air transport undertakings. Tramp shipping companies tend to be larger in number and smaller in size. Large tanker fleets are operated by the British Petroleum and Shell Oil companies, but there is also a large number of independently owned and operated tramp tanker companies.

2. Denmark.—The Danish merchant fleet, which totaled 705 vessels of 1,174,944 tons gross in 1939, suffered heavy losses during World War II, but by 1950 the strength of the fleet had more than recovered, and it reached 718 vessels of 1,269,011 tons gross. By 1964 the number of vessels had increased to 901 and the gross tonnage had advanced to 2,431,020 tons gross, of which 36% consisted of tankers. Most Danish ships are propelled by diesel engines, the Danish firm of Burmeister and Wain having built the motor ship "Selandia," the first oceangoing motor ship in the world. There is a substantial ferry traffic between Denmark and its neighbouring countries, but only about one-third of the foreign trade passing through Danish ports is carried in Danish ships. Danish shipowners, however, find much employment in overseas trading and in 1963 carried more than 9,800,000 tons of cargo between foreign ports, time charter not included. In 1963 gross earnings of Danish shipping in foreign currency totaled 1,211,000,000 Kr., excluding 736,000,000 Kr. in time-charter earnings. All except 231,000,000 Kr. was earned in purely foreign trade. By far the largest shipowning company in Denmark is the A. P. Møller concern, which operates around-the-world services as well as tramps and tankers, with a fleet totaling, in 1964, 963,000 tons gross. Next largest is the East Asiatic Company, running regular liner services to India, Australia, the Far East, and North America, with a fleet of nearly 326,000 tons gross. The United Steamship Company operates cargo liners to all parts of Europe, as well as to North and South America, while the J. Lauritzen concern operates refrigerated cargo-liner services to North and South America and to the Mediterranean, and also provides ships specially built for service in polar waters.

3. France.—The principal activity of the French mercantile marine is the carrying of trade from metropolitan France to other parts of the French Union, notably to those in North Africa, West Africa, and the West Indies, and to former French-controlled territories in southeast Asia. A worldwide network of liner services, however, is operated by the state-controlled Compagnie Générale Transatlantique (French Line) and other lines such as Messageries Maritimes. A considerable amount of coasting trade is carried in French vessels, and traffic between French territories in Africa and metropolitan France is largely reserved to vessels of French registry. The French liner fleet includes a substantial number of large passenger liners, like those of the French Line which operate in the North Atlantic trade, and a high proportion of specially designed banana carriers and fruit ships. The passenger liner fleet, however, experienced a marked reduction as former overseas possessions achieved political independence. France runs proportionately fewer dry-cargo tramp ships than do many other European nations.

In 1939 the French merchant fleet numbered 1,231 vessels totaling 2,933,033 tons gross. About half of this fleet was lost during World War II, but by 1949 its previous strength was restored, partly in consequence of a prompt rebuilding program and partly through the purchase of surplus U.S. ships and the acquisition of vessels as war reparations from defeated nations. Regular planned and subsidized shipbuilding programs enlarged the French-flag fleet until by 1964 it had reached the total of 5,116,232 tons gross, of which 43% consisted of tankers. Much of the postwar expansion was due to the building up of a tanker fleet much larger than that considered necessary before the war.

4. Germany.—German merchant shipping reached its peak in 1914 after a decade of highly successful competition on the high seas. In that year the German merchant fleet totaled nearly 5,500,000 tons gross, and the flags of such liner companies as Norddeutscher Lloyd, the Hamburg-America Line, and the Hamburg South America Line, were well known throughout the world. The defeat sustained in World War I, however, caused the disruption of this fleet. By 1920 less than 700,000 tons gross remained under the German flag. But by persistent effort, the merchant fleet was rebuilt, and although it had not quite reached 4,500,000 tons gross by 1939, its quality and efficiency were above average. A worldwide network of liner services had been reestablished, as well as a tramp fleet which flourished particularly in the Baltic and short-sea trades.

World War II resulted in a second eclipse of the German merchant fleet, and postwar restrictions on the size and speed of the ships that the Allies permitted to be built deferred the rebuilding of the fleet for several years. By 1949 the West German fleet had been reduced to 300,000 tons gross. A rapid expansion began in 1950, however, the bulk of the building program being confined to fast modern cargo liners of moderate size. By 1964 the fleet of the Federal Republic of Germany had grown to 2,504 vessels of 5,159,186 tons gross, nearly all of them constructed after 1950. The nucleus of a Federal Republic tanker fleet emerged, and plans were made for the resumption of the passenger trade.

5. Greece.—The Greek-flag merchant fleet expanded steadily between World Wars I and II, from 496,996 tons gross in 1920 to 1,790,666 tons gross in 1939, but this was due not so much to new construction as to the purchase of second-hand ships from other flags. As a result the average age of the ships of the Greek fleet was high, and in 1939 some three-quarters of the total tonnage was more than 20 years of age. For the most part this fleet was engaged in international dry-cargo tramping. Severe losses were suffered during World War II, but after the war large numbers of war-built Liberty ships were purchased from the United States and by 1950 the Greek-flag fleet totaled 1,348,874 tons gross. Thereafter a decline set in, for many Greek owners preferred to register their ships under foreign flags, especially under flags of convenience such as those of Panama and Liberia. Indeed, even before World War II, many Greek shipowners felt their gift for shipping commerce restricted by successive Greek legislatures, and transferred their operations to other flags. By the end of the 1950s, however, more favourable Greek maritime legislation brought about a large-scale repatriation of Greek-owned tonnage from flags of convenience to the Greek national flag. By 1964 the Greek flag fleet totaled 1,290 vessels of 6,887,624 tons gross. Owners of Greek nationality achieved a predominant position in world dry-cargo and tanker tramping in the years following World War II. They operated mainly from London and New York.

6. Italy.—Between World Wars I and II the principal Italian liner companies were liquidated and reconstituted as a state-owned organization known as the Finmare group, comprising four major lines, the Italia Line, the Lloyd Triestino Line, the Adriatica Line and the Tirrenia Line. Capital was provided by the state, which in 1937 also assumed control of the major shipbuilding yards, and in 1939 the Italian merchant fleet consisted of 1,227 vessels, totaling 3,424,804 tons gross. Only about 12% of this tonnage remained at the end of World War II, but by purchase, salvage and new construction the prewar total was passed by 1953. The state-owned liner group continued in existence and was responsible for the building of several large passenger liners, fast cargo liners, and specialized ships. The privately owned section of the

industry, engaged in liner, tramping, and tanker operations, also succeeded in rebuilding and expanding, and by 1964 the Italian-flag fleet had reached the record total of 5,707,817 tons gross, of which about one-third were tankers.

7. Japan.—The maritime expansion of Japan was extremely rapid between World Wars I and II, tonnage having almost doubled between 1920 and 1939, when the peak figure of 5,629,845 tons gross was reached. The geographical situation of Japan is favourable to the natural development of a large shipping industry, which was fostered between the wars by comparatively low wage rates and a measure of state support. A great deal of the tramp and liner trades between Japan, China, and other Pacific countries fell to Japan during this period, and in addition such large liner groups as Nippon Yusen Kaisha and Osaka Shosen Kaisha built up a worldwide network of passenger and cargo-liner services. Defeat in World War II left Japan with a largely obsolescent fleet totaling little more than 1,000,000 tons gross, but by 1950 postwar restrictions had been for the most part removed and the Japanese fleet again began to increase, largely as a result of government-sponsored shipbuilding programs. By 1964 the Japanese merchant fleet had reached a total of 10,813,228 tons gross.

8. Liberia.—The rise of a merchant fleet registered under the flag of Liberia is entirely a development of the decade following World War II. Whereas in 1939 the tonnage registered in Liberia was too negligible to be noted separately, by 1964 it totaled 14,549,645 tons gross, and the Liberian fleet was the third largest merchant fleet in the world. This phenomenal increase was due entirely to the postwar tendency of shipowners of other nationalities, especially Greeks and Americans, to register their new ships under flags of convenience, thereby evading onerous taxation burdens or other restrictive measures imposed by their national legislatures. Most of the ships registered in Liberia are vessels built after World War II, and in 1964, 8,619,449 tons consisted of tankers, almost all the remainder being dry-cargo ships engaged in international tramping.

9. The Netherlands.—There are three principal branches of the Netherlands shipping industry, which consists mainly of oceangoing passenger and cargo liners, tankers owned by the major oil companies, particularly the Royal Dutch Shell group, and coasting and short-sea vessels engaged mainly in trade with the United Kingdom and neighbouring and Baltic Sea countries. The granting of independence to Indonesia resulted in the Dutch liner companies losing much of their traditional network of services, but they established themselves in other trades, notably to the Great Lakes in North America, to Australia, and to the Far East. The Netherlands merchant fleet in 1939 totaled nearly 3,000,000 tons gross, a figure restored, despite war losses, by 1949. By 1964 the fleet had grown to 5,110,022 tons gross, the tanker fleet growing especially.

10. Norway.—The Norwegian merchant fleet expanded to a marked extent during the 20th century, owing to the traditional skill and seafaring bent of the Norwegians and to the dependence of Norway's economy on its income from international shipping services. Although the overseas trade of Norway itself is not sufficient to maintain a larger merchant fleet, the latter grew to comparatively large proportions mainly through the enterprise of Norwegian shipowners in the worldwide tramping trades. From mainly tramp origins, a number of liner groups developed, operating worldwide services, but the expansion of the fleet after World War II was largely due to the enterprise of tramp tanker owners. The Norwegian merchant fleet grew from 2,219,388 tons gross in 1920 to 4,883,813 tons in 1939, but about 40% of this tonnage was lost during World War II in the service of the Allied powers. Substantial purchases of U.S. war-built tonnage, together with an intensive postwar reconstruction program, helped to restore the prewar strength by 1949; but the potentialities of the international tanker market were such that Norwegian shipowners made even greater efforts to build. As a result, by 1964 the fleet had reached the figure of 14,477,112 tons gross, of which more than half consisted of tankers.

11. Panama.—Statistics of ships registered under the flag of Panama were not kept separately by Lloyd's Register of Shipping until 1924, when 15 vessels totaling 85,593 tons gross were listed. From 1931 onward there was a gradual increase in the merchant shipping registered in Panama, although the beneficial ownership was vested almost entirely in the United States, until the total of 717,525 tons was reached in 1939. After World War II the practice grew for U.S. oil companies and Greek and other independent tramp shipowners to register their ships under the flag of Panama, and by 1949 the merchant fleet registered in Panama exceeded 3,000,000 tons gross. By 1954 the figure of 4,000,000 tons gross had been exceeded, and this increased to 4,269,462 tons by 1964, when 2,253,418 tons gross consisted of tankers.

12. Sweden.—The Swedish merchant fleet reached 1,000,000 tons gross in 1913 and gradually increased its tonnage until by 1939 it reached 1,577,120 tons gross. This expansion continued after World War II and by 1949 the total exceeded 2,000,000 tons gross. In 1964 the total reached 4,308,042 tons gross, of which about 34% consisted of tankers. In addition to a substantial coasting and short-sea traffic, and a fair-sized oceangoing tramp fleet, Sweden has a number of well-established liner groups, such as the Brostrom concern, which includes the passenger liners of the Swedish American Line, the Transrederierna group, operating worldwide cargo-liner services, and the Johnson Line, which specializes in high-speed cargo-liner services. Iron ore is one of Sweden's major exports, and the Grangesberg-Oxelosund concern built up a large fleet of ore carriers to cope with this trade, many of the larger units being able to carry iron ore exports on the outward voyage and cargoes of imported oil products on the return voyage.

13. United States.—In 1913 the United States oceangoing merchant fleet, including the coastwise fleet, totaled nearly 3,000,000 tons gross. As a result of an emergency shipbuilding program toward the end of World War I, the U.S. oceangoing fleet reached a total of 14,738,506 tons gross by 1922, but a gradual decline then set in until by 1939 the total oceangoing fleet amounted to 8,909,892 tons gross, of which only about two-thirds was active. High wage and shipbuilding costs made it impossible for the U.S. privately owned merchant fleet to compete effectively with foreign flags. In an endeavour to stem the decline of the U.S. foreign-going fleet, the Merchant Marine Act of 1936 introduced a system of differential subsidies for the building and operating of liner services on specified "essential trade routes." These helped to sustain the cargo-liner companies, although they failed to provide any incentive for private investment in shipping. They did not apply to the tramp companies, which almost disappeared, or to the tanker companies, which found the obligations of the shipbuilding subsidy provisions irksome and sought refuge in registration in Panama and other countries abroad, or to the intercoastal and coastwise shipping companies which, despite the protection afforded them by the reservation of their trades to U.S. flag ships, were unable to cope with the competition of inland transport systems.

World War II saw another intensive shipbuilding effort in the United States, which agreed with Great Britain to build chiefly cargo ships of emergency type, while Britain for the most part built warships. As a result the United States merchant fleet in 1946 totaled 36,000,000 tons gross. This was more than half the world total tonnage at the end of World War II, because ships of other flags were often lost, and few losses were replaced during the war years. Disposal of the U.S. surplus tonnage to the other Allied powers began in 1946 and many ships were sold at favourable prices. By 1950 the United States-flag merchant fleet had been reduced to 25,000,000 tons gross, but of these 14,000,000 tons gross had been permanently laid up in the emergency defense reserve fleet. The Merchant Marine Act of 1936 continued in force and various other legislative measures were taken in an attempt to maintain the volume of U.S.-flag shipping in commercial service. At least half the ships carrying government-sponsored cargoes had to be of the U.S. flag. United States-controlled tanker tonnage under foreign flags continued to increase, and the subsidized passenger and cargo liners maintained their tonnage, but the decline of coastwise and intercoastal shipping continued.

By 1964 the total United States seagoing merchant fleet, including inactive ships but excluding Great Lakes vessels, comprised

3,222 ships totaling 20,351,334 tons gross. Of the oceangoing fleet 390 vessels, totaling 4,428,272 tons gross, were tankers.

United States shipping on the Great Lakes continued busy. Tonnage reached a peak of 2,611,040 tons in 1934. By 1956 the total was still just over 2,500,000 tons gross, but prospects began to decline with the decrease of iron ore deposits in the area. The completion of the St. Lawrence Seaway in 1959 opened up the Great Lakes to intense competition from foreign-flag vessels engaged in oceangoing trade, and by 1964 the U.S. registration on the Great Lakes had fallen to 2,078,915 gross tons.

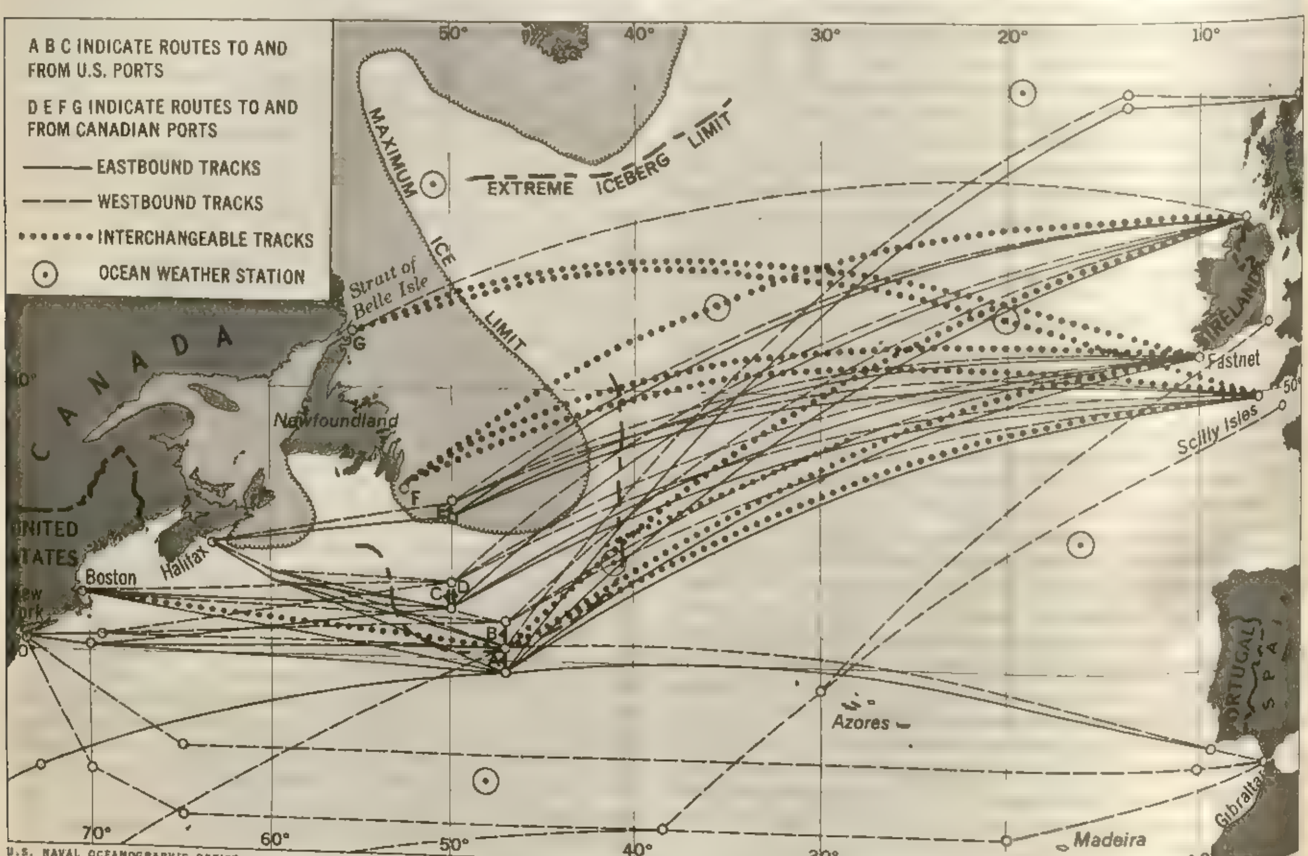
BIBLIOGRAPHY.—W. S. Lindsay, *History of Merchant Shipping and Ancient Commerce* (1874–76); *Lloyd's Register of Shipping* (issued annually and periodically within the year); the American Bureau of Shipping also issues periodical statistics; *Report of the Royal Commission on Shipping Rings* (1909); Chamber of Shipping of the United Kingdom *Annual Reports* (1920 *et seq.*); D. Marx, Jr., *International Shipping Cartels* (1953); H. Osborne Mance and J. E. Wheeler, *International Sea Transport* (1946); C. E. McDowell and H. M. Gibbs, *Ocean Transportation* (1954); J. Bes, *Chartering and Shipping Terms*, 3rd ed. (1954); C. E. Golding and D. King-Page, *Lloyd's* (1952); R. H. Thornton, *British Shipping* (1939); Peter Duff, *British Ships and Shipping* (1949); Kaare Petersen, *The Saga of Norwegian Shipping* (1956). (P. DF.)

SHIPPING ROUTES. When man first began to venture on the sea, his route was determined by the landmarks along the shore. As the science of navigation developed and he could determine his latitude by the heavenly bodies, he was no longer bound to the coast. It was not, however, until the perfection of the chronometer by John Harrison in 1762 that a shipmaster could accurately plot both his latitude and longitude. Then with exact positions he became more aware of the effects of winds and currents on his ship and began shaping his course to take advantage of them. The shortest route was not always the quickest one: for example, the westbound clipper ships from Europe to America made much faster passages by following a southerly route to avoid the influence of the Gulf stream and the prevailing westerly winds.

No systematic effort was made to study regular ship routes until the decade preceding the American Civil War when Lieut. Mathew Fontaine Maury, then superintendent of the depot of charts and instruments, the forerunner of the present naval oceanographic office, enlisted the aid of shipmasters generally to furnish him with copies of their log books. From these he produced his *Pilot Charts* with recommended routes that by estimate saved the merchants and shipmasters of the United States more than $2,000,000 annually, and those of Great Britain about $10,000,000 each year. His contributions to the maritime world earned him the title of "Pathfinder of the Seas."

The development of steam navigation ended the shipmaster's dependence on wind as the sole means of propulsion and led to the gradual adoption of the ship lanes of today. The shortest distance between two ports is a great circle on the earth's surface and the modern high-powered steamer follows such routes whenever practicable. However, deviations are made to avoid intervening land masses, ice-infested waters and regions of continuous adverse meteorological conditions. Low-powered steamers travel routes taking advantage of favourable winds and currents. The leading hydrographic offices of the world have made studies of the various routes and publish in their volumes of sailing directions advice on the routes to follow. While the high speed of planes has led to international agreements on air lanes and a rigid control on plane flights, there are at present no mandatory international rules governing the ocean lanes, and the selection of the routes to be followed is left to the discretion of the steamship owners and shipmasters.

North Atlantic.—The combination of frequent fog, high density of travel and the annual incursion of ice has led to the establishment of definite ship lanes that are recognized internationally in the North Atlantic between the United States and Europe; these are traveled by the largest and fastest liners. The danger of collision in the North Atlantic was first recognized by Maury, following the sinking of the U.S. mail steamer "Arctic" in Oct. 1854 by the French steamer "Vesta" in a thick fog about 50 mi. (80 km.) east of Cape Race, Nfd., which resulted in the loss of about 300 lives. This disaster inspired him to include in his *Sailing Directions* (1855) a section "Steamer Lanes Across the Atlantic," wherein were graphically depicted and recommended for the first



NORTH ATLANTIC SHIPPING ROUTES

time separate lanes for steamers eastbound and westbound that would not only lessen the danger of collision between ships but would also provide nearby help in case of disaster. The lane to Europe crossed the 50th meridian of west longitude in latitude 42° and was from 15 to 20 mi. (24 to 32 km.) wide; the lane from Europe crossed the 50th meridian of west longitude 200 mi. (322 km.) to the northward and was from 20 to 25 mi. (32 to 40 km.) wide, the latter being made wider on account of the greater percentage of fog, the greatest width in both lanes being given where most fog was to be expected.

The pioneer work of Maury was continued by the U.S. navy hydrographic office which emphasized the necessity of lanes across the Atlantic, not only to obviate collisions, but also to avoid the danger from ice. In 1875 the Cunard Steamship company ordered its shipmasters to follow ship lanes that were laid south of the zone into which icebergs normally drifted. Several of the other large and more progressive passenger lines followed this policy, and as a result the number of accidents due to ice showed an encouraging decline. In the meantime, through the medium of the monthly *Pilot Chart of the North Atlantic Ocean*, the hydrographic office pointed out to the maritime world in general the need to adopt steamer lanes.

The International Marine conference, held at Washington, D.C., in 1889, and attended by delegates from 26 maritime countries, decided: "Steamer lanes for trans-Atlantic navigation are not adopted, although the various steamship companies are urged to adopt regular routes for vessels of their own line." At the urging of the hydrographic office, a conference was held in 1891 between representatives of five principal trans-Atlantic steamship companies, the Cunard, White Star, Inman, National and Guion lines, to consider the lanes as recommended on the *Pilot Chart*. These routes were formally adopted to be followed by all ships of the lines concerned and in 1898 the North Atlantic Track agreement came into being with the adherence of all the major passenger lines of that day. The agreement, although purely voluntary in nature, has remained in effect into the second half of the 20th century with seasonal shifts of the lanes being made as recommended by the commandant of the U.S. coast guard and the U.S. hydrographer. This adoption of recognized steamer lanes contributed greatly to the safety of navigation in the North Atlantic and was especially gratifying to the hydrographic office, which had so long advocated the lanes as shown on the *Pilot Charts*.

As a result of the sinking of the British steamer "Titanic" on April 15, 1912, with a loss of over 1,500 lives, the first International Convention for Safety of Life at Sea was convened at London in 1913. The steamer lanes were considered and the following resolution adopted: "The selection of the routes across the North Atlantic in both directions is left to the responsibility of the steamship companies, nevertheless the High Contracting Parties undertake to impose on these companies the obligation to give public notice of the regular routes which they propose their vessels should follow, and of any changes which they make in them. The High Contracting Parties undertake, further, to use their influence to induce owners of all vessels crossing the Atlantic to follow, as far as possible, the routes adopted by the principal companies." Similar resolutions were adopted at the subsequent conventions of 1929, 1948 and 1960.

With but minor changes, the lanes of the North Atlantic Track agreement, used by the principal steamship companies prior to the London conference, remained unchanged until 1924 when the tracks A, B, C, D, E, F and G were adopted. These, with slight modifications, are the tracks in use today and are shown on the accompanying chart. The routes are seasonal and provide, as far as possible, for safety from the danger of ice, fog and collision with fishing vessels on the Grand Banks. Tracks A, B and C are the United States-European lanes, while D, E, F and G relate to vessels bound to and from Canada. The seasonal United States track employed is the one that is ice-free.

An additional safeguard to the North Atlantic lanes resulting from the conference of 1913 was the establishment of the International ice patrol, which, during the ice season, guards the region in the vicinity of the Grand Banks of Newfoundland, warns ships of dangerous ice, recommends safe tracks and makes studies of ice conditions. The cost of the patrol is borne by those nations whose vessels traverse the area. The management of the ice patrol is entrusted to the United States government and is carried out by the U.S. coast guard.

Originally conducted with surface craft, the ice patrol today is maintained principally by reconnaissance aircraft based at the headquarters of the commander of the international ice patrol in Argentia, Nfd. Only in serious ice years is it necessary to dispatch a cutter to stand by ice known to be drifting into the steamer lanes. Since the inauguration of the International ice patrol services, not a single life has been lost or a vessel sunk on the United States-European lanes as the result of collision with icebergs. *See* also COMMERCE, HISTORY OF. (W. G. WT.)

SHIP'S BELLS, in use as early as the 15th century, make possible the distinctive method of sounding the time on board ship. The mariner's 24-hour day is divided into six watches, each four hours long, except that the 1600 to 2000 watch (4 P.M. to 8 P.M.) may be "dogged" or shortened into the first and second dogwatches, each two hours long, in order to rotate the watches and to allow the men on duty to have their evening meal. Through the 18th century, time was ordinarily measured on board ship by using a sandglass which approximated 30 min. The quartermaster or ship's boy turned the glass when the sand ran through, and it became customary for him to strike the bell as he did this. Eight times in each watch the glass was turned and the number of strokes on the bell indicated the number of half hours elapsed after the men came on deck. These strokes are sounded in pairs, with an interval following each pair.

The time and place of origin of this seagoing custom is unknown but it was nearly universal among Europeans and sailors of the Mediterranean area by the 18th century. After the mutiny at the Nore (1797), British ships followed a special numbering in the dogwatch. From 1600 to 1800 the usual bells are struck, but at 1830 (6:30 P.M.) only one bell is struck instead of five; two at 1900 (7 P.M.); three at 1930 (7:30 P.M.); and eight bells at 2000 (8 P.M.). Thus, the signal for the mutiny, five bells in the second dogwatch, has never been given.

International law required that a ship at anchor sound a series of rapid successive strokes on the bell as a fog signal, while at other times this is a fire signal. (J. B. HN.)

SHIPS' FIGUREHEADS: *see* FIGUREHEADS, SHIPS'.

SHIRAKAWA (1053–1129), Japanese emperor who succeeded his father Gosanjō II on the throne in 1072. The times were troubled. The encroachment of private estates (*shōen*) on the public domain seriously undermined the economic foundations of the imperial government. The warrior monks of the nearby temples threatened the capital city of Kyōto, and the weakening of the Fujiwara family, which had dominated the emperors for over two centuries, made for bitter factionalism within the court. Shirakawa left the throne in 1086 and as retired emperor (*jōkō*) succeeded in retaining power in opposition to the Fujiwara regent (*kampaku*). Drawing around him a group of courtiers, many of whom were from the Minamoto and other non-Fujiwara families, he established an administrative centre replete with judicial functions and a military guard. This was the cloister government (*in-no-chō*) through which former emperors exercised power until 1183. Shirakawa himself held the reins of government until 1129. Although at first Shirakawa sought to reduce private estates, he soon gave up the effort and became instrumental in converting large tracts of public domain into imperial *shōen*. With these sources of wealth he lavishly patronized Buddhism. He failed to strengthen the imperial government and prevent the rise of the provincial warrior gentry. *See* also JAPAN: *History*. (J. W. H.)

SHIRAZ, a city of Iran and capital of Fars (ancient Persis) *ostan* (province), lies at an altitude of 5,200 ft. (1,585 m.), 115 mi. (185 km.) ENE of Bushire, its port on the Persian Gulf. Pop. (1964) 215,000 city; 325,919 environs. Famous for its poets and its wine, it is both a historic site and an attractive modern town (the first on the plateau to have piped water), with wide, tree-lined boulevards, gardens, shrines, mosques, and other monuments. It is the birthplace of the two Persian poets, Sa'di and

Hafiz (*qq.v.*), whose garden tombs, both resplendently renovated, lie on the outskirts.

Inscribed clay tablets discovered in 1935 reveal that to build his palace at Persepolis Xerxes I, in 466 B.C., employed workers from Shiraz, which retained its importance during the Seleucid, Parthian, and Sassanid periods, swelling into a town when it absorbed the population of the holy city of Istakhr on its final destruction by the Arabs in the mid-11th century. The Mongols in the 13th century spared the city and built the New Mosque (1218), the Abesh Khatun (tomb of the last Atabeg ruler), and the fortress Bagh-e-Takht. In 1387 and again in 1393 Timur (Tamerlane) occupied Shiraz which, with its Masjid-e-Jume or Congregational Mosque (894), Shah Cheragh shrine (1344–49), and great library, had become a Muslim centre rivaling Baghdad. In prosperous Safavid times (1501–1736) the striking tiled Madrasseh (theological school; 1615) and the avenue on the road to Isfahan were constructed, but in 1724 the Afghan invaders, and in 1744 their vanquisher Nadir Shah himself, both sacked the city. From the calamitous floods (1630, 1668), pestilences, famines, and earthquakes (chiefly 1824, 1853) it is remarkable how much of Shiraz has survived. The remains of the old capital and the buildings constructed by the *vakil*, or regent, Karim Khan Zand (1750–79), who made Shiraz his capital, lie within the old city, whose western walls are marked by the Sa'di and Darius avenues. These buildings include the regent's mausoleum, an octagonal tiled kiosk, now the museum; the Arg, or citadel, now the prison; the Vakil Bazaar; and the Vakil Mosque. The two principal avenues, Zand and Lotf-e-Ali Zand, run east–west from the old city to the new. Buildings in the new city include the Persian Church of St. Simon the Zealot, the army headquarters, the *ostandari* (governor-general's office and residence), several hospitals including the Nemazee, Khalili, and Christian hospitals, and Pahlavi University (1948) with its faculties of arts and science, agriculture, medicine, and engineering.

Shiraz is still the trading centre for the Kashgai (Qashgai) and Khamseh tribes, and carpets and rugs of excellent quality are made in the surrounding area. There are cement, sugar, and fertilizer factories and several textile mills, but the crafts of gold, silver, and khatam inlay work still flourish. The local lambskins, well known for their quality, are exported. Main roads connect the city with Bushire, Isfahan (via the ancient sites of Persepolis and Pasargadae), and Teheran, and with the southwestern towns of Abadan, Khorramshahr, and Ahvaz. Daily bus and air services operate to Isfahan and Teheran.

BIBLIOGRAPHY.—Ali Sami, *Shiraz* (1958); A. J. Arberry, *Shiraz, Persian City of Saints and Poets* (1960); L. Lockhart, *Persian Cities* (1960); J. I. Clarke, *The Iranian City of Shiraz* (1963). (P. B. Go.)

SHIRE, a river of Africa which flows for about 300 mi. (480 km.) through southern Malawi (Nyasaland) to join the Zambezi River in Mozambique (Portuguese East Africa). Together with a left-bank tributary, the Ruo, part of its lower course forms the southwestern boundary of Malawi with Mozambique. The Shire is mainly the overflow channel from Lake Nyasa to the Zambezi, although it has also a comparatively small local drainage from the southern part of the Nyasa Valley. The river issues from the southeastern end of Lake Nyasa at Monkey Bay and flows slowly past Fort Johnston before entering the shallow Lake Malombe, from which it flows, still slowly, for 22 mi. (35 km.) to Liwonde. At Matope it begins its 50-mi. (80-km.) stretch known as the Murchison Cataracts or rapids. These end about 12 mi. (19 km.) above Chikwawa, 12 mi. below which the river spreads out into the vast Elephant Marsh, about 10 mi. (16 km.) wide, which continues, alternately narrowing and widening, to near the Portuguese town of Vila Bocage; below Port Herald it is known as Ndindi Marsh. The excessive flooding of this part of the river valley is caused largely by the gradient of the lower Shire being lower than that of the Zambezi, whose waters flood back along the Shire channel and also reach the Ndindi marshes by the Ziu Ziu overflow from the Zambezi.

The course of the Shire divides naturally into three tracts. The first, from Monkey Bay to Matope is 82 mi. (132 km.) long with a fall of 23 ft. (7 m.), the gradient over considerable stretches being as low as 1:100,000. It is thought that this tract may be a part of the former southward extension of the lake. There are no rock outcrops in the riverbed. The second tract extends from Matope for 50 mi. (80 km.) over rugged gneiss outcrops. Five main rapids occur: at Matope Bridge, Nkula Falls, Tedzani Falls, Mpatamanga Gorge, and Hamilton Falls (12 mi. upstream from Chikwawa). Together these constitute the Murchison Cataracts in which the river descends 1,260 ft. (384 m.). The third tract is the Elephant and Ndindi marsh area, where the gradient is about two feet to the mile and the floodbasin is covered with reeds, elephant grass, and papyrus. The discharge of the river is determined very largely by the level of the water in Lake Nyasa. From 1915 (lowest recorded level) to 1937 (highest recorded level) there was no outflow from the lake. In 1937 the waters broke through the obstructions in the riverbed and forced their way again to the Zambezi.

Plans for development include the control of the water level of Lake Nyasa; the construction of a barrage and low-head power station at Matope; the clearing and draining of the marshes; and pilot experiments on various forms of irrigation.

See J. G. Pike and G. T. Rimmington, *Malawi: a Geographical Study* (1965); N. J. Cochrane, "Lake Nyasa and the Shire River," *Proc. Instn Civ. Engrs,* vol. 8 (1957). (J. H. WN.)

SHIRE COURT, in England, a medieval court for the conduct of administrative, financial, and judicial business, based on the shires or counties, which were the most important administrative units in Anglo-Saxon local government. The shire, with its court, appears to have been fully established by the reign of Edgar (959–975). The shire court met in Saxon days in its traditional meeting place in the open air twice a year for specially full meetings attended by the great landowners of the shire. The king sent his instructions to these meetings by his ministers and all the business of government, both judicial and financial, was dealt with. Between those two meetings routine business was conducted in smaller meetings, probably held every four or five weeks. The sheriff presided and called on the suitors (in theory all the great landowners and representatives of the tenants, but eventually few besides the holders of particular plots of land to which the burden of suit of court had become attached) to declare the law. The Norman and Angevin kings threw more work on the court by the issue of many "viscontiel" writs, ordering the sheriff to do justice in particular cases. But the increasing complexity of the law, royal distrust of powerful sheriffs, the rapid development of the central courts of justice, and the increasing frequency of judicial eyres all combined to lessen the judicial importance of the old shire courts. Nevertheless, although no actions of trespass, debt, or detinue could be brought which involved a greater sum than 40*s.*, a considerable amount of litigation was done in the shire courts as late as the mid-13th century. Apart from proceedings in outlawry the shire court was not a court of record.

BIBLIOGRAPHY.—T. F. T. Plucknett, *A Concise History of the Common Law*, 5th ed. (1956); Sir F. Pollock and F. W. Maitland, *The History of English Law*, 2 vol., 2nd ed. (1898); *Brevia Placitata*, ed. by G. J. Turner for the Selden Society, vol. lxvi for 1947 (1951). (D. M. S.)

SHIRLEY, JAMES (1596–1666), English poet and dramatist, a leading playwright in the decade before the closing of the theatres by Parliament in 1642, is best remembered for the poem beginning

> The glories of our blood and state
> Are shadows, not substantial things;
> There is no armour against Fate;
> Death lays his icy hand on Kings . . .

Born in London and baptized on Sept. 7, 1596, he was educated at Merchant Taylors' School (1608–12), and, after service in a moneylender's household, matriculated at Catharine Hall, Cambridge, in 1615. Graduating in 1617, he took orders, and went to St. Albans, where he married (1618) and (*c.* 1621) became Master of the Grammar School. In *c.* 1624, having become a Catholic, he moved to London and turned playwright. His first play, *Love Tricks, or the School of Compliments*, was performed in 1625 at the Phoenix, Drury Lane, for which theatre most of his plays until 1636—about two a year—were written. When the theatres

closed in May 1636 as a plague precaution, he went to Dublin, where he was dramatist for St. Werburgh's Theatre. He returned to London in 1640, to succeed Massinger as dramatist for the King's Men at the Blackfriars. The theatres having closed, in the Civil War he may have served with the king's forces, and then returned to teaching, at Whitefriars, London, publishing two Latin grammars, some nondramatic verse, and four masques. Anthony à Wood (*q.v.*) reports that Shirley and his wife died the day after they were forced by the Great Fire to leave their house in Fleet Street; they were buried on Oct. 29, 1666, at St. Giles-in-the-Fields.

Thirty-one plays, five masques, and a five-act moral allegory are extant. The most elaborate of the masques was the Inns of Court *The Triumph of Peace* (1634), with scenery by Inigo Jones and music by William Lawes. *The Contention of Ajax and Ulysses for the Armour of Achilles* (printed 1658, but performed earlier) ends with the dirge quoted above. He wrote plays in most of the current modes, but never became completely subordinate to Caroline courtly fads. Among the best are the mildly satirical comedies of fashionable London life: *The Witty Fair One*, a model of inventive plotting and fresh use of traditional material, *Hyde Park*, *The Ball*, and *The Lady of Pleasure*, his most polished comedy of manners; all performed between 1626 and 1635. *The Young Admiral* (1633) and *St. Patrick for Ireland* (?1639) are notable, the first for its ingenious and genteel tragicomic heroics, and the second for its astonishing mixture of farce, romance, miracle, and show. His best tragedies, both on dark, Italianate themes, are *The Traitor* (1631) and *The Cardinal* (1641), a play of love, politics, revenge, and poison. This, like a number of Shirley's other plays, was revived at the Restoration and was twice enjoyed by Pepys.

Bibliography.—*Editions: Dramatic Works and Poems*, ed. by A. Dyce, with notes by W. Gifford, 6 vol. (1833); *James Shirley*, six plays, ed. by Sir Edmund Gosse, for the "Mermaid Series" (1888); *The Poems of Shirley*, ed. by R. L. Armstrong (1941).

Biography and Criticism.—A. C. Swinburne, "James Shirley," in the *Fortnightly Review*, xlvii (1890; reprinted in *Contemporaries of Shakespeare*, 1919); R. S. Forsythe, *The Relation of Shirley's Plays to the Elizabethan Drama* (1914); A. H. Nason, *James Shirley, Dramatist* (1915); A. B. Harbage, *Cavalier Drama* (1936); F. T. Bowers, *Elizabethan Revenge Tragedy, 1587–1642* (1940); F. S. Boas, *An Introduction to Stuart Drama* (1946); G. E. Bentley, *The Jacobean and Caroline Stage*, vol. v (1956). (P. Ur.)

SHIRLEY, WILLIAM (1694–1771), colonial governor of Massachusetts, was born at Preston in Sussex, Eng., Dec. 2, 1694. He studied at Cambridge and was later admitted to the bar. In 1731 after 11 years of law practice in England, he migrated to Boston, Mass. Shirley was a constant but not always successful office seeker. His loyalty to the imperial prerogatives, his ability, and his friendship with the duke of Newcastle resulted in his being appointed admiralty judge in 1733 and the king's advocate general in 1734. In 1741 he was appointed to the governorship of Massachusetts and proved to be a popular and successful governor in a rather difficult period. He gave Massachusetts a relatively sound currency.

Shirley's most important achievements as governor were in the realm of colonial and imperial defense. He was quite successful in building up the Massachusetts fortifications and unusually astute in his analysis of the nature and importance of the struggle between England and France for control of North America. During King George's War (1740–48) he organized and planned Britain's one great victory, the capture in 1745 of Louisburg on Cape Breton Island. In 1753 Shirley renewed his recommendations for united action against the French and drew up a plan for offensive warfare. After the defeat and death of Gen. Edward Braddock in western Pennsylvania in 1755, Shirley became commander in chief of the English forces in America but neither in America nor in England could he command the respect and cooperation necessary to carry out his well-conceived plans. After the failure of his expedition against Ft. Niagara he was replaced as commander and governor and was recalled to England where he was charged with treason. After being vindicated he was appointed governor of the Bahamas in 1761. In 1770 he returned to Roxbury, Mass., and died there on March 24, 1771.

See G. A. Wood, *William Shirley, Governor of Massachusetts, 1741–1756* (1920); C. H. Lincoln, *Correspondence of Wm. Shirley* (1912). (Ra. Mu.)

SHISHKOV, ALEKSANDR SEMENOVICH (1754–1841), Russian writer and statesman noted for his Slavophil, nationalistic religious fervour. A naval officer by training, he rose to the rank of vice-admiral and was appointed a member of the Admiralty College under the emperor Paul I. Disagreeing, however, with the reforms instituted on Alexander I's accession, he retired from official life and devoted himself to literature. He became engrossed in the study of Old (Church) Slavonic and regarded himself as a philologist, but his linguistic studies became a vehicle for nationalistic propaganda. He attacked everything new, whether administrative or literary. A member of the Russian Academy from 1796, he sought to preserve the purity of the Russian language, praised 18th-century Russian writings, and criticized the literary innovations of Nikolai Karamzin and French influence on Russian vocabulary, thought, and manners. He asserted that Old Slavonic was the root of modern Russian, and demanded that words of foreign origin be replaced with words derived from Old Slavonic. To encourage and train young *literati* in pure Russian literature, Shishkov organized literary meetings that were designed to instill in Russian society greater patriotism through Russian language and literature. The works read at these meetings were published, among them Shishkov's own *Discourse on Love for One's Country* (1811).

As the life-and-death struggle with France approached (*see* Napoleonic Wars), Shishkov's ultrapatriotism was useful, and Alexander I, impressed by the *Discourse*, recalled Shishkov to public service as secretary of state. Accompanying the emperor to Vilna, he wrote all the major orders and rescripts in his usual patriotic style. Nominated president of the Russian Academy in 1813 and a member of the State Council in 1814, Shishkov in 1824 became minister of education and director of non-Orthodox religious affairs. He opposed mass education, instituted strict censorship, and persecuted the biblical societies for spreading revolutionary books. Shishkov retired from the ministry of education in 1828 but remained head of the Academy. On his death in 1841 the Russian Academy was closed and its functions were assumed by the newly created Division of Russian Language and Literature of the Academy of Sciences. The ideas of Shishkov and his associates found amplification in the highly romantic nationalism of the Slavophils (*q.v.*) of the 1840s. (G. A. Ln.)

SHIVAISM (Shaivism), the cult of the god Shiva (Śiva), which, with Vishnuism and Shaktism (*qq.v.*), is one of the main forms taken by modern Hinduism (*q.v.*). Shiva originated in the Vedic god Rudra (*see* Vedic Religion), who became identified with him. Although Rudra is celebrated in only a few hymns of the Rigveda, it is clear that from the beginning he was one of the chief gods. A beautiful litany, the *Satarudriya*, addressed to Rudra, forms part of the Yajurveda. In the Atharvaveda and the Brahmanas there are allusions to Rudra as well as descriptions of rituals connected with his worship. The fact that the word *rudra* sometimes occurs in the Rigveda as an epithet meaning "terrible" has led some scholars to deduce that Rudra was a completely malevolent God who was referred to as *shiva* ("auspicious") only euphemistically, but this view cannot be reconciled with the passages in which prominence is given to Rudra's benevolent aspect. According to one derivation of the word *rudra*, Rudra is so called because he drives away sin and suffering. The name has also been interpreted to mean "he who attenuates sin." Some of the other names by which Rudra-Shiva is known are Sambhu ("benignant"), Samkara ("beneficent"), Pasupati ("lord of cattle"; *i.e.*, souls), and Mahadeva ("great god").

In the Epic period (*c.* 600 B.C.–*c.* A.D. 200) the function of cosmic destruction is assigned to Shiva, but since destruction of the world only precedes re-creation, this is considered to be an act of grace in Hinduism. The consort of Shiva is Ambika, the Mother, known by various names—Parvati, Uma, Durga, and Kali. This divine couple has two sons, Ganesha and Karttikeya (*q.v.*).

In modern India Shiva is the supreme God to the devotees of


ROYAL ACADEMY OF ARTS, LONDON

DANCING SHIVA (NATARAJA) SOUTH INDIAN BRONZE, 12TH–13TH CENTURY A.D. IN THE MUSEUM OF ASIATIC ART, AMSTERDAM

Shaivism, who are found all over India. There are several schools of Shaiva thought, ranging from pluralistic realism to absolute monism, but they all agree in recognizing three principles: *pati* or God, *pasu* or individual soul, and *pasa* or bonds that confine the soul to earthly existence. The goal set for the soul is to get rid of its bonds and gain *shivatva* (the nature of Shiva). The paths leading to this goal are *carya* (external acts of worship), *kriya* (acts of intimate service to God), *yoga* (meditation), and *jnana* (knowledge). Among the classical schools of Shaivism are the Trika (or Pratyabhijna) of Kashmir, Virashaivism of the Kanarese and Telugu areas, Shaiva-Siddhanta of the Tamils, and Shivadvaita of Shrikantha and his followers. Shaivism, like some of the other forms of Hinduism, spread in the past to other parts of southeast Asia, including Java and Bali, Champa and Cambodia.

In popular Hinduism the god retains his character as the destroyer of the world, the third member of the Hindu trinity (with Brahma, the creator, and Vishnu, the preserver), and he is a somewhat remote though by no means always terrifying figure. Shaiva temples always contain an image of Nandi, the sacred bull, the mount of Shiva, symbolic of the soul. He is represented anthropomorphically in a variety of ways, often as a dancer (Nataraja), a form which has inspired many beautiful works of art. His most familiar symbol is the linga (lingam), representing that which has a shape and yet no definite shape, an object of great sanctity to which offerings of water, flowers, fruit, and rice are made.

See S. Dasgupta, *History of Indian Philosophy*, vol. 5 (1955); T. M. P. Mahadevan, *The History and Culture of the Indian People* (1954).
(T. M. P. M.)

SHIVPURI, a town in the Gwalior division of Madhya Pradesh, India, and the administrative headquarters of the district of the same name, lies 75 mi. (121 km.) SW of Gwalior at the terminus of the Gwalior-Shivpuri Light Railway. Pop. (1961) 28,681. It is situated on an elevated watershed from where streams radiate in all directions. Formerly called Sipri, the town was the headquarters of the Narwar District of Gwalior state. In the 17th century it was granted to Amar Singh Kachwaha of Narwar. In 1804 it was seized by the Sindhias, whose palace is one of the town's notable features. Shivpuri increased in importance as a trade mart following the opening of the railway and is a centre for the distribution of forest produce.

Shivpuri District, originally known as Narwar Zila, has an area of 3,986 sq.mi. (10,324 sq.km.) and a population (1961) of 557,954. The greater part is cut up by jungle-covered ridges striking from northeast to southwest, but the area lying west of the Vindhyan spurs is fairly level and fertile, and jowar, wheat, and oilseeds are grown there. The district also possesses a national wildlife sanctuary. Narwar, the former district headquarters, is of considerable historical and archaeological interest. It has an old fort picturesquely situated on the steep scarp of the Vindhyas.
(S. M. A.)

SHIZUOKA, prefecture (*Ken*) on the Pacific coast in central Japan. Area 3,000 sq.mi.; pop. (1960) 2,756,271. Favoured by a warm climate, Shizuoka has developed a varied agriculture. Oranges and tea are the chief products, and fishing is highly developed at Shimizu, the principal port. The processing of food and other industries are important. Hamamatsu (*q.v.*) is famous for the manufacture of musical instruments. Topographically and geographically the eastern and western sections differ. Volcanoes and hot springs dominate the east, while large river valleys occur in the west. Shizuoka is served by the Tōkaidō central trunk railway line.

Shizuoka, the prefectural capital (pop. [1960] 328,819) is situated on the delta of the Abegawa River. The climate is mild. Shunpei, as Shizuoka was called until 1869, was a castle town as well as one of the 53 stage towns on the Tōkaidō highway. About 70% of Japan's tea is raised in its environs and the city is famous as a trading and processing centre for green tea. Shizuoka is the site of Shizuoka University, a women's pharmaceutical college, and several industrial and agricultural experimental stations.
(R. B. H.)

SHKODËR (ancient Scodra; Italian Scutari), a town of northwestern Albania and administrative centre of Shkodër *rreth* (district), 75 mi. (120 km.) NNW of Tiranë by road. It lies at the southeast end of Lake Shkodër (Skadarsko Jezero), where the Buenë (Bojana) River leaves the lake for the Adriatic Sea, and at the edge of a wide plain surrounded southeast, east, and north by high mountains; it is overlooked by Mt. Tarabosh to the west across the Buenë. Pop. (1960) 43,305, more than half Roman Catholics, with many Muslims and a few Orthodox.

Shkodër is the most historic town in Albania. It is dominated by the old citadel on an isolated rock at the south, past which the Kiri flows into the Buenë. The Drinasë, the new channel cut by the Drin in 1858, joins the Kiri just above the Buenë and periodically causes flooding in the old town near the citadel by raising the water level in the lake. The newer part of the town spreads to the north. The substantial remains of the fortified citadel are mainly Venetian (14th century), built on the site of an earlier walled town erected (or restored on an Illyrian site) by the Romans and repaired by Byzantines and Slavs; the Venetian walls were later strengthened by the Turks. The old part with its market and narrow streets is a typical old Balkan town. There is a Roman Catholic cathedral (19th century) and modern administrative and other buildings. Shkodër has been the seat of an archbishop (though not continuously) since the 4th century. The national poet Gjergj Fishta (Franciscan, 1871–1940) lived, taught, and wrote at Shkodër, and his community especially contributed much to the country's literature and national spirit.

The town is the traditional market centre for the whole northern mountain area, which is mainly Catholic but includes the Muslim clans of the northeast. Before World War II it was the most colourful town in Albania. Surfaced roads now link Shkodër with the main northern centres as well as with the south. Industries include the production of cement, cotton textiles, glass, and shoes, and there is tanning and food canning.

Shkodër was probably an Illyrian capital during the 1st millennium B.C. and is mentioned as such by Livy. At the partition of the Roman Empire the town passed to Byzantium, and later to the Bulgars, Serbs, and Venetians. After the Serb Stephen Dushan's death in the 14th century it became the capital of George Balsha, prince of Zeta (or Zenta), and was sold by his family in 1396 to Venice. It was a stronghold of Skanderbeg in the 15th century. The Turks laid violent siege to it in vain in 1473 and again in 1478; but after the latter siege Venice ceded it to Turkey and the valiant citizens of Shkodër were evacuated to Italy. The Bushati family took control in the 18th century, but in 1829 the Turks made it the capital of a vilayet (administration division). In 1912–13 it was besieged by the Montenegrins but was then taken over by an international brigade. Occupied by Austria in 1916–18, it was then controlled by Allied forces until reunited with free Albania in 1922. After 1944 it was for a time the main centre of resistance to the Communist government.

Lake Shkodër, 143 sq.mi. (370 sq.km.) in extent and mainly shallow, is divided between Albania and Yugoslavia. It has a fine rugged mountain setting, and fishing is locally important.
(D. R. O.-H.)

SHMIDT, OTTO YULIEVICH (1891–1956), Russian scientist and polar explorer, was born on Sept. 30 (new style Sept. 18, old style), 1891, at Mogilev. He was a university teacher of mathematics before and immediately after the Revolution of 1917. His support for the new regime and his administrative ability caused him to be appointed to various senior civil service jobs: he was manager of the state publishing house, deputy head of the central statistical administration and a member of the state

planning commission. After taking part in the Soviet-German Pamir expedition of 1928, he became interested in the arctic and went on an expedition to Franz Josef Land in 1929. Next year he was made director of the Arctic institute and in 1932 first head of the chief administration of the northern sea route (Glavsevmorput), the newly formed government department responsible for arctic development. During his six years in this job he directed the extensive Soviet program of exploration and exploitation of arctic resources, and generally spent the whole of each summer season aboard an icebreaker off the north Siberian coast. He led the expeditions in the "Sibiryakov" (1932) and "Chelyuskin" (1933–34), the establishment of the north pole drifting station (1937) and its relief (1938). In 1939 he left arctic work and devoted himself to science for the rest of his life. He had been elected an academician in 1935 and vice-president of the Academy of Sciences in 1939. On his initiative the academy created the Institute of Theoretical Geophysics in 1938. He was its director until 1948, advancing at this time a new theory of the origin of the earth. He was chief editor of the *Great Soviet Encyclopaedia* from 1924 to 1941 and of the journal of popular science, *Priroda*. To this remarkable record of activity he added a professorship of mathematics at Moscow University from 1926 until his death, becoming in this capacity the founder and leader of the Moscow school of algebra. His publications were mainly mathematical papers and articles on polar work. His influence on Soviet life and thought was both considerable and remarkably diverse. He died in Moscow on Sept. 7, 1956. (T. E. A.)

SHOA, the central province of Ethiopia, and the site of residences of the Ethiopian kings. Pop. (1961 est.) 2,500,000, largely of mixed Galla-Amhara stock. The province comprises mostly high plateau country rising to 13,123 ft. (4,000 m.) in Abuya Myeda, but the eastern and southeastern boundaries are in the Great Rift Valley along the Awash River. The hydroelectric plant at Koka on that river, 60 mi. (97 km.) S of Addis Ababa, was designed to provide power for the industrial development of the province. Stock is raised; cereals, vegetables, coffee, and tobacco are the chief crops. Roads from all parts of the country converge on the province; and a metre-gauge railway runs eastward from Addis Ababa, linking it with Djibouti in French Somaliland. The important centres are Addis Ababa (*q.v.*), the Ethiopian and provincial capital; Debra Birhan; Fiche; Nazareth (Hadama); Debra Libanos, a monastic centre; and Ankober, the former capital of the Shoa kingdom.

Shoa was the residence of the Ethiopian kings from the mid-10th century to the end of the 14th century. The actual sites of the royal residences have not yet been accurately located although they are named and described on contemporary maps. In 1528 Shoa was overrun by Muslim invaders from the state of Adal to the east and the ancient cities were destroyed. For over a century afterward Shoa was penetrated by the Galla peoples from the south who moved into the political vacuum left in this region and established themselves in farming communities as far north as the Blue Nile Valley. In 1856 Shoa was incorporated in the Ethiopian Empire by the emperor Theodore II, but Menelik II regained his kingdom of Shoa and in 1886 chose the site for Addis Ababa, his future capital. Building began at once and when Menelik became emperor of the whole country in 1889 Addis Ababa became the capital of Ethiopia. *See* also ETHIOPIA: *History*. (G. C. L.)

SHOCK is the medical term for an acute condition usually marked by weakness, profound depression, and pallor. The skin becomes moist, cold, and grayish. The pulse is rapid and weak. The blood pressure usually is below normal. Breathing is shallow. Urination is suppressed or stopped. If the patient has sufficient strength, he may be restless, but frequently he can be aroused only with great difficulty. The metabolic rate is low and there is suppression of all vital functions, so that the patient is obviously gravely ill.

The term shock was first used in the 19th century to characterize the alarming symptoms known to result from a wide variety of causes, both physical and psychic. Such an uncritical use of the term robbed it of significance, but the term persisted. Qualifying words indicating the probable cause of the condition are now usually employed; *e.g.*, traumatic, postoperative, anaphylactic, histamine, transfusion, etc. In the last analysis, death from almost any cause may be preceded by a number of symptoms characteristic of shock; therefore consideration of the condition usually is limited to collapse resulting from hemorrhage and from various forms of physical trauma (injury).

Traumatic shock was classified by W. B. Cannon into two types, primary and secondary. The primary type was defined as the collapse occurring immediately after severe injury and was considered as indistinguishable from syncope (fainting) and collapse. Secondary shock was defined as collapse occurring after a variable period of time following severe injury.

Medical authors make little distinction between shock and collapse or syncope, except to recognize that if collapse or syncope persists, shock will result. According to V. H. Moon, the attempt to distinguish between shock and collapse is like distinguishing between breeze and wind. He defined shock as "a disturbance of fluid balance resulting in a peripheral circulatory deficiency which is manifested by decreased volume of blood, reduced volume flow, hemoconcentration and by renal functional deficiency." N. E. Freeman defined it as "the clinical condition characterized by progressive reduction in circulating blood volume due to increased capillary permeability." H. N. Harkins defined shock as "an oligemia [decreased volume of blood] *initiated by* traumatic local fluid loss, either whole blood, plasma, or both; *accompanied by* decreased cardiac output, diminished volume flow, lowered venous pressure, decreased oxygen consumption, arteriolar vasoconstriction, progressive hemoconcentration, capillary congestion, acapnia [diminished carbon dioxide in the blood], and secondary blood pressure fall; and *perpetuated by* a summation of these factors and possible hyperpotassemia [increased blood potassium], increased generalized capillary permeability, anoxia [lack of oxygen], action of tissue metabolites, and deficiency of adrenal cortical hormone." Harkins also pointed out that "other changes, both chemical and pathologic, may occur in shock, including increased blood nonprotein nitrogen, decreased coagulability of the blood, and in some instances increase in plasma magnesium."

In a summation of a 1961 conference on shock, J. E. Rhoads concluded that the definitions offered by most of the participants would be encompassed by considering shock as "a default in the transport mechanism of the body generally affecting vital cells." The default, when absolute, is the irreversible shock of the physiologist; when relative it is the reversible shock syndrome so important for the clinician. This definition would allow the inclusion of hypovolemic (decreased volume of circulating fluid) and cardiac shock as well as that due to partial respiratory obstruction, uremia, and at least the late phases of neurogenic and septic shock (associated with vasodilatation resulting from, respectively, nerve action and toxic conditions).

Cannon and W. M. Bayliss developed the so-called traumatic toxemia theory to account for the initiating factor of shock. According to this theory, a toxic substance is absorbed from the injured tissues and distributed to all parts of the body, where it injures the capillaries which in turn allow fluid to leak through them to such an extent that the volume of circulating blood becomes inadequate. In view of the fact that histamine (*q.v.*) or histaminelike substances had been demonstrated in animal tissues and that shock had been produced experimentally by injections of this substance, the toxemia theory was readily accepted. It was called into question, however, by A. Blalock and by E. Parsons and D. B. Phemister; though weakened, it was not abandoned.

There is substantial agreement among authorities that loss of blood, blood plasma, or body fluids by whatever means is the most important initiating factor in shock of most types. Hemorrhage, excessive sweating, vomiting, dehydration, diarrhea, injury to the body, low blood pressure, cold, pain, and psychic stimuli such as fear and grief predispose the individual to shock, since any one of these, if sufficiently severe and prolonged, may result in reduction of the volume of circulating blood. Some of the results of this are known, such as increased capillary permeability after anoxia, but others have been only surmised. Clearly, reduction of the circulating blood volume causes anoxia of the tissues, in

turn causing injury to the capillaries which permits further loss of plasma volume. This cycle, if uninterrupted, becomes irreversible, and death is certain.

Treatment.—Treatment of shock depends to some extent on the nature of the injury or disease responsible for its development. Attention is first directed to the loss of blood or other body fluids, and steps are taken to prevent further loss. Hemorrhage, vomiting, sweating, and diarrhea are controlled as soon as possible, and pain is relieved. In case of severe injury of the extremities, the part is immobilized. Rest and quiet are recommended as soon as the patient has been made reasonably comfortable. Prevention of further loss of body heat is accomplished by use of covers, hot-water bottles, and heating pads, but overtreatment should be avoided. Elevation of the foot of the bed may be helpful. Vasoconstrictor drugs are not indicated as a routine treatment but may be of benefit in exceptional cases such as so-called spinal shock. In cases in which pulmonary ventilation is inadequate for saturation of the blood, oxygen is administered. Operations in the presence of shock are done when possible with local-acting or short-acting drugs such as pentothal sodium or hexobarbital. If the presence of shock is thought to be owing to bacterial infection, antitoxins or antibiotics are employed.

Since the fundamental factor in the production of shock is decreased volume of circulating blood, the best treatment of the condition is the transfusion of whole blood or blood plasma, but if these are not available appropriate doses of solution of acacia, pectin, dextran, or other blood substitute should be injected. Isotonic solutions of sodium chloride and glucose are recommended for temporary use until more adequate measures can be instituted.

Prevention.—In hospitals prevention of shock is of paramount importance. Patients who have lost large amounts of blood are prepared for operation by adequate transfusion. Those who have become dehydrated receive fluids by infusion or are given proper quantities of fluid by mouth. Preparation of intravenous administration of fluids is routinely made before the operation begins; at the first sign of deficient circulation, fluids are supplied to the patient. In case of severe hemorrhage during the operation, blood is administered in sufficient quantities to prevent the occurrence of shock.

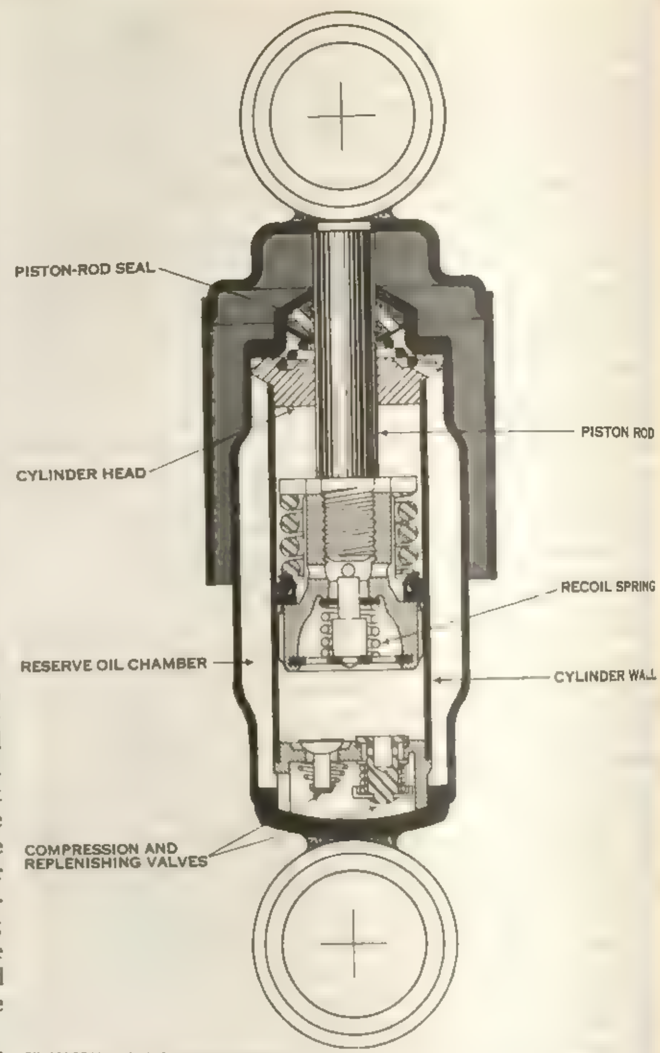
For "shell shock"—a use of the term outside the scope of this article—*see* COMBAT FATIGUE. *See* also references under "Shock" in the Index.

BIBLIOGRAPHY.—W. B. Cannon, *Traumatic Shock* (1923); C. J. Wiggers, "The Present Status of the Shock Problem," *Physiol. Rev.*, 22:74–123 (1942); V. H. Moon, *Shock: Its Dynamics, Occurrence and Management* (1942); Norman E. Freeman, "The Mechanism of Shock," *Military Surgical Manuals*, vol. v (1943); H. A. Davis, *Shock and Allied Forms of Failure of the Circulation* (1948); "Proceedings of a Conference on Recent Progress and Present Problems in the Field of Shock," *Federation Proceedings*, Supplement no. 9, p. 235 (July 1961).

(H. E. Ex.)

SHOCK ABSORBER, a device for controlling the motion of an elastic suspension system such as that of a vehicle. The spring elements of such a suspension system are deformed by an impulse applied to the wheels and tend to react after the force of the impact is spent. Shock absorbers slow down the vibratory motion of the suspended portion and quickly restore equilibrium.

Modern shock absorbers are hydraulic devices and are double acting in that they oppose both the compression of the springs and the rebound. The direct-acting or strut type (*see* figure) is attached to the vehicle frame and axle by two eyes. The central tubular cylinder is fitted with a piston that is attached by a piston rod to the upper eye. The cylinder moves with the lower eye and an annular fluid reservoir surrounds it. A spring-loaded one-way seal prevents fluid flow downward past the piston, but permits controlled flow upward. A valve in the under side of the piston controls the rate at which fluid can flow from the chamber during rebound. The fluid reservoir communicates with the compression chamber through two spring-loaded valves in the base of the cylinder. The discharge valve at the right controls the rate at which fluid can escape from the cylinder during a compression stroke and the valve at the left later refills the cylinder. Relative motion of the sprung and unsprung portions of the vehicle



BY COURTESY OF THE GABRIEL COMPANY

CROSS SECTION OF HYDRAULIC SHOCK ABSORBER

can be only as rapid as the reciprocation of the piston in the shock absorber, which can be no faster than the fluid flow from one compartment to another permits. Controlled resistance to fluid flow limits the rate at which the suspension can vibrate.

Shock absorbing mountings are used to isolate the vibration of machines, such as diesel engines. These mountings usually employ some resilient member, such as blocks of cork or rubber, interposed between the engine and the floor on which it rests.

Large calibre guns are fitted with recoil mechanisms that permit the gun barrel or the breech to move back under the reaction to the acceleration of the missile. This movement is gradually stopped by a hydraulically damped spring that returns the parts to their operating position.

See VIBRATION CONTROL. (O. C. C.)

SHOCKLEY, WILLIAM BRADFORD (1910–), U.S. physicist, Nobel Prize co-winner for physics in 1956 with John Bardeen and Walter Houser Brattain (*qq.v.*) for work on the transistor (*q.v.*). Shockley was born in London of U.S. parents Feb. 13, 1910. Reared in California, he was educated at California Institute of Technology (Pasadena) and Massachusetts Institute of Technology (Ph.D., 1936).

He joined the technical staff at the Bell Telephone Laboratories in 1936, and there began experiments that led to his invention and development of the junction transistor. During World War II he served as director of research for the Antisubmarine Warfare Operations Group of the U.S. Navy and as expert consultant for the office of the secretary of war. After the war he served as director of transistor physics research for Bell. He was visiting professor of physics at the California Institute of Technology in 1954, deputy director of the Weapons Systems Evaluation Group

of the Department of Defense in 1954–55, and joined Beckman Instruments, Inc., to establish the Shockley Semiconductor Laboratory in 1955. In 1958 he became lecturer at Stanford University, and in the 1960s was director of the Shockley Transistor Unit of the Clevite Corporation in Palo Alto, Calif.

See S. Thomas, *Men of Space*, vol. 4 (1962).

SHODDY, originally probably a mill term referring to the wool waste thrown off during the processing of wool, now a term given to the fibres reclaimed from wool cloths. The reclamation consists of reducing the cloth to the fibrous state by a tearing-up process called "shoddy picking" or "garnetting." The resulting product is deficient in fibre length and strength and is usually mixed or blended with fleece wool or some other fibre before being respun. In the United States reclaimed wool used in consumer products must be identified on the label as either "reused" or "reprocessed." (C. P. MA.)

SHOEBURYNESS, a promontory on the Essex coast of England, 4 mi. E of Southend-on-Sea, and the point at which the coastline trends northeast from the estuary of the Thames. It has a gunnery testing establishment. On the seaward side of the Ness is an earthwork attributed to the Norsemen.

SHOES and boots, worn as protective footwear, are ancient items of dress. In early civilizations shoemaking skills were developed to a high degree; styles and variety were produced by the artisan, but the process was slow and expensive. Modern mass production depended upon the development of shoe machinery, which began about the middle of the 19th century, and the industry logically developed most extensively in the more highly industrialized nations. In the 1960s the United States alone produced and consumed more than 25% of the world's leather footwear—an average of 3½ pairs per person annually. It is estimated that purchase of about half of these was for replacement or utility; that of the other half was motivated by the appeal of fashion. Women's dress shoes are influenced by fashion to a greater extent than are other types, and in the United States about 47% of all shoes manufactured are women's shoes. (The remainder consists of men's shoes, 17%; children's shoes, 23%; slippers, 12%; athletic and other specialized types, 1%.) Somewhat similar figures apply in Great Britain, another leading producer of leather footwear, although the consumption rate is slightly lower—just over three pairs per person. Per capita consumption and spending for footwear is also directly related to personal or family income (except that high-income families spend proportionately less on all clothing, including footwear, than middle- and low-income families).

CONSTRUCTION AND DESIGN

History.—The first type of footwear was a simple wraparound of leather (the basic construction of a moccasin), held together on the foot with rawhide lacings. Gradually this design was improved upon and assumed many variations. Another early design was the simple sandal, consisting of a sole held on the foot by a leather thong. The oldest shoe now known is a sandal made of woven papyrus, found in Egypt and dating from about 2000 B.C.

Moccasins, sandals, and simple boots were the prevailing types of footwear almost up until the Middle Ages. Wooden clogs or shoes, commonly used by the poorer classes, were also important, though distinctly different from the more desirable leather types. Until modern times footwear was largely homemade, though a considerable share was produced by artisans and craftsmen; large guilds of shoemaking specialists date back to ancient Rome. While there were small shoemaking shops consisting of a few craftsmen, the first signs of the "mass-production" factory did not appear until the 18th century. The first shoe factory in the United States is credited to John Adams Dagyr, a Welshman who opened a shop in Lynn, Massachusetts colony, in 1760. He employed a number of shoemakers, each performing only one operation, who produced shoes that could be purchased directly from stock. Previously, most purchased shoes were custom-made.

In 1818 lasts (wooden forms) for shaping shoes were made in rights and lefts. Another milestone was reached in 1846 when Elias Howe invented the sewing machine. Previously, the majority of shoes were made by attaching the sole to the upper part with tacks, nails, or wooden pegs; any sewing was done by hand. The sewing machine not only greatly speeded production but made a better shoe. Shortly after, and over the next 60 or 70 years, a steady flow of new and improved shoe machinery was introduced, and by the beginning of the 20th century the shoemaking industry was on a full mass-production basis.

Materials.—Since time immemorial the universal basic material for shoes and boots has been leather, used for uppers, linings, counters, inner soles, outer soles, welting, and other parts. Though still dominant, leather is being replaced by other materials, particularly in the component parts of shoes. For example, most inner and outer soles are now made of nonleather materials, such as rubber, man-made fibres, or composition materials. Other parts of the shoe are also being made of nonleather materials, such as plastic welting and plastic heels, coated fabrics for linings, natural and coated fabrics for uppers, plastic counters or back-part stiffeners, and chemically compounded box toes. Most of these materials, the products of modern chemistry, are priced lower than leather and at the same time meet all required performance standards.

Each type of leather commonly used in shoe uppers and linings has its distinctive characteristics, and each tends to find its own specialized applications in footwear. Calf is the "luxury" leather, used in the finest grade men's and women's shoes. Side leather (made from cattle hide, and called "side" because the large hide is cut down the middle, lengthwise, into two sides for easier handling) is the most versatile of all shoe leathers. It comes in several grades and is used for a great variety of footwear—women's lightweight casual footwear, children's school shoes, and other kinds. An estimated 70% of all leather used in shoe uppers is side leather.

Kid leather, made from goatskin, also has wide use, ranging from women's high-grade dress shoes to men's comfort shoes. Sheepskin is used largely in slippers and linings. Reptile leathers (alligator, lizard, and snake) are used in some women's and a few men's shoes. Cordovan (from horsehide—a small "muscle" layer in the hide) is a heavy leather popular in men's shoes. Certain specialized leathers such as kangaroo (used in comfort and athletic shoes), ostrich (used in a few women's shoes), and pigskin also find application.

Patent leather, usually made from cattle hide, is given a special surface finish. Suede is made from any of several leathers (calf, kid, cattle hide) by buffing the inner surface to create a napped finish.

Certain fabrics, including linen, satin, and silk, have found increasing use in footwear. Another modern development has been coated fabrics, consisting of a fabric base coated with a chemical surface finish to give a vast variety of textures and designs, many of them simulating leather grains to a remarkable degree. These find wide use in less expensive shoes, especially in women's and girls' footwear. Man-made fibres are producing new families of materials, such as nylon mesh and nylon velvet. Much synthetic patent and suede is also produced for shoes.

Some of these leathers and other materials are used for linings as well as for outside uppers of shoes.

Styling and Design.—Many kinds of footwear—men's work shoes, oxfords, and riding boots, for example—remain essentially unchanged from year to year, but in other types fashion plays a considerable role. Each year new styles are prepared to conform with the seasonal turnover and the general trends of fashion. In the U.S. it is estimated that more than 100,000 new sample styles are prepared annually by the manufacturers. After the retailers have made their selections, about 20 to 25% of the original samples will go into production.

Most manufacturers design their lines twice yearly, for spring-summer and fall-winter. Some designing is done for one brief season, such as summer, or for resort-cruise wear. The greatest demand for variety of style is in women's shoes, followed by men's and girls' shoes. Seasonal turnover in women's styles is quite pronounced, while in men's shoes it usually requires several seasons before distinct changes take place.

Determining the styling features to be incorporated in a new season's line is a highly complicated procedure. In women's shoes, for example, consideration must be given to colours, materials, lasts, patterns, ornamentation, silhouettes, heel shapes and heights, textures, and surface finishes. These must be welded into a "line" of scores of different styles for the new season, and they must be attuned to general fashion trends in women's apparel, for shoes are regarded as the most important accessory in a woman's ensemble.

Styling a new line involves an expert team of specialists: the designer, fashion coordinator, style man, sales manager, sales staff, and pattern man. New lines are usually prepared from 4 to 12 months ahead of season. New styles or style trends in footwear are usually initiated by the makers of higher-grade, higher-priced shoes. Once a new shoe is launched and receives the first signs of consumer interest, it is immediately adopted by the volume, or medium- and lower-priced, shoe lines.

After about 1945 in the U.S., juvenile footwear, particularly in girls' shoes, moved toward "fashion." While the conventional school oxford remained a basic shoe, a wide variety of new types and styles of shoes won increasing favour.

Methods of Construction.—There are said to be more than 800 ways in which to make a shoe, but only a few of these are actually used. Fundamentally, shoe construction consists of attaching the sole to the upper in such a way as to create a bond sufficiently durable to withstand the wearing conditions for which the shoe was designed. There are four basic construction methods:

1. Sewing, which includes such types as the Goodyear welt process, in which the sole is sewed to a welt strip that has previously been sewed to the upper and insole, and the stitchdown, in which the upper is turned out and stitched down to the sole.
2. Cementing, in which the outer sole is attached to the insole and upper by adhesives.
3. Molding, in which the sole is vulcanized or "molded" to the upper.
4. The use of fasteners, as in nailed or stapled shoes.

There are numerous variations within each classification, and special features of one process may be combined with special features of another to provide a sort of "crossbreed" construction.

Each kind of shoemaking has distinctive features that make it particularly adaptable to certain kinds of wear. For example, the Goodyear welt process, known for its sturdiness, is commonly used in men's and boys' dress shoes as well as in better-grade children's shoes of certain types. Cement construction enables shoes to be built on lighter, sleeker lines and is therefore used for most women's and many girls' shoes and for some lighter types of men's shoes. Molding, as in the vulcanized process, combines economy of construction with rugged durability and is used in men's work shoes and in some juvenile footwear. It is also used in all rubber-canvas shoes.

About 80% of all shoes manufactured in the U.S. are made by three processes: cement (54%), Goodyear welt (19%), and stitchdown (7%). Trends or shifts in the use of certain shoe constructions are motivated by technological changes, economics, and fashion. The cement process was launched commercially in the early 1930s and by the 1960s was used in over half of all shoes made in the U.S. The Goodyear welt was still regarded as one of the best methods but had steadily declined from its once dominant position. Market demand for lighter shoes and the proportionately higher number of women's shoes manufactured partially account for this. Differences in manufacture are also significant. In most basic constructions about 150 different operations are required to make a shoe, and as many as 200 may be required. Certain modern methods of construction, such as the vulcanized and injection-molding methods, have been well received because they reduce both the number of operations required and the number of component parts. In Great Britain, for example, one-fifth of all the footwear made at the beginning of the 1960s was produced by these methods, and the proportion was increasing.

Steps in Production.—Shoe manufacturing is essentially an assembly process of fusing together a number of components: heel, sole, counter, upper (vamp and quarter), lining, toe box, welting, insole, ornamentation, and the like. A great variety of machines and devices, each for a specialized operation, is utilized. Each style or type of shoe has its own preplanned schedule in moving through the production line. Operations and machines must be adjusted differently for different styles or types of shoes or even when different materials are used.

The individual components of a shoe are usually purchased by the shoe manufacturer from outside suppliers. These components (linings, soles, insoles, counters, heels, etc.) are purchased according to specifications for size, weight, grade, colour, finish, etc., all to be assigned to a particular style or type of shoe.

Each specialized construction has its distinctive procedures, but there are certain essential features common to all constructions. The major steps in shoe construction may be briefly outlined as follows:

The cutting room is the launching point in the shoemaking process. Here the upper materials and linings are cut out with dies that have been designed especially for that particular shoe, but for short production runs the older method of hand cutting around templates is still retained. The dies are used with clicking or cutting machines, and the weight and cutting edge of the die must be accommodated to the particular type of material being cut. Special machines are also used to perforate the upper material or emboss it with a particular design. Splitting machines (for reducing the thickness of a material) or crimping machines (for preforming leather parts) may also be used in the cutting room.

The shoe then moves to the fitting (or closing) room where the precut parts of the upper are fitted together, usually by sewing or stitching, although some cementing is done. This assembly includes the upper portions of the shoe, linings, stays, facings, eyelets, etc. Many types of sewing machines are used, some of them attaining speeds of more than 4,000 rpm and able to sew from one to eight rows of stitching at one time.

As the shoe moves out of the fitting room in preparation for the important lasting operations, it begins to take shape. The assembled upper, in loose form, is ready to be fitted over the wooden last for which the patterns were precisely designed. First, however, the insole is placed on the bottom of the last. This insole, of a shape and size to conform exactly to the bottom of the last, is a sort of "security layer" between upper and outer sole to fuse the shoe parts into a firm unit. The counter or stiffener at the back part of the shoe is placed between the lining and outside of the upper. These parts—insole, counter, and upper—are now on the last, and the entire unit is presented to the assembling machine.

In lasting operations ingenious mechanisms grip the upper at the toe and sides with pincers, pulling the leather tightly over the wood last so that the material conforms in minute detail to the whole fore part of the last. The heel or back part of the shoe is also lasted.

After lasting, however, a number of operations are required to prepare the shoe for bottoming. It is at this point that some of the essential differences between the various types of shoe constructions are evidenced. The Goodyear welt, cement, stitchdown, and vulcanized constructions enter the sole-laying stage under different sets of prepared conditions. Also, the method by which the sole itself is attached is distinctly different. Once the sole is attached it must still receive finishing touches, such as trimming, staining, and burnishing the edges. The bottom of the sole must be cleaned or buffed and polished.

Applying the heel is an equally meticulous operation, done with special heeling machines. These attach the heel with nails; in the case of high heels, a long screw goes down into the heel for added security. Once the heel is attached, it also undergoes trimming and finishing operations.

The final series of operations consists of applying heel or sock linings, inserting laces, attaching buckles to a strap, and brushing or polishing. Finally there is a thorough inspection, and the shoes are then ready for packing and shipping.

Boots are made in much the same way, the major differences being in the lasts, which generally have a bootlike shape, and in the patterns for the uppers. The type of construction depends upon the type of boot; for example, a work shoe or boot might

be made by a vulcanized process while a riding boot would be made by a Goodyear welt process.

THE FOOTWEAR INDUSTRY

United States.—*Industry Structure.*—In the 1960s there were about 850 shoe-manufacturing firms in the U.S. The leading 10 firms accounted for 30% of the total production; the leading 30 firms for 42%; and the top 80 for more than 54%. Though shoe production has shown a steady gain (averaging about 10,000,000 pairs a year increase), the number of shoe-manufacturing firms continues to decline, marginal firms being eliminated under the duress of competition and capital requirements.

New England accounted for about one-third of the nation's shoe production, with Massachusetts the leading state, representing 17% of all shoe output. Other major shoe-producing states were New York and Pennsylvania, each with 14%; Missouri, 10%; and Ohio, Tennessee, Wisconsin, and Illinois, each averaging between 3% and 5% of the total.

Unlike shoe factories in most other countries, American plants are highly specialized, making only men's or women's or children's shoes or slippers, with further specialization in grades or price lines produced. The American shoe worker averages an output of about 2,900 pairs a year, the highest productivity level in the world. Factories tend to be medium sized. The best efficiency level is found to be at a production rate of between 4,000 and 10,000 pairs a day. The largest shoe-manufacturing firm produces over 200,000 pairs a day, but this is distributed over some 50 different plants. Most shoe factories are located in smaller communities rather than in large cities or heavily industrialized areas.

Retailing and Marketing.—Approximately 850,000,000 or more pairs of footwear (including rubber footwear and imports) are sold in retail stores annually in the U.S. Slightly over half of these are sold through shoe stores. The rest are distributed through various types of retail establishments—department stores, mail-order and variety store chains, clothing stores, general merchandise stores, discount houses, and even drugstores. About 71% of all shoes are sold in only 13 states, and 44% are sold in only 5 states (New York, California, Ohio, Illinois, and Pennsylvania). New York alone accounts for 14% of all shoe sales.

The annual dollar volume for the average shoe store is below $75,000. Shoe retailing is a relatively complex function, involving not only the basic inventory and operating requirements of any retail establishment but also consideration of fashion, foot health, fitting, and other specialized requirements. For this reason it is a higher-risk business than comparable retail stores in other fields. The seasonal turnover of styles, for instance, necessitates astute judgment to hold losses on wrong selections to a minimum. In women's shoes, careful selections must be made in terms of colours, heel heights, patterns, etc.

Shoe fitting is an extremely important function in the shoe store and requires experience and judgment combined with highly specialized knowledge and training; badly fitted shoes do not wear well, feel uncomfortable, and constitute a health hazard (*see* FOOT, DISEASES AND DISABILITIES OF). In addition to size (length and width), fitting involves the proper adaptation of shoe to foot in terms of leathers (some have more "yield" than others), lasts, heel heights, gait characteristics, etc. The store must carry a large and costly inventory to fit the vast variety of feet it must service, to say nothing of the customer's desires in style. A shoe factory may produce as many as 140 different sizes on a single style of shoe. A shoe store may carry as many as 65 sizes on a single style and have only one pair in each size. It would require duplication of sizes in the more popular ranges. A store may carry 50 to 100 different types and styles of shoes, and it can readily be seen that the problem of inventory and risk is extremely complex.

After 1945 there were some important changes in shoe retailing trends in the U.S. One change was in the locations of shoe stores, many of which are now situated in suburbs and shopping centres and even as "free-standing" stores on highways. A considerable amount of shoe business is done outside the downtown districts. Certain types of outlets formerly not associated with shoes have gone increasingly into the business. Among these are supermarkets, discount stores, variety stores, some drugstores, and various general merchandise stores. Another trend has been toward "self-service" shoe stores with interior layouts where customers can browse and shop. These are called self-selection layouts but are accompanied by fitting services.

Another change, particularly after the mid-1950s, was in the amount of footwear imported into the U.S. In 1954 leather shoe imports amounted to about 7,700,000 but by 1962 reached over 55,000,000 pairs. Rubber footwear imports showed an even more spectacular gain—from less than 1,000,000 pairs in 1954 to 100,000,000 pairs in 1961. These gains in footwear imports were expected to continue through the 1960s, though perhaps at a slower rate. Principal sources were the United Kingdom, Italy, Japan, and Hong Kong; except for rubber footwear, most of the imports were men's and women's shoes, chiefly in the lower price lines.

U.S. manufacturers have given only token effort to the exporting of shoes. For many years before and after World War II shoe exports remained around 4,000,000 pairs a year and by the early 1960s had dropped to around 3,000,000 pairs.

The U.S. shoe industry as a whole continues to show steady though rarely spectacular growth. It is known as a "stabilized" industry, not subject to sharp increases in times of prosperity or sharp declines in times of recession or slump. Its production and sales growth pattern is closely related to population growth and to a lesser extent to income growth, with some proportionate increase in consumer spending for shoes. (WI. A. R.)

Great Britain.—In the early 1960s there were some 1,200 shoe-manufacturing firms in the United Kingdom, but a third of these were very small, employing under 20 workers. As in the U.S., there was a tendency for manufacturing units to become fewer and larger, but shoemaking, as compared with other industries, was still one of small-size factories, only five employing more than 1,000 workers. Almost equal numbers of men and women were employed, and piecework was common.

After the U.S., Great Britain is one of the world's largest footwear producers. The main centres of production are Leicestershire and Northamptonshire, traditional shoemaking areas where originally cattle and sheep were plentiful (for hides and skins) and the country was well wooded with oak (for bark-tanning). Northampton is world renowned for its high grade of men's footwear, and Leicester has developed into a centre for medium-grade women's shoes. High-grade fashion shoes are made in Norwich and London, and these two centres also manufacture children's shoes. Southwest England, once the centre for heavy boots for miners and agricultural workers, has developed as the main centre for molded footwear and good-quality women's and children's shoes. The Rossendale Valley in Lancashire is noted for its slipper production. The tendency for larger manufacturing units accentuated the trend toward greater specialization as regards type and price of shoe produced in different firms and areas.

Although the labour force in the industry steadily decreased (in the early 1960s it was about 100,000 compared with 115,000 a decade earlier), improved techniques and machinery raised productivity, and after World War II total output rose steadily by an average of 3,000,000 pairs a year. In 1960 the industry produced nearly 160,000,000 pairs of leather footwear (*i.e.,* excluding rubber), of which some 40% was women's shoes while men's and children's shoes and slippers each accounted for about 20%.

With financial support from the government, the industry maintains an efficient research station (British Boot, Shoe and Allied Trades Research Association). Besides laboratory research it offers to its members consultancy services on practical problems and acts as independent arbiter on matters concerning consumer complaints.

Retailing and Marketing.—The end of the 1950s saw a spectacular rise in consumption of footwear in Great Britain, largely because of the growing "teen-age" market and a general increase in consumer expenditure on clothing and footwear. Speedier changes in fashion, particularly in women's shoes, also caused the buying public to spend more on shoes, but these changes brought new problems for manufacturers and retailers alike. Total consump-

tion of leather footwear in 1960 amounted to more than 160,000,000 pairs (or more than 3 pairs a person), and rubber footwear purchases made up a further 40,000,000 pairs, the overall value being £290,000,000.

Multiple retail organizations (covering 7,700 outlets) sold 45% of this footwear, and independent retailers (comprising 5,900 outlets), 26%. Cooperatives, department stores, mail-order and credit trading concerns accounted for 5–6% each. Some 20–25% of British footwear was retailed under manufacturers' brand names, but the great majority of shoemaking firms produced anonymously to orders from wholesalers and retailers.

Domestic production only partially satisfied the increased consumption, and imports of leather footwear rose from 2,500,000 pairs in 1955 to 14,500,000 in 1960, when one in every 11 pairs sold came from abroad. Most of these came from Europe, with Italy and France predominating, and were mainly women's shoes and slippers. Imports of rubber footwear also rose, with Hong Kong traditionally the largest supplier.

British footwear was exported to more than 100 overseas markets, and in 1960 some 9,000,000 pairs of leather footwear worth £12,000,000 were sent abroad, nearly half of them being men's shoes. The biggest single market was the U.S., taking practically one-third of Britain's footwear exports; one-half went to the Commonwealth and one-sixth to Europe. The Commonwealth countries, however, were a diminishing market for British shoes, as quantitative restrictions were imposed to protect the growth of the young, local footwear manufacturing industries. As old markets contracted, new ones opened up, and 1959 saw the start of a flow of British footwear exports to the U.S.S.R.

Western Europe.—Besides Great Britain, the main footwear producing countries in Europe are Germany, France, and Italy. Italy's output rose conspicuously during the 1950s, trebling itself between 1954 and 1959. Italy also became the leading European exporter of footwear, accounting for 40% of the continent's exports of leather footwear. This success was due to the development in Italy of an elongated, lightweight styling of shoes that proved extremely popular. All the European countries experienced a rise in consumption of footwear in the 1950s, and all, except Norway, increased their output. The emergence of the European Economic Community (EEC) and the European Free Trade Association (EFTA) led to greater trading in footwear between the countries concerned. On the retail side the main outlets were independent shops and department stores; multiple organizations were not developed to the degree they had in Great Britain, but buying groups began to be established by retailers. Both the EEC and the EFTA had their own liaison committees of manufacturers' organizations, and the Organization for Economic Cooperation and Development hides and skins committee compiled production and trading statistics on footwear in their member countries to encourage greater coordination of footwear production and marketing in this area. (P. G.-Sm.)

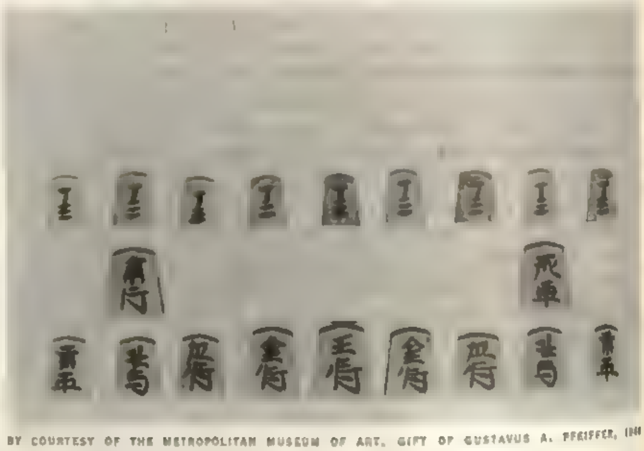
U.S.S.R.—Footwear is produced in large factories, usually with long runs on the same style. Workmanship and finish have often been poor, but in the 1960s research on automatic transfer lines and multioperation machines for mass-produced, cheap, serviceable footwear was being carried out. In 1964, 474,000,000 pairs of leather footwear were produced, utilitarian styles being sold with guaranteed lengths of life. In the early 1960s more than half the imported footwear was from Czechoslovakia and most of the remainder from other European countries. Smaller quantities were imported from India and the People's Republic of China.

See also references under "Shoes" in the Index. (H. E. L. F.)

SHOFAR, the ancient ram's horn of the Israelites, is a natural horn that has been bent upward at the wide end and partially flattened from end to end. It can be made to produce sobbing, wailing, and sustained sounds in sequences that are varied strictly according to ritual. In the Pentateuch the use of the shofar is prescribed for the New Moon and other festivals. In modern Jewish liturgy it is sounded in the synagogue on Rosh Hashana (New Year) and Yom Kippur (Day of Atonement) as a call to repentance and sacrifice, love of the Torah, and hope in the messianic future.

SHŌGI is a Japanese form of chess (*q.v.*), the history of which is obscure. Traditionally it is thought to have originated in India and to have been transmitted to Japan via China and Korea.

Shōgi, like chess, is played on a squared board with pieces of varying powers, and the ultimate object is the checkmate of the king. Two distinctive features, however, differentiate shōgi sharply from European chess: (1) captured men are not "dead" but may be replayed as part of the captor's forces; (2) there is no pawn chain or blocked pawn position. Each player has 20 men, which oppose each other on a board composed of nine horizontal and nine vertical rows. There are nine pawns (*fu*), a rook (*hisha*) and a bishop (*kaku*), one king (*ōshō*), two gold generals (*kinshō*), two silver generals (*ginshō*), two knights (*keima*) and two spearmen (*kyōsha*). From the standpoint of mobility, *hisha* and *kaku* roughly correspond in chess to the queen, *kyōsha* to the rook, *kinshō* and *ginshō* to the bishop; the others are the same in both games. Rook and bishop oppose each other diagonally, the rest on the same file. The pieces are flat, blunt and slightly tapering toward the front. Each bears identi-



BY COURTESY OF THE METROPOLITAN MUSEUM OF ART, GIFT OF GUSTAVUS A. PFEIFFER, 1948

WOODEN SHŌGI CHESSMEN: JAPANESE, 20TH CENTURY

fying characters. Captive men are held at reserve bases until they are needed on the field again, hence the number of pieces in play remains approximately the same from the beginning to the end of the game. All men except the king and gold generals are promoted on entering the last three rows of enemy territory.

See E. Ohara, *Japanese Chess: the Game of Shogi* (1958); Stewart Culin, *Games of the Orient* (1958). (E. D. S.)

SHOGUN, abbreviation of *sei-i-tai-shogun* (barbarian-subduing generalissimo), was one of several titles under which early Japanese emperors commissioned supreme military commanders for campaigns against the Ainu of northern Japan. The title first appeared in A.D. 720 but lost its original meaning with the pacification of the north country in the 9th century. Minamoto Yoritomo (1148–99) first used the title as a basis for asserting military and political authority over the country. Yoritomo won military hegemony of Japan in 1185 and acquired imperial appointment as shogun in 1192. Together with other inheritable high court ranks and titles it provided the legal foundation from which his military headquarters (*bakufu*) exercised control over all military families in Japan. As the military class acquired increased powers over national and local affairs, the shogun became *de facto* ruler of Japan, though the emperor continued to retain formal sovereignty. Yoritomo's successors as shogun ruled from Kamakura (1192–1333). The second line of shogunal succession, begun by the Ashikaga family, ruled from Kyōto (1338–1573). The third line, established by the Tokugawa house, used Edo as its capital (1603–1867). With the intrusion of western powers into Japanese affairs in the 19th century, the question of definition of powers between the shogun (called "tycoon") and emperor (mikado) became an issue. In 1867 the Tokugawa shogun resigned his title and "returned" his prerogatives of civil and military administration to the emperor.

See also Japan: *History*. (J. W. H.)

SHOLAPUR, a town and district in Maharashtra state (till April 30, 1960, in Bombay state), India, lies 140 mi. S.E. of Poona by road. Pop. (1961) 337,583. The town has two arts colleges and a college of commerce affiliated to Poona university. Its chief industry is the manufacture of cotton cloth in textile mills and on power and hand looms. Sholapur's convenient situation between Poona and Hyderabad, with which it is connected by the Central railway, has made it the centre for the collection and distribution of goods over a large area.

SHOLAPUR DISTRICT comprises an area of 5,809 sq.mi. with a pop. (1961) of 1,860,119. The population, predominantly rural, is engaged mainly in agriculture. The district is situated in the middle of an extensive plain 1,800 ft. above sea level and has a generally flat or undulating terrain. The main rivers are the Bhima and its tributaries, the Man, the Nira and the Seena (Sina). Sholapur is an important jowar-producing district. Paddy and bajra are the other principal food crops, although peanuts and cotton are slowly gaining in importance. The chief industries are spinning, weaving and dyeing. The main trade centre for cotton, oilseed and oil is Barsi (pop. 50,389), 37 mi. N. of Sholapur.

Pandharpur (*q.v.*), 33 mi. W. of Sholapur, is a famous place of pilgrimage and a popular centre for the worship of the god Vithoba, believed to be an incarnation of Vishnu. It has association with the great poet-saints of Maharashtra devoted to the Bhakti cult. Places of interest are the Hemadpanti temples at Barsi, Mohol, Malsiras, Pandharpur and elsewhere. There is also the tomb of a daughter of the emperor Aurangzeb at the village of Begampur. (M. R. P.)

SHOLEM ALEICHEM, pseudonym of SOLOMON RABINOWITZ (1859–1916), popular Yiddish classical author. Born in Pereyaslav, in the Poltava district of the Ukraine, on Feb. 18, 1859, he was educated for the rabbinate at nearby Varonkov. Drawn to writing as a youth, he became a private tutor of Russian at the age of 17, and at 18 went to the estate of Elimelech Loeff near Kiev to tutor his 12-year-old daughter Olga. He fell in love with her, an event that cost him his position. He served as a government rabbi in Lubin and when he was 21 married Olga. Her father's death in 1885 brought him a considerable fortune. His first writing had been in Russian and Hebrew, but between 1883, when his first story in Yiddish appeared, and his death he published over 40 volumes of novels, stories and plays in Yiddish. He used part of his fortune to encourage Yiddish writers and edited the annual, *Die Yiddishe Folksbibliotek* (1888–89); he lost the rest of it in business. His works were widely translated, and he became known in the United States as the "Jewish Mark Twain." He left Russia in 1905, established his family in Switzerland, and lectured in Europe and United States. At the outbreak of World War I he went to New York city where he died on May 13, 1916.

English translations from his *Verk*, 14 volumes (1908–14), include *Jewish Children*, translated by Hannah Berman, 3rd edition (1937); *The Old Country*, trans. by Julius and Frances Butwin, 3rd edition (1954); and *Adventures of Mottel, the Cantor's Son*, translated by Tamara Kahana (1953). He was the first to write in Yiddish for children. Adaptations of his work were important in the founding of the Yiddish Art theatre in New York.

See Maurice Samuel, *The World of Sholom Aleichem* (1943).

SHOLOKHOV, MIKHAIL ALEKSANDROVICH (1905–), Russian novelist, winner of the 1965 Nobel Prize in Literature for his "artistic strength and honesty when depicting a historical epoch in the life of the Russian people." He was born on May 24 (new style; May 11, old style), 1905, on a farm (Kruzhilin) in Veshenskaya, a Cossack village in the Don region of southern Russia. After joining the Red Army in 1920 and spending two years in Moscow, he returned in 1924 to his native village. He made several trips to Western Europe, and in 1959 accompanied Khrushchev to the United States.

Sholokhov began writing at 17, his first published book being *Donskie rasskazy* (1926; Eng. trans., *Tales of the Don*, 1961), a collection of realistic short stories. In 1925 he began his famous novel *Tikhi Don* ("The Silent Don"). The slow maturing of his work is remarkable: it took him 15 years to write *Tikhi Don* (four volumes, 1928–40; translated in two parts as *And Quiet Flows the Don*, 1934, and *The Don Flows Home to the Sea*, 1940) and 28 years to complete his other major novel, *Podnyataya tselina* (1932–60; translated in two parts as *Virgin Soil Upturned* [U.S. title, *Seeds of Tomorrow*], 1935, and *Harvest on the Don*, 1960). *Tikhi Don*, about the Don Cossacks before and after the Revolution, became the most widely read novel in the U.S.S.R.

See I. Lezhnev, *Mikhail Sholokhov* (1948); E. J. Simmons, *Russian Fiction and Soviet Ideology: Introduction to Fedin, Leonov, and Sholokhov* (1958).

SHONA is the collective name of more than 1,000,000 (in the 1960s) culturally similar Bantu-speakers of Negroid-Armenoid origin primarily inhabiting the eastern half of Southern Rhodesia north of the Lundi river (*see* BANTU LANGUAGES). Main Shona tribal groupings are the Zezuru, Karanga, Manyika, Tonga-Korekore and Ndau, the last three overlapping neighbouring territories. The leading tribe previous to European arrival was the Rozwi, associated with the Zimbabwe culture (*see* ZIMBABWE).

The Shona are farmers of millet and maize (corn), and occasionally hunt. Their cattle and other stock are important for marriage and ritual purposes. Their villages consist of clustered mud huts, granaries and common cattle corrals and accommodate one or more interrelated families, to a total of perhaps several score persons. Modern legislation is causing them to abandon the habit of shifting their village sites when the land is exhausted. Shona traditional culture, now fast declining, shows excellent ironwork, good pottery, but unimaginative basket, skin and woodwork. They are musically gifted, and their expert drumming and playing on the *mbira* (a flat hand instrument with curved, tuned iron bars), to accompany a rich variety of dances for old and young, are as much a part of ritual as of leisure time.

Shona personal and political relations are largely governed by a classificatory kinship system with exogamous clans (*mutupo*) and corporate patrilineages (*rudzi*). Chiefdom (*nyika*), ward (*dunhu*) and village (*musha*) are the territorial units, under hereditary leadership. Marriage is normally validated by *roworo* (cattle or cash) given by the groom in exchange for the bride. Polygyny and levirate (*qq.v.*) are allowed; but the position of women is fairly secure and they may hold and dispose of property in their own right.

Shona religious practices are concerned less with the creator-god Mwari than with the propitiation of ancestral, tribal and other minor spirits to ensure good health, rain and success in enterprise. Elementary European education, Christian missions and partial urbanization, while weakening traditional institutions and leadership, have barely affected the belief in magic, witchcraft and sorcery.

See also AFRICA: *Ethnography (Anthropology): Southern Africa.*

BIBLIOGRAPHY.—C. Bullock, *The Mashona and the Matabele*, 2nd ed. (1950); M. Gelfand, *Shona Ritual* (1959), *Shona Religion* (1962); J. F. Holleman, *Shona Customary Law* (1952), *Accommodating the Spirit* (1952); H. Kuper, *The Shona* (1955). (Jo. F. H.)

SHOOTING, as described in this article, is the sport of discharging firearms at targets as a test of marksmanship. Three basic types of firearm are used in this sport—rifles (including modern rifles, muzzle-loading rifles, and benchrest rifles), shotguns, and handguns (pistols and revolvers). (*See* GUNS, SPORTING AND TARGET.) Two of these are also the most commonly used hunting weapons—the modern rifle for shooting game animals and the shotgun for small game and wing shooting (shooting game birds in flight). The others are primarily target firearms, with the exception of some light benchrest rifles which are also used for shooting vermin (animals harmful to agriculture) at long ranges. For the sport of hunting with firearms and the shooting of game birds *see* HUNTING.

RIFLE SHOOTING

MODERN RIFLE

Firing Technique.—When mounted in shooting position a modern rifle is held with the heel of the stock (butt) forced firmly against the shooter's shoulder, which absorbs the recoil or "kick" of the weapon. When the weapon is fitted with iron sights, the

head is held low with the cheek pressed against the stock, so the eye can align the sights of the target; with telescopic sights, the head is necessarily held higher. Either shoulder can be used, but right-shoulder shooting is the most common. A few bolt-action rifles are made with the bolt handle, etc., on the left side for left-handed shooters who find it too awkward to operate the conventional kind. The right-handed shooter supports the fore-end of the rifle on the palm of his extended left hand with his left elbow usually directly below the barrel. It is this hand that directs the rifle in aiming it. At the same time the right hand grasps the grip, or "small," of the stock behind the trigger guard and pulls the rifle back against the shoulder while the tip joint of the forefinger is crooked around the trigger. The rifle is fired by squeezing the trigger rearward with a steadily increasing pressure, not by pulling or jerking it, since a sudden or violent motion can swing the sights from the target. To increase steadiness, the shooter holds his breath when firing.

The Shooting Eye.—The shooter's stronger (dominant) eye, his shooting eye, is used for best accuracy in sighting. To determine his shooting eye, the shooter holds the forefinger of one hand before him at arm's length and, with both eyes open, aligns its tip with a distant object. Then, without moving his finger, he closes his left eye. If his fingertip remains aligned with the object, his right eye is dominant and is his shooting eye. If his fingertip appears to shift from the object, he may have to use his left eye for sighting.

Positions.—Four specific body positions are used: offhand (erect), kneeling, sitting, and prone.

Offhand is the standing position common in hunting because it is the most quickly assumed and elevates the rifle above anything that might obstruct it in the other, lower positions. In the classic position, often modified, the shooter stands with his left side toward the target (when shooting from the right shoulder), his feet planted firmly about 18 in. (46 cm.) apart and his right foot to the rear, in line with his body and the target.

The kneeling position permits greater steadiness for shots at longer distances. The shooter sits on the side of his right foot or ankle with his right knee pointed at a 45° angle to the right of the target. His left elbow is rested on his left knee which is raised and pointed toward the target. Both this knee and elbow are directly below the rifle barrel and the body is inclined slightly forward for balance.

In the sitting position, which provides even steadier sighting, the shooter sits at an angle of about 45° to the target, his knees raised and his heels dug into the ground. His elbows or upper arms rest on his knees or slightly below them, with his left elbow and knee directly below the rifle barrel. In a variation of this position, the legs are crossed below the knees.

The prone position provides the steadiest sighting. To assume this position the shooter lies on his stomach with his body at a 45° angle to the target, both elbows resting on the ground with the left one directly below the barrel. The legs are spread wide apart with the toes turned outward and the ankles flat on the ground. This again is the classic position, which is greatly modified by many modern marksmen.

Muzzle-Loading Rifle

Although virtually obsolete since the American Civil War, the muzzle-loading rifle still has much romantic interest for many modern shooters. The American design for the "Kentucky," or long, rifle was developed in the 18th century by immigrant German and Swiss gunsmiths of Pennsylvania ("Kentucky" at that time being the name usually given to the largely uncharted hunting wilderness west and southwest of Pennsylvania). The Kentucky rifle has been credited with helping to win the American Revolution because its long-range accuracy was superior to that of the British "Brown Bess" smoothbore musket. Indeed, some British records of the war refer to the Kentucky rifle as an unethical secret weapon that could kill farther than the eye could see. Actually, its effective range was only about 300 yd. (274 m.) but this was nearly three times that of the Brown Bess. It was not used in large numbers by the Americans.

The muzzle-loader is held and fired like any rifle from the conventional positions, with one interesting addition in which the shooter lies on his back, resting on one elbow, and supports the barrel between his feet. This position (and numerous variations), made possible by the barrel's extreme length (40 in. [101.6 cm.] or more), came into vogue in the 19th century. The similar dimensions of the rifles used by some Oriental tribesmen, notably in the Pakistan-Afghanistan-Iran area, permit the barrel to be supported at about the point of the front sight between the shooter's toes. High accuracy results.

Benchrest Rifles

The benchrest rifle is fired in competition from a sandbag rest on a solid platform called a benchrest. These competitions are relatively new and largely confined to the United States, having been organized on a national basis with national competition in 1948. Participants are shooters who seek extreme accuracy, the ideal being a number of shots passing through the same hole in the target (*see* Guns, Sporting and Target: *Modern Target Firearms: Rifles*).

Shooting from a rest of one kind or another goes back at least to the late 18th century in colonial America. A later and highly organized phase occurred about 1840–90, chiefly in New England, using heavy percussion cap rifles, some weighing 30–40 lb.

SHOTGUN SHOOTING

Gun Handling.—The shotgun is mounted to the shoulder in somewhat the same manner as the rifle and is fired in the same way, the left hand (for right-handed shooters) supporting and directing the barrel and the right hand forcing the gun against the shoulder and squeezing the trigger. However, only the offhand body position is used since the shotgun is fired at flying targets (*see* Trapshooting and Skeet Shooting). (In the field, there are exceptions: sometimes single bullets called rifled slugs, instead of scatter shot, are used in the shotgun for hunting large animals such as deer, and the shotgun is sighted like a rifle; sometimes wildfowl are shot by sportsmen sitting or lying in boats.) Because the shotgun is pointed rather than sighted, it usually has only a single, bead-type front sight; but some shotguns also have a sight in the middle of the barrel. When a target flies in front of the shooter, he brings the gun to his shoulder as in the offhand position—though facing more toward the target than in rifle shooting—swings the gun's muzzle toward the target, and squeezes the trigger.

Swinging.—Proper swinging of the gun muzzle is a requisite for successful wing shooting since the charge of shot must be fired ahead of (lead) the flying clay pigeon to allow for the movement of the target, if it is to score a hit. One method is to swing the muzzle from behind the bird and through it to a position ahead of it, at which point the gun is fired. The swing is continued even beyond the point of firing with the follow-through motion in the direction in which the bird is flying. Moreover, a hit is more probable when the lead distance seems too great rather than too small because the shot pellets do not travel in a solid mass but form a long line called a shot string and, should the leading pellets pass ahead, those following them will probably hit the target.

Fit.—Another requirement for accurate shooting is proper fit of the shotgun. Since the gun must be mounted and fired quickly the shape and dimensions of its stock should be such that it can be mounted in perfect shooting position every time with no conscious adjustment of the shooter's body or head. When the gun is raised quickly into firing position, the shooter's cheek pressed against the stock, his eye should be looking directly down the top of the barrel in alignment with the front bead sight as he points the gun and swings on the target. To test his shotgun's fit, the shooter mounts the gun quickly with both eyes closed, then opens them. If his shooting eye is not in perfect alignment as described above, the gunstock should probably be altered.

HANDGUN SHOOTING

Stance.—The shooter's stance or body position is as important in handgun shooting as in other types but is not rigidly defined

and varies more with the individual. Each shooter can determine the one which suits him best by facing the target with his feet together and arms at his sides, moving the left foot outward from 12 to 15 in. (in the case of a right-handed shooter), raising the pistol to arm's length, and aiming at the target. If the arm feels strained, the left foot should be shifted backward until the strained feeling is eliminated. Once this most comfortable body angle is determined, it should be used at all times.

Grip.—The shooter's grip on the pistol (or revolver) may vary according to size of his palm and length of fingers. Sometimes the handgun must be fitted with a special stock, although this is uncommon except among competition shooters. In general, the stock should be cradled in the hollow between the thumb and fingers with the palm flat against the side of the stock, the thumb resting along the opposite side of the frame, and the remaining three fingers curled around the front of the stock under or behind the trigger guard. When the weapon is held correctly, the barrel, the wrist, and the forearm are in a direct line and the forefinger is in direct forward alignment with the trigger so that its only possible direction of motion is rearward. Only the tip joint of the forefinger should bend across the trigger. All fingers should exert firm and equal pressure on the weapon to hold it. The thumb, however, should exert no pressure in any direction. To keep the sights aligned on the target this equal finger-pressure must be maintained while the forefinger gradually squeezes the trigger rearward for firing. For accurate, consistent shooting the handgun must be held the same way each time and with the same pressure. The thumb is not used in handgun shooting except to cock the hammer of a revolver. In resting between shots during slow-fire shooting the handgun is lowered to rest the shooter's arm, but to maintain the grip for accuracy the positions of the fingers on the gun are not changed until an entire five-shot group has been fired.

COMPETITION

Competition with firearms is conducted on many levels, from the intramural matches held periodically by local shooting clubs or police forces to the Olympic Games, which draw the finest shooters from all over the world. Even the most stay-at-home shooter, through postal matches, can often engage in national and international competitions, firing at a local range and mailing his targets and scores to judges who compare them with results obtained by other shooters firing under similar conditions.

Although most shooting competitions are designed to test the marksmanship of the shooter, in some, such as the modern turkey shoot (usually a rifle contest based on a U.S. folk tradition) and in benchrest competition, the design or accuracy of the weapon may be an important, even decisive, factor. In other types of competition, accuracy may be secondary to some other point of judging. This section, however, deals primarily with formal competitions with rifles, pistols, and shotguns.

Rifle

Modern Rifle.—In the United States, rifle-shooting competitions are held throughout the year by various local (nonmilitary) shooting clubs and organizations under the sponsorship of the National Rifle Association of America (NRA) and by the branches of the armed services. The culmination of such competitions—to some extent they may be considered elimination trials—are the annual national matches, usually held during August, at Camp Perry, O. These events (which include handgun competitions) include the Small Arms Firing schools, the National Trophy matches, and the NRA championship matches. They are the responsibility of the secretary of the army and are under the supervision of the National Board for the Promotion of Rifle Practice (NBPRP). The NRA, which represents the great body of civilian shooters and whose championships are an integral part of the national matches, acts in an advisory capacity on policy matters and cooperates with the national match director and the NBPRP in scheduling and administering the matches. The rifle-shooting portion of the national matches includes competition with the small-bore rifle (.22 calibre) and high-power rifle (.30 calibre).

In the NRA championships, small-bore competition is divided into civilian, service (including both regular and reserve), and junior categories while high-power competition with service rifles includes civilian, regular service, and reserve categories. Competition with the high-power bolt (sporting) rifle is technically for civilians only, although military personnel may compete in this category.

The National Trophy rifle matches, conducted by the NBPRP, are fired with high-power service rifles. Competitors are divided into regular service, reserve, National Guard, ROTC (Reserve Officers' Training Corps), and civilian categories.

Each category in the national matches is further subdivided into various classes, based on previous performance, in order to equalize competition. There are awards for the highest scores in each class. Special awards also are made to ranking women, collegiate, senior (over 55 years of age), junior, and National Guard shooters. Rifle competition in the national matches is conducted at ranges varying from 50 to 100 yd. (45.72 to 91.44 m.; small-bore) or 200 to 1,000 yd. (182.88 to 914.4 m.; high-power). The type of sight and the shooting positions employed vary with the requirements and limitations of each match.

In Great Britain high-power rifle shooting and pistol shooting are organized and administered by the British National Rifle Association, founded in 1860. Similar to the U.S. national matches—with schools and competition for civilian and military shooters—is the annual Imperial meeting at Bisley Camp, Surrey, to which all countries of the Commonwealth may send competitors. The winner of the Queen's Prize, the chief individual award, wins not only a cup and honours but is traditionally seated in an armchair that is carried to the pavilion on the shoulders of his rivals.

Small-bore rifle programs in Great Britain are administered by county rifle associations and the National Small-Bore Rifle Association. Indoor ranges (15–25 yd. [13.72–22.86 m.]) or open ranges (up to 200 yd.) are used for competitions and there is an annual meet at Bisley.

In Scotland, at Barry, dedicated amateurs of professional skill fire at a range of 2,100 yd. (1,920.24 m.), hitting a 30-in. (76.2-cm.) bull's-eye. In this shooting the supine position (like that in 19th-century U.S. competition, and variously called "Texas," "Creedmoor," and "Fulton") is often used. Barrels of exceptional length (over 30 in.) are sometimes fitted to standard bolt-action rifles, and rear sights are placed on the heel of the stock rather than near the breech, to bring them closer to the shooting eye.

In other countries national shooting associations conduct special programs which may be colourful or of popular interest. In Sweden, "combat shoots" are a practical feature of paramilitary emphasis in the civilian rifle program. In this competition five-man teams, consisting of two riflemen, two submachine gunners, and an automatic rifleman, patrol through an area where they are unexpectedly confronted with silhouette targets showing enemy soldiers firing at the competitors. A "running dog" target also is used. Similar events are held with the shooters traveling on skis, as, for example, the biathlon in the winter Olympic Games (*see* Skiing).

In Switzerland the national shooting matches are held every four years, a different canton playing host each time. These shoots are noted for their pageantry and are well attended (over 50,000 at some shoots) because each male citizen, after completing military training, is given a military rifle with which he must pass annual qualifying competitions and inspections. In Italy the *carabinièri* conduct regular competitions with military arms.

Probably the largest shooting club in the world is the DOSAAF of the U.S.S.R. It is estimated that several million members have a basic familiarity with small-arms shooting, many of whom have qualified as "Voroshilov marksmen," skilled with rifle, pistol, and light machine gun. Soviet successes in Olympic and other international events are ascribed to their having so vast a training pool from which to draw. Rifle and other shooting sports are similarly encouraged in other Communist countries.

In the Federal Republic of Germany small-arms competition and training was de-emphasized after World War II, but com-

petitions are still carried on at ranges established by the *Sturmabteilung* (SA) between 1933 and 1945. Many of these ranges were taken over by units of the U.S. armed forces in the Association of American Rod and Gun Clubs in Europe, comprising about 100 clubs in Germany and about 30 more in other parts of Europe, with more than 30,000 members engaging in off-duty recreational shooting. Club charters were drawn up to admit German nationals to membership. Many club grounds formerly belonged to the old *Schützenbund,* for centuries social as well as sports groups in German life. Traditional German shooting events include running boar and deer, two quick shots at crossing targets at 100 m. Sometimes a poacher target is used, tracing back to the social upheavals of the Revolution of 1848; preserving the game and the hunting privileges was symbolic of stability in German society. In Austria, shooting sports follow the German *Schützenbund* system, with rules and regulations of the International Shooting Union observed in major competition.

Muzzle-Loading Rifles.—Current competition among muzzle-loading rifle shooters in the United States is a revival of the traditional American frontier sport of shooting at targets for prizes of money or food. The so-called Kentucky riflemen of early U.S. history and their feats of marksmanship with rifled muzzle-loaders are legendary. In exhibitions, one rifleman would shoot at a small board held between the knees of another at 60 yd. (54.86 m.). In the original version of the turkey shoot, a live turkey would be tied behind a rock some distance away, and competing riflemen would try to shoot off its projecting head.

The revival of target-shooting matches with these old firearms began in the 1920s. The National Muzzle Loading Rifle Association (NMLRA) was organized in 1937 and the sport has since grown steadily. Many of the rifles are originals, preserved by antique collectors or gathered from dusty attics and closets. Some have been reconditioned and the barrels of many have been relined with new rifling to restore their original accuracy. Flintlock-ignition types are rare and the majority are percussion-cap models. In competition either peep sights or open sights are used for the most part, the latter ordinarily being a square-post front sight with a square-notch rear sight. Some telescopic-sight matches also are held. Frequently sight shades are added to prevent glare from the sun.

Since the round lead ball fired by the muzzle-loader is ballistically poor, hence sensitive to crosswinds, large calibres from .45 to .55 are preferred in competition since their heavier weights have better crosswind resistance. Shooting matches test not only the accuracy of the individual rifle and the shooter's marksmanship but also his skill in loading the powder, patch, and ball and in judging the effect of the prevailing wind. Tournament competition includes open-sight matches at 50 and 100 yd., peep-sight matches at the same distances, and offhand matches at 25, 50, and 100 yd. Other competitions include matches in which the rifle is fired from a benchrest at 100 and 200 yd., muzzle-loading pistol matches, and shotgun matches.

The NMLRA conducts three large matches annually at Friendship, Ind.—a spring shoot, a fall shoot (national matches), and a turkey shoot in October. Competition is open to men, women, and junior shooters. Contestants frequently dress in old-style Kentucky rifleman costumes.

Benchrest Rifles.—The perfect target score in benchrest competition, with every bullet passing exactly through the same hole, may never be achieved because of certain uncontrollable factors such as wind drift and mirage. The latter, in this context, is a shimmering of the target image when seen through the rifle sights, caused by heat waves rising between eye and target or from the hot barrel. Many shooters, however, attain near-perfect scores in which all five bullets in a round pass almost exactly through the same hole in the target, with a spread as small as the cross section of a lead pencil. For this reason a paper strip, driven horizontally by a motor, moves slowly behind each target to prove that a complete round has been fired. Thus five shots may show on a target as a single ragged hole, but on the moving paper-backing they are strung out in a horizontal line.

In the United States two benchrest rifle championship matches are conducted annually under sponsorship of the National Bench Rest Shooters Association (NBRSA). In firing the national match course in the unrestricted category, any firearm may be used, even the most unconventional. This competition usually consists of ten 10-shot matches at 100 yd. and ten 10-shot matches at 200 yd., shot over several days. The aggregate of these matches determines the national match course grand aggregate winner. At these matches, shooters may also compete for separate awards with conventional rifles. The other national benchrest course (in the conventional rifle category), held separately, usually consists of five 5-shot groups at 100 and 200 yd. in each of three rifle classes—heavy varminter (total weight under 13.5 lb. [6.12 kg.]), light varminter (limited to 10.5 lb. [4.76 kg.] and a 16-power telescopic sight), and sporter (limited to 10.5 pounds and an 8-power telescopic sight). Championships are awarded the shooters with the smallest aggregate group size in each division.

Shotgun

Games.—Some current forms of shotgun-shooting competition in local U.S. clubs can be classified as games. One of these is the turkey shoot in which each competitor shoots at a small stationary square of cardboard at a specified distance. The shooter placing the most shot pellets in his card wins the turkey. This is really a test of the ability of the shotgun to fire a dense pattern of shot since such a pattern will normally score more hits than a sparse pattern. Similar games involve shooting at playing cards.

Clay Pigeons.—Since the shotgun is primarily a wing-shooting firearm, the most popular competition is shooting at targets in flight. These targets are circular disks of clay called clay pigeons because they have generally replaced the live pigeons which were formerly widely used in America and are still popular in parts of Europe and Latin America. Live-pigeon shoots, in which thousands of pen-reared birds were flighted and slaughtered in a single match, were curtailed in the U.S. and many other nations in the early 20th century because of the cruel and wasteful aspect of the sport. In the Principality of Monaco, live-pigeon shoots are a featured competition at the Casino Shooting Field, which is also the site of international skeet and trap clay-pigeon events.

Clay pigeons are usually thrown or scaled through the air to resemble flying birds, although in a variation of this method they are projected low across flat ground to simulate running rabbits and other small game animals which are also hunted with shotguns. The targets are propelled from "traps," named after the device from which the live pigeons were formerly released to be shot. These traps vary from simple hand-swung instruments to elaborate self-loading, electrically operated, and remotely controlled machines that can throw several clay pigeons at one time.

One type of clay-pigeon shoot is the grouse or quail walk, designed to simulate such bird shooting. The shooter walks along a woodland path and shoots both singles and doubles (two targets thrown at the same time) which are released unexpectedly from hidden traps.

In Great Britain, traditional tests chiefly governed by simulation of natural targets include driven partridge (shooter stationed behind opaque fence and targets are flung in pairs without warning from the other side), high pheasant (target flung in pairs from a 90-ft. [27.43-m.] tower), springing teal (targets flung vertically from hidden trap at ground level), running rabbit (target bowled along ground), and many others. However, trapshooting and skeet shooting are the most widely practised clay-pigeon sports, with both national and international competition.

For history and description of these sports *see* Trapshooting and Skeet Shooting.

Handgun

The sporting use of the handgun is largely confined to target shooting since this type of firearm has inherent disadvantages as a hunting weapon. Its relatively low-power ammunition limits its range and its short barrel limits its accuracy at longer ranges. Further, the relatively short distance between the front and rear sights (the sighting radius) makes it more difficult than a rifle to aim precisely. Some high-power (magnum) handguns equipped

with a telescopic sight, however, make fairly effective hunting weapons.

As with rifle shooting, the NRA National Pistol championships and the National Trophy Pistol matches, both part of the U.S. national matches at Camp Perry, are the culmination of local and regional competitions held throughout the year.

The NRA championships are conducted with all calibres of automatics (autoloading pistols), classified as .22-calibre, .45-calibre, and centre-fire (other calibres). Competitor categories are regular service, civilian, reserve, and police. These are further divided into classes and there are class and category awards as well as the national championship which goes to the shooter with the highest overall score.

The National Trophy matches, conducted by the NBPRP, are fired with the calibre-.45 pistol by competitors in regular service, reserve, National Guard, police, ROTC, and civilian categories.

Slow-fire pistol competition at the national matches is conducted at a range of 50 yd. Timed-fire (five shots in 20 sec.) and rapid-fire (five shots in 10 sec.) strings are fired at a range of 25 yd.

Muzzle-loading pistol matches, at 25 and 50 yd., are conducted by the National Muzzle Loading Rifle Association.

Police Shooting.—In almost every country where police are armed with handguns or, for special assignments, other weapons such as shotguns (riot guns) or submachine guns, training programs have been set up for the use of these weapons. As a result, many police officers compete in local, national, and international sporting matches, both open competition and matches for police only.

Police marksmanship training is divided, not always formally, into two categories. The first of these is normal target shooting at bull's-eye targets to sharpen marksmanship ability and familiarize the shooter with his weapon. Silhouette man targets are also used with the vital parts of the body outlined on them.

The second category, shooting at targets representing criminals, is often called "combat" or "practical" shooting. This is designed to test the officer's reflexes and ability to draw, aim, or point, and fire his weapon accurately under various conditions. In the United States practical shooting is conducted usually at ranges from 7 to 60 yd. (6.4 to 54.86 m.) and from various positions (standing, crouched, sitting, kneeling, or prone) with the shooter using either or both hands and firing in the open or from behind a barricade. In Sweden a "running thief" target is used. Hits on this target above the legs count against the shooter as it is his job to capture a criminal, not kill him. In other variations used in the U.S., motion pictures of running men are fired upon or surprise pop-up targets appear singly or in groups as the shooter moves along a firing line. In this latter form of training the shooter may be penalized if he fires upon a target representing an unarmed man or an armed man using another person as a shield. Further, in the case of more than one target, the shooter must decide almost instantaneously which should be fired upon first, *e.g.*, one representing a criminal with a drawn and aimed gun as opposed to another in the group showing a criminal who is just drawing his gun.

In this type of shooting, often called "instinct" shooting, there is little or no time for aiming. Instead, the shooter simply points the gun at the target and fires. Practice can lead to a high degree of accuracy in this type of shooting.

International

The heritage of international rifle shooting is strongly Germanic, hence the weapons used, particularly in the "free" categories have some semblance to traditional German target arms. For the offhand position deeply curved butt plates with long underarm hooks or prongs are fitted to muzzle-heavy rifles. Trigger pulls adjustable down to a few ounces are permitted and some rifles have miniature roller sunshades which fasten near the front sight and can be pulled back to the breech. Nonreflective webbing or fabric is placed over the barrel to break up heat waves and avoid "mirage." (*See* also Guns, Sporting and Target: *Modern Target Firearms: Free or Unrestricted Firearms.*)

World Championships.—The world championship shooting matches are held in a different country each year under the auspices of the International Shooting Union (ISU), which was founded in 1907 and reorganized after World War I. Competitions are with the small-bore rifle, free rifle, centre-fire pistol, free pistol, rapid-fire pistol (.22-calibre), and shotgun. The army rifle match and the national match are fired with a military weapon designated by the ISU.

The small-bore rifle competition consists of firing from three positions (standing, prone, kneeling) at a range of 50 m. Three-position matches are held at 300 m. in the free-rifle (unrestricted as to rifle design and accessories) and army-rifle competitions.

Free-pistol matches are conducted at 50 m. while centre-fire and rapid-fire competition is at 25 m.

Running-deer competition is conducted at 100 m. with the shooters firing at rapidly moving silhouette targets. Clay-pigeon (trapshooting) and skeet matches are also conducted.

NATO.—Among the nations of the North Atlantic Treaty Organization an annual military shoot is conducted during the summer. Although the competitors are full-time military personnel on duty assignment the program of the Prix Leclerc, as the matches are called, has taken on the characteristics of a sporting event. Special matches for light-machine-gun and submachine-gun shooters, however, leave no doubt that this is basically a military exercise.

Olympic Games.—Shooting has been a recognized Olympic sport since the first modern Olympic Games in 1896. In addition to the more regular events of shooting competition, however, Olympic shooters have also taken part in such events as live-pigeon shooting and competition with dueling pistols. In the winter Olympic Games the biathlon is a combination of cross-country skiing and rifle shooting. *See* Olympic Games.

AIDS TO ACCURACY

With hunting activity being limited by shorter seasons and diminishing game populations, target competition among modern hunting-rifle shooters has shown a rapid increase. This has stimulated attempts to obtain a degree of accuracy from both firearms and ammunition far beyond that needed in hunting. One result has been an increased production of extremely precise sighting equipment and accessories which the shooter can add to his basic rifle to increase his accuracy. Another result has been a growing interest in handloading, the refilling of fired rifle-handgun cases with more exact weights of powder and bullet than are obtainable in some production-line ammunition, in order to obtain the consistent ballistic performance needed for near-perfect accuracy. An increasing number of shotgunners are also handloading their fired shotshells, but the principal reason is economy rather than precision since empty shells can be reloaded for less than half the cost of new ammunition.

Shooting Accessories.—Since rifles are usually delivered from the factory with a minimum of equipment, shooters often install accessories to help them handle the weapon more efficiently and shoot more accurately. Factory-installed open rear sights are simple and cheap but optically inefficient. In most cases they are difficult to adjust precisely for windage (lateral adjustment to compensate for the wind) or for elevation (vertical adjustment for different distances) and are usually replaced by adjustable peep sights or telescopic sights. The peep sight is preferred over the telescope for shooting animals, especially running animals, at distances up to 100 yd. because it provides a wider field of vision. The telescopic sight is better adapted to long-range shooting at both stationary animals and targets. Since shotgun sights are not as critical, they are less frequently replaced. For more accurate handgun shooting an adjustable rear sight is frequently added; and many handguns are factory equipped with such sights.

For shooting with a rifle, an adjustable leather or web sling is often used. One end is fastened to the rifle's fore-end, the other to the stock. By looping the sling tightly above the elbow of his aiming arm, the shooter can steady the gun for more accurate sighting. In both shotguns and high-power rifles a rubber recoil pad, fitted to the butt of the stock, protects the shooter's shoulder from excessive recoil and helps prevent flinch, an involuntary reaction at the moment of firing which pulls the sights

from the target. Earplugs can also help prevent flinch. For single-barrel shotguns, variable choking devices which fit over the muzzle are often used to vary the spread of shot pellets. Gas vents in some of these devices also help to reduce recoil. In match handguns, special accessory triggers and fitted handgrips are used. Some factory-supplied grips are replaced by grips of carved ivory or exotic woods. A common accessory for target shooters is the spotting scope, a powerful, tripod-mounted telescope (up to 40 power) used for determining the positions of bullet holes in targets at long ranges.

Handloading.—The reloading of fired centre-fire brass cartridge cases by hand (rimfire cartridges are seldom reloaded) and of fired shotgun shells increased in popularity after 1950 with the availability of efficient and inexpensive hand-operated reloading tools, called presses. In cartridge reloading for both rifle and handgun the press ejects the fired primer from the empty case and simultaneously forces the case into a resizing die to restore the stretched brass to its original shape and to adjust the case mouth to receive the bullet. Another device on the tool inserts the new primer. Then the shooter measures or weighs the correct amount of powder necessary, determined from a chart, and fills each primed case. Finally the press inserts the bullet in the case neck. A shotgun-shell reloading press operates similarly: depriming, resizing, and repriming. Then the powder is added and the gas-sealing wads, forced against the powder at a predetermined pressure by the press, are inserted. Next a measuring device on the press meters the desired weight of shot into the shell and another device applies the final crimp to seal it.

Handloading is safe when simple rules are followed. Smoking is prohibited during the process. All cases or shells to be reloaded must be in good condition. Primers must be handled carefully since they can be detonated by a sharp blow. Special care must be exercised to use only the exact amounts of the powders recommended by the chart which accompanies all reloading handbooks. Maximum powder loads should be used with caution.

SAFETY

Statistically, a shooter has a greater chance of being injured accidentally by a fall in a bathtub than he has by a gun while out hunting or shooting, but to decrease the number of firearm accidents that do occur, certain safety rules are recommended. Every firearm, even when apparently empty, should be treated as though it were loaded. Even an apparently empty firearm should never be pointed at anything the bearer does not intend to shoot at. Only empty disassembled firearms should be carried in a car, or kept in the home or in a hunting camp. On the range or in the field, the firearm must be carried so that the shooter can always control the direction of its muzzle, even should he stumble, and its "safety," a device which prevents the firing mechanism from operating, should be in the "on" position. It should be remembered, however, that with very few exceptions standard-design safety systems lock the triggers only, and not the sear (the catch that holds the hammer at cock or half cock). Before firing, it must be ascertained that the barrel is free from obstructions such as snow or mud or a forgotten cleaning patch. With a rifle or handgun, shooting into the air should be avoided because such shots endanger distant people, livestock, or buildings. Even with low-power cartridges, shooting at hard flat surfaces, or at water, is dangerous because ricocheting bullets can travel long distances. Before any shot with a rifle or handgun, the shooter should make certain the backstop (dense trees, a hill, etc.) will stop his bullet should it miss the target aimed at. The drinking of intoxicating beverages or the use of any drug, which might dull the shooter's judgment and caution, can be dangerous. In the U.S., many states require new hunters to take National Rifle Association training and safety courses before a hunting licence will be issued.

See also Small Arms, Military; Pistol and Revolver.

(G. H. P.; W. B. Ed.; W. S.)

SHOOTING STAR (American Cowslip), a plant of the genus *Dodecatheon* of the primrose family (Primulaceae), comprising species native to North America and northeastern Asia. They are stemless, perennial herbs, with basal leaves and a naked flower stalk, which bears at its summit a cluster (umbel) of singularly shaped flowers similar to those of the garden cyclamen (*q.v.*). The shooting star commonly grown in gardens, *D. meadia*, with pink or white flowers, is native in open grounds in eastern North America. Throughout the Middle West the native *D. meadia* is sometimes called prairie pointer because of its pointed and apparently inverted flower, which blooms in late April or early May. Numerous species occur in the western states, seven being found in California, among which are the Sierra shooting star (*D. jeffreyi*), the lowland shooting star (*D. patulum*), the upland shooting star (*D. hendersoni*), often called mosquito bills or sailor caps, and *D. clevelandi*, with handsome purple and yellow-spotted flowers. All are cultivated for ornament, but western species do not grow well in eastern gardens. Shooting stars do best in wild flower gardens partially shaded and in well-drained soil. (N. Tr.)

SHOP FRONT DESIGN. The purpose of a shop front is to identify the name, character and nature of the store or shop, and, usually, to give a full or token display of the goods or services offered inside. Of the many devices and techniques employed by the merchant to attract the attention of customers, two of the commonest are a sign or some graphic identification of the store name, and a show window or windows.

Before the concept of a specialized store architecture evolved, merchants occupied space in such buildings as were available; a sign was added to the building or store front and the owner did what he could to enlarge and give importance to existing show window space. Later, as modern retailing developed, stores occupying old buildings broke out the street wall and created new fronts to gain greater display space, provide surfaces for large building front signs and enlarge their entrances. These practices continue wherever it is not practical or desirable to erect a new building. In modern buildings utilizing steel framing and non-load-bearing walls, store architecture is offered the potential of a completely transparent main floor selling area and maximum shop front display.

Signs.—Some of the commonest types of signs are raised lettering backlighted against the store façade; neon lights; letters applied on a luminous glass lintel; and bronze or metal letters applied to the face of the building. Usually a clear script or block lettering is utilized and the sign is executed as an architectural element of the store front.

In modern retailing techniques a store's sign serves as its trademark, used in all its advertising and in its packages, labels, letterheads, truck panels, etc. Before the development of local and national advertising, the sign was often required to describe the store. It was then not so much a trademark as an informative poster, and the store's name was often presented on the sign in conjunction with such descriptive phrases as "dry goods store," "ladies clothing and furnishing goods," etc. Later, as advertising made the name synonymous with the services and goods offered, the use of such phrases went out of fashion.

Windows.—Store fronts can be classified on the basis of the types of show window they employ: (1) a conventional show window or windows spaced out across the face of the building; (2) unit displays set within the store behind a total glass front interrupting but not obscuring the view of the interior of the store; and (3) continuous glass which exposes the main floor interior completely. Display methods relate directly to the way a store conducts its business. Supermarkets and some chain stores, for instance, have found that the all-glass front best serves merchandising based on offering a great mass of goods at low cost. Thus, the store itself is the display. Specialty shops, on the other hand, prefer to use key displays set behind an open glass front, presenting samplings of the total merchandise.

Size also may affect the appearance of a specialty shop front. Seldom occupying the space of a chain store or supermarket, such a shop must use some of its interior space to increase the perimeter of the show window from the street. Thus a great many shop fronts are recessed into arcades cutting into the interior store space. The effect on the appearance of the shop front is one of a

diagonally placed show window which can be seen as the shopper approaches the store along the street. The arcade, further, offers a place for window-shopping under shelter. Many shops of this kind have ceiling lighting patterns which start in the ceiling of the arcade and continue into the store proper.

Major department stores still make profitable use of the show window framed in the building wall, which serves to intensify the appeal of selected items. Window displays are changed as often as the store's promotional needs require and are designed by display specialists. Every theatrical device is employed: special sets, lighting, background treatments, properties, etc. (*See* also DISPLAY DESIGN.)

Modern Developments in the United States.—Among considerations affecting the design of store fronts in the United States, modern structural techniques and materials are of the foremost importance. But second only to these is the location of the store, particularly as it is influenced by increased use of the automobile. Special mention may be made of the shopping centre and multiple-unit retail operations—branch stores and chain stores.

Shopping Centres.—In city locations buildings occupied by stores may vary in size, height, age, style of architecture and materials; and a venerable brick building adjoining a new porcelain enamel shop front render each other distinctive by contrast. In the shopping centre, stores conform generally to the modular character of the centre, and the problem of achieving individuality rests chiefly on window display and signs. The total centre has a character, in effect that of a shop front; a major sign, dramatically lighted, usually identifies the unit as a whole. The sheer intensity of the light attracts and, as customers come close to the centre, makes all high-placed signs legible. The all-glass open front is associated primarily with supermarket design; variations in treatment must be effected so that two or more supermarkets occupying the same shopping centre and using the same display technique can be distinguished from each other.

Multiple-Unit Operations.—Branch stores are generally housed in one- or two-story buildings with customer approaches from one of several parking lots, and all the faces of such stores bear a part of the burden of identification and display. Usually removed from a steady stream of pedestrian traffic, the store cannot hope to attract customers primarily with enticing street window displays, and the show window thus diminishes in importance. The store generally makes its initial impression on a customer arriving by automobile, and the entire building may thus serve as a signboard. Signs are enlarged, placed high and lighted so that they can be seen from a considerable distance. A downtown store usually depends on one or two major entrances, seldom more; a branch store may have as many as there are approaches from parking lots. Since many branch stores are located in residential communities in which the land has never been leveled for commercial use, store designers are capitalizing on sloping or variable sites, often utilizing lower- and upper-level entrances. The architectural treatment usually reflects the prevailing domestic architecture of the community, and the exterior treatment is much less flamboyant than it would be on a city shopping street. Neon or other spectacular sign treatments are seldom used, or needed, because of the store's isolation from competing shops.

Chain stores use prototypes for signs and in some instances for building design. Among notable successes in achieving a distinctive character in sign, architecture and shop front are the Howard Johnson chain, which houses its restaurants in pitched roof, gable, quasi-colonial buildings with a white, coral and turquoise colour scheme; and Shell gas stations, with their strong, formalized shell trademark and startling colour combination of yellow and orange. (W. SN.)

Modern Developments in Europe.—The rise in living standards and the change in shopping habits after World War II have contributed to a complete overhaul of store design and, in particular, the display planning of the shop front. The open front has become the accepted formula except in a few cases of stores of an exclusive nature, *e.g.*, a jeweler, and even here the all-glass or toughened plate glass door, gives some visibility of the interior.

The most important development in postwar Europe has been the "doorless" entrance or air curtain doorway; the earliest examples were used in the Globus store at Basel and the Jelmoli store at Zürich in the early 1950s. The doorway has a lobby with a low, louvred ceiling and a closely divided metal grid for a floor; warm air passes through the ceiling and is extracted through the floor, creating a warm air curtain which retains the heat of the interior without impeding entrance into the shop. The entrance is closed at night by a sliding or rising wall which may be a glass screen, can incorporate doors, or may even be a display window or showcase. Later examples effected a compromise by a system of toughened plate glass swing-doors which operate during adverse weather conditions but will fold away, concertina-wise, giving an unobstructed entrance. In the late 1950s France and Germany introduced the use of adhesive and "clip" joints in the glass, eliminating even metal frames and surrounds, and increasing visibility. A development of this system around 1960 led to "suspended" glazing where huge sheets of plate glass were hung by metal clips from a structural member (instead of being bedded in frames). This, combined with adhesive joints, allowed the glazing of immense areas—often the whole frontage of a large store—without any interruption of visibility.

Self-service and self-selection, which help to turn the "window-shopper" into a customer, have played their part in encouraging the open front and unobstructed entrance. The open front has raised new design problems such as reflection, sun glare and sun control, artificial lighting by both day and night, a general reorganization of merchandise within the store, now completely exposed, together with a new approach to the design of display fixtures and fittings. Freer handling of levels within the shop makes it possible to expose more than one floor of the interior to the street. Many split-level shops exist in Germany and the idea was developing rapidly in Europe in the 1950s and 1960s.

Another important factor is the imaginative use of colour and lighting. Cleverly used, colour can enhance the merchandise on display. It can also modify undesirable architectural features which cannot otherwise be corrected; for instance, disguising the length of the long narrow store. Different harmonizing or blending colours also identify and separate adjacent displays, and relate them, by the use of similar colours, to the departments within the store. Lighting has become an integral part of design. Intensities are five or six times greater than prewar lights, but because this is common to all shops, a degree of lighting monotony has been created. Effective lighting therefore needs contrast, perspective and colour as well as brightness. To bring out correct colour values, a combination of fluorescent and incandescent fixtures is common, as each stresses a different end of the colour spectrum. Much recent progress has been made in the use of the low voltage spot lighting which can focus beams of great intensity on particular items thus punctuating and dramatizing the display even against the greatly increased general brightness. The fittings incorporating the low voltage lamps are very small and unobtrusive, and are placed in the windows at points to give the greatest effect. (*See* also LIGHTING: *Electroluminescence: Retail-Shop Lighting.*)

Design and layout of lettering on signs had improved by the 1960s, and the increased use of plastic coverings, particularly in England, was eliminating the naked neon tube type of sign, although the latter still predominated on the European continent. (E. So.)

SHORE, JANE (d. 1527), mistress of the English king Edward IV, is said to have been the only child of Thomas Wainstead, a prosperous London mercer. She was married at an early age to William Shore, a goldsmith, and became mistress to the king *c.* 1470. Although she wielded great influence over the king, Sir Thomas More (*Richard III*) asserted that his favour "she never abused to any man's hurt, but to many a man's comfort and relief." After Edward's death (1483) she was mistress of Thomas Grey, 1st marquess of Dorset, son of Queen Elizabeth Woodville by her first marriage, and she was also a concubine of Lord Hastings. She probably acted as an intermediary between the Woodville party and Hastings in opposition to Richard, duke of Gloucester, and therefore incurred the latter's hostility. In

June 1483 Richard (who became King Richard III later that month) accused her of sorcery, had her arrested, and afterward got the bishop of London's court to make her do public penance as a harlot; but all the beholders pitied her for her great beauty and patience. Her husband was now dead and she was imprisoned in Ludgate. There she captivated the king's solicitor, Thomas Lynom, who wished to marry her. Richard dissuaded him and Jane's fortunes never recovered. She died as a beggar in 1527.

See Sir Thomas More, *History of King Richard III,* ed. by J. R. Lumby (1883). (A. R. M.)

SHOREHAM-BY-SEA, a seaside resort and urban district in the Arundel and Shoreham parliamentary division of West Sussex, Eng., 6 mi. W. of Brighton. Pop. (1961) 17,410. Area 4.8 sq.mi. It lies at the mouth of the Adur river and 1 mi. upstream is the village of Old Shoreham, the original port, whose importance declined with the eastward shift of the river mouth, now 1 mi. from New Shoreham. The modern harbour embraces this lower reach of the river and also a semiartificial waterway that extends eastward from the river mouth for 2 mi. to Portslade and where ships can lie afloat at all states of the tide.

Shoreham was called a borough in 1236 and from 1272 to 1885 returned two members to parliament. Weekly markets and an annual fair dating from the time of Edward I were held for centuries. Shipbuilding has been carried on since the 13th century, and there is trade in coal, grain, timber and cement.

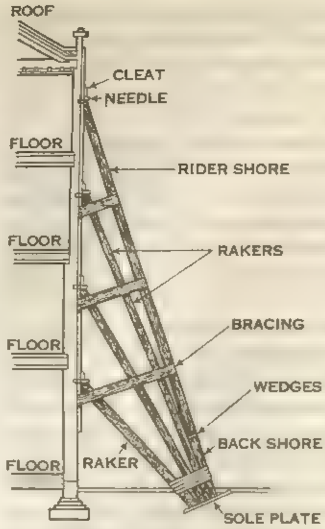
On the west side of the Adur opposite Shoreham is the Brighton-Worthing municipal airport; to the northwest of it is Lancing college, a public school for boys. At Shoreham Gap, 2 mi. N.E., 596 ac. of downland were presented to the National trust in 1946.

SHOREY, PAUL (1857–1934), outstanding U.S. classical scholar and humanist, was born at Davenport, Ia., on Aug. 3, 1857. Having graduated from Harvard in 1878 he was admitted to the Chicago bar in 1880. Later he studied at Leipzig, Ger. (1881–82), Bonn, Ger. (1882), the American School of Classical Studies at Athens (1882–83) and the University of Munich, from which he received the Ph.D. in 1884. He was professor of Greek at Bryn Mawr college (1885–92), then went to The University of Chicago as professor of Greek and chairman of the department of Greek. He was Roosevelt professor in the University of Berlin, 1913–14. A man of vast erudition, Shorey was a brilliant teacher and lecturer, especially in the fields of Greek poetry and philosophy.

Shorey's writings include *The Idea of Good in Plato's Republic* (1895); *The Odes and Epodes of Horace* (1898); *The Unity of Plato's Thought* (1903), his most significant work in that it summed up his whole concept of Plato; *The Assault on Humanism* (1917); *What Plato Said* (1933); and his edition (with English translation) of Plato's *Republic* (Loeb Classical Library, 1930–35). He was a constant contributor to *Classical Philology*, of which he was managing editor from 1908 until his death in Chicago on April 24, 1934. (G. Sm.)

SHORING, a form of prop or support, generally temporary in nature, used during repair or original construction of buildings, and in excavations. Temporary support may be required, for example, to relieve the load on a masonry wall while it is repaired or reinforced. The support may be supplied by shoring the wall with heavy timbers sloping upward at about 60°–75°. The top of the timber is so arranged that part of the wall load will be transferred to the shore. The lower end of the shore will frame onto a base in order to transfer the load to the ground with a minimum of deformation. Wedges may be used to bring the shore snuggly into contact with the wall. If the wall is several stories high, as in the figure, more than one shore is required. The spacing between the shores depends upon the condition of the wall, magnitude of the load, structural quality of the shores, and the character of supporting base and ground.

Shores are used to support the forms for cast-in-place concrete slabs, beams and girders in reinforced concrete frames. Such supports may be 4 × 4's or larger timbers cut to appropriate length and placed to true grade by wedges. The size of the timber depends upon the load and length of the shore. Lateral bracing is needed in both directions to prevent the buckling of



SYSTEM OF RAKING SHORES

the shore. In some cases timbers may be framed into towerlike arrangements to provide capacity for greater loads. There are several patented shores that may be used in place of timber shores for concrete forms. These have the great advantage of easier placement and easier adjustment for length. The adjustment may be made by means of a threaded member or by a gripping device.

One method of constructing a pier in water is to use a cofferdam, which may be a rectangular enclosure made by driving sheet piling, usually steel, around a horizontal frame system. The outside members of the frame against which the sheet piles bear are called wales. The transverse and longitudinal members between the wales are shores. The wales and shores may be constructed of heavy timbers, structural steel, or a combination of the two materials.

The shores must be sturdy and well braced to withstand the shock of being hit by a heavy dredging bucket; they must also be able to withstand the horizontal forces resulting when the water is pumped from the cofferdam. The number of horizontal frames required may vary from one to five. (F. W. St.)

SHORT, SIR FRANK (Francis Job) (1857–1945), English topographical engraver and water-colourist, was born at Wollaston, Worcestershire, on June 19, 1857. He was educated to be a civil engineer, and in 1881 came to London. In 1883 he studied at the Royal College of Art, the Westminster School of Art and the Royal Institute of Painters in Water-Colours. Short's style was at first closely modeled on that of Seymour Hayden's, and earlier successes included mezzotints after J. M. W. Turner's *Liber Studiorum* (1885 *seq.*) and his best original works are mezzotints and aquatints such as "The Night Picket at Hammersmith" (1916) and "Morning Flare in Chichester Harbour" (1922). Elected A.R.A. in 1906, he was knighted and made a royal academician in 1911. As director of engraving at the Royal College of Art (1891–1924), he greatly influenced younger engravers. He published *On The Making of Etchings* (1888). He was president of the Royal Society of Painter-Etchers and Engravers (1910–38). He died April 22, 1945, at Ditchling, Sussex. (D. L. Fr.)

SHORTHAND is any system of rapid writing using symbols or shortcuts that can be made quickly to represent letters of the alphabet, words, or phrases. It has been variously known as stenography (close, little, or narrow writing), tachygraphy (swift writing), or brachygraphy (short writing). Shorthand is now employed extensively in reporting the proceedings of legislative bodies, the trial of cases in courts of law, and for taking dictated business correspondence.

The earliest record of an organized system of shorthand dates from the year 63 B.C., the age of eloquence in Rome. At that time Marcus Tullius Tiro, a freedman and friend of Cicero, invented a system of *notae* that was used in recording the speeches of Cicero, Seneca, and others of the Roman Senate. The system invented by Tiro was taught in the Roman schools, was learned by emperors, and was widely used. In his life of Cato the Younger, Plutarch stated that the speeches of Caesar and Cato in connection with the conspiracy of Catiline were taken down verbatim by *notarii* who had been placed by Cicero in various parts of the Senate. After the fall of the Roman Empire the use of the Tironian system survived for several centuries.

Beginnings of Modern Shorthand.—England was the birthplace of modern shorthand. The publication by Dr. Timothy Bright in 1588 of his *Characterie: an Arte of Shorte, Swifte, and Secrete Writing by Character* marked the beginning of this development there. In the next 50 years, thirteen systems were

known to have been published, a number of them passing through several editions. They were orthographic, *i.e.*, they followed normal spelling, although many omitted silent letters. One of the best known of these early systems was Thomas Shelton's *Short Writing* (1626), in which Samuel Pepys kept his famous diary. Nearly two centuries later another famous English author, Charles Dickens, learned shorthand and earned his living for a time as a reporter in the law courts and in Parliament.

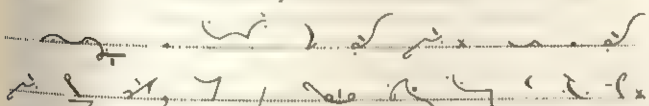
Although most of the early systems of shorthand in England were orthographic or alphabetic, the idea of writing according to sound continued to gain in favour. The Rev. Phillip Gibbs (1736) was the first to break away from an alphabetic basis to a phonetic one by distinguishing between long and short vowels.

Pitman System.—The publication by Isaac Pitman in 1837 of *Stenographic Sound Hand* marked a new era in the development of phonetic systems. Not only did he classify the sounds of the language scientifically and arrange his material for writing accordingly but he also introduced simple expedients of abbreviation that made for rapidity. A short summary of the principles underlying the system is given.

Since the system is phonetic, all words are written according to their sounds. The words *lain, deal, may, knife* would therefore be written as if they were spelled *lān, dēl, mā, nīf.* The consonants are drawn from simple geometrical forms, straight lines, and shallow curves. As far as possible all consonants are paired: a lightly written stroke represents the sound of F, for example, and the same stroke, written slightly heavier, stands for the companion heavy sound of V. To illustrate: P ... B ... T ... D ... Vowels are indicated by disjoined dots and dashes, which are placed in proper relation to the consonants and by writing words in position with relation to the line of writing. To illustrate: bay ... tea ... ode ... bug ...

The Pitman system makes use of circles, loops, and hooks for sounds of frequently used consonant combinations and syllables (*i.e.*, S, ST, STR, SPR, -ter, -der, -tion, etc.). Syllables are also added by halving or doubling the length of a consonant stroke.

Illustration of Pitman Shorthand



Many years ago the family was the social unit. Now the social unit has become the world, in which each person's welfare affects that of every other.

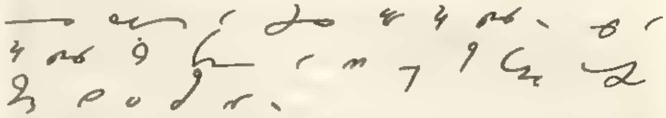
The Pitman system was introduced into the United States a few years after its publication. In the Eastern Hemisphere, including Australia, New Zealand, and India, Pitman is the predominant system of shorthand; more than 250,000 people in these areas learn Pitman every year.

Just as Pitman and some of his contemporaries rejected the orthographic principle in shorthand and adopted the phonetic basis, another group of authors discarded the use of geometrical signs in favour of cursive symbols. As early as 1620 William Folkingham produced *Brachygraphie Post-Writt* in which he anticipated Gregg and the more modern "script" systems and used blended consonant signs for such combinations as FL, FR, TR, etc. He also introduced position writing to indicate vowels.

Gregg System.—In 1888 John Robert Gregg published his *Light-Line Phonography* in England and soon thereafter brought his system to the United States. It is taught in more schools and practised by more stenographers in the United States than any other system. It is published and taught in most of the nations of the world.

The Gregg system is based on the following principles: (1) The system is phonetic; all words are written by sound. (2) There is a total absence of shading or thickening; all symbols are written light-line. To illustrate: r l t d n m . (3) The characters are based on the elements of ordinary longhand; the strokes are familiar and the motion is uniform. (4) Vowels are expressed by circles and hooks; the words are not placed on, above, or through the line of writing to imply the omission of certain vowels. To illustrate: a e o u . (5) Vowels are inserted in word outlines in their natural order without lifting the pen. To illustrate: reel late den name . (6) There is a preponderance of curve motion to aid writing. The Gregg system makes use of brief forms for some of the commonest words, blended consonants, suffix and prefix forms, and an abbreviation principle in order to enable the writer to gain speed in writing.

Illustration of Gregg Shorthand



Many years ago the family was the social unit. Now the social unit has become the world, in which each person's welfare affects that of every other.

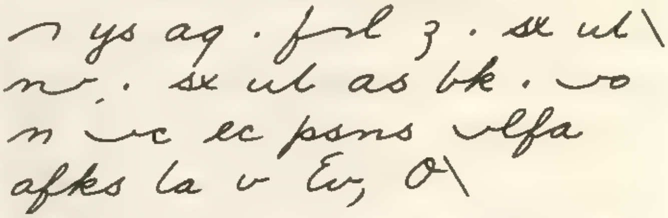
Longhand Shorthand Systems.—A recent trend in the development of shorthand systems makes use of longhand symbols for most, if not all, of the letters of the alphabet. Illustrations are: Abbreviatrix (1945), Baine's Typed Shorthand (1917), Carter Briefhand (1957), Forkner Alphabet (1955), Gersten's Alphabet Shorthand (1949), HySpeed Longhand (1932), Quickhand (1953), Speedhand (1952), "Speedwriting" (1923 and 1951), Stenoscript (1955), Stenospeed (1953), and Zinman (1950).

The first "longhand shorthand" is found in *Cadmus Britannicus* by Simon George Bordley (1787), which comprised three systems. The first, called "Slower Shorthand," used letters of the alphabet; the second, "Swifter Shorthand," was a script shorthand; and the third was "Shorthand for Music." The main difference between Bordley's "Slower Shorthand" and modern systems of this type is that Bordley's letters were "printed" longhand and could not be joined.

The advantages claimed for alphabetic systems are ease of learning because of the learner's familiarity with the forms, a shorter learning period than is needed with symbol systems, and the ease of transcription. The chief disadvantage is limited speed.

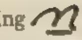
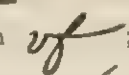
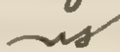
The system known by the registered trademark "Speedwriting" was devised by Emma Dearborn in 1923. It is written by sound; only long vowel sounds are expressed. To illustrate: knew *nu* bill *bl* file *fl* did *dd*. Some letters are modified for speed. To illustrate: t (crossmark omitted), i (dot omitted), m, w (humps omitted).

Illustration of "Speedwriting" Shorthand



Many years ago the family was the social unit. Now the social unit has become the world, in which each person's welfare affects that of every other.

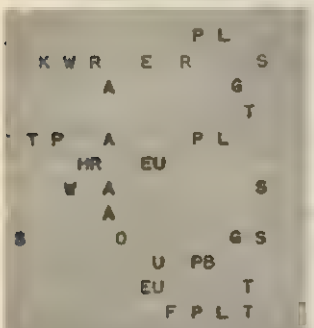
In addition, brief forms, phrasing, abbreviating principles, and flourishes are used. A few of the flourishes follow: (1) a line drawn under the final letter expresses final *ing* and *thing:* billing

, nothing ; (2) the *f*, extended upward, expresses *ification:* verification ; (3) the *s*, extended upward, *self, selves:* myself , ourselves .

In Britain there has been a steadily increasing following in recent years for *Speedhand* by C. T. Rutherford (1952). It is a system that can be used either by hand or on the typewriter, and it is claimed that it can be learned in 15 hours.

Machine Shorthand.—Machine shorthand is the method of recording speech by the use of machines, a method developed in the last quarter of the 19th century. The Stenotype and Stenograph shorthand machines are common examples. These two machines have identical keyboards of 22 keys. All fingers and thumbs are used; any number of keys can be struck simultaneously; a strip of folded paper moves ahead with each stroke, carrying the imprint of as many keys as were struck.

Illustration of Stenograph Machine Shorthand



Many years ago the family was the social unit. Now the social unit has become the world, in which each person's welfare affects that of every other.

Each key prints a letter of the English alphabet (except one which prints an asterisk). The basic principle is the recording of a syllable at a stroke, phonetically. The fingers of the left hand control keys that print consonants occurring before vowels; these print at the left side of the tape. The thumbs control vowel keys, which print in the middle of the tape. Fingers of the right hand control consonants occurring after vowels and print at the right. Thus most consonants can be written on either side of the keyboard.

Letters for which there are no keys are represented by combinations of other letters. Arrangement of the consonants on both sides corresponds with the sequence of sounds as they occur in the English language. Abbreviated forms are used for some of the most frequent words, enabling the operator many times to write two or three words in one stroke. Completed notes are in the form of a folded pad.

Advantages claimed for machine shorthand over manual systems are greater speed and the interchangeability of the notes. Disadvantages reported are inconvenience (machine instead of pen) and difficulty of revising notes already taken.

Two machines that have been in use in Europe are the Grandjean (France) and the Palantype (Britain). The latter takes notes in phonetic print rather than by means of a "code" and the majority of words are instantly recognizable to anyone who has never operated a machine, provided the keys are depressed correctly.

Shorthand in Other Languages.—Soon after pen shorthand was introduced in England, other systems appeared on the continent of Europe. The first French system was Jacques Cossard's *Méthode pour escrire aussie vite qu'on parle* (1651). Coulon de Thévenot published his *Tachygraphie* in 1787, and in 1826 a system was invented by Hippolyte Prévost; this system was improved by Albert Delauney in 1878. Today it is known as the Prévost-Delauney system and is one of the two most widely used among French-speaking peoples, the other being that of Abbé Duployé (1862). They are based on geometric principles with joined circles, hooks, and loops for the vowel sounds. In Germany the most notable systems have been those of Franz Xavier Gabelsberger (1834) and H. Wilhelm Stolze (1841). In Italy the adaptation of Gabelsberger by Noë, subsequently altered by various authors, is widely used. There is hardly any language in the world that has not had its own shorthand system at some time or other, including Eskimo and American Indian languages.

Bibliography.—J. W. Beers, *Modern Shorthand* (1948, privately published); E. H. Butler, *The Story of British Shorthand* (1951); J. W. Carlton, *Bibliotheca Pepysiana*, part IV, *Shorthand* (1940); H. Glatte, *Shorthand Systems of the World* (1959); J. R. Gregg, *Gregg Shorthand Manual Simplified*, 2nd ed. (1949) and *Selections from the Story of Shorthand* (1941); O. W. Melin, *Stenografiens Historia* (1927); Pitman Publishing Corporation, *Basic Course in Pitman Shorthand* (1946); J. E. Rockwell, *Shorthand Instruction and Practice* (1893); A. L. Sheff, *Speedwriting Shorthand: Basic Course* (1951).

(S. J. Wa.; E. H. B.)

SHORTHOUSE, JOSEPH HENRY (1834–1903), English novelist, influenced by the Oxford movement (*q.v.*), whose "philosophical romance," ***John Inglesant***, is one of the few English examples of the novel of ideas. The eldest son of Quaker parents, he was born in Birmingham on Sept. 9, 1834. His education was interrupted by illness and at 16 he entered the family chemical works, but continued to read widely. Influenced by Ruskin and the Pre-Raphaelites, and attracted by Anglican worship and liturgy, in 1861 he joined the Church of England; the dichotomy between the quietist view of religion in which he was brought up and the sacramental view he adopted gives a particular colouring to *John Inglesant*. This, his first novel, begun in 1866, is set in 17th-century England and Italy, and discusses the conflicts between church and state, the Church of England and the Church of Rome, ritualism and simplicity, different views of the sacraments, etc., in long Platonic dialogues and disquisitions. It incorporates unacknowledged quotations from such 17th-century writers as Thomas Hobbes, John Aubrey, Miguel de Molinos and John Evelyn, and depicts with scholarly detail but Tractarian bias the atmosphere of the period, giving a notable picture of the community founded at Little Gidding by Nicholas Ferrar (*q.v.*).

It remained in manuscript until 1880, when 100 copies were privately printed. One came to the notice of Mrs Humphry Ward on whose recommendation Macmillan published it in 1881. Its success (9,000 copies were sold within a year) was due partly to the praise it received from T. H. Huxley, Cardinal Manning and Gladstone, and partly to its treatment of topical themes, but, by its quiet charm, its dignity and the devotional, mystical tone of its religious feeling, it continues to hold a higher place in the esteem of many readers than its purely literary quality perhaps merits.

Shorthouse's other novels (*Sir Percival*, 1886, etc.) were less successful, although revealing the same blend of Platonism, aestheticism and mysticism.

Shorthouse died at Edgbaston on March 4, 1903.

Bibliography.—*Life, Letters and Literary Remains*, ed. by his wife, 2 vol. (1905); J. E. Baker, *The Novel and the Oxford Movement* (1932); M. Polak, *The Historical, Philosophical and Religious Aspects of "John Inglesant"* (1933).

SHORT STORY, a kind of prose fiction, distinguished from the novel (*roman*) and the novelette (*nouvelle* or *novella*) by its compression and intensity of effect. More prescriptive definitions than this have been advanced, some of them so arbitrary as to place unattractive and unnecessary limitations upon a developing art form. For the short story, like the novel, is still in a process of development and perhaps ultimate definition. The serious short story of the 20th century might be visualized as occupying a square, the four corners of which are marked by the narrative essay or sketch, the lyric poem, the prose drama and the unit of local social history. Some short stories would be located in the centre of the square; others, although within the square, would be closer to one corner than to the others. Washington Irving's 19th-century "The Legend of Sleepy Hollow," with its slight, uncomplicated action, its embellishments of time and place and the immediate presence of the voice and personality of the

author, can only with difficulty be distinguished from a narrative essay or sketch. James Joyce's "The Dead," with its luxuriant evocation of sensuous images and its intensity of mood, differs from a lyric poem only in its expansiveness and its failing to fully utilize rhythm and sound. Ernest Hemingway's "The Killers," stripped as it is of overt comment or analysis, and making its effect primarily by the implications of dialogue and a few simple actions within a limited scene, could easily be produced as a one-act play. Finally, Theodore Dreiser's "Old Rogaum and His Theresa," concentrating as it does upon a conflict within an immigrant family in special economic and social surroundings, is akin to a chapter in a social history of New York city.

Early Forms of Short Fiction.—If it has been largely in the 20th century that the lyric and dramatic possibilities of the short story have been explored, the similarities of the short story to particularized history and to the expanded anecdote are traditional. These likenesses may be suggested by the origins of the words used to denote a short prose fiction: "story," with its roots in the Latin *historia*, suggests the relation of fictional accounts to the presentation of historical events; the English "tale" and the French "conte" suggest something told or recounted, and by implication the felt presence of the teller or narrator. Short stories, in the sense of short tales told by an oral teller, of course antedate the records of human history. The most primitive man undoubtedly could arrange a series of remembered events in a time sequence and thus have the rudiments of a story. Among the earliest writings of man are *Tales of the Magicians*, a collection of stories from ancient Egypt which probably date from about 4000 B.C.; similar collections may be found in the ancient writings of Arabian, Hindu, Greek and Hebrew cultures. It has often been pointed out that the Bible contains tales that would fit the present conception of the short story: the books of Ruth and Jonah, for example, or such parables in the Gospels as that of the Prodigal Son. The literature of the middle ages and the Renaissance abounds in short narratives in both prose and verse, of which the *Decameron* of Giovanni Boccaccio is the foremost example. Western European literature after the Renaissance has produced numerous fables, romances, tales and sketches, which kept a tradition of short prose narration alive.

Emergence of the Short Story.—It was not until the beginning of the 19th century, however, that the short story as a distinct literary form began to attract the attention of serious writers in large numbers. Almost simultaneously in Germany, the United States, Russia and France collections of short fiction began to appear. In Germany, E. T. A. Hoffmann published his exotic tales between 1814 and 1821; Johann Ludwig Tieck, who had begun publishing stories in the 1790s, published several collections in the 1800s. In the United States the publication of Irving's *The Sketch Book* in 1819 and 1820 marked the beginning of a long tradition in the American short story. One of the three tales in that collection, "Rip Van Winkle," has generally been regarded as the first American short story. In 1832 the first stories of Edgar Allan Poe and Nathaniel Hawthorne began to appear. With Irving, these two authors were to become the best American writers of the short story in the first half of the century. At almost exactly the same time, two Russians, Aleksandr Pushkin and Nikolai Gogol, turned from the writing of novels and the drama to the short story, where their attention to the details of ordinary life contrasted sharply with that of the fantastic and the legendary which had been exploited by the earliest German and American story writers.

Finally by 1830, three writers in France—Prosper Mérimée, Honoré de Balzac and Théophile Gautier—had begun a distinguished tradition of the conte which was to continue throughout the century. What is remarkable about this sudden flourishing of a literary form in several countries at approximately the same time is that all of these writers, in spite of their varying techniques and subjects, were united in their conception of the short prose narrative as a potentially important form of literary art and in their consequent exploration of the possibilities and limits of the short story as a genre. Indeed the term short story was not to appear very frequently until much later in the century. Even story was quite infrequent, although as early as 1824 Irving used "Strange Stories by a Nervous Gentleman" as a subtitle for an installment of his *Tales of a Traveller*. Tieck called one of his collections *Die Gemälde* (pictures); Irving called his stories sketches or tales; Poe called his tales or articles; Hawthorne used tales, sketches, even parables.

Why the conception of the short story as a new form should have come at just this time, relatively so late in the development of western literature, is somewhat difficult to explain. Possibly the short story could not be developed until the novel itself had reached the status of serious art and some degree of maturity in the latter half of the 18th century. The widespread experimentation with new forms as part of the romantic movement at the beginning of the 19th century may also be suggested as a contributing factor. In the United States the form found special favour, not only as a literary medium in the refinement of which the U.S. writer could participate from the very beginning but also as a type of writing admirably suited to publishing conditions in the new republic. There the flood of annuals or "gift books," which followed the first number of *The Atlantic Souvenir* in 1826, provided a ready market for short works which the U.S. writer, handicapped in the publication of novels by the international copyright situation, could supply on equal terms with British writers. From the annual gift books, which provided a first publishing outlet for Hawthorne, evolved the popular magazines, in which Poe's work most often appeared. Because magazine editors competed with each other for short fiction to fill the pages of their monthly issues, a whole new market was created. These editors probably adopted the sentiment expressed facetiously by Irving in a preface to his *Tales of a Traveller:* "If the tales I have furnished should prove to be bad, they will at least be found short; so that no one will be wearied long on the same theme." Later Poe expressed the importance of these economic circumstances when he argued that the short story was the child of the American magazine. Certainly the popular magazine, both in the United States and in Europe, has continued to be by far the largest outlet for short stories, consuming hundreds of thousands of them during a century and a half and creating various formulas for "the popular story," which provides ephemeral entertainment for readers without disturbing them unnecessarily, and conceals old plots and hackneyed ideas in ingeniously invented novelties of detail.

The importance of Poe in the history of the short story is that he used the magazine to insist that a short story was potentially a superb art form. It is true that such writers as Hoffmann, Irving and Gogol had recognized the rewards of attention to the artistry of the tale; but Poe in his famous review of Hawthorne's *Twice-Told Tales* in *Graham's Magazine* for May 1842 claimed that "the tale proper, in our opinion, affords unquestionably the fairest field for the exercise of the loftiest talent, which can be afforded by the wide domains of mere prose," reserving the highest order of excellence for the lyric poem. Furthermore, he formulated the first set of principles for the art of the short story in a famous paragraph:

> A skillful literary artist has constructed a tale. If wise, he has not fashioned his thoughts to accommodate his incidents, but having conceived, with deliberate care, a certain unique or single *effect* to be wrought out, he then invents such incidents—he then combines such events as may best aid him in establishing this preconceived effect. If his very initial sentence tend not to the outbringing of this effect, then he has failed in his first step. In the whole composition there should be no word written, of which the tendency, direct or indirect, is not to the one pre-established design.

Some of Poe's principles—coherence and design, for example—were hardly new, being those assumed for literary art from the time of Aristotle onward; yet their application to the short story indicated that it had grown by 1842 into a significant literary form. Moreover, if it is true that Poe's somewhat mechanical conception of composition and his emphasis on a "single effect" could in the hands of critics like Brander Matthews 60 years later lead to a falsely rigid reduction of the short story to certain "essential ingredients" which could be concocted for the popular magazines, it must be remembered that Poe achieved a notable

variety of works within his definition: the classic horror tale "The Cask of Amontillado"; the detective tale "The Purloined Letter"; and the tales of psychological exploration "Ligeia," "William Wilson" and "The Fall of the House of Usher."

The two other U.S. writers of this period whose work deserves mention, Hawthorne and Herman Melville, were more important as novelists than as writers of shorter works. Hawthorne in his *Twice-Told Tales* (1837) and Melville in *The Piazza Tales* (1856) added a moral ingredient to the short story which Poe's tales of effect lacked. Rich in symbol, proceeding from the depths of human experience, and appealing to the intellect as well as to the emotions, their work has an integrity and complexity which for many modern readers Poe's and Irving's have seemed to lack. Each of Hawthorne's stories, like the black cloth in his "The Minister's Black Veil," holds a "mystery which it obscurely typifies," and it examines the ambiguities of human life in a manner which Poe's "single effect" does not comprehend.

The Short Story in Europe.—While U.S. writers were taking advantage of their special circumstances to build a strong tradition of the short story, writers in France and Russia were no less active. In contrast to the romantic fantasy and eccentric distortions of Tieck and Hoffmann in Germany, or the U.S. variations of the tale of romance and terror in the comic Gothic of Irving, the psychological interest of Poe and the moral concern of Hawthorne, such French writers as Mérimée and Balzac were grounding their work in realism and emphasizing such clarity of observation, vividness of detail and precision of statement that the stories "stand and speak before the reader," the writer remaining impersonal and objective. This tradition of objectivity, which led the short story away from its similarities to the narrative essay or sketch, where the presence of the author's personality continuously is felt, toward the direct representation of the drama, where the incidents must reveal the meaning of the work implicitly rather than explicitly, has been the major contribution of the French story writers, not the early masters, but also those who continued this tradition much later in the century—Gustave Flaubert and Guy de Maupassant.

The Russian tradition, beginning with the stories of Gogol in *Dead Souls*, reached an early maturity in the work of Ivan Turgenev, whose first and perhaps finest collection of stories, *A Sportsman's Sketches* (1852), indicated the direction which the Russian story was mainly to take. These stories, grounded in the ordinary circumstances of peasant life as Gogol's had been, have little of the appearance of the striking combination of incidents leading to a climax which Poe had recommended in his review of Hawthorne's tales ten years before. Each story catches a character in a particularly revealing incident, one which illuminates the whole life of the individual in an impression which seems deceptively casual. In one of his best stories, "The Country Doctor," Turgenev seems to express wonder at the very method which he invented: "It's strange how things happen in life: you live with someone for a long time, you are on the best of terms, yet you never once speak to them frankly and from the heart; with someone else, you've hardly even got acquainted—and there you are: as if at confession, one or other of you is blurting out all his intimate secrets." The overheard revelation of the "intimate secrets" of a character or characters is the primary intention of Turgenev's stories, and the representation of character rather than the combination of striking incidents is his primary method.

Thus in the contrasting emphases of Poe and Turgenev two different approaches to the short story appeared, the one assuming that the proper method of the story was to invent incidents which when plotted to a suspenseful climax offered a sweeping total effect, and the other assuming that the proper method was to begin with the representation of characters who themselves seemed to suggest incidents that would reveal the quality of their lives.

This basic division can be seen again in the contrasting methods of the two most influential short story writers of the beginning of the 20th century—one French and the other Russian. Guy de Maupassant mastered the method of Poe, bringing to it the compression, the economy of detail and the vivid precision of style which were typical of the French story almost from the beginning. His ingeniously constructed plots and striking climaxes were imitated by writers on both sides of the Atlantic, to whom such a story as "The Necklace" seemed a perfect model. That the mannerisms and tricks of the Maupassant story should have been adopted by great numbers of writers of the popular magazine story and converted into a sterile formula does not deny the achievement of Maupassant himself, whose ironic pessimism and realistic examination of life set him apart from the writers who packaged his wares for mass consumption. The influence of Anton Chekhov, on the other hand, was not so much to be found among the writers of the popular story as among the large number of young writers of the 1920s in England and the United States, to whom the plotted story of Poe and Maupassant, as brought to a superficial polish in the tales of O. Henry, Jack London and Rudyard Kipling, seemed to eliminate the possibilities of individual experiment.

Chekhov's stories, like Turgenev's, focused on revealing moments in the lives of his characters, and their climax comes not so much in a striking incident as in a growing understanding of the nature of a character and his situation. A story he argued, should have neither beginning or end, for it was at these points that the author was tempted to impose an artificial neatness and certainty on the complexity of life. The unity for which he strove was not an external resolution of a complicated action, but rather an internal unity of revelation of character.

Each tradition, the Poe-Maupassant story emphasizing plotted action or the Turgenev-Chekhov story emphasizing revelation of character, could of course provide successful stories, for plot and characters are inextricably intertwined in a significant story. The best stories of the 20th century proceed from an organic theory of the form, for which Henry James became not only a tireless explicator but also a distinguished practitioner. In his important essay "The Art of Fiction" (1884) James argued that a piece of fiction "is a living thing, all one and continuous, like any other organism, and in proportion as it lives will it be found . . . that in each of the parts there is something of each of the other parts . . . What is character but the determination of incident? What is incident but the illustration of character? . . . It is an incident for a woman to stand up with her hand resting on a table and look at you in a certain way; or if it be not an incident I think it will be hard to say what it is. At the same time it is an expression of character." James's preoccupation with discovering the proper form for each work, in which each of its elements contributes fully to its effect and in which the total structure embodies its meaning, is reflected less in his immediate influence on his contemporaries than in the influence which his fiction and his criticism had on English and American writers after World War I.

The Short Story in English.—James's own contemporaries in the U.S. were developing two other schools of the short story, both of which seemed more promising at that time than they did several decades later. Near the beginning of James's career, the "local colour" movement produced hundreds of stories aimed at exploiting the regional peculiarities of the different sections of the United States. Among the best of the local colourists were Bret Harte and Mark Twain in the west, Hamlin Garland in the midwest, George Washington Cable and Joel Chandler Harris in the south, and Sarah Orne Jewett and Mary E. Wilkins Freeman in New England. At best, the movement provided its readers with genuinely imagined representations of the regional life of a nation only recently torn by civil war, and it extended the range of short fiction into localities not before explored by the short story writer. At worst, the local colourists merely dwelt upon the quaintness of local characters and dialect, and marketed an "atmosphere" applied with a broad brush. Near the end of James's career "the O. Henry manner" was in vogue. The stories of O. Henry, with their rush of incident, crackling style and startling trick endings, achieved for their author an immense popularity and a host of imitators who filled the popular magazines with stories written to his formula. Later critics have found his work to be lacking in intellectual depth and his central theme—the effect of coincidence

on character—scarcely worthy of the endless variations he played upon it; yet for many ordinary readers in the United States and England, O. Henry is pleasantly remembered as "a man who could tell a story."

Although storytelling has been a folk pastime in England for centuries and short fiction has been written in England from the time of "The Green Knight" of Arthurian romance on to the present, the English short story was quite undistinguished until late in the 19th century. The masters of the English novel—Sir Walter Scott, Charles Dickens, George Meredith and Thomas Hardy—occasionally attempted short works, but they gave little evidence of comprehending the potentialities of the form which was maturing in Russia, France and the United States. Short fiction filled the English magazines, but most of the stories were merely amusing anecdotes or sentimental and pious recountings of minor domestic crises.

During the 1880s, however, the influence of such American writers as Irving, Poe, Bret Harte and Ambrose Bierce began to be felt in the romances and fantasies of Robert Louis Stevenson, Oscar Wilde and the contributors to *The Yellow Book*. It was in Rudyard Kipling, however, that the short story in England found a talent comparable to Maupassant in France. Kipling's stories abounded in flamboyant effects and striking incidents of British colonial life; well plotted and colourful, they provided the models for many English writers during the first two decades of the 20th century. Yet it was not until after World War I that the short story in Great Britain really flourished. The major novelists—Joseph Conrad, James Joyce, E. M. Forster, Virginia Woolf and D. H. Lawrence—showed an awareness of the short story form that their 19th-century counterparts had lacked. Ironically, the two writers whose influence has been strongest on the 20th-century story were Joyce of Ireland and Katherine Mansfield of New Zealand. Joyce's *Dubliners* (1914) showed a remarkable ability to transmute the incidents of ordinary life into classic art through richness of imagery, objectivity and restraint. Katherine Mansfield's delicate stories, rendered in poetic prose and focusing upon internal, psychological conflicts, have an obliqueness of narration and a subtlety of observation that reveal plainly the influence of Chekhov.

Though probably not so influential in determining new courses for the short story as Joyce and Mansfield, W. Somerset Maugham undoubtedly had the widest public of any English short story writer throughout almost half of the 20th century. A sleek objectivity and an incisive dissection of human frailties are the marks of his stories; their sardonic comedy is an acid variation on the wry humour that characterizes many English stories. A later generation of writers—V. S. Pritchett, Graham Greene, H. E. Bates, Elizabeth Bowen and William Sansom—have demonstrated that the short story in England reached maturity by the middle of the 20th century, after scarcely 50 years of distinguished achievement.

The 20th-century short story in the United States has been equally rich, justifying Poe's prediction in 1842 that the form could encompass "a vast variety of modes or inflections of thought and expression." As in England, the major novelists—Willa Cather, F. Scott Fitzgerald, Hemingway, William Faulkner, John Steinbeck and Robert Penn Warren—were likely to be equally at home in the short story and the novel.

From Stephen Crane to Eudora Welty, the U.S. writer, whose only tradition sometimes seems to be to break with tradition, was constantly evolving new individual forms for his vision of life. His work has been marked by an intense devotion to craft, a thorough grounding in actuality without false sentiment or moralizing and an ability to discover the universal in the regional. His American "voice," a robust colloquial speech, has gone through successive stages of refinement in the style of Mark Twain, Sherwood Anderson and Hemingway. His preoccupation with the regional qualities of U.S. life, particularly the south (as in Faulkner, Warren, Katherine Anne Porter and Eudora Welty) and the midwest and west (as in Anderson, Steinbeck and Walter Van Tilburg Clark), has come not merely from the attraction of local colour but rather from the hope that the human spirit could best be caught at close range. His concern for his native scene has not prevented influences from Europe from being absorbed into the American short story, particularly the work of Thomas Mann, Franz Kafka, Katherine Mansfield, Joyce, Chekhov, Maupassant and Flaubert, and in turn the impact of the U.S. story has been widely felt in Europe.

See also NOVEL and biographies of major short story writers.

BIBLIOGRAPHY.—F. L. Pattee, *The Development of the American Short Story* (1923); H. E. Bates, *The Modern Short Story* (1941); Cleanth Brooks and Robert Penn Warren, *Understanding Fiction* (1943); Eugene Current-García and Walton R. Patrick (eds.), *American Short Stories* (1952); Ray B. West, Jr., *The Short Story in America, 1900–1950* (1952). (W. L. Ps.)

SHOSHONE (SHOSHONI), a North American Indian group that in historic times occupied the territory from southeastern California across central and eastern Nevada and northwestern Utah into southern Idaho and western Wyoming. Comanche (*q.v.*) are a comparatively recent offshoot of Wyoming Shoshone. Shoshone, together with Southern Paiute-Ute and Northern Paiute, comprised the three divisions of the Basin-Plateau branch of the Shoshonean language. Shoshone dialects are so similar that speakers from Death Valley, Calif., have no difficulty conversing with Comanche.

Shoshone of historic times may be roughly divided into four groups: Western (unmounted) Shoshone, centering in Nevada, lacking horses, and early designated "Diggers" along with other far western Indians; Northern (mounted, or horse) Shoshone of northern Utah and Idaho; Wind River Shoshone in western Wyoming; Comanche in west Texas.

Western Shoshone had no band organization, and were divided into loosely affiliated family units that subsisted on wild seeds, small mammals, and insects. Each family was independently nomadic during most of the year and joined other families only briefly for rabbit drives, antelope hunts, or dancing. Hostilities were confined to feuds between families. A few Western Shoshone obtained horses after Europeans settled Nevada and Utah.

Wind River Shoshone and Northern Shoshone probably acquired horses as early as 1700, before Europeans occupied their lands. They formed bands of mounted buffalo hunters and warriors, and acquired such Plains Indian traits as tepees, skin clothing, and counting coup as war honours (*see* PLAINS INDIANS).

After acquiring horses, the Comanche split off from the Wind River Shoshone and moved south into Texas. Since raids on the Spaniards of the Southwest were the principal source of horses for Plains Indians, the Comanche developed predatory bands which were feared because they subsisted as much by plunder as by buffalo hunting.

Wind River Shoshone population is estimated to have been 2,500 in the early 19th century, and that of the Comanche 7,000. Northern and Western Shoshone together did not exceed 10,000 at that time. In the 1960s the combined total was about 7,500 for all four groups; most lived on reservations in territory they once held.

See also HOPI; PAIUTE; UTE; MONO.

BIBLIOGRAPHY.—J. H. Steward, *Basin-Plateau Aboriginal Sociopolitical Groups* (1938); V. C. Trenholm and M. Carley, *The Shoshonis, Sentinels of the Rockies* (1964); R. F. Murphy and Y. Murphy, *Shoshone-Bannock Subsistence and Society* (1960). (J. H. SD.)

SHOSTAKOVICH, DMITRI DMITRIEVICH (1906–), the leading Russian composer of the Soviet era, was born in St. Petersburg on Sept. 25 (new style; 12, old style), 1906. During 1916–18 he studied at Glasser's School of Music and in 1919 entered the Petrograd conservatory, where for four years he studied the piano under Leonid Nikolaev and composition under Maximilian Steinberg. After leaving the conservatory he found it difficult to decide between the careers of virtuoso and of creative artist, but the remarkable success of his First Symphony (1926), not only in the Soviet Union but throughout Europe and the U.S., set the seal on his decision to concentrate on composition. His next works, the Second Symphony ("Dedication to October," 1927) and an opera based on Gogol's *The Nose* (Leningrad, 1930), were thoroughly contemporary in feeling—*The Nose* fully revealed Shostakovich's penchant for parody and the grotesque—

and were sharply criticized for their decadent "western" sophistication. In the Third Symphony ("First of May," 1931) and three ballets—*The Golden Age* (1930), *Bolt* (1931), and *The Limpid Stream* (1935)—Shostakovich chose Soviet themes and modified his musical idiom, though without appeasing his critics. His second opera, *The Lady Macbeth of Mtsensk* (after N. S. Leskov's story, later renamed *Katerina Izmailova*), after being hailed as a masterpiece on its production at Leningrad on Jan. 22, 1934, was two years later denounced by *Pravda* as "Muddle instead of Music." Awed by this severe official criticism, he withdrew and suppressed his Fourth Symphony while it was in rehearsal.

After almost two years of nearly complete silence Shostakovich produced "a Soviet artist's practical creative reply to just criticism" in the form of a Fifth Symphony, which was received with approval. The period immediately before the German invasion was marked by the first of his series of string quartets (1938), the Sixth Symphony (1939), the popular Piano Quintet (1940), and a large quantity of film music, a genre in which Shostakovich was prolific from 1929 onward. Within a few months after Russia's entry into World War II Leningrad was beleaguered and Shostakovich wrote the first three movements of the quasi-programmatic Seventh Symphony ("Leningrad") during the siege (1941). Monumental and heroic in conception, the symphony, first performed at Kuibyshev (1942), enjoyed enormous popularity at the time both in the U.S.S.R. and in the other Allied countries. Shostakovich turned from this score to another Gogol opera, *The Gamblers*, which he abandoned unfinished. Neither the similarly monumental Eighth Symphony (1943) nor the relatively flippant Ninth (1945) won anything like the approval accorded to the Seventh and, after another Communist pronouncement on musical policy in 1948, Shostakovich turned for a time from chamber music (quartets No. 2 and 3, Piano Trio) mainly to choral works, such as *Song of the Woods* (1949), and to film music.

Shostakovich visited the U.S. in 1949 and in 1958 he made an extended tour of western Europe, including Italy (where he had already been elected an honorary member of the Accademia nazionale di Sta. Cecilia) and Britain (where he received an honorary doctorate of music at Oxford). His most important later compositions include the problematic Tenth Symphony (1953), the monumental Eleventh and Twelfth (1957 and 1961), commemorating the insurrection of 1905 and the Revolution of 1917, and the Thirteenth (1962) consisting of choral settings of poems by Y. Yevtushenko; the 24 Preludes and Fugues for piano (1951); a Violin Concerto (performed 1955, though written much earlier), a second piano concerto (1957; the first, accompanied by trumpet and strings, dates from 1933); a cello concerto (1959); and five more quartets.

Although brilliantly gifted, Shostakovich from 1930 onward revealed in his work a conflict between, on the one hand, his natural inclinations toward the witty and lighthearted expressed in a pungent harmonic idiom and, on the other, the effort to produce music that, according to Soviet standards, should be optimistic, monumental, and "democratically acceptable."

BIBLIOGRAPHY.—I. Martynov, *Dmitri Shostakovich* (1947); D. Rabinovich, *Dmitry Shostakovich* (1959); V. I. Seroff, *Dmitri Shostakovich: the Life and Background of a Soviet Composer* (1943). (G. AB.)

SHŌTOKU TAISHI (574–622), regent of Japan, one of the most important contributors to development of Japanese civilization, who brought to it new political, religious, and artistic institutions, was a nephew of the empress Suikō. He became crown prince and regent in 593. By means of persuasion and political maneuver, he expanded the power of the imperial house, bringing back into its hands powers that had been delegated to the nobles.

A brilliant scholar, Shōtoku was instrumental in introducing Chinese culture to Japan. His motives were not entirely intellectual and aesthetic, however, for he saw in Chinese political institutions and technology a means of building a stronger central government in Japan. His 17-article constitution expressed the ideals of a centralized bureaucratic state on the Chinese order; he was thus the forerunner of the Taika reform. Shōtoku was a devout believer in Buddhism, and in order to spread that faith through the country he began the building of temples in areas far beyond the capital city. After his death, he was looked upon in popular belief as a Buddhist saint. *See* also JAPAN: *History*.

See J. and R. K. Reischauer, *Early Japanese History* (1937); G. B. Sansom, *History of Japan*, vol. 1 (1958). (T. C. SH.)

SHOT PEENING: *see* BLAST CLEANING AND SHOT PEENING.

SHOT-PUT, a sport in which a spherical weight is thrown or put from the shoulder for distance. It derives from the ancient sport of putting the stone. First to use a shot instead of a stone competitively were British military sports groups, and later the idea was adopted by civilians. While the weight varied from 8 to 24 lb., the 16-lb. (7.257-kg.) shot was adopted for men in Olympic Games, national and international competitions. Weight for women's competition is 8 lb. 13 oz. (4 kg.), and for high school girls 8 lb. A 12-lb. shot is used for high school boys' competition, ages 15–19.

Constant improvements in technique resulted in the doubling of record distances. The International Amateur Athletic Federation recognized the first official world record as 30 ft. 11½ in., by J. M. Mann of the United States in 1876. E. J. Borr of Great Britain surpassed 40 ft. in 1880 and Ralph Rose (U.S.) in 1909 became the first to exceed 50 ft. Speculation over man's ability to attain 60 ft. paralleled, in a lesser way, the topic of man's ability to run the 4-min. mile. Interestingly, the 60-ft. shot-put was achieved two days after Roger Bannister, the British runner, broke through the 4-min. barrier. Parry O'Brien (U.S.) put the shot 60 ft. 5¼ in. on May 8, 1954. In 1960 Bill Nieder (U.S. Air Force), who beat O'Brien in the Olympics that year, raised the world record to more than 65 ft. D. C. Long (Southern California Striders) won the 13th Olympic title for the U.S. with 66 ft. 8¼ in. at Tokyo in 1964 and on May 8, 1965, Randy Matson (Texas A. & M.) broke the 70 ft. barrier with a throw of 70 ft. 7 in. (21.718 m.).

In women's competition with the 4-kg. shot, Galina Zybina of the U.S.S.R. brought the record above 50 ft. when she won the 1952 Olympic championship. In 1960 Tamara Press of the U.S.S.R. pushed beyond the 55-ft. barrier with an Olympic record of more than 56 ft. and a world record of more than 58 ft. In 1962 Miss Press extended the world record to 60 ft. 10¼ in.; and in 1964 she raised the Olympic record to 59 ft. 6$\frac{3}{16}$ in. (18.307 m.).

The shot generally is made of solid iron or brass, although any metal not softer than brass may be used. It is put from a circle 7 ft. (2.135 m.) in diameter.

Technique.—The put is started at the back of the circle by shifting the weight of the body to the right foot with good balance, left knee bent and the left toes raised so that the spikes clear the ground. Next, the left knee is brought close to the right knee and the right leg bent gradually as the hips cross the circle. This method of starting the put enables the shot-putter to be in a correct position when he reaches the centre of the circle and thus able to start lifting the shot immediately. The shot is lifted via the medium of the leg muscles, and the put must always be made upward and not outward by a rhythmic movement of the arm muscles, which can be acquired by constant repetition.

Whereas it was conventional to start from a position at right angles to the direction of the put, O'Brien developed a style of beginning from a backward position. Thus he brought the implement around 180°, rather than 90°. He found that the longer one pushed the shot, applying force against a moving object, the longer the throw would be. *See* OLYMPIC GAMES.

BIBLIOGRAPHY.—Frank G. Menke, *The Encyclopedia of Sports* (1960); Amateur Athletic Union, *Official Track and Field Handbook* (1963); National Collegiate Athletic Association, *Track and Field Guide* (1963); G. T. Bresnahan, W. W. Tuttle and F. X. Cretzmeyer, *Track and Field Athletics*, 4th ed. (1956). (H. AR. RN.; SY. S.)

SHOTTS, a town of Lanark, Scot., 18 mi. E of Glasgow by road, and on the Glasgow-Edinburgh Railway. Pop. (1961) 18,924. There are iron works, quarries, coal mines, hosiery factories, and engineering works. About 2 mi. NW, at Kirk of Shotts, is a BBC Scottish television transmitter.

SHOWBOAT, an American folk institution which for more than a century did much to relieve cultural starvation on river frontiers of the middle west and the south. The earliest of these entertainment boats penetrated regions where churches, schools,

newspapers and theatres had not gone.

The British-born actor William Chapman (1760–1841), built the first showboat, the "Floating Theatre" (14 by 100 ft.), at Pittsburgh, Pa., in 1831. He and his family floated from landing to landing, playing such dramas as August Friedrich Ferdinand von Kotzebue's *Menschenhass und Reue* (translated as *The Stranger*) and Shakespeare's *The Taming of the Shrew*, with music and dance specialties added. Upon reaching New Orleans late in winter, they junked the boat and returned by steamer to Pittsburgh, to repeat the pattern the next year. After Chapman's death, showboats first degenerated and then, because of the Civil War, disappeared entirely.

When showboats were revived (1878), they specialized in vaudeville and melodrama. Steamer tows and the calliope greatly increased territory and audiences, and Stephen Foster's songs added sentimental charm to their programs. Such boats as A. B. French's "New Sensation," McNair's "New Era," E. A. Price's "Water Queen," Thom's "Princess" and dozens of others, with seating capacities of 100 to 300, carried their rich cargoes of humour, music and simple emotion on every river of the Ohio-Mississippi system, from the narrow Monongahela in the northeast to the Atchafalaya bayous (creeks) in the south.

With the disappearance of river frontier conditions in the 1900s, and with the coming of better roads, automobiles and motion pictures, the decline of showboats was inevitable. To compete with land entertainment, they became larger and more elaborate. Menke's "Goldenrod" seated 1,400 persons and cost more than $35,000. The "Cotton Blossom," the "Sunny South" and Bryant's "New Showboat" were floating theatre palaces. All emphasized melodrama. With the burlesquing of these programs in the 1930s to attract sophisticated audiences, showboats ceased to perform their original function. The last one to travel the rivers in authentic pattern was the "Goldenrod" in 1943.

See Philip Graham, *Showboats: the History of an American Institution* (1951). (PH. G.)

SHOW JUMPING: *see* HORSEMANSHIP AND RIDING.

SHREVEPORT, a city of northwestern Louisiana, U.S., on the Red River, 317 mi. (510 km.) NW of New Orleans and 185 mi. (298 km.) E of Dallas, Tex. Shreveport is the centre of a standard metropolitan statistical area that covers Caddo and Bossier parishes. Pop. (1960) city 164,372; standard metropolitan statistical area, 281,481. (For comparative population figures *see* table in LOUISIANA: *Population*.) The area is slightly hilly except for parts of the river valley, has several lakes, and is forested with hardwood and pine.

Henry Miller Shreve, after clearing the Red River of a huge natural timber raft, founded the city in 1835 on lands obtained from the Caddo Indians at the head of navigation on the river. Cotton and river traffic formed the economic base, and Shreveport's association with nearby Texas is indicated by the many city streets named for heroes of the Texas revolution. At the end of the Civil War, Shreveport was the Confederate state capital and headquarters for the Trans-Mississippi forces of the Confederacy. Using logs to simulate artillery, its defenders constructed Fort Humbug to command the river; but the only major attempt to take the city was halted 40 mi. S at Mansfield. By 1900 the river was no longer used for commerce but railroads and cotton trade kept the city growing. Discovery of oil in the area in 1906 gave the city new impetus, and oil and gas continue to be of major economic significance.

A police jury administers Caddo parish except for the incorporated places of Shreveport, Vivian, Mooringsport, and Oil City. Of the 1960 population of the metropolitan area, 77.8% was urban, 18.5% rural nonfarm, and 3.7% rural farm; about 65% was white, 34% nonwhite, and 1% foreign born.

The area produces lumber, glass, metals, machinery, foods, and chemical and petroleum products. Cotton remains the chief farm crop, and the city is the wholesale and retail trade centre for the region known as the Arklatex. Educational facilities include a modern parish-wide public school system, a state trade school, and Centenary College of Louisiana, a Methodist liberal arts institution founded in 1825. Shreveport has a public library system, a symphony orchestra, a repertory opera company, and three amateur theatres. The Louisiana State Fair and Holiday in Dixie, a spring festival, are major annual events. Parks, golf courses, water sports and fishing on Cross, Caddo, and Wallace lakes, a municipal auditorium, a civic theatre and convention centre, a large stadium, and an indoor arena are available. Three highway bridges connect Shreveport with Bossier City (pop. [1960] 32,776), the location of Barksdale Air Force Base. (W. M. Lo.)

SHREW, a high-strung rodentlike mammal of the family Soricidae, order Insectivora. The common shrews or shrewmice of the northern hemisphere, belonging to the genus *Sorex*, are about the size of mice. They are distinguished externally by the long, pointed muzzle, which projects far beyond the lower lip. The small eyes are almost concealed by fur; the ears are short, broad, and provided internally with deep folds; the tail is usually slightly shorter than the body, generally quadrangular in section, and thinly haired, often with a small tuft at the tip. The fur is short and velvety, grayish or brownish in colour. On the side of the body there is a gland secreting a fluid with an unpleasant odour, a fact which renders shrews unpalatable to many larger predators. The short-tailed shrews, genus *Blarina*, common to eastern North America, are reputed to have a poisonous bite; the toxin, however, is only powerful enough to affect the shrews' favourite prey.

JOHN H. GERARD

AMERICAN SHORT-TAILED SHREW (BLARINA BREVICAUDA) FEEDING ON A SNAIL

Shrews feed largely on insects, snails, worms, and small mammals and are active day and night and in all seasons. They usually live in moist situations, burrowing in leaf mold in search of their food. They must feed almost continuously because of their very rapid metabolic rate and are said to consume their own weight in food about every three hours. In the sorecine shrews all teeth are tipped with red pigment, while crocidurine shrews of Africa, Asia, and part of Europe have white teeth. Skulls of shrews have no cheek arch. Several shrews are minute, among the smallest of mammals: the pygmy shrew (*Suncus etruscus*) may weigh as little as 2 g. and be less than 2 in. long. The largest species is probably one of the Asiatic water shrews, *Chimarrogale*, some 6 in. long, the tail 3 in.

The tree shrews are small, arboreal, insectivorous mammals found in southern Asia and in the islands of the Malay archipelago. They are squirrellike but have long pointed muzzles; their teeth are like those of the Insectivora. The common tree shrews (*Tupaia*) are dark olive brown with bushy tails. Pen-tailed tree shrews (*Ptilocercus*) are dark gray-brown with a black facial mask and narrow tail, ending in a featherlike tuft. Tree shrews, active during the day, feed on insects, some vegetable matter and the young and eggs of birds. Mammals of this type doubtless gave rise to the Primates (*q.v.*), and some authorities classify these forms with that order, others consign them to the Insectivora (*see* INSECTIVORE). (J. E. HL.; X.)

SHREWSBURY, EARLS OF. The title of earl of Shrewsbury was first held by the Norman baron, ROGER DE MONTGOMERY (d. 1094), who married the heiress Mabel of Belesme. In 1066 he remained behind in Normandy to assist in the government of the duchy, but was granted Arundel, Chichester, and other Sussex estates when he went to England in 1067. As a trusted councilor of King William, he was created earl of Shropshire in December 1074, a position which gave him palatine control of that county and placed him among the greatest of the Marcher lords; but he and his successors were usually styled earls of Shrewsbury. A great patron of monasticism, he became a monk in his newly founded abbey of Shrewsbury just before he died in 1094. His Norman inheritance passed to his eldest surviving son, Robert of Belesme (*q.v.*), and the title and the English lands went to Robert's younger brother, HUGH (d. 1098). With the earl of Chester, Hugh led the Norman conquest of north Wales and was killed by Viking raiders in Anglesey (1098). ROBERT OF BELESME

(*c.* 1052–*c.* 1131) succeeded as 3rd earl, but when he was deprived of all his honours for rebellion in 1102, the earldom became extinct. It was revived in 1442 for JOHN, LORD TALBOT (d. 1453) whose descendants have borne the title to the present day (*see* TALBOT; SHREWSBURY, JOHN TALBOT, 1st Earl of). As the possessor of the most ancient English earldom not merged in a higher dignity, the earl of Shrewsbury now ranks as the premier earl of England. (C. D. R.)

SHREWSBURY, CHARLES TALBOT, DUKE OF (1660–1718), English statesman who played a leading part in the Revolution of 1688 and was largely responsible for securing the Hanoverian succession in 1714, was born on July 24, 1660, the son of Francis (d. 1668), 11th earl of Shrewsbury. Brought up as a Roman Catholic, he was educated mainly in France during the years 1674–78. In 1679, influenced by John Tillotson, then dean of Canterbury, he joined the Church of England, thus qualifying himself, although under age, to take his seat in the House of Lords and to accept appointment as lord lieutenant of Staffordshire.

He first became active in politics in the reign of James II, when he resisted every temptation to return to the Catholic Church and became a leader in the conspiracy which led to the Revolution of 1688. The early meetings of the conspirators were held in his house in London; he was one of the famous seven who signed the letter of June 30, 1688, inviting William of Orange to come over with an armed force to England. In September he joined William in Holland and in November returned to England with his expedition, securing Bristol and Gloucester for him. In the Convention Parliament which assembled on Jan. 22, 1689, he advocated the elevation of William and Mary to the throne, and as soon as they had been proclaimed as William III and Mary II, he became a privy councilor and the secretary of state for the southern department.

Although moderately successful in office Shrewsbury was never happy there. Responsibility weighed upon him, and in the party rivalries following the Revolution he was unable either fully to support the aims of his Whig associates or to repudiate them altogether. On June 2, 1690, disgusted at the growth of Tory influence in the government, he resigned his secretaryship, and for four years, on the plea of ill-health, refused to resume office. This obstinacy exposed him to being called a Jacobite, an imputation somewhat supported by his association with the earl of Marlborough, whose conduct was giving great dissatisfaction to the king. On June 23, 1692, accordingly, he and Marlborough were struck off the Privy Council, and for a year thereafter he opposed the government, criticizing their conduct of the war and in 1693 advocating in the House of Lords a Triennial Bill, the first of a series which led to the Triennial Act of 1694. Nevertheless, William still liked and trusted him, and on March 2, 1694, when the Whigs were regaining influence at court, he accepted office as secretary of state for the northern department, transferring shortly afterward to the southern department. On April 25, 1694, he was made a knight of the garter and on April 30 was created duke of Shrewsbury.

BY COURTESY OF THE NATIONAL PORTRAIT GALLERY, LONDON

CHARLES TALBOT, DUKE OF SHREWSBURY. PORTRAIT BY THE SCHOOL OF SIR GODFREY KNELLER, ABOUT 1685

His second period of office, though longer, was no more satisfactory than his first. Ill-health, real or imaginary, continued to worry him, and his chief energies were directed toward securing permission to retire. Although appointed one of the lords justices for the administration of the country during the king's absences on the continent in 1695, 1696, and 1697, he was accused during these years by the conspirator Sir John Fenwick and by the informer Matthew Smith of complicity in Jacobite intrigues, and he practically ceased to take any part in the government. On Dec. 12, 1698, he was allowed to resign his secretaryship, accepting in October 1699 the less responsible post of lord chamberlain; but this too he resigned on June 20, 1700.

Leaving England in November 1700, Shrewsbury spent the next five years in travel on the continent. On his return early in 1706 he became increasingly dissatisfied with his former Whig associates, and in 1710 went over to the Tories. He was appointed lord chamberlain on April 14 and a privy councilor on April 16, 1710, lord lieutenant of Ireland on Sept. 22, 1713, and lord high treasurer on July 30, 1714, when Queen Anne was on her deathbed. This last great office enabled him to exercise a decisive influence in the peaceful recognition of George I, who appointed him groom of the stole and keeper of the privy purse on Sept. 26, a privy councilor on Oct. 1, and lord chamberlain on Oct. 17, 1714. Shrewsbury died in London on Feb. 1, 1718, when his dukedom became extinct and his earldom passed to a cousin.

Shrewsbury was regarded as one of the finest and most cultured gentlemen of his day, but an excessive sensitiveness and a reluctance to face the harsh realities of life greatly impaired his effectiveness as a statesman. At the Revolution and on the death of Anne he pursued a bold and determined policy, but as a rule his attitude was weak and vacillating. His declaration that if he had a son he "would sooner breed him a cobbler than a courtier, and a hangman than a statesman," is a reflection on himself no less than on his political world.

See T. C. Nicholson and A. S. Turberville, *Charles Talbot, Duke of Shrewsbury* (1930); D. H. Somerville, *The King of Hearts* (1962). (A. Bo.; X.)

SHREWSBURY, JOHN TALBOT, 1ST EARL OF (*c.* 1384–1453), English soldier, celebrated for his exploits in the Hundred Years' War, was the second son of Richard, 4th Baron Talbot, and Ankaret, heiress of the last Lord Strange of Blackmere. By the death of his elder brother Gilbert's daughter Ankaret (1421) he acquired the baronies of Talbot and of Strange. In or before 1433 he married his second wife, Margaret (d. 1467), eldest daughter of Richard Beauchamp, earl of Warwick, and coheiress of the barony of Lisle.

After campaigns in Wales between 1404 and 1413, Talbot served as lieutenant of Ireland from 1414 to 1419. He was later to be justiciar of Ireland (1425) and lieutenant again (1445–47). In these offices he showed great vigour: the Irish said that "there came not from the time of Herod any one so wicked in evil deeds."

After 1419, however, most of Talbot's career was spent in France, where he became one of the foremost English captains. He fought at Verneuil in 1424 and took part in the siege of Orléans in 1429. His rashness was chiefly to blame for the English defeat at Patay in June 1429, when he was captured by the French. Released in 1433, he took Clermont in 1434 (the countship was then assigned to him), was present at the siege of Saint-Denis in 1435 and suppressed the revolt of the Pays de Caux in 1436. Henry VI made him a marshal of France. The mainstay of the English cause in France for the next five years, Talbot defeated the Burgundians before Le Crotoy in 1437 and took Harfleur in 1440. In May 1442, on a visit to England, he was created earl of Salop (Shrewsbury was the name that he himself used for the title), but the next winter, before Dieppe, he "fared so foul with his men that they would no longer abide with him" and was forced to raise the siege. His fighting qualities made him something of a popular hero: in the rhymes of the day he was "Talbot our good dog," whose valour was frustrated by the duke of Suffolk's treason.

After the fall of Rouen in October 1449, Shrewsbury was held for a time as a hostage by the French (1449–50); a pilgrimage to Rome was stipulated as a condition of his release. In command of the expedition for the relief of Gascony, he landed in the Médoc in October 1452; the Bordelais rose in his favour and Fronsac fell to him in the summer of 1453. The French, however, laid siege to Castillon. Shrewsbury hurried to its relief and on July 17, 1453, attacked the enemy in their entrenched camp without waiting for his artillery. He and the flower of his troops were killed, and English rule in Aquitaine ended with them. (T. B. P.)

SHREWSBURY, an ancient market centre, municipal borough and the county town of Shropshire (Salop), Eng., in the Shrewsbury parliamentary division, 153 mi. N.W. of London by road. Pop. (1961) 49,566. Area 12.7 sq.mi. (32.9 sq.km.). Its position in relation to the routes leading into Wales and along the border has made it a town of great importance. The old and central part of the town lies almost entirely on a peninsula within the remarkable southward loop described by the river Severn. The strategic position of this peninsula rising above the Severn flood lands was recognized from the time of the princes of Powys, who made it their seat, called Pengwern, in the 5th and 6th centuries. At the end of the 8th century it was engulfed in the provinces of Mercia and given its Saxon name Scrobbesbyrig, from which comes Shrewsbury. In the reign of Edward the Elder it had a mint, and in Domesday Book it is styled a city. After the Norman conquest, Shropshire was granted to Roger de Montgomery, who made Shrewsbury his headquarters and founded the abbey. At about this time, alternative names arose, Sloppesbury or Salopsbury, from which comes the modern Salop. In the next 200 years it was frequently involved in wars with the Welsh, and in 1283 Edward I called an assembly of parliament to meet at Shrewsbury. In 1403 the battle of Shrewsbury was fought close to the town on the north side, when Henry IV by defeating the Percys consolidated his position on the throne. During the late middle ages and again in Tudor and Elizabethan times the establishment of law and order in the Marches and trade with the Welsh in wool and flax resulted in a period of great prosperity. In the Great Rebellion Shrewsbury supported the king, who took up his quarters there in 1642, but the town was captured by parliament in 1645. Subsequent development was continuous and Shrewsbury became the cultural, business and transport centre for a large area of surrounding country.

Shrewsbury is a borough by prescription. The town possesses more than 30 royal charters from the charter of Richard I (1189) to that of James II (1685) and earlier charters were granted by Henry I and II. The charter of Charles I gave to the town a new constitution, and under its authority the first mayor was elected in 1638. The borough returned two members to parliament from 1394 till 1885, when the number was reduced to one. It is now included in the Shrewsbury division of Shropshire.

There are many buildings, streets, lanes and "shuts" of architectural and historic merit, including 15th- and 16th-century timber framed houses of which Ireland's mansion, Owen's mansion, the Abbot's house and Rowley's house are outstanding examples; the latter contains a museum of Roman antiquities from the nearby Roman-British city of Uriconium. The old Market hall is a fine stone building dating from *c.* 1596; there are also many Georgian and regency houses. A considerable length of the old town wall, including one of the watchtowers, still remains. Within the loop of the river (but outside the town walls) is a public park called the Quarry. Over its two ancient bridges, the English and the Welsh, the town spreads east into Abbey Foregate and around Abbey church and west into Frankwell on the Welsh side. In the Mount, at the top of Frankwell, Charles Darwin was born. St. Mary's church, on high ground south of the castle, is a noble building with a lofty tower and spire dating from early Norman to Perpendicular. It has some remarkable stained and painted glass, including the great Jesse window of English glass dating from about 1345; the St. Bernard glass from the Abbey church of Altenburg, attributed by some to Albrecht Dürer (1471–1528); and 15th-century glass from the cathedral of Treves. St. Alkmund's, St. Julian's and St. Giles' are old foundations, much altered subsequently. Old St. Chad's church, dating mainly from the 12th century, was largely destroyed by the fall of the tower in 1788, and only the 15th-century Lady chapel remains. The new St. Chad's church erected in 1790–92 on another site is built to a circular plan. The restored Abbey church (Holy Cross) retains its massive Norman nave, built of deep-red sandstone, and two Early English arches and a Decorated western tower; of the monastic buildings little is left, save a remarkable roofed outdoor pulpit of ornate Decorated work. Shrewsbury is the seat of a Roman Catholic bishop.

A fortification of some sort probably existed on the site of Shrewsbury castle before the Norman conquest, but the first reliable evidence of a castle being constructed there is in 1067. In 1071 it was given to Roger de Montgomery, then made earl of Shrewsbury, who continued working on it, and it was entirely rebuilt by Edward I. It was held for the king in the Great Rebellion but fell to parliament in 1645. It continued fortified until the reign of James II, but subsequently came into private ownership. The castle was converted to residential use by Thomas Telford toward the end of the 18th century. In 1924 it was bought by the Shropshire Horticultural society, restored as nearly as practicable to its 14th-century condition, and given to the corporation; it is used as a council chamber.

Shrewsbury school (founded by King Edward VI in 1552) occupied buildings opposite the castle until 1882, when it was moved to Kingsland, south of the river, formerly the scene of the Shrewsbury show, a pageant held during the festival of the Trinity. The old school buildings in the town centre were taken over by the corporation and are now used as a library, museum and art gallery. The cross at the junction of Pride hill and St. Mary's street, on the site of the original High cross, was given by the school to the town in 1952 to commemorate the 400th anniversary of the school's foundation.

The cattle market, one of the busiest in England, was enlarged and moved to the north of the town in 1959. Also in the north are most of the main industries which include locomotive rolling stock and general engineering; the manufacture of machine tools, safes and electrical equipment; and malting. (S. R. H. L.)

SHRIKE (BUTCHER-BIRD), a robust, predacious bird, about thrush size, with a heavy bill hooked at the tip. Shrikes belong to the songbird family Laniidae, which contains about 73 species. Their plumage is usually boldly patterned and varies from gray, white, and black to browns, greens, and reds. Their call notes are harsh but their song may be pleasing. Their nests are rather bulky structures, with many twigs in them, and the eggs are spotted. The distribution of the family covers Africa, where most species live; Europe and Asia to New Guinea (one species); and North America (two species).

The loggerhead shrike (*Lanius ludovicianus*) is the common American species, found from southern Canada to Mexico. It is gray above, white below, with a black mask, and is about nine inches long. The more northern birds retire southward in winter. The northern shrike (*L. excubitor*) is similar but larger (ten inches long), with fine barring on the breast. It breeds in northern Canada and Alaska, as well as in the Old World. Some winter in the northern United States.

Typical shrikes sit upon vantage points, hawklike, watching for their prey of large insects, small reptiles, mammals, or birds. Though predacious in habit, they lack strong feet for holding their prey and some species impale their food on thorns, sometimes leaving part of it uneaten. This fancied resemblance to meat in a butcher shop is the reason for one of their common names. Certain bush shrikes and helmet shrikes of Africa are only doubtfully placed in this family. Birds of an unrelated Australian family of butcherbirds and bell magpies also have the habit of impaling food on thorns. The vanga shrikes of Madagascar and the cuckoo shrikes, widespread in the warmer parts of the Old World, form two other unrelated groups of songbirds. (A. L. RD.)

SHRIMP, a crustacean distinguished from such other larger long-tailed members of their order (Decapoda) as lobsters and crayfishes. The most obvious features of shrimp are the long legs, weak claws, and laterally compressed abdomen.

Although most species are quite small, some even microscopic, a few reach a length of up to eight inches. Commercial fishermen usually refer to larger individuals as prawns and to smaller ones as shrimp. Shrimps (suborder Macrura) differ in greater development of the paddlelike limbs of the abdomen used in swimming and in their thin and sometimes fragile shell or exoskeleton. Like other crustaceans, shrimps wear their skeleton on the outside of the body and, in order to grow, must cast off this shell and replace it with a new and larger one. In the process of shedding, all of the hard structures of the shrimp are cast off and renewed.

The common shrimp swims in a forward direction by the use of the pleopods or abdominal feet. When frightened or when rapid movement is desired, the shrimp with a flexing of the powerful, muscular abdomen, can propel itself backward with remarkable speed and also can leap clear of the water.

Shrimp occur on mud bottoms of inshore and offshore waters in many parts of the world. Important fisheries operate in European, North American and Asian waters. Although often thought of as a warm-water shellfish, shrimp are also found in northern seas, and there are commercial shrimp fisheries in the waters off Norway, Greenland and Alaska. Considerable effort was expended in the search for new shrimp grounds in many parts of the world following World War II, and important new fisheries for the brown-grooved shrimp (*Penaeus aztecus*) and the pink-grooved shrimp (*Penaeus duorarum*) were developed in the Gulf of Mexico. A red shrimp, *Hymenopenaeus robustus*, was found in the deeper waters of the gulf at depths of 180 to 350 fathoms.

There are several hundred species of shrimp. The catch of species of *Penaeus* probably exceeds that of all other crustacean fisheries in magnitude and importance. The world's greatest shrimp fishery, based principally on species of *Penaeus*, is in the Gulf of Mexico, where several hundred million pounds of these shellfish are taken annually by United States, Mexican and Cuban fishermen.

Shrimp are taken in a variety of ways—with hand or cast nets, baited traps, haul seines, stake or channel nets set in tideways and with boat-drawn beam and otter trawls. The trawls, which account for the major portion of the world catch, consist of large, baglike nets which are dragged over the floor of the ocean, scooping up the shrimp in their path.

Shrimp are marketed fresh, frozen, dried, canned and cooked-and-peeled. Shrimp bran, a by-product, is manufactured from dried heads and hulls and sold for animal feed. Development of machines for removing the hulls from the shrimp meat, and the production of fresh-frozen and frozen-breaded shrimp, have assisted in a remarkable expansion in the market for shrimp in the U.S., the world's largest producer and consumer of shrimp.

River shrimp or prawns of the genus *Macrobrachium* (*Palaemon*), found in most tropical countries, are also much used for food. (E. A. Pr.)

SHRINE, THE (Ancient Arabic Order of Nobles of the Mystic Shrine): *see* Freemasonry: *Appendant Orders*.

SHROPSHIRE or Salop, an English county bordering on Wales, with Cheshire to the north, Staffordshire to the east, Worcestershire and Herefordshire to the south. The geographical area is 1,347 sq.mi. Both names, Shropshire and Salop, derive from O. E. Scrobbesbyrigscir (shire with Shrewsbury as its head), Salop being an abbreviation of a Normanized form, Salopescira.

Physical Features.—Shropshire is crossed by the Severn, which divides the hilly south and west from the undulating plain to the north and east. The upland is composed of a series of ridges including Wenlock Edge, View Edge, Stiperstones, and Stapeley Hill, and hogbacks (Ragleth, Caradoc, Lawley, and the Wrekin), all running from northeast to southwest with deep valleys between them. Among them lies the Longmynd Plateau (1,696 ft.); south of Corvedale is a triangular plateau with two tabletopped heights, Brown Clee (1,790 ft.) and Titterstone Clee (1,749 ft.). A third mass of high ground, Clun Forest, lies between the rivers Clun and Teme. The northern plain (about 200 ft. or less) is broken by a band of sandstone knobs (Nesscliff, Grinshill, Hawkstone). Streams, such as the Tern, have little fall, and large areas were formerly marsh (*e.g.*, the Wealdmoors). Peat mosses (Whixall) and meres, especially near Ellesmere, Whitchurch, and Baschurch, give variety to the landscape. The eastern plain has been cut by rivers—the Worf, Bowhill Brook, Claverley Brook, etc.—into a series of valleys and low ridges (Shatterford, Tuckhill, High Rock) running NNW–SSE.

The geological pattern runs on similar lines, older rocks being found in the uplands and later rocks mainly under glacial drift in the plain, but this pattern was broken by volcanic upheaval. The main series of rocks are: Precambrian in Longmynd, Wrekin, and Ercall districts; Cambrian at Church Stretton; Ordovician in the Shelve area and a strip north of Wenlock Edge; Silurian in Wenlock Edge, Corvedale, View Edge, Clun Forest, and the Ludlow districts; Carboniferous limestone poorly represented but seen near the Wrekin, at Lilleshall and Llanymynech (four small Coal Measure Basins: Chirk, Hanwood, with continuation northwest, Coalbrookdale, and Wyre Forest). Trias is found principally north of the Severn above Shrewsbury, north of Wellington, and east of Shifnal and Bridgnorth. A small area of Lias is exposed around Prees; and glacial drift covers most of the plain. The soil and vegetation are as varied as this geological pattern would suggest. The climate is usually mild and humid, though very low winter temperatures have been registered at Shawbury.

History.—Considerable prehistoric traffic through Shropshire is indicated by portable antiquities of late Neolithic and Bronze ages (collections in Rowley's House, Shrewsbury; Clun and Ludlow museums) and by field monuments of the Bronze Age, *e.g.*, round barrows on the Longmynd and Oldfield near Ludlow, stone circles on Stapeley Hill, and remains in the Clun region. Hill forts of Early Iron Age character abound, especially the Bury ditches, Hopesay Burrow Camp, the two Caer Caradocs, the Wrekin, Bury walls, and Old Oswestry. Related island sites—the Berth, Baschurch, and Kynnersley Wall—were possibly occupied later.

A Roman legionary fortress existed in the 1st century at Viroconium (Uriconium), later the cantorial capital of the Cornovii and one of the largest towns in Britain. There were small towns or posting stations; *e.g.*, Uxacona (Oakengates). The country appears to have been sparsely cultivated, with farms at Acton Scott, Linley Hall, Cruckton, Yarchester, etc., but silverbearing lead ores and outcrop coal were exploited.

The Saxon conquest was marked by the construction of Watt's Dyke and Offa's Dyke. Anglo-Saxon villages are recognizable by their names, terminating in "bury," "ton," "ley," etc., while Celtic place-names include elements like "pentre," "llan," "tre"; and Offa's Dyke is generally the linguistic boundary. Some Welsh elements, however, may be attributed to later infiltration which has gone on since the 11th century.

After Danish incursions had been finally repulsed, Edward the Elder divided Mercia into shires, and Shrewsbury became a mint town and coins were struck there until the reign of Henry III. At the Conquest, large areas of Shropshire were set aside as forests, or hunting grounds, subject to special forest jurisdiction such as those of Morfe, Mount Gilbert, Shirlett, Clee, and Stiperstones. Roger de Montgomery was given extensive estates for the defense of the border. A double line of castles against the Welsh was established during the 12th and 13th centuries, and the history of medieval Shropshire is a chronicle of Welsh incursions and baronial rebellions.

All Shropshire's religious houses were for men. The most important (with foundation dates) were: Benedictine at Shrewsbury (1083) and Bromfield (refounded 1155); Cluniac at Wenlock (refounded 1080); Cistercian, originally Savigniac, at Buildwas (1135); Augustinian canons at Haughmond (*c.* 1135), Wombridge (*c.* 1130), Chirbury (from Snead) in the late 12th century, Lilleshall (Arrouaisians, from Donnington Wood; *c.* 1144); Grandmontines at Alberbury (*c.* 1220–1441); Templars at Lydley Heys (*c.* 1158–1308); Hospitallers at Halston (*c.* 1221); Dominican friars at Shrewsbury (1222); Franciscan friars, also at Shrewsbury (1245–46) and at Bridgnorth (1244); Augustinian friars at Ludlow (1282), Shrewsbury (1298), and Woodhouse near Cleobury Mortimer (1250); Carmelites at Ludlow (1349).

The Midland system of three open fields was general in Shropshire, but some traces of Welsh land custom are found.

The high quality of Shropshire wool brought prosperity in the 13th century, Ludlow, Shrewsbury, and Bridgnorth being the main centres. As trade grew, Shrewsbury became established as the principal market for an area including much of North Wales. There was plenty of heath available for sheep, and therefore no outcry was made against conversion of common field to sheepwalk. But Welsh raids, plague depopulation, and bad seasons brought a decline in arable farming. In areas of high rainfall, cattle raising became important. From the 15th to the 17th centuries there are indications of a general increase in the size of

holdings, piecemeal inclosure of common fields, and some attempts to drain and enclose marshes and wastes.

A prerogative court, the council in the Marches of Wales, was established about 1473, with headquarters at Ludlow and with jurisdiction in Wales and the four border counties to check disorder and provide quick, cheap, effective remedies for local litigants. It was suppressed in 1642, reestablished in 1660, and abolished in 1689.

On the outbreak of the Civil War many families declared for the king, who came in person to recruit, but Puritan influence divided the north and extreme south of the county. While Shrewsbury and Bridgnorth stood siege for the king, Wem and Hopton Castle stood siege for Parliament. Fortified country houses were skirmishing centres. After the Battle of Worcester in 1651, Charles II in his escape found many Shropshire friends.

In 1708 Abraham Darby I went to Coalbrookdale and developed his new method of smelting ironstone with coke from pit-coal. He found a tradition of ironworking, local supplies of ironstone, sulfur-free coal, limestone, and charcoal, an ideal site with water power in the Dale, and water transport on the Severn. In the 18th century, the Darbys, the Reynolds, John Wilkinson, and others introduced many improvements and made Shropshire the greatest iron-producing area in England. The first cast-iron bridge was erected at Ironbridge in 1779; the first iron-built boat floated on the Severn in 1787; one of the first experimental locomotive engines for use on a railway was built by the Coalbrookdale Company for Richard Trevithick in 1801. In the same period, Caughley and Coalport were producing fine china, and Broseley became noted for churchwarden pipes.

The county is celebrated in the famous anthology of poems, *A Shropshire Lad*, by A. E. Housman (1859–1936).

Architecture.—Good examples of ecclesiastical, military, and domestic architecture are to be found in Shropshire. The development of castle-building techniques from the 12th to 14th centuries may be studied at Ludlow, Clun, Bridgnorth, Shrewsbury, and Hopton castles, where considerable portions survive. Ludlow and Shrewsbury retain parts of their town walls. In domestic architecture, Shropshire possesses outstanding examples of town and country houses of every period from the 13th century. Shropshire had no dominant noble family but many ancient county families, no great seat but many fine country houses. Stokesay (*c.* 1260–80) and Acton Burnell (1283) show the fortified manor house; the Old Mint, Shrewsbury, and Forester's Lodge, Millichope (*c.* 1280), are comparable town and country dwelling houses; the Provost's House, Edgmond, is 14th century, and the Abbot's House, Shrewsbury, dates from about 1450. Many Shrewsbury houses are partly medieval. There are numerous Elizabethan country houses, either half-timbered (Pitchford, *c.* 1570) or stone, for example, Benthall (1580), Wilderhope (1586), Condover (*c.* 1590). Moreton Corbet (early 17th century) is a magnificent ruin. A great building period (*c.* 1670–1730) produced such houses as Longnor (1670), Court of Hill (1693), Ludstone (1685), Cound (1704), and Davenport (1726). Ludlow and Shrewsbury are rich in 18th-century townhouses.

Population and Administration.—The population of the county in 1961 was 297,466. There are six municipal boroughs: Shrewsbury, the county town (pop. 1961, 49,566), Wenlock (14,935), Oswestry (11,215), Ludlow (6,796), Bridgnorth (7,552), Bishop's Castle (1,228) (*qq.v.*); and nine urban districts. In 1963 a New Town (*see* New Towns) was planned at Dawley to accommodate surplus population from Birmingham.

Shropshire is divided ecclesiastically between Lichfield and Hereford dioceses. It is in the Oxford circuit, and assizes are held at Shrewsbury. It has one court of quarter sessions, Shrewsbury alone retaining its commission of the peace, and eight petty sessional divisions. There are four parliamentary divisions: Ludlow, Oswestry, Shrewsbury, and the Wrekin.

Shropshire has three pre-Reformation grammar schools (Ludlow, Oswestry, and Bridgnorth) and three more founded in the 16th century (Whitchurch, Shrewsbury, and Market Drayton).

The Royal Salop Infirmary was founded in 1745. The Sir Robert Jones and Dame Agnes Hunt Orthopaedic Hospital at Gobowen, the first of its kind and a model for others, grew from small beginnings at Baschurch in 1904.

The Economy.—Shropshire is mainly an agricultural county. Holdings vary from farms of 1,000 ac. or more to highly efficient small holdings. About 80% of the area is under crops and grass; the chief crops are wheat, barley, oats, sugar beet, and potatoes, and cattle, sheep, pigs, and poultry are kept.

The southwestern uplands are devoted mainly to cattle and sheep raising, the native sheep breeds being Clun Forest and Kerry Hill. Shropshire sheep are bred in the lowlands. The indigenous beef breed is the Hereford, of which there are many famous herds. In the northern plain, dairy farming is the main activity. The predominant dairy breed is the British Friesian, followed by the Ayrshire and Shorthorn. The main cattle market towns are Shrewsbury with a fine new market, Oswestry, Wellington, Bridgnorth, Ludlow, and Craven Arms; sheep sales at Craven Arms are among the biggest in the country.

The most heavily wooded district is the southwest, where pine and spruce are found on higher ground, larch and mixed woods on slopes, and hardwoods in valleys. Northeastern Shropshire lacks woodland but is heavily timbered with hedgerow trees. Some fine woods are found in the Severn Valley below Buildwas, ending with a large remnant of Wyre Forest, as yet oak and coppice. The National Trust owned 1,445 ac. and protected 485 ac. in 1964.

Ironfounding and engineering are major Shropshire industries. In addition to constructional engineering, typical products are large castings, machine tools, locomotives, motor vehicles, wheels and chassis, steel furniture, safes, agricultural machinery, grates, and cookers. Other factories produce radios, toys, mats, lenses, aluminum pans, carpets, and clothing. The principal collieries working in the 1960s were Granville, Madeley Wood (Coalbrookdale coalfield), Highley (Wyre coalfield) and Ifton Heath (Chirk coalfield); some National Coal Board opencast working; and small private pits. Fireclay is excavated, and a large number of refractory products are made. Agricultural limestone is quarried in the Wenlock area, dolomite at Llanyblodwel, and road stone at Titterstone Clee, Squilver, Pontesbury, Maddocks Hill, etc.; but the best Grinshill freestone and Stiperstones lead and barites veins are exhausted. Maltings, breweries, creameries, and a sugar factory use local agricultural products.

The Severn was once navigable for barges up to Welshpool, and in the 18th century a brisk trade was carried on between Shrewsbury and Bristol, while a canal network connected the Severn with the Dee, Mersey, and Stour. Both rivers and canals fell into disuse in favour first of railways and then of road transport.

Bibliography.—*Transactions* of the Shropshire Archaeological and Natural History Society (1878–); *Transactions* of the Caradoc and Severn Valley Field Club (1905–); Publications of the Shropshire Parish Register Society (1900–1924) and of the Shropshire Archaeological and Parish Register Society (1924–); *Victoria County History of Shropshire* (1908–); R. W. Eyton, *Antiquities of Shropshire* (1853–60); H. Owen and J. B. Blakeway, *A History of Shrewsbury*, 2 vol. (1925); W. J. Slack, *The Lordship of Oswestry, 1393–1607 . . .* (1952); A. Raistrick, *Dynasty of Iron Founders; the Darbys and Coalbrookdale* (1953); W. W. Watts, *Shropshire: The Geography of the County . . .* (1919); G. C. Cowling, *Printed Maps of Shropshire A.D. 1577–1900* (1959); *British Regional Geology Handbook, Welsh Borderland*, 2nd ed. (1948). (M. C. Hl.)

SHROUD OF TURIN, a linen cloth about 14 ft. 3 in. long and 3 ft. 7 in. wide, preserved since 1578 in the cathedral of Turin, Italy, and purported to be the burial cloth of Jesus. In 1898, at a public veneration, the first photographic plates were made of it, and these indicated that the image on the shroud—the front and back of a human body, marked with the stigmata—was a negative. After studying this evidence, two professors of biology presented to the Académie des Sciences in 1902 their conclusions that the image on the shroud is not a painting, that it is actually the imprint of a human body and that the image is that of the body of Jesus Christ. The authenticity of the shroud, however, has always been a matter of controversy. Its history cannot be traced further than the middle of the 14th century.

See Werner Bulst, S.J., *The Shroud of Turin* (1957); P. Scotti-A. Vaccari, S.J., in *Enciclopedia Cattolica*, vol. xi, col. 692–697, with detailed bibliography (1953). (E. J. L.)

SHROVE TUESDAY, the day immediately preceding Ash Wednesday (with which Lent begins in the Western church), may fall on any date between Feb. 2 and March 9, according to the date of Easter. Shrove, derived from *shrive*, refers to the confession of sins usual in the middle ages as a preparation for Lent. This origin is clear from many medieval documents. Thus an Anglo-Saxon translation of Theodulf of Orléans' *Capitula* says: "In the week immediately before Lent everyone shall go to his confessor and confess his deeds and the confessor shall so shrive him as he then may hear . . . what he is to do" (that is, as penance). Many customs connected with Shrove Tuesday were so deeply embedded in popular life that they continued in Protestant countries long after the Reformation. The most widely known of these customs is that of eating pancakes (the day is often known as pancake day). It is almost the sole relic of the merrymaking customary before the fast of Lent began, but there was originally a practical reason for the popularity of the dish on this day: it served to use up eggs and fat which were prohibited foods during Lent. Though Shrovetide corresponded with the carnival (*q.v.*) usual in European countries at this time (*Mardi gras* in French, *fetter Dienstag* in German), it never attained in England the proportions of the carnival. The day was usually kept as a holiday; games of football were common, together with throwing at cocks, and all sorts of horseplay took place in schools, universities and among apprentices. "They presently (like Prentises vpon Shroue-tuesday) take the lawe into their owne handes and doe what they list" (Thomas Dekker, *The Seuen deadly Sinnes of London*, 1606). The previous day was known in England as Collop Monday from the practice of eating collops of bacon and eggs.

See Herbert Thurston, *Lent and Holy Week* (1904). (L. C. S.)

SHUBERT, LEE (1875–1953), U.S. theatre manager and producer, was born at Syracuse, N.Y., on March 15, 1875. With his brothers SAM S. (1876–1905) and JACOB J. (1880–1963), he formed the Shubert Theatre Corporation, which came to control a large chain of theatres throughout the United States. The Shuberts began as theatre managers in Syracuse, where they organized several touring companies for the comedies of Charles A. Hoyt. In 1900 they leased the Herald Square Theatre in New York City and proceeded to acquire other theatres, from which grew the vast theatrical empire that once was estimated to be worth $400,000,000. Impetus for the growth of this empire came early from David Belasco and other independent producers during the struggle against the theatre trust, headed by Marc Klaw, A. L. Erlanger, and others. The Shuberts were in a position to rent theatres to producers discriminated against by the syndicate. Soon, however, the Shubert Corp. became a producing organization also. As its productions increased in number, fewer theatres were available to other producers. The Shuberts were then criticized for their harsh contracts and were said to have a stifling effect upon the American theatre. Lee Shubert died in New York City on Dec. 25, 1953. (M. Rs.)

SHUDRA (SUDRA), the fourth and lowest of the traditional varna or classes of Hindu society. The term does not appear in the earliest Vedic literature. In its first application it probably included all conquered peoples of the Indus civilization (*q.v.*) as they were assimilated as craftsmen and menials to the three-class Aryan society.

During centuries of social and economic diversification, the Shudra varna has become a wide spectrum of endogamous status groups, ranking from artisans down to untouchables (*q.v.*). These variations derive from the Hindu belief that certain behaviour patterns and occupations are polluting. This has given rise to a distinction between "clean" and "unclean" Shudra groups; for example, washermen, tanners, shoemakers, sweepers, and scavengers being relegated to untouchability.

See also CASTE (INDIAN); HINDUISM: *Social and Ethical Aspects of Hinduism.* (H. N. C. S.)

SHUFFLEBOARD (SHOVELBOARD; originally SHOVEBOARD), a game in which plastic or metal disks are shoved by the hand or with an implement so that they come to a stop on or within certain lines or compartments marked on the "board" or court (on a table, floor, or outdoor hard surface like concrete). It was popular in England as early as the 15th century, especially with the aristocracy, under the names shove-groat, slide-groat or shovel-penny. Some of the great country houses had boards of exquisite workmanship; that at Chartley hall, in Staffordshire, was over 30 ft. long, comprising 260 pieces. Shove-ha'penny, a later version of shovel-penny, in which a coin or disk is pushed along a polished board so that it stops between closely ruled lines, is still a popular game in English pubs.

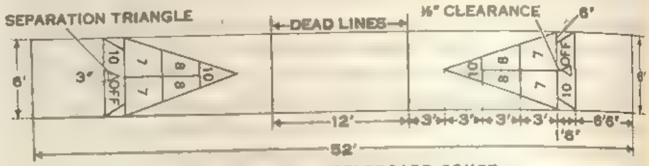


DIAGRAM OF SHUFFLEBOARD COURT

In modern times, a modified form of the old indoor game became popular among travelers on ocean liners as a deck game. For the shipboard version, called shuffleboard, courts of various designs were marked on the deck, with lined sections at either end, numbered one to ten; the section nearest the player, called ten-off, reduced scores by ten. The winning score was 50 points.

U.S. Game.—Shuffleboard was introduced about 1913 at Daytona Beach, Fla. So popular was the game that it spread rapidly through the United States, with each community devising its own rules of play. The modern form of shuffleboard was defined at St. Petersburg, Fla., in 1924, when Pierce V. Gahan, city recreation director, called a conference of all state shuffleboard clubs.

The rules adopted at that meeting defined the size and shape of courts (concrete or terrazzo, 6 × 52 ft.); the maximum length of the cues (6 ft. 3 in.) and the disks (either wood or composition, 1 × 6 in.; four red, four black). The rules also set forth the method of choosing partners, rotation of play, method of scoring and a list of penalties for violations of good sportsmanship, and introduced methods of strategic play. Though it is played by persons of all ages, and is a popular family game, shuffleboard is especially well-suited for elderly persons and others interested in light physical exercise.

Modern shuffleboard may be defined as a competitive game in which disks are propelled by means of cues onto a scoring diagram at the opposite end of a court, in order to score, to prevent opponent's scoring or both. It may be played by two persons (singles) or four (doubles). Interest is sustained as shots are made alternately by "red" and "black."

Strategy may play an important part in the game; to start, each player places his four disks in and not touching the lines of his half of the ten-off area, and a player's first shot (the red disk is shot first) may place a "pilot" or blocking disk in opponent's line of play just outside the scoring section, for example; or he may place a "sleeper" in the part of the scoring section nearest himself so that later he may "sneak" behind it a disk his opponent cannot reach. The blocking play may be met with a carom shot to remove the disk; or the opponent may elect to place a sleeper shot on his side of the court. In singles, when eight shots have been made from the head of the court, players move to the opposite end or foot of the court and continue play in turn until game score is made. In doubles, team players remain at the ends they occupy at the beginning of the game, though play alternates as in singles. Game may be 50, 75 or 100, as players desire. To count, disks must be entirely within scoring sections, clearing all lines. In match play (best two out of three games) the second game is started with a black disk.

During the depression of the 1930s, there was a revival of table shuffleboard in the United States. Played usually in amusement halls, it was often accompanied by gambling, and, being largely a game of chance, soon declined in popularity. (G. B. W.)

SHUMEN: *see* KOLAROVGRAD.

SHURUPPAK, an ancient Sumerian city, the site of which is at Tall Fa'rah in southern Iraq, about 12 mi. (19 km.) S of Nippur and now about 40 mi. (64 km.) E of the Euphrates, though in antiquity the river flowed past it. Brief excavations were made there at the beginning of the 20th century and again in 1931

Three levels of habitation were found, extending in time from the late prehistoric period to the 3rd dynasty of Ur (*c.* 2130–2021 B.C.). The most distinctive finds were ruins of well-built houses, and archaic cuneiform tablets with administrative records and lists of words, indicating a highly developed society already in being toward the end of the 4th millennium B.C.

Shuruppak was most celebrated, however, in legend as the scene of the Deluge, which destroyed all mankind except one survivor, Ziusudra, son of Ubar-Tutu, who reigned, according to the Sumerian king lists, for "18,600" (or "36,000") years in Shuruppak as the last king "before the Deluge." He was commanded by a favouring god to build an "ark" in which he rode out the disaster, afterward re-creating man and living things upon the earth, being himself removed from the haunts of mankind and endowed with eternal life. Ziusudra, whose name was preserved in Greek as Xisuthros, corresponds with Utnapishtim in the Gilgamesh epic (*q.v.*) and with Noah. *See* also FLOOD (IN RELIGION AND MYTH).

BIBLIOGRAPHY.—S. H. Langdon in *Cambridge Ancient History*, vol. i, pp. 377 ff. (1924). For excavations *see* A. Parrot, *Archéologie mésopotamienne*, vol. i (1946). *Tablets:* A. Deimel, *Die Inschriften von Fara*, i–iii (1922–24). *Legends:* J. B. Pritchard (ed.), *Ancient Near Eastern Texts Relating to the Old Testament*, 2nd ed., pp. 42 ff; 93 ff., 265 (king list) (1955). (C. J. G.)

SHUSHTAR, a town in Khuzistan *ostan*, Iran, about 50 mi. (80 km.) SE of Dezful on the Karun River about 15 mi. (24 km.) below the point where it debouches into the plains from the Bakhtiari Mountains. Pop. (1956) 18,527. The town is picturesquely situated on rising ground above the division of the Karun into two branches: the Ab-i-Shutait (west) and the Ab-i-Gargar (east). It is in decline, the inhabitants being attracted by the more prosperous oil towns; large parts of the town are deserted and ruinous. The bazaar is poor and the mosques are lacking in special architectural features except the oldest one dating from the time of the Abbasids. The citadel of Qal'eh-i-Salasil (Salasib), which formerly crowned the cliff above the river, has been demolished to make way for new urban development. Many of the stately houses of stone and brick have cellars, called *shewadan* or *zir zamin*, in which the inhabitants take shelter in the excessive summer heat which may reach 128° F (53° C). The town was once an important trade centre on the ancient road along the foot of the Zagros Mountains, dominating an extensive area rich in *deimi* (dry) and irrigated cultivation.

Water Works.—Shushtar is most famed for the great works constructed in ancient times for the disposal and use of the abundant water of the Karun River. The Band-i-Muhammad 'Ali Mirza, a stone dam with nine sluices dating from Sasanian times, which controls the flow of water, allows two-thirds to flow into the Ab-i-Shutait (the main arm of the Karun) and one-third into the Ab-i-Gargar (called Masruqan in the Middle Ages), an artificial channel through a cut more than 100 ft. deep. This is dammed by the Band-i-Gargar, a dam supporting a bridge which connects Shushtar with the suburb of Bulaiti to the east of the canal. It supplies water through tunnels on each side to drive flour mills and to generate electricity. The great Band-i-Qaisar barrage (the emperor's dike), on the Ab-i-Shutait west of the town, was about 600 yd. long and supported Valerian's Bridge, or Pul-i-Dezful. It first fell into ruin in the time of the Omayyads but was restored by the governor-general of Khuzistan, Muhammad 'Ali Mirza, the son of Fath Ali Shah, early in the 19th century. It was damaged by floods in 1885 and is now in ruins. There is also the Nahr-i-Dariyan (or Minau) Canal, which starts above the barrage in tunnels cut out of the rock on the western side of the town underneath the citadel, the purpose of which was to irrigate the Miyanab (the land between the two arms of the Karun).

According to tradition the Minau Canal was built by Darius I (the Great; d. 486 B.C.). Ardashir I (d. *c.* A.D. 240) began to construct the barrage after the mouth of the canal had dried up, probably because of river erosion. The barrage was completed under Ardashir's son Shapur I (d. 272) who is said to have used Roman prisoners captured with the emperor Valerian. The Ab-i-Gargar was first dug to divert the water of the main river; the Band-i-Qaisar was then constructed and the bed of the river above it was paved with huge stone slabs bound with iron to prevent further erosion. This paving was called *shadurwan*, a term also applied to the barrage itself.

See Laurence Lockhart, *Persian Cities* (1960). (H. Bo.)

SHUSTER, WILLIAM MORGAN (1877–), U.S. lawyer, civil servant, financial expert and publisher, who served as treasurer-general to the Persian government (1911–12), was born in Washington, D.C., on Feb. 23, 1877. He entered the Cuban customs service in 1899 but resigned in 1901 to become collector of customs at Manila. In 1906 he was appointed secretary of public instruction in the Philippines and a member of the Philippine commission (*see* PHILIPPINES, REPUBLIC OF THE: *History*).

When the Persian government appealed in 1910 to the U.S. government for advice and help in the reorganization of its financial system, Shuster, on the recommendation of Pres. William Howard Taft (under whom he had served in the Philippines), was chosen to head a small party of U.S. financial experts to achieve this end. On reaching Teheran in May 1911 Shuster and his colleagues were welcomed by the Persians and given every facility. Their endeavours, however, bore but little fruit, as they soon aroused the vehement opposition of Russia which went so far as to threaten armed intervention; in consequence the mission had to leave Persia in Jan. 1912. Shuster recounted the history of his mission in his book *The Strangling of Persia* (1912). In subsequent years Shuster, besides practising as an attorney, turned to publishing. He was president of the Century company of New York city (1915–33) and of Appleton-Century-Crofts Inc. from 1933 onward. He subsequently became chairman and chief executive officer of the latter company. *See* also PERSIAN HISTORY. (L. Lo.)

SHUVALOV, PETR ANDREEVICH, COUNT (1827–1889), Russian government official and ambassador, more successful as police official than as diplomat, was born in St. Petersburg on July 27 (new style; 15, old style), 1827, a member of an old noble family which had risen to imperial favour in the mid-18th century. He began his military service in 1845 and fought in the Crimean War as major of cavalry before entering upon a diplomatic career as a member of the Russian delegation to the Paris peace conference of 1856.

In 1857 Shuvalov was put in charge of the St. Petersburg police. His success there brought him to the post of director of the political police in the ministry of the interior (1861–64). After two years as governor general of the three Baltic provinces and commander of the Riga army corps, he returned to St. Petersburg in 1866 to be chief of staff of the gendarmerie corps and head of the political police, or "3rd section" of the imperial chancery. In this capacity he became the closest and most influential adviser of Alexander II. He wielded vast power as the leader of a conservative clique opposed to further liberal reform.

In 1873 Shuvalov was sent to London to reassure Great Britain over Russian advances in central Asia and to prepare the way for the marriage of the grand duchess Maria Aleksandrovna with the duke of Edinburgh (Alfred of Saxe-Coburg). His appointment as ambassador to London in 1874 apparently resulted from intrigues by his enemies at court and represented semidisgrace. Despite his easy charm, the count at first failed to distinguish himself in London. Before the Russo-Turkish War he weakened Russia's position by making unnecessary concessions to the British, whom he often found too clever for him. After the publication of the treaty of San Stefano (*see* EASTERN QUESTION) he concluded a secret convention with Lord Salisbury settling the principal Russo-British differences (May 29, 1878), making concessions for which he subsequently obtained his emperor's consent. At the congress of Berlin (*q.v.*) he was the principal Russian delegate, and the public storm in Russia which greeted the Berlin treaty induced the emperor to recall him in 1879. After living for nearly ten years in retirement, Shuvalov died in St. Petersburg on March 22 (N.S.; 10, O.S.), 1889. (D. MACK.)

SHWEBO, a town and district in the Sagaing Division of Burma. The town is situated on the railway about 50 mi. (80 km.) NNW of Mandalay. Pop. (1953) 17,842. It is of historic interest as the birthplace and capital of Alompra (Alaungpaya; d. 1760),

the founder of the last Burmese dynasty.

SHWEBO DISTRICT has an area of 7,605 sq.mi. (19,697 sq.km.) and a population (1962 est.) of 676,625. The district lies on the northern margins of the Dry Belt and stretches to and beyond the Upper Irrawaddy on the east where it now includes the old ruby mining area of Mogok. The main part of the district west of the Irrawaddy consists of a broad central plain drained by the Mu Chaung and extensively irrigated so that much rice is grown, with millets, sesamum, peanuts (groundnuts), peas, and cotton on the nonirrigated portions. The irrigation works are very old but fell into disrepair in King Thibaw's time (the 1880s), to be restored and extended from 1906 onward. The Mandalay-Myitkyina railway runs through the heart of this plain; elsewhere the country is one of wooded hills where the rainfall rises above the 25–40 in. (650–1,000 mm.) of the plain. (L. D. S.)

SHYOK, a large tributary of the upper Indus, which rises behind the Karakoram range near the Karakoram pass and after cutting across the range first flows toward the southeast and then takes a sharp turn to the northwest joining the Indus near Kiris, about 20 mi. S.E. of Skardu. The trans-Karakoram basin of the Shyok contains the plains of Depsang. From its source to its junction with the Indus the length of the river is 310 mi. and the total fall is 11,000 ft. Its catchment basin has an area of 13,000 sq.mi. The Gasherbrum glacier in the Karakoram region (*c.* 21 mi. in length) and the Khundam glacier drain into it. Its upper course is rushing and turbulent down a gorge but its middle course is broad and in places divides into many channels in an open valley. There the river is usually fordable although with some difficulty. Between Tertse and Unmaru there are seven distinct channels with an average depth of two feet. At the Turtuk bridge the river narrows to 70 ft. and in its lower course is again a furious rapid between precipitous cliffs. The Shyok's chief tributary, the Nubra, has its source in a large glacier, the Siachen, which lies across the Karakoram range. The principal peaks of this range between the Nubra and the Shyok are from 24,000 to more than 25,000 ft. in height. (K. S. AD.)

SIALKOT, a town and cantonment, and a district in the Lahore division of West Pakistan. The town lies on the northern bank of the Aik Nala, south of the Jammu hills, about 67 mi. N.E. of Lahore. Pop. (1961) 164,346 (including the cantonment). Popular legends attribute its foundation to Raja Sala, uncle of the Pandavas (of the epic, *Mahabharata*), and claim that it was refounded in the time of Vikramaditya by Raja Salivahan who built the fort, the remains of which, standing on a mound, are still to be seen. It has also been suggested that Sialkot is the site of ancient Sakala or Sagal, capital of the Indo-Greek Menander (Melinda) and Mihiragula, the Hun (d. 540). Originally the town proper was built around the fort. As a result of later expansion, however, several *abadis* (townships) sprang up nearby, of which the most important are Rangpur, Nekipura, Hajipura, Mianapura, Mubarikpura and Puran Nagar. Notable monuments include the tombs of several well-known Muslim scholars and religious leaders. The town also contains a shrine of the founder of the Sikh faith, Nanak, a temple erected by Raja Tej Singh. A Sikh temple established by Nanak himself lies in the eastern outskirts of the town. Of the modern buildings the municipal offices and town hall are built on the site of the fort. Educational institutions include Murray college, Jinnah Islamia college and the Government College for Women. There are two notable libraries and several parks. The town was the birthplace of the philosopher and poet Mohammed Iqbal. Sialkot is known for its manufactures of sports goods, surgical instruments, cutlery, rubberware and ceramics.

SIALKOT DISTRICT has an area of 2,067 sq.mi. and had a population (1961) of 1,596,383. It is an oblong tract of country occupying the submontane portion of the Rechna doab. The northern part is very productive, and the southern part, though less fertile, is irrigated by the upper Chenab canal. About nine-tenths of the cultivable area is under crops, the highest proportion of any district in West Pakistan. The principal crops are wheat, barley, maize (corn), millets and sugar cane.

The early history of Sialkot is closely interwoven with that of the rest of the Punjab. It was annexed by the British after the second Sikh war in 1849; and thereafter its area was considerably reduced, its present size dating from 1867. (K. S. AD.)

SIAM: *see* THAILAND.

SIAMESE LANGUAGE: *see* THAI LANGUAGE.

SIAMESE LITERATURE: *see* THAI LITERATURE.

SIAN (HSI-AN), capital of China's northwest province of Shensi and in early history the site of China's capital for a total of 970 years during the western Chou, the Ch'in, the Western Han, the Sui and the T'ang dynasties. Its earlier names were Changan and Siking and it reverted to its Manchu name Sian (meaning "western peace") in 1943. The city is situated on a broad loess terrace at the northern foot of the Tsinling Shan not far from the south bank of the Wei Ho. In the 14th century, Marco Polo, who called it Kenjanfu or Quengianfu, described it as a thriving trade centre. Its significant position derives from the productivity of the Wei plain and its natural topographic defenses, as well as its strategic command over radiating communications routes to Kansu and Turkistan in the west, Szechwan in the south, the Ordos desert in the north, Shansi and the Yellow plain in the northeast and east. It was there in 1936 that Chang Hsüeh-liang kidnaped Chiang Kai-shek in the "Sian incident" leading to a united front between Nationalist and Communist forces to resist Japan (*see* CHINA: *History*). However, during the Chinese-Japanese war, a strong Nationalist army was garrisoned at Sian, not only against the Japanese, but also to contain the Communists who had established a base at Yenan in north Shensi after 1935 under Gen. Hu Tsung-nan.

The city is surrounded by a rectangular wall, (approximately 3.1 mi. by 1.8 mi.) with a city gate on each side except the east which has two gates. Suburbs extend from each gate. The 1939 population of about 209,000 grew to 1,310,000 by 1957, because of industrial developments extending from the western suburbs. The expansion has resulted from new economic developments in China's northwest, including the completion of the trans-Tsinling railroad and the extension of the now doubled-tracked Lung-hai railroad passing Sian into Sinkiang.

Sian was one of the key cities in the Communist First Five-year plan. A small iron and steel mill was constructed, a cotton textile combine set up, and a new thermal electric power plant built which supplies nearby cities and Sian with power. Also constructed were a cement plant, with a seamless steel tube rolling mill, and a moderate-size chemical plant.

Historically important, the city contains temples, tombs and monuments. The collection in the Shensi provincial museum (formerly the Pei-lin) is noteworthy, and the Nestorian tablet discovered in 1625 was housed there. Sian is the seat of Northwestern university and Medical college, and the location of the Northwestern Institutes of Art and Music. (H. J. WS.)

SIANG FRONTIER DIVISION is the central of the five divisions constituting the North East Frontier Agency of the Republic of India. It is through the heart of this division that the Dihang river (a section of the Brahmaputra) passes in a series of stupendous gorges from the plateau of Tibet to disgorge on the valley plain of upper Assam less than 1,000 ft. above sea level. Navigability on the Brahmaputra ceases at Pasighat; from there a rough track runs alongside of the valley west of the river, eventually passing into Tibet and linking with the motor road to Lhasa. The Abor is the principal tribe living in this division. They were first visited by the British in 1826. From 1848 they carried out numerous raids and outrages on the neighbouring territory and several expeditions were sent against them. From 1912 to 1913 the territory known as the Abor hills formed part of the Sadiya frontier tract (from Sadiya, near the confluence of the rivers Lohit and Dibang with the Dihang) and was loosely administrated by the government of India. Eventually it was included in the North East Frontier Agency (*q.v.*). (L. D. S.)

ŠIAULIAI, the fourth largest town in Lithuania (now the Lithuanian Soviet Socialist Republic, U.S.S.R.), and administrative centre of a district of the same name. Pop. (1959) 59,722. A Christian church was erected there in 1445, but the town began to expand only during the latter part of the 19th century when a

large leather and footwear industry was established there. After 1918, when Lithuania regained its independence, Šiauliai became a major economic centre, and it contains 85% of Lithuania's leather industry (tanning) and 60% of the footwear industry, as well as two-thirds of the flax processing. Chocolate and sweetmeats, meat and fish conserves, and other food products and beverages are produced. The town is a major rail junction and is on the Riga-Kaliningrad road. During World War II Šiauliai was partly destroyed, but it has been rebuilt, with wide avenues and modern buildings.

SIBAWAYH, the nickname of Abu Bishr 'Amr ibn 'Uthman ibn Qanbar (*c.* 760–793), the most celebrated Arabic grammarian. He was a Persian client of an Arab tribe, and studied under Khalil (*q.v.*) in Basra. He wrote the first known full-scale Arabic grammar, called *al-Kitab fi'l nahwi,* on which all subsequent Arabic grammars were based. The book defines three parts of speech—noun, verb, and particle—and explains *i'rab,* or accidence, as applied to nouns and verbs. Then the parts of speech and their use are dealt with in great detail, with supporting quotations from the Koran and from Arabic poetry.

Editions of the *Kitab* are by H. Derenbourg, *Le Livre de Sibawaihi* (1883); and by G. Jahn, with commentary and German translation, two volumes (1895–1900).

See also G. L. Flügel, *Grammatische Schulen der Araber* (1862); C. Brockelmann, *Geschichte der arabischen Litteratur,* vol. 1 (1898), suppl. vol. 1 (1937). (J. A. Hd.)

SIBELIUS, JEAN (Johan Julius Christian Sibelius) (1865–1957), the greatest symphonic composer of Scandinavia and one of the most original figures in 20th-century music, was born at Hämeenlinna, Finland, on Dec. 8, 1865. He became a pupil at the Suomalainen Normaalilyseo, the first Finnish-speaking school, where he came into contact with Finnish literature and in particular the *Kalevala,* the mythological epic of Finland, which remained for him a constant source of inspiration. Many of his symphonic poems, such as *Pohjola's Daughter* and *Luonnotar,* draw on this source.

Although intended for a legal career he soon abandoned his law studies at Helsinki, devoting himself entirely to music. At first he planned to become a violinist. Under the guidance of Martin Wegelius he composed much chamber and instrumental music. He adopted the name Jean, which he used throughout his professional career in preference to his baptismal names. In his mid-twenties he left Finland to continue his studies in Berlin and Vienna, where his teachers included Robert Fuchs and Karl Goldmark.

On his return to Finland a performance of his first large-scale orchestral work, the *Kullervo* Symphony (1892), created something of a sensation. This and succeeding works, *En Saga,* the *Karelia* music and the *Four Legends,* established him as Finland's leading composer. In 1897, before the appearance of his First Symphony, the Finnish Senate voted him a small life pension in recognition of his genius. His tone poem *Finlandia* was written in 1899 and revised in 1900.

© KARSH, OTTAWA

JEAN SIBELIUS, PHOTOGRAPHED IN 1949

In the first decade of the 20th century Sibelius' fame penetrated the continent. Busoni, whose friendship he had made in Helsinki as a student, conducted his Second Symphony in Berlin and Henry Wood and Granville Bantock introduced his works to English audiences, Bantock commissioning his Third Symphony. With this work Sibelius turned his back on the national romanticism of the Second Symphony and Violin Concerto and moved toward the more searching and uncompromising mode of utterance of *En Saga* and the Fourth Symphony. After World War I he published his greatest works, the last three symphonies and *Tapiola,* but then lapsed into the long silence of his last years. Rumours of an Eighth Symphony (promised for performance in the early 1930s) and even a Ninth Symphony were unfounded. No manuscripts survived his death at Järvenpää on Sept. 20, 1957.

The 1930s saw a vogue for Sibelius prompted by such writers as Cecil Gray and Constant Lambert in England and Olin Downes in the U.S. Despite a reaction against this vogue in the following generation, Sibelius retained his firm hold over the musical public. Although his inspiration is intimately connected with the Scandinavian landscape, it is not primarily as a nature poet that he is remembered. His achievement both in the symphonic poems and the seven symphonies lies principally in his remarkable mastery of form. As Gerald Abraham pointed out, the first movement of the Third Symphony has the clarity of construction of a Haydn or Mozart first movement, yet its organic unity and architecture even surpasses its models. It was in this capacity for organic growth that the secret of his genius lay.

Bibliography.—C. Gray, *Sibelius* (1931); K. Ekman, *Jean Sibelius,* Eng. trans. by E. Birse (1935); G. Abraham (ed.), *Sibelius: a Symposium* (1947); S. Parmet, *The Symphonies of Sibelius* (1959); H. Johnson, *Sibelius* (1960). (Rt. L.)

SIBENIK (Italian Sebenico), a port of the Socialist Republic of Croatia and headquarters of Sibenik *srez* (district), Yugoslavia, lies at the mouth of the river Krka on the Adriatic Sea, 190 mi. (306 km.) SE of Rijeka by road. Pop. (1961) 24,800. The town is overlooked by fortifications and has some fine Renaissance architecture including the loggia and part of the beautiful cruciform Roman Catholic cathedral (1431–1536) built entirely of stone. Sibenik is the seat of a Roman Catholic bishop (since 1298) and of an Orthodox bishop. It is a naval base and steamship station and is connected with Zagreb by rail. It exports bauxite, timber, and wine. Electric power is supplied by the celebrated falls of Krka (11 mi. [18 km.] inland).. Insecticide powder and calcium carbide are manufactured, there are woolen mills and oil refineries, and the wood of the smoke tree (*Cotinus coggygria*) is prepared for dyeing purposes. Fishing and sponge and coral gathering are other occupations.

Sibenik is first mentioned in 1066 and it was in Hungarian hands from 1180 to 1322 and from 1351 to 1412 when it passed to Venice. By the peace of Campo Formio in 1797 it was awarded to Austria and in 1805, under Napoleon I, became attached to the kingdom of Italy. From 1814 to 1918 it was part of the Austrian Empire, after which it was incorporated into Yugoslavia. During World War II Sibenik was occupied by the Italians and then by the Germans. (V. De.)

SIBERIA (Russian Sibir'), a vast region of the U.S.S.R. composing the whole of northern Asia between the Ural Mountains (west) and the Pacific Ocean (east) and extending southward from the Arctic Ocean (Kara Sea, Laptev Sea, and East Siberian Sea) to the hills of central Kazakhstan and the frontiers of the U.S.S.R. with China and Mongolia. Within the Soviet Union a small part of Siberia in the southwest (Tselinnyy [Tselinny] Kray and Semipalatinskaya [Semipalatinsk] and Vostochno-Kazakhstanskaya [East Kazakhstan] *oblasts*) is administratively incorporated into the Kazakh Soviet Socialist Republic, while the rest of the region belongs to the Russian Soviet Federated Socialist Republic. The administrative areas adjacent to the Ural Mountains (Tyumenskaya [Tyumen], Sverdlovskaya [Sverdlovsk], Chelyabinskaya [Chelyabinsk], and Kurganskaya [Kurgan] *oblasts*) and those along the Pacific coast (Khabarovskiy [Khabarovsk] and Primorskiy [Primorski] *krays* and Amurskaya [Amur], Magadanskaya [Magadan], Kamchatskaya [Kamchatka], and Sakhalinskaya [Sakhalin] *oblasts*) are not now officially considered as part of Siberia, the former group belonging to the Urals region and the latter being known as the Far East. The rest is divided into western and eastern Siberia by a line running meridionally somewhat west of the Yenisey (Yenisei) River. The constituent administrative divisions are: Altayskiy (Altai) Kray, Kemerovskaya (Kemerovo), Novosibirskaya (Novosibirsk), Omskaya (Omsk), and Tomskaya (Tomsk) *oblasts* in western Siberia; and Krasno-

yarskiy (Krasnoyarsk) Kray, Irkutskaya (Irkutsk) and Chitinskaya (Chita) *oblasts*, and the Yakut, Buryat, and Tuva Autonomous Soviet Socialist republics in eastern Siberia. The total area of Siberia in the wider sense is 5,330,896 sq.mi. (13,807,037 sq.km.) (or more than a quarter of the Asian continent and more than 60% of the territory of the U.S.S.R.), in the narrower sense 4,374,475 sq.mi. (11,330,600 sq.km.). There were 32,471,380 inhabitants in the whole region in 1959, including 21,466,775 in Siberia in the narrower sense (western 10,159,437, eastern 6,960,535, far eastern 4,346,803), 9,112,337 in the Urals part, and 1,892,268 in the Kazakh part. The population is sparse, chiefly concentrated in the west and south, more than half urban and overwhelmingly Russian. This article deals with the historical development of the whole region. (*See* also UNION OF SOVIET SOCIALIST REPUBLICS and the separate articles on the administrative divisions.)

Early History.—It is still not established whether man came to Siberia from Europe or from central and eastern Asia. The earliest Paleolithic finds, in the Altai area of southern Siberia, are of the Mousterian type. Finds of the later Paleolithic Age are numerous and are concentrated, apart from the Altai area, along the upper reaches of the Yenisei (the Minusinsk Basin) and near Lake Baikal. The Neolithic Age has been mainly studied in two areas: near Lake Baikal, which was then the centre of a hunting culture stretching from the Yenisei to the middle reaches of the Amur; and along the lower Amur and the Pacific coast, where a totally different fishing culture existed, connected with Manchuria, China, and Korea. About 2000 B.C. the inhabitants of the Altai-Minusinsk area first began to make use of copper. The people of this culture were of the Europeoid race, while those living to the east and to the north were predominantly Mongoloid. Soon the tribes of the Baikal-Lena area, who are regarded as the ancestors of the Paleo-Asiatic Yukaghir, also started using copper and bronze—possibly under Chinese influence—though they remained primarily hunters. They were later partly displaced and partly assimilated by the Tungus tribes, which, coming from the Amur area, spread over a vast territory in eastern Siberia and probably brought with them the use of reindeer. The bronze culture which centred in the Altai-Minusinsk area and reached as far west as the Urals shows not only a further development of animal husbandry but also the beginnings of agriculture. The working of bronze was stimulated by the existence of rich local ores and attained high technological and artistic levels. Around 1000 B.C. Chinese influence began to be felt. Subject to strong influences from China and central Asia, southern Siberia developed rapidly both economically and politically, while the Paleo-Asiatic and Tungus tribes of the northern forests and tundra lagged far behind. In the 7th–2nd centuries B.C. the Minusinsk Basin was the home of an advanced sedentary culture involving complicated systems of irrigation developed by the people who are tentatively identified as the Ting-Ling of the Chinese chronicles; they were the first users of iron in Siberia. The rest of southern Siberia was at this time inhabited by nomadic tribes, probably Iranian (Scythian) in the west, belonging to the large belt of similar cultures stretching from the steppes north of the Black Sea to Mongolia and characterized by the "animal style" of ornamentation. The famous Pazyryk burial mounds of the Altai area, on the right bank of the Bol'shoy Ulagan (Bolshoi Ulagan) River, show strong signs of Persian influence.

From the 3rd century B.C. the whole of southern Siberia belonged in turn to the Turkic-Mongol Huns and to the various Turkic states centred in Mongolia. From the 6th century A.D., when the local Altai Turks established their khanate, the Turks also predominated ethnically, partly assimilating the neighbouring Samoyed and partly driving them north. The central part of southern Siberia, around the Minusinsk Basin, was the home of the Kirgiz, ancestors of the present-day Khakass and Altai peoples as well as of the Kirgiz of central Asia. Further east, in the Baikal area, lived the Kurykans, perhaps the ancestors of the Yakuts, while the steppes of northern Kazakhstan were inhabited by the nomadic Kipchaks (Kumans), who later became the chief ethnic component of the Kazakh people. From the 10th century the Turks were overpowered by the Mongols, whose states, however, remained unstable until the creation of the great Mongol Empire by Genghis Khan. Almost the whole of Siberia was incorporated into this empire, at least nominally. The primitive peoples in the far north were scarcely affected, but for the comparatively high civilization of southern Siberia the Mongol conquest had fatal consequences. The larger part of Siberia belonged to the domain of the great khan himself, but western Siberia was included in the domain of his son, Juchi, and commonly known as the Golden Horde (*q.v.*). The Mongol conquest strengthened the Turkic element in the Ob' (Ob) River area, hitherto predominantly Ugrian, where the Turks later became known as Siberian Tatars. They were the dominant element in the Sibir khanate which emerged after the breakup of the Golden Horde in the mid-15th century. (*See* MONGOL EMPIRES.)

Russian Colonization (1581–1890).—Siberia is first mentioned in Russian sources in the 11th century, and from the 12th century the merchants of Novgorod traded with the tribes of the lower Ob area, the main attraction being furs. In 1581 a small band of Cossacks under Ermak Timofeevich, a mercenary in the service of the Stroganov family (merchants who had large property in the western Urals), undertook an expedition to western Siberia and within a year conquered the Siberian khanate. From then on the opening up and annexation of Siberia by the Russians proceeded rapidly. Small Cossack units, using the river systems of the Ob, Yenisei, and Lena rivers, spread throughout the whole of northern and parts of southern Siberia, establishing fortified towns in strategic positions (Tyumen in 1585, Tobolsk in 1587, Tomsk in 1604, Kuznetsk in 1617, Krasnoyarsk in 1628, Yakutsk in 1632, Okhotsk on the Pacific coast in 1649, Albazino [Albazin] on the Amur in 1651, and Irkutsk in 1652) and imposing their administration upon the surrounding areas for the purpose of collecting tribute for Muscovy. There was a major reverse in the Amur area, where the Russian advance met with Chinese resistance; and, according to the Treaty of Nerchinsk in 1689, most of the area was declared Chinese, though in fact it remained a no-man's-land between Russia and China.

Kamchatka Peninsula was annexed in 1699, but only by the Aigun and Peking treaties in 1858 and 1860 respectively did China renounce all claims to the Amur area and the Pacific coast between the mouth of the Amur and Korea. In the southwest the khan of the Kazakh Middle Hundred exchanged Dzungarian (Kalmyk) suzerainty for Russian in 1740, and the Altai area was annexed in 1756, when the Dzungarian state, to which it had belonged since the breakup of the Mongol Empire, passed under Chinese rule.

Cossack officers, Muscovite officials, monks, and merchants were the first Russian explorers of Siberia, and the Cossack Semen Dezhnev was the first to sail through the Bering Straits in 1648. The exploration was put on a systematic and more scientific basis in the 18th century, when the Russian Academy of Sciences organized several expeditions, mainly to northern Siberia and the Far East. In the 19th century, the task of studying Siberia's physical and human geography was largely taken over by the Imperial Russian Geographical Society, assisted by Siberian businessmen.

The administration of annexed Siberia was directed by the Siberian office in Moscow (later in St. Petersburg) from 1614 till 1763. Several provinces were formed during the 18th century, and as a result of the administrative reform of 1822, prepared by Count M. M. Speranski, two governors-general were appointed: one for eastern Siberia with his seat in Irkutsk and another for western Siberia with his seat first in Tobolsk, then, from 1838 in Omsk. The western Siberian provinces of Tobolsk and Tomsk were in 1882 exempted from the jurisdiction of the governor-general in Omsk, who retained authority only over the Kazakh steppe. Finally a viceroy was appointed in 1884 for the newly annexed territories in the Far East together with the old Russian possessions on the Pacific coast.

At first Russian economic activity was limited to the collection of tribute, which was paid by the local inhabitants in furs as it had been paid to the Mongols; Siberian furs were one of the main assets of the Muscovite state's trade with Western Europe

However, Russian agricultural colonization on a small scale went on parallel with the military occupation in the late 16th century and throughout the 17th, its primary aim being to feed the military and administrative personnel. An extensive network of stagecoach and mail routes was developed by the government and many settlers were employed in this service.

Silver mining on a small scale was started in 1698 and in the next century two important industrial areas developed in Siberia: the Altai, where silver and copper were mined and smelted; and the Nerchinsk area in Transbaikalia (Zabaykal'ye), with silver and lead mining. With the decline of the fur trade, mining became the main economic activity in Siberia, the principal entrepreneurs being the state and the imperial household. The supply of labour was largely guaranteed by a system of bondage which tied workers permanently to the mines and metal plants. Forced labour by convicts was also persistently practised by the government, though it was highly unproductive and the majority of convicts managed to escape, usually to become bandits. Gold mining, in contrast, which developed in many areas from the 1830s, was mainly in the hands of private enterprise employing free labour.

The emergence of an indigenous class of industrialists and the banishment to Siberia of large numbers of revolutionary intellectuals (beginning with the Dekabrists in 1826 and participants of the Polish insurrection of 1830–31) combined to produce a small but vocal Siberian intelligentsia and the development of a Siberian regional consciousness. The regionalists demanded the removal of all discriminatory measures arising from Siberia's colonial status, with regional self-government.

Most indigenous Siberian peoples had resisted Russian conquest and submitted only after a long struggle and attempts to move away from the Russian centres into the interior. Indeed, the most warlike and least accessible of them, the Chukchi, considered themselves independent until they were finally subdued in the 1930s. The impact of the Russians upon the indigenous peoples was twofold. The smaller and more primitive tribes succumbed to exploitation by the merchants and corrupt officials, to poverty and to diseases brought by the colonists. On the other hand, larger peoples such as the Yakuts, Buryats, and Kazakhs adjusted themselves to the new situation and began to profit from the material benefits of the colonization. The internal institutions and way of life of the indigenous peoples were generally not interfered with by the Russian authorities, and there was remarkable religious tolerance, which led to the spread in the 17th and 18th centuries of Buddhism among the Buryats and Islam among the Kazakhs, while most of the other peoples became at least nominally Christian.

The Modern Period.—The great event that marked the transition of Siberia (with its vast natural resources) from a traditionalist to a modern economy was the construction of the Trans-Siberian Railway. The original line was begun at both ends, at Chelyabinsk and at Vladivostok, in 1891 and was completed in 1905. It facilitated the influx of new colonists, first in connection with the construction itself, then under the colonization plan introduced by P. A. Stolypin, president of the Council of Ministers with the portfolio of the interior), to reduce the rural overpopulation in European Russia. Siberian agriculture, which had never known peasant serfdom or large private estates, quickly developed American-type farming methods, specializing in grain and particularly in butter, the production and marketing of which were largely in the hands of cooperatives. Coal mining was started in several places along the line to supply the needs of the railway, and the railway repair shops were the first engineering plants in Siberia.

Siberia took an active part in both the 1905 and 1917 revolutions, and events of the Russian civil war there were most dramatic. An autonomous Siberian government was formed after the overthrow of the Soviet power in the region early in 1918 but was soon superseded by the government of Adm. Aleksandr Vasilievich Kolchak (*q.v.*), in Omsk, who was proclaimed the supreme ruler of Russia by the leaders of the "White" armies. After the defeat of Kolchak, the Far Eastern Republic was set up in 1920 in the area east of Lake Baikal as a buffer state between Soviet Russia and Japan; in reality, however, this was one of the first "people's democracies" and was incorporated into the Soviet state as soon as the Japanese interventionists left Vladivostok in 1922. Indigenous authorities set up by the Kazakhs, Buryats, and Yakuts during the civil war were suppressed and several ostensibly autonomous republics and regions were established instead with the purpose of remolding the population's way of life according to the general Soviet pattern.

Industrial development, interrupted by revolution and civil war, was resumed during the first five-year plan period (1928–32), when the Ural-Kuznetsk coal mining and iron and steel combine was built in southwestern Siberia, largely by the forced labour of deported peasants. Forced-labour camps spread throughout Siberia during the 1930s, the most important being the Dalstroi (Far Eastern Construction Trust) system in the extreme northeast, and the Noril'sk (Norilsk) area on the lower Yenisei, both concerned with the mining of nonferrous and precious metals. World War II gave a new impetus to the industrial development; many enterprises were evacuated from the west and Siberia, together with the Urals, became the industrial backbone of the Soviet war effort.

Agriculture, on the contrary, greatly suffered from collectivization in 1930–33 and was largely neglected until the Virgin Land Campaign of 1954–56, when the southwest of Siberia (including northern Kazakhstan) was the principal area to be cultivated. Siberia figured prominently in the seven-year plan (1959–65), with the main emphasis on the construction of large thermal and hydroelectric power stations, electrification of the Trans-Siberian Railway as far as Irkutsk, and a further increase in the production of steel.

See articles on indigenous peoples: TUNGUS; SAMOYED; KIRGHIZ; YAKUT; KAZAKH; CHUKCHI; TURKIC PEOPLES; ALTAIC PEOPLES; UGRIC PEOPLES; *see* also BURYAT AUTONOMOUS SOVIET SOCIALIST REPUBLIC and references under "Siberia" in the Index.

BIBLIOGRAPHY.—R. J. Kerner, *Northeastern Asia: a Selected Bibliography*, 2 vol. (1939), *The Urge to the Sea* (1942); G. Kennan, *Siberia and the Exile System*, 2 vol. (1891); M. A. Czaplicka, *Aboriginal Siberia* (1914); F. A. Golder, *Russian Expansion on the Pacific, 1641–1850* (1914); G. K. Gins, *Sibir, soyuzniki i Kolchak*, 2 vol. (1921); V. Jochelson, *Peoples of Asiatic Russia* (1928); E. Varneck and H. H. Fisher (eds.), *The Testimony of Kolchak and Other Siberian Materials* (1935); G. H. Fisher, *The Russian Fur Trade, 1550–1700* (1943); J. A. White, *The Siberian Intervention* (1950); E. Lipper, *Eleven Years in Soviet Prison Camps* (1951); B. Nolde, *La Formation de l'empire russe*, vol. 1 (1952); C. Krypton, *The Northern Sea Route . . .*, 2 vol. (1953–56); W. Kolarz, *The Peoples of the Soviet Far East* (1954); M. Raeff, *Siberia and the Reforms of 1822* (1956); D. W. Treadgold, *The Great Siberian Migration* (1957). (S. V. U.)

SIBIU (Ger. HERMANNSTADT, Hung. NAGYSZEBEN), a town of the Rumanian People's Republic, in the Braşov region, situated north of the Turnu Roşu pass at the foot of the Transylvanian Alps, 1,361 ft. above sea level and 72 mi. S.S.E. of Cluj. Pop. (1963 est.) 100,659, including a high proportion of German and Hungarian origin.

The old town centre, which retains a medieval appearance, is divided into the upper town on a terrace and the lower town on the banks of the river Cibin, a tributary of the Olt, the two parts of the town being connected by steep alleys and flights of steps. There are remains of the town walls and towers. The Lutheran church dates from the 14th century and the town hall was built in the 15th century. The Brukenthal museum, with the largest art collection in Transylvania, is in a handsome baroque building (1778–85) containing a library and paintings which include Dutch, Flemish and German masters as well as Rumanian. In addition there are important collections of ecclesiastical antiquities, ethnography, folk art and natural science. Sibiu is the seat of a Lutheran bishop and of a Rumanian Orthodox metropolitan. It has teachers' training colleges and secondary schools including the Brukenthal school dating from the Reformation. The town is a cultural centre and has a state theatre and a philharmonic orchestra.

Sibiu is connected by road and rail to Braşov and to Bucharest (through the Turnu Roşu pass). It is a commercial and manufacturing centre producing machinery, textiles, chemicals, inks, leatherwork, footwear, and building materials.

Originally a Roman colony (Cibinium), Sibiu was refounded as Hermannsdorf by Saxon (German) colonists in the 12th century. It was destroyed by the Tatars in 1241. In the 14th century it became an important administrative and commercial centre of the German communities in Transylvania, and suffered much in the wars with the Turks. In 1699 it became subject to Austria.

SIBLEY, HIRAM (1807–1888), a founder and second president of the Western Union Telegraph company, was born at North Adams, Mass., Feb. 6, 1807. When he was 16 his family moved to western New York where he later operated a foundry and machine shop and was in 1843 elected sheriff of Monroe county.

Visiting Washington, D.C., Sibley met Samuel F. B. Morse, the telegraph inventor, and helped obtain congressional backing for construction of the first telegraph line in 1844. Recognizing the need for a national telegraph system, Sibley and other Rochester, N.Y., citizens, in 1851, formed the New York and Mississippi Valley Printing Telegraph company, which bought 11 small lines north of the Ohio river. In 1856 the company was renamed The Western Union Telegraph company.

Sibley became president of Western Union later that year. Under his leadership, the first transcontinental telegraph line was built in 1861, to help hold the western states in the Union in the Civil War and to develop the west.

Sibley began building a line to Europe via Russian America, the Bering strait and Siberia in 1865. When he was in St. Petersburg, negotiating to buy right-of-way, the tsar's nephew offered to sell what is now Alaska to Western Union. Sibley refused, but launched a campaign to persuade the United States to buy it, which was done in 1867. The expedition was abandoned when the first permanently successful transatlantic cable was laid in 1866, but it had mapped and reported Alaska's vast resources.

Sibley retired as president of Western Union in 1865 and became a builder of railroads in the middle west and south and owner of vast farm holdings. Part of his fortune was used to establish the Sibley College of Mechanic Arts Engineering (later Mechanical Engineering) at Cornell university, of which with Ezra Cornell (*q.v.*) he was one of the incorporators. He died in Rochester on July 12, 1888. (G. P. O.)

SIBSAGAR, a town and district of Assam, India. The town lies on the Dikhou river, about 10 mi. from its confluence with the Brahmaputra. Pop. (1961) 15,106. Established by the Ahom king Siba Singha (1714–44), Sibsagar was named after the large water tank built by Queen Ambika Devi in 1720. Among notable buildings are the temple of the god Shiva and the Vishnu and Gouri Devi temple, both erected during Siba Singha's reign. Sibsagar college is affiliated to Gauhati university. The town is a station on the North-East Frontier railway and lies on the national highway.

Sibsagar District, has an area of 3,453 sq.mi. and had a population (1961) of 1,508,390. It is situated between the south bank of the Brahmaputra, the western boundary of Nagaland and the Mikir and North Cachar hills on the south and west. The alluvial plains of the mid-Brahmaputra valley are dissected by a large number of tributaries. One-third of the district is forested and, besides providing valuable timber, the forests are abundant in wild life. Tea, rice, pulses and mustard-seed are grown on the plains. Oil deposits are exploited near Rudra Sagar and Moran in the northeastern part of the district; tertiary coal is mined at Naginimara on the Nagaland border. Sibsagar district was the heart of the Ahom kingdom and Gargaon (Nazira; 8 mi. from Sibsagar), the original capital of the kingdom, contains the ruins of forts and royal palaces. Jorhat, the last capital of Ahom, has agriculture, engineering and other colleges. An earthquake on Aug. 15, 1950, followed by large-scale floods, caused many deaths and widespread destruction throughout the district. (M. Ba.)

SIBYL (Sibylla), a prophetess in Greek legend and literature. The word is of unknown, probably Asiatic, derivation and appears first in the Greek philosopher Heraclitus (*c.* 540–*c.* 480 B.C.) of Ephesus as a proper name. The typical legend represents her as a woman of prodigious old age uttering predictions in ecstatic frenzy, but she is always a figure of the mythical past, and her prophecies, in Greek hexameters, are handed down in writing. In the 5th and early 4th centuries B.C. she is always referred to in the singular Sibylla is treated as her proper name and she is apparently located in Asia Minor (at Erythrae in Ionia or Marpessus near Troy) in uncertain antiquity. From the late 4th century the number of sibyls is multiplied; they are localized traditionally at all the famous oracle centres and elsewhere, particularly in association with Apollo, and are distinguished by individual names, sibyl being treated as a title. Varro in the 1st century B.C. lists ten (Persian, Libyan, Delphic, Cimmerian [in Italy], Erythraean, Samian Cumaean, Hellespontine, Phrygian, Tiburtine) and others could be added from other sources. Virgil in the *Aeneid* gives a vivid description of the consultation of Deiphobe, the sibyl of Cumae, by Aeneas, on the lines of a traditional consultation of the Delphic oracle. According to legend, this sibyl was granted by Apollo the gift of prophecy and a life of as many years as the number of grains of dust in her hand, but she did not ask for youth as well, so she gradually withered away almost to nothing. Finally she was hung up in a bottle, saying only that she wished to die.

A famous collection of sibylline prophecies, usually called the Sibylline Books, was according to legend offered for sale to Tarquinius Superbus, the last of the seven kings of Rome, by the Cumaean sibyl. He refused to pay the sum demanded for the nine books, so the sibyl burned three books, and then three more, before finally selling him the remaining three at the price she had originally asked for all nine. The books were thereafter kept in the temple of Jupiter on the Capitoline hill under the care of a priestly college, to be consulted in emergencies only by order of the senate. After the burning of the Capitol in 83 B.C. a new collection of Sibylline Books was compiled. The last recorded consultation of them was ordered by Julian the Apostate in A.D. 363, and they were officially destroyed under Stilicho (d. 408).

A Judaean or Babylonian sibyl was credited with writing the Judaeo-Christian Sibylline Oracles (*q.v.*). The sibyl came thus to be regarded by some Christians as a prophetic authority comparable to the Old Testament: hence the references to her in Christian literature, such as the *Dies irae*, where she is mentioned in the same breath as King David (*teste David cum sibylla*). The most notable examples of sibyls in art are those by Michelangelo in the Sistine chapel of the Vatican (*see* Fresco Painting).

See Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft*, 2nd series, vol. 2, col. 2073–2183 (1923). (H. W. Pa.)

SIBYLLINE ORACLES. During the Hellenistic age and under the early Roman empire, literary-minded Jews and Christians were impressed by the respect paid to the various sibyls (*see* Sibyl) as seeresses predicting the future, and they proceeded to create a collection of oracles, written in Greek hexameters in which the sibyls would first prove their reliability by predicting the recent past and then go on to predict the downfall of kings or empires in the immediate future; along the way they would set forth doctrines peculiar to Hellenistic Judaism or to Christianity. Not unnaturally, the Jewish apologist Josephus and such Christian apologists as Justin and Theophilus were greatly impressed by the way in which their doctrines were confirmed by such external testimonies. Both Theophilus and Clement of Alexandria referred to the sibyl as a prophetess apparently no less inspired than the Old Testament prophets; their pagan contemporary Celsus scathingly referred to Christians as "sibyllists." Those who criticized sibylline inspiration often attacked the crudity of the hexameters, relying on the theory that the gods could speak good Greek. In the Byzantine period no fewer than 12 of these compositions were collected in a single manuscript (in 14 books; ix and x are lost); an incomplete text of this collection was first published in 1545.

Modern scholars have been able to assign dates to the various oracles, largely by comparing the known sequence of events with what the oracles predicted. When real events are being described the oracle-writer is describing the past; when errors begin, he is predicting the future. This principle of dating has the merit of simplicity, though it can be too rigidly applied.

In any event, the oldest oracle is probably found in book iii (lines 97–294 come from 150–100 B.C.); books iv and v are Jewish works of the reign of Domitian (A.D. 81–96); books i–ii (originally a single book) and vi–vii are Christian; book viii reflects Christian

antagonism to the empire in the reign of Marcus Aurelius (A.D. 161–180). Books xi–xiv, considerably later, are Jewish; the Church Fathers do not quote from any of them.

BIBLIOGRAPHY.—M. P. Nilsson, *Geschichte des griechischen Religion*, vol. ii, pp. 105–107, 458 465 (1950); J. Geffcken, *Die Oracula Sibyllina* (1902); A. Kurfess, *Sibyllinische Weissagungen*, with German translation (1951); M. S. Terry, *The Sibylline Oracles* (1899).

(R. McQ. G.)

SICANI (English SICANS), the ancient inhabitants of western Sicily. Sikanie, the Greek name for Sicily in the *Odyssey* (xxiv, 307), appears to be a doublet of Sikelia, the usual Greek name (from Sikeloi, the Siculi or Sicels). Archaeologically there is no substantial difference between Sicans and Sicels in historical times; but ancient authorities distinguish the two peoples and assign a different origin to them. The Sicans, who were believed to be Iberians from Spain, were driven by the invading Sicels into the western and southern parts of the island (Thucydides, vi, 2). The Greek colonists of Gela and Acragas (Agrigento) on the south coast had to fight Sican wars; the people of this part of the interior retained more cultural independence of the Greeks than the Sicels of eastern Sicily did. In the Roman period no distinction is drawn between Sicans and other inhabitants of Sicily. (T. J. DN.)

SICILIAN VESPERS, the massacre of the French with which the Sicilians began their revolt of 1282 against Charles I (*q.v.*), the first king of the Angevin dynasty of Naples-Sicily, so-called because it began with a riot in a church outside Palermo at the hour of vespers on Easter Monday, March 30, 1282. The Sicilians were aggrieved by oppressive taxation and by the new regime's neglect of their interests; and the Aragonese king Peter III, whose wife Constance was heiress of the former Hohenstaufen kings of Sicily, was sponsoring a conspiracy, organized with Byzantine subsidies by Giovanni da Procida, a Sicilian exile at his court, for a rising in Sicily to break out when Charles launched his long-expected attack on the Byzantine Empire. The rising, however, broke out prematurely: some French soldiers at vespers in the church of S. Spirito behaved disrespectfully to a Sicilian woman and were killed by the congregation; the people of Palermo followed suit, massacring 2,000 French men, women, and children in the night of March 30–31; and the other towns of Sicily rose likewise (Messina not till April 28). Communes were proclaimed, and the Sicilians tried to put themselves under the protection of Pope Martin IV, suzerain of the Sicilian kingdom, who, however, refused to approve them. The Aragonese, to whom the Sicilians next appealed, landed at Trapani on Aug. 30.

The War of the Sicilian Vespers ensued. The Angevins, Charles I and his successor Charles II (*q.v.*), were supported by the papacy, by the Italian Guelphs, and by Philip III of France, who launched his disastrous campaign against Aragon in 1285 in an attempt to put Charles of Valois on Peter III's throne. The excommunicated Aragonese were helped by the Italian Ghibellines and profited especially from the naval genius of Ruggiero di Lauria (*q.v.*). The separation of the Aragonese and Sicilian crowns on Peter's death (1285) between his sons Alfonso III and James (I of Sicily) was followed, after the former's death (1291), by the latter's accession to Aragon as James II; and James, by the Treaty of Anagni (June 1295), made peace with the papacy, France, and the Angevin, to whom he renounced Sicily. The Sicilians, however, then took James's younger brother Frederick III (*q.v.*) as their king; and, despite Aragon's adhesion to his enemies, Frederick maintained himself in his kingdom, which he finally secured by the Treaty of Caltabellotta (Aug. 31, 1302). The papacy's authority was impaired by the long insistence of Martin IV and his successors on identifying themselves with the Angevin cause in the vain hope of preventing partition of the original Sicilian kingdom.

See S. Runciman, *The Sicilian Vespers* (1958).

SICILY (SICILIA), the largest island (9,830 sq.mi. [25,460 sq.km.]) in the Mediterranean sea, is separated from the "toe" of Italy by the Strait of Messina (2–10 mi. wide). It is shaped like a triangle, and on its shape its ancient name, Trinacria, was based. Its three points are Cape Peloro (Punta del Faro) in the north in the Strait of Messina; Cape Passero in the southeast; and Cape Boeo near Marsala in the west. Administratively the Lipari, Aegadian and Pelagie islands and the isles of Pantelleria and Ustica are part of Sicily, making the total area 9,926 sq.mi. (25,708 sq.km.). Under the Italian constitution of 1948 Sicily became a semiautonomous region; its capital is Palermo.

This article is divided into the following sections and subsections:

I. PHYSICAL GEOGRAPHY

1. Physical Features.—The mountains of Sicily as part of the Tertiary Alpine-Himalayan fold system are a continuation of the Atlas mountains of northwest Africa and the Apennines of Italy.

Northern Sicily is dominated by a group of mountain chains that extend from the Strait of Messina to the Torto river. The easternmost of these, the Peloritani mountains (Pizzo di Polo 4,219 ft. [1,286 m.]), is closely connected structurally with the Calabrian Apennines and is crystalline in composition. The Nebrodi mountains, farther east (Mt. Soro 6,060 ft. [1,847 m.]), and the Madonie mountains (Pizzo Carbonara 6,486 ft. [1,977 m.]) are limestone; the latter range is known for its underground drainage systems and provides drinking water for the city of Palermo. These mountains stand like a high wall between the north shore of Sicily and the interior and are traversed only by roads. Western Sicily is characterized by limestone massifs, some shaped like promontories or peninsulas (*e.g.* Mt. Pellegrino outside Palermo). The central and southwest part of Sicily consists of the "sulfurous plateau" between the Platani and the Salso rivers, a landscape of gypsum, sulfur and clay, desolate and poor; east of the Salso the plateau changes to low uplands between the towns of Gangi and Caltagirone, known as the Erei mountains. Their highest point (3,109 ft.) is near Enna. In the southeast the low Iblei mountains (Mt. Lauro 3,236 ft.), a complex tableland with low peaks, slope gradually to the sea. Eastern Sicily is dominated by the great volcanic cone, Mt. Etna (*q.v.*; 10,705 ft.); to the south the Catania plain extends as far as the foothills of the Iblei mountains. Along the coastline there are numerous small plains and low terraces.

2. Climate.—Sicily is warm in winter and summer. Near the coast the mean January temperatures are more than 10° C. (50° F.) and even inland more than 4° C. (40° F.). In summer the temperature rises considerably: near sea level the mean temperatures are 24° to 26° C. (75° to 79° F.) with extremes of 38° C. (100° F.). The effects of the heat are accentuated by the low

rainfall. The rainy season is confined to the late autumn and winter; the total rainfall on the north and south coasts and in much of the interior is less than 25 in. (635 mm.), in the extreme south only 15 in. (380 mm.). The result of this low rainfall, combined with deforestation and failure to systematize the mountain torrents, is that most of Sicily suffers from severe lack of water. The frequent, dry sirocco wind from the south is a scourge. (G. Kh.)

3. Vegetation and Animal Life.—Man's influence has greatly changed the natural vegetation. In the lower belt *Stipa tortilis* steppe and carob and olive-tree shrub, *maquis, garigue,* and woods of holm oak, cork oak, or aleppo pine cover the places free from cultivation. Papyrus grows along the Anapo and Ciane rivers. The submontane belt is broken up by cereal cultivation into scattered woods of pubescent oak, chestnut or mixed mesophilous woods. Native plane tree occurs along eastern rivers. Beech woods in the montane belt pass gradually into the formations of the alpine belt with juniper (*Juniperus hemisphaerica*), *Berberis aetnensis* or *Astragalus siculus* and the volcanic ash vegetation of Mt. Etna. (R. E. G. P.-S.)

The animal life has much in common with southern Italy, which is to be expected, but it also has some animals in common with Sardinia, for example, the ringed plover, black vulture and griffon vulture, which are not found in Corsica or southern Italy. Sicily shares the pratincole and the black and bearded vultures with Spain and the Balkans, and also shares the black wheatear with Sardinia and Spain. (Ma. Bu.)

II. ARCHAEOLOGY

The pre-Hellenic inhabitants of Sicily are called by classical authors Sicani or Siculi, and these two names have been used as conventional designations for those early cultures which have been revealed since Paolo Orsi began investigation in 1889. The term Sicanian is applied to that period of the Stone Age which followed the Paleolithic exemplified in the remarkable rock engravings of the cave of Addaura near Palermo (first published in 1953); the term Siculan is applied to the Chalcolithic, to the Bronze and to the Iron Ages, the first Siculan period being Chalcolithic, the second being Bronze Age, the third, from 900 to 700 B.C., being Iron Age and the fourth from *c.* 700 B.C., being the period when the native civilization was hybridized with the Greek. The earlier cultures tended to persist in "retarded" communities.

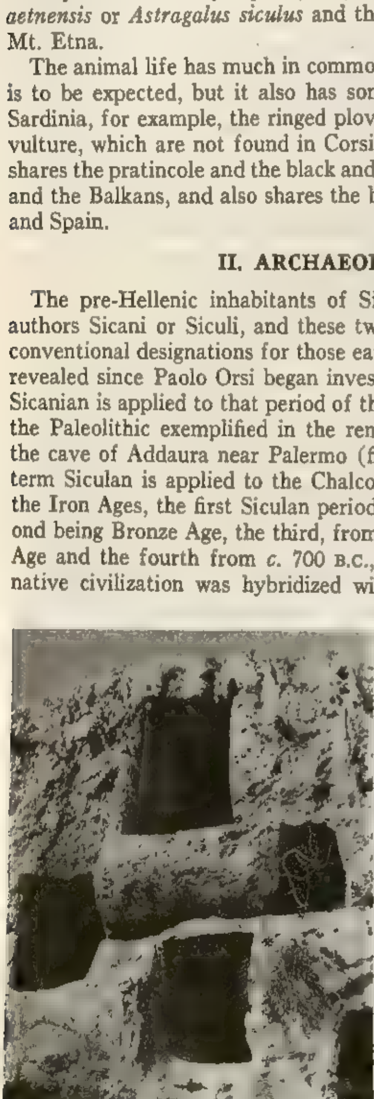
1. Sicanian Period.—The Sicanian period is known principally from fortified villages such as those at Stentinello and Matrensa. Stone implements from these places were of poor quality, and there were no other objects of interest except the pottery. This was handmade and baked in an open fire. Its surface is blackish-gray, and the geometrical ornament upon it, which appears merely incised, has in parts a strange regularity produced by real stamps—not by puncture by bone awls—a phenomenon unique at that early time. Later various styles of painted pottery are found; apparently later still are the incised wares of the Conca d'Oro and of Moarda, which are associated with the bell-beaker culture.

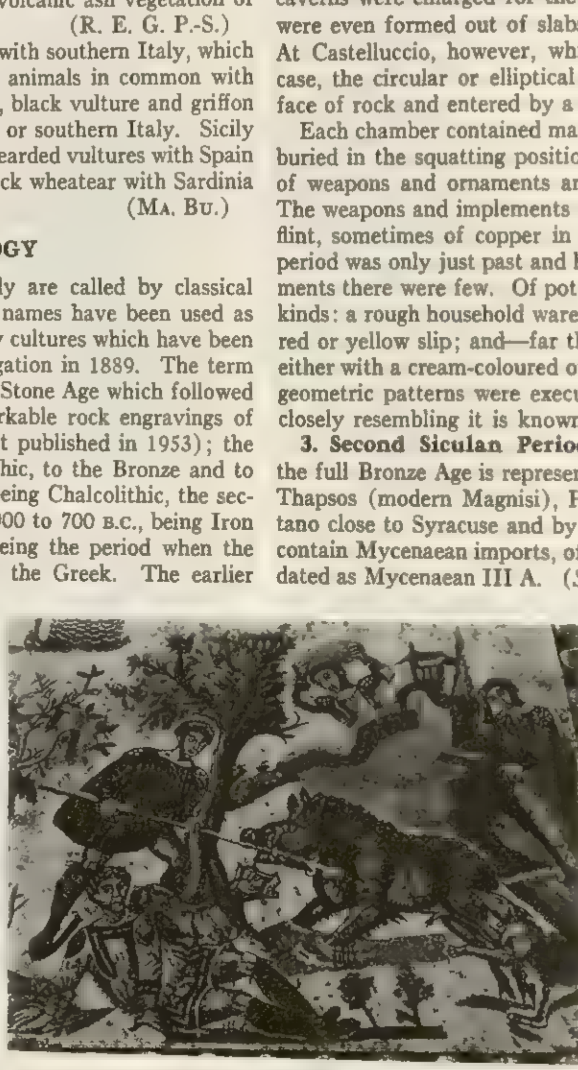
2. First Siculan Period.—The principal sites of this period are those of Castelluccio, Melilli, Monte Racello, Monte Tabuto, Vallelunga and Monte Salia. The typical burial was in a rock-hewn chamber, the construction varying according to the nature of the ground, so that at Monte Tabuto the dead were buried in disused flint-mines, while at Monte Racello natural caverns were enlarged for the purpose, and actual surface graves were even formed out of slabs on the top of the broken ground. At Castelluccio, however, which may be taken as the standard case, the circular or elliptical chambers were hewn in a vertical face of rock and entered by a short horizontal corridor.

Each chamber contained many skeletons. They were invariably buried in the squatting position, accompanied by a small number of weapons and ornaments and a regular equipment of pottery. The weapons and implements were sometimes of stone, especially flint, sometimes of copper in primitive forms as if the neolithic period was only just past and hardly forgotten. Of personal ornaments there were few. Of pottery, all handmade, there were three kinds: a rough household ware; pots of a better clay coated with a red or yellow slip; and—far the most important—a ware covered either with a cream-coloured or dark-red slip on which very simple geometric patterns were executed in dark brown. Nothing very closely resembling it is known anywhere else.

3. Second Siculan Period.—The second Siculan period of the full Bronze Age is represented at its best by the cemeteries of Thapsos (modern Magnisi), Plemmirio, Milocca and Cozzo Pantano close to Syracuse and by Caldare near Agrigento. All these contain Mycenaean imports, of which the earliest are clearly to be dated as Mycenaean III A. (*See* Aegean Civilization.) Bronze swords from the same sites are also of Mycenaean origin or influence and prove the existence of a considerable direct trade with the Aegean. All the cemeteries consisted of rock-hewn tombs, a natural evolution from the rock tombs of the first Siculan. The roof of the chamber, however, was sometimes of tholos (beehive) form, which again recalls Aegean precedents. Inside it was usually elliptical; a raised bench cut in the rock ran round it, and niches were often hewn in the walls. Within this chamber the dead were buried as if seated at a stately banquet. The size of the great food basins was remarkable; they were often 2½ or 3 ft. in height.

The native pottery was made principally in a gray-faced ware ornamented with molded strips or with a few sparsely incised lines; there was also a yellow-faced ware. Painted ware is unknown. The shapes are very few—a high stand, a biconical cup with side handles, a conical cup, a jug and one or two large water jars. For-





(BOTTOM RIGHT) BY COURTESY OF THE MUSEO NAZIONALE ARCHEOLOGICO, SYRACUSE; ALL PHOTOS BY LEONARD VON MATT

(ABOVE) TOMBS FROM THE GREAT SICEL BURIAL GROUND OF PANTALICA; (TOP RIGHT) MOSAIC OF A BOAR HUNT FROM THE LATE 3RD CENTURY VILLA NEAR PIAZZA ARMERINA; (RIGHT) TETRADRACHM OF THE WINGED FIGURE OF NIKE, COINED TO COMMEMORATE AGATHOCLES' VICTORY OVER THE CARTHAGINIANS, 310 B.C.; (BOTTOM RIGHT) BRONZE AGE VESSELS IN THE CASTELLUCCINO STYLE, FOUND NEAR AGRIGENTO

eign trade added the amphora, pyxis, pedestaled cup and stirrup-handled vase and gave a great impetus to metalworking. Swords, daggers and even basins may have been imported, but the various hoards of bronze objects show that a great quantity of native work must have existed. The tombs, however, were systematically ransacked for metals after the Roman time. Another proof of foreign influence was the introduction of the fibula (large safety pin), which appears at Cozzo Pantano for the first time. Early fibulae were of two types, the plain violin bow with bamboo knots and the harp-shaped or elbowed fibula, each made in a massive form.

The four cemeteries near Syracuse must be assigned on the evidence of the foreign imports to the 14th and 13th centuries B.C. Caldare, farther inland, may be later.

4. **Intermediary Period.**—Next in chronological order come the large necropolises like Pantalica, Grammichele, Caltagirone and Monte Dessueri which span the gap between the Bronze and Iron Ages, the later graves in them belonging entirely to the Early Iron Age. On Orsi's system they are to be classed as intermediate between the second and the third Siculan. The form of the tomb had already begun to be modified, the number of burials in each grave was smaller, and the tomb furniture was later in character. This was especially perceptible in the personal ornaments; the fibulae changed considerably, for in place of the primitive violin bow the simple rounded bow was predominant and more sophisticated forms, like the eyed harp fibula, began to appear.

In addition to arm rings and finger rings and mirrors of bronze there were also found, though rarely, gold rings, silver armlets and silver rings. Rectangular bronze razors came into use, like those known in Italy but of different origin. There were no weapons, perhaps because of the rifling of the graves, but flame-shaped and leaf-shaped bronze knives occurred. The native pottery was very varied in form at that time, the best of it being in a ware faced with red hematite.

5. **Third and Fourth Siculan Periods.**—On a site like Pantalica, with a range of several centuries, the Iron Age (Orsi's third period) begins without any striking changes. Modifications occurred in the types of the fibulae and there was a gradual, but quite perceptible, deterioration of the pottery. All the Bronze Age types of fibulae gradually gave way to later forms, of which the most general and popular was the two-eyed serpentine, which continued in use down to 500 B.C.

The first appearance of this fibula marks the beginning of the Early Iron Age, and if the contents of tombs belonging to this stage are isolated and examined separately from the rest, then certain definite characteristics begin to appear. The old Siculan civilization of the great days survived in a much impoverished form. The architecture had lost all its beauty and elaboration of detail. Within a very simple chamber the dead were no longer seated at a banquet but extended at full length on the ground with their heads resting on a block of stone; and the objects buried with them consisted of little but a few small water jars and trumpery pots. Everything of interest in Siculan life from the 9th to the 5th century B.C. is either a Greek importation or the direct imitation of a Greek original, known either from traders or, after *c.* 730, from Greek settlers.

From a number of sites, the most important of which were Finocchito, Lentini (the ancient Leontini) and Licodia Eubea, were obtained examples of geometrically painted vases, the earliest of which are of pure Dipylon style, while the latest are a hybridized product which may be termed Greco-Siculan.

The introduction of these new models led to the imitation of purely Greek shapes by the native potters, so that *oenochoai* and *askoi* (vases) were copied in the rough country ware, and to a closer study of decorative designs, which resulted in the production of a new kind of white-faced ware with geometrical patterns painted upon it. This gradually improved in technique until it reached its high-water mark in the late third and early fourth Siculan periods. The Greco-Siculan ware continued in use until it was finally replaced about 500 B.C. by purely Greek imported vases. For this whole period, the stratigraphic investigations at Megara Hyblaea (1949 onward) yielded valuable results.

(D. R.-M.; A. W. V. B.; X.)

6. **Classical Archaeology.**—After having long been relegated to secondary rank in art as merely a provincial area in relation to Greece proper, Sicily in recent years has steadily emerged as an artistic entity in its own right. The earliest works of sculpture confirm the literary tradition of a "Daedalic" school (*see* GREEK ART: *The Greek Archaic Period*); and the art of the later archaic period, including clay statues and statuettes, pursued a course of development parallel to that of Greece, with which it clearly remained in contact while achieving considerable freedom in style and technique. "Nowhere more forcibly than in its mouldings does the architecture of the West stand forth as a vigorous, original architecture, based on, but independent from, the forms of Greece, the Aegean, and Asia Minor" (L. T. Shoe, *Profiles of Western Greek Mouldings*, American Academy in Rome, 1952). Down to the beginning of the 5th century B.C., when the traditional terracotta revetments were superseded by marble, the coloured adornment of the temples showed features distinctly characteristic of this area. Throughout the 5th and the early 4th centuries, the issues of coins exhibited an exuberance and a standard of artistic excellence all their own; clearly the die cutters stood in close relation to the gem engravers and, in some cases, were identical with them. Although the rich ceramic yields of the cemeteries consist largely of imported products, first of the Corinthian and later of the Athenian potteries, there is an abundance of local wares as well, revealing several clearly defined schools of the 4th century and later; some of the painted decoration suggests the influence of the Western Greek theatrical plays. Excavations at the Phoenician town of Motya (*q.v.*) and many Sicel towns throw light on the background of the Greek cities. Sicily is rich in remains of villas and other buildings of the Roman period; and the architecture and mosaic art of late antiquity are nobly represented at Piazza Armerina (*see* below: *Roman Sicily*). *See* further GREEK ARCHITECTURE; GREEK ART; ROMAN ART; NUMISMATICS: *Greek Coins*. For some of the outstanding architectural monuments, *see* AGRIGENTO; GELA; SEGESTA; SELINUS; SYRACUSE.

(A. W. V. B.; X.)

III. HISTORY

At the coming of the Greeks three peoples occupied the island of Sicily: in the east the Sicels or Siculi (*q.v.*), who gave their name to the island but were reputed to be late-comers from Italy; to the west of the Gelas river, the Sicani (*q.v.*); and in the extreme west the Elymians, a people to whom a Trojan origin was assigned, with their chief centres at Segesta (*q.v.*) and at Eryx (Erice) with its temple of Aphrodite. The Sicels spoke an Indo-European language; there are no remains of the languages of the other peoples.

Thucydides says (vi, 2) that the Greeks were preceded in Sicily by Phoenicians, who established trading stations on promontories and small islands all round the coast. But no material remains of such early settlements have been found. Thucydides adds that when the Greek colonization began, the Phoenicians withdrew to the western corner of the island, where they had the Elymians as allies and the support of Carthage. There they had three towns, Motya, Soloeis and Panormos (Palermo). Hostilities between Greeks and Phoenicians are not heard of earlier than the Cnidian attempt to settle on the westernmost point of Sicily *c.* 580 B.C. (*see* below).

1. **Greek Sicily, to 413 B.C.**—Mycenaean traders had already visited Sicily and the Lipari Islands between the 15th and 13th centuries B.C., and the fibula was introduced by direct or indirect importation from the Aegean (*see* above, *Archaeology*). There was then a long period when Sicily was beyond the horizon of Greek sailors (the few references in the *Odyssey* need be no older than the early period of colonization). Toward the middle of the 8th century intercourse was resumed, and the first colony was planted *c.* 734 B.C. at Naxos on the east coast by Chalcidians from Euboea, who had already founded Cumae near Naples. (Precise dates for the foundation of the Sicilian colonies are given by Thucydides; and though some modern scholars suppose that these are based only on calculations made about 300 years after the era of the first colonies, they may be accepted as approximately correct.) In the year after Naxos, according to Thucydides, the Corinthians

founded Syracuse, the greatest city of ancient Sicily, and at the same time occupied Corcyra (Corfu) which commanded the way thither. The east coast was rapidly occupied: the Chalcidians founded other colonies at Leontini and Catania (both *c.* 729) and held the Straits of Messina by the foundation of Zancle (later Messana; modern Messina) on the Sicilian side and Rhegium (Reggio di Calabria) on the Italian side (both probably in the last quarter of the 8th century). The Megarians founded Megara Hyblaea (*q.v.*) at about the same time as Leontini. On the south coast the earliest colony was Gela, a joint Cretan and Rhodian settlement (*c.* 688). The southeastern corner was occupied by the secondary Syracusan foundations of Acrae (*c.* 663), Casmenae (*c.* 643) and Camarina (*q.v.; c.* 599), the last of which was the most important. In or about 628 the Megarians founded Selinus (*q.v.*), the westernmost Greek city of Sicily, facing the coast of Africa. Between Gela and Selinus was Akragas or Acragas (Roman Agrigentum), a Geloan colony, the last of the foundations of the great age of colonization (580), which rapidly rose to be the second city of Sicily. On the north coast there was long only a single Greek city, the Chalcidian Himera (*q.v.*) (*c.* 649). About 580 the Cnidians founded Lipara, after an unsuccessful attempt to settle in the neighbourhood of Eryx. The defeat of this attempt by the Elymians and Phoenicians marks the first recession of the tide of Greek settlement, and the Greeks never succeeded in occupying the western part of the island. The mountainous centre remained in the hands of Sicels and Sicani, who were increasingly hellenized in ideas and material culture.

After the foundation of the colonies, little is known of them until the 5th century. They prospered materially, building temples whose remains, at Syracuse, Akragas and Selinus, are among the finest monuments of archaic and classical architecture. Their sculpture and other arts were vigorous, though provincial and largely dependent on impulses from mainland Greece. There is some reason to believe that they had a colonial economy, producing food (corn, sheep and cattle, fish) and importing manufactured objects (of which clay vases are the chief survivors) from Greece —in the period before 550 B.C. mainly from Corinth, afterward from Athens. The broad acres of the colonies and; at Syracuse and probably also elsewhere, the use of Sicel serfs gave rise to a life of easy circumstances and, in the 5th century, to the great wealth of the tyrants (*see* below). The fine series of Sicilian coins began toward the middle of the 6th century (*see* NUMISMATICS: *Greek Coins*). The poet Stesichorus of Himera, who retold many of the stories of epic with a new spirit reflected in archaic vase paintings, belonged to this period.

There were many tyrants—unconstitutional monarchs—in the 6th century, especially in those towns which were pressed by the growing hostility of the Carthaginians; the most famous was Phalaris (*q.v.*) of Akragas; in general, however, this was a period of aristocratic or oligarchical constitutions. The early 5th century saw tyrants in most cities. The tyranny of Gela was founded in 505, and Hippocrates (498–491) carried his arms over most of eastern Sicily; his successor Gelon (*q.v.*) took Syracuse in 485 and transferred his seat thither, enlarging the city with men from other colonies and settlers from old Greece. This is the beginning of Syracusan rule over Greek Sicily. The families of Gelon and of Theron (tyrant of Akragas, 488–472) were united by marriage. Theron seized Himera, expelling its tyrant Terillus, who with his son-in-law, Anaxilas of Rhegium (494–476), ruler also of Zancle, formed a coalition against Gelon and Theron and invited Carthaginian support. The Carthaginian invasion was defeated in a single battle at Himera (480), and its commander, Hamilcar Barca, son of Hanno (his mother was a Syracusan), was killed. The victory opened a short period of brilliant prosperity for Greek Sicily. The courts of Gelon and of his brother and successor Hieron I (*q.v.;* 478–467) and of Theron were visited by Pindar, Aeschylus, Simonides of Ceos and Bacchylides; the leading artists of Greece were commissioned to make rich offerings at Olympia and Delphi; fine temples were built as thank offerings for the victory of Himera. Hieron extended his activity to Italy, where he protected Locri against Anaxilas of Rhegium and defeated the Etruscans at sea off Cumae (474). The darker side of the tyrants' rule is seen in forced movements of population from city to city and in the destruction of many Greek cities (*e.g.,* Megara Hyblaea). Hieron refounded Catania with great splendour, under the name of Etna, after expelling the Chalcidian inhabitants; he set up his son Deinomenes as king and hoped that this would be a refuge for his family; but within a few years of his death the original inhabitants returned, and the new settlers carried the name Etna inland to Inessa, on the slopes of Mt. Etna.

These tyrannies were short-lived. Theron's power fell to pieces when his son Thrasydaeus succeeded. Hieron's brother Thrasybulus was overthrown by a combined movement of Greeks and Sicels (466). They were succeeded by uneasy democracies in most cities, Syracuse still taking the lead. The Sicels then for the first time formed a united power, under Ducetius, who founded a new city at Palici in the plain of Catania and became a serious threat to Syracuse and to Akragas. He was defeated and his political work largely undone, except for the foundation of Cale Acte (Caronia) on the north coast. The Sicels did not again attempt to combine as a force, but their hellenization continued; many Sicel cities began to coin on the Greek model.

The extent of the native contribution to Sicilian culture is disputed. The Sicels may have had a worship akin to the cult of Demeter and Persephone, which was widely spread over both Greek and native Sicily; but the localization of the rape of Persephone at Enna is due to Greek influence. The material culture of the Greek colonies shows no Sicel elements. Some words of Italic origin were taken into the Greek dialects of Sicily; and the racy comedies of Epicharmus and the mimes of Sophron may show some kinship of spirit to Italian comedy. The main Sicilian contribution to Greek literature and thought in the 5th century was in the development of rhetoric and its transference to Athens by Gorgias (*q.v.*). A more original philosopher was, it appears, Empedocles (*q.v.*) of Akragas, who was poet, physician, politician and perhaps something of a charlatan.

The Athenians looked for allies in Sicily at least as early as 458/457, when they made a treaty, still extant, with Segesta. At the beginning of the Peloponnesian War they sought for means to cut off supplies from Sicily to their enemies and allied themselves with the Ionian cities, Leontini and Rhegium. In 427–426 they sent a small squadron to Sicily but accomplished little. In 415, during the "peace of Nicias," they were led by expansionist ideas, expressed by Alcibiades, to send a large force in answer to an appeal for help from Segesta and Leontini. The ultimate object of this expedition was the subjection of the whole of Sicily. The Athenians were received with suspicion, even by some of their allies, and Naxos and Catania were the only Greek cities to support them; they also had help from some of the Sicels. The utter defeat of the Athenian force under the walls of Syracuse in 413 (*see* PELOPONNESIAN WAR; SYRACUSE) was achieved in part by the arrival of the Spartan leader Gylippus and was followed by the dispatch of small forces from Syracuse and other Sicilian cities to help the Spartans against Athens.

2 Carthaginian Wars.—The failure of Athens left the field open for Carthage, with whom the Athenians had sought an alliance. In 409 an army under Hannibal, son of Gisgo, landed to support Segesta against Selinus. Selinus and Himera were taken and destroyed; Himera, where Hannibal sacrificed 3,000 of the citizens to avenge his grandfather Hamilcar, was never rebuilt but was replaced by a new town, Thermae Himerenses. In 406 Akragas fell also. Dionysius (*q.v.*) the Elder rose to power at Syracuse as its military leader but was forced to evacuate the populations of Gela and Camarina before he could obtain a peace with the Carthaginians which left him in control of Syracuse. For the rest of his life he fought a series of inconclusive wars with the Carthaginians. In 397 (possibly 398) he took Motya, the chief Phoenician city in Sicily (which had a considerable Greek element within its walls) and seemed on the point of driving the Phoenicians out of the island. But in the following year the tables were turned by the arrival of a Carthaginian fleet under Himilco in the harbour of Syracuse. A pestilence, spread among Himilco's troops from the marshes of Lysimelia, assisted Dionysius in winning a complete victory, which confirmed his position as ruler of Syracuse and mas-

ter of most of Sicily. But he had to fight other Carthaginian wars in 392–391, 383–378, and 368–367, the year of his death. In the peace made in 378 the boundary between Greek and Carthaginian Sicily was fixed at the river Halycus (Platani). Selinus and Himera remained on the Carthaginian side, Akragas became a frontier town. Lilybaeum (the modern Marsala) replaced Motya in 396 and became the chief centre of the Phoenicians in Sicily.

Dionysius used his position as defender of Greek Sicily to build up a personal power which anticipates the Hellenistic monarchies in many ways. The transplantations of population and refoundations of cities which had been a feature of the rule of Gelon and Hieron continued. New foundations were Tauromenium (modern Taormina), which replaced Naxos, and Tyndaris on the still rather empty north coast. The Sicel towns of Hadranum (Adrano) near Mt. Etna and Halaesa on the north coast were also refounded, one by Dionysius, the other by his Sicel ally Archonides of Herbita. Dionysius made considerable use of the forces of Sicel allies, and the distinction of barbarian from Greek Sicily begins to disappear. He also employed large mercenary forces, some of them from Gaul and Spain.

After 390 Dionysius intervened in the affairs of the Greeks in Italy, in alliance with the Lucanians, who had recently pressed down into the "toe" of Italy. Locri was his base in Italy. He took Rhegium and Caulonia and defeated the forces of Crotona and Thurii, thus winning control over most of Greek Italy, except Tarentum (modern Taranto). He sent a fleet into Etruscan waters which plundered the wealthy temple at Caere. His activities in the Adriatic were of more permanent importance; he planted a colony at Lissus (modern Lezhe in Albania) to control the crossing of the Ionian sea; he helped the Parians to colonize the Dalmatian island of Pharos (Hvar); and Syracusan exiles founded Ancona, perhaps with his consent and backing.

Dionysius had as ally the Molossian Alcetas and engaged Illyrian mercenaries. These activities appear to have been in preparation for intervention in the affairs of Greece.

The rule of Dionysius at Syracuse depended on foreign mercenaries and on secret police, and many of the typical features of Plato's and Aristotle's accounts of Greek tyrants are no doubt derived from him. But he was well served by efficient ministers, such as Philistus the historian (who however went into exile). He was himself a poet and, like earlier tyrants, kept a court, which was visited by Plato and other philosophers and poets. He made Syracuse the greatest power in the Greek world, and it became the largest and, probably, the most populous of Greek cities, with its extensive fortifications (*see* SYRACUSE). His power, however, did not long survive him. Most of his conquests in Italy fell to the Lucanians and Bruttii, and at Syracuse his son Dionysius the Younger, after ruling from 367 to 356, was expelled by his uncle Dion (*q.v.*). Dion was killed in 354, and there followed ten years' confusion: fighting in Syracuse between Dionysius the Younger and the citizens; tyrannies in other cities; and renewed danger from Carthage, allied with Hicetas, tyrant of Leontini and rival of Dionysius. The Syracusans appealed to their mother city Corinth, and Timoleon (*q.v.*) came as a deliverer. He freed the cities from tyrants and defeated the Carthaginians at the battle of the Crimisus (probably 341); but the boundary remained at the river Halycus. Timoleon's reputation was high because he restored the democracy at Syracuse and retired after his work of liberation was done; but his settlement did not last long after his death (*c.* 336). The Carthaginians, who had already played off one city against another, continued this policy, while the Greeks consumed their energies in struggles between would-be tyrants. They found another leader in Agathocles (*q.v.*), one of the ablest soldiers of the time, who made himself tyrant of Syracuse in 317. Akragas, strengthened by Syracusan exiles, stood out again as the rival of Syracuse; and Hamilcar son of Gisgo won many Greek cities to the Carthaginian alliance and blockaded Agathocles in Syracuse. Agathocles broke through and carried the war into Africa, where he won many successes (310–307), but was finally completely defeated and had to flee back to Sicily. In spite of this defeat he maintained his position at Syracuse and made peace on the old terms with Carthage. He formed marriage alliances with Ptolemy I of Egypt and with Pyrrhus of Epirus, who married his daughter. He was the first of the Sicilian tyrants to take the title of king. He died in 289. In spite of his reputation for treachery and massacre, his rule was remembered as a period of prosperity.

In the troubles which followed Agathocles' death, his disbanded Campanian mercenaries seized Messana and called themselves the Mamertini (*q.v.*), children of Mamers or Mars. The fortunes of Sicily were thus linked with Rome. Another new foundation of that time was Phintias (Licata) at the mouth of the southern Himeras, named after the Agrigentine tyrant Phintias (289–279). New Carthaginian attacks called forth another liberator from overseas, Pyrrhus (*q.v.*) of Epirus, but his Sicilian war (278–276) was a mere interlude between the two acts of his war with Rome. When he left Sicily to fight for Tarentum against Rome, he had to fight his way out through Carthaginians and Mamertini, the latter already allied with Rome; he said, in true prophecy, that he left Sicily as a wrestling ground for Romans and Carthaginians.

The Syracusans chose as leader Hieron II (*q.v.*), who defeated the Mamertini and came near to capturing Messana; he won the title of king of Syracuse (*c.* 270). In 264 his attack on the Mamertini led to the intervention of Rome and the First Punic War (264–241). The war began as a three-cornered event between Rome, the Carthaginians and Hieron; but in 263 Hieron turned from the Carthaginians to Rome and formed an alliance to which he remained loyal for the rest of his long life. The Romans were thus free to use Greek Sicily as a base for war with Carthage. The western part of the island, both Greek and Phoenician, suffered greatly in this long war (for its course *see* PUNIC WARS). By the treaty which ended it Carthage ceded to Rome all its possessions in Sicily, which thus became the first Roman province (241). Hieron retained possession of eastern Sicily, south of Messana. His rule was able and enlightened, and his financial enactments, particularly his corn laws, were taken over when Rome incorporated his kingdom. This period of peace was the last golden age of free Sicily: the great theatre at Syracuse was rebuilt by Hieron, and Theocritus sang the pastoral pleasures of his native island and offered an encomium to Hieron (*c.* 275).

3. Roman Sicily.—At the outbreak of the Second Punic War Hieron held firm to the Roman alliance, but after his death in 216 his grandson Hieronymus repudiated it. Hieronymus was overthrown by revolution at Syracuse (215), but the city had to stand a siege from the Romans. The great fortress of Euryelus (*see* SYRACUSE: *Archaeology*) perhaps took its final form then, under the inspiration of Archimedes, though there is no reason to doubt that its construction began in the time of Dionysius I. Syracuse was taken and sacked in 212, Akragas after a further campaign in 210. The whole of Sicily then became Roman.

Little is known of the early organization of the Roman province. It was governed by a praetor sent out yearly from Rome, who after the annexation of Syracuse had his capital there. Two quaestors were appointed, one with his office at Syracuse, the other at Lilybaeum. The province included a number of free cities: Messana, Tauromenium and Netum (Noto) were allied cities (the two latter had probably taken the Roman side in the Second Punic War); a number of others, including allies from the First Punic War, were also free—Segesta, Halicyae, Panormos, Halaesa and Centuripe. The rest paid tithes to the Roman people according to the law of Hieron, which was extended to the whole island. Sicily had long had a surplus of corn for export, and Livy records occasional dispatch to Rome as early as the 5th century B.C.; the island now became the granary of the Roman people. The rolling country of the central part of the island was suitable for pasture and cultivation on a large scale, in *latifundia* (landed estates), and slave gangs were introduced on the estates both of rich Sicilians and of Roman citizens. Hence rose the two great slave revolts of the second half of the 2nd century B.C., the first, led by Eunus, from 135 to 132, with Enna and Tauromenium as its centres, the second from 104 to 100; both periods of internal and external stress at Rome. The settlements after these two wars by Publius Rupilius (131) and Manius Aquilius (99) modified the organization of the province.

In spite of slave wars and the burden of Roman provincial

governors and tax farmers, Sicily was not unprosperous under the Roman republic. It was free from external dangers; and even the unprivileged cities kept their own laws, magistrates and assemblies, and provision was made for lawsuits between Romans and Sicilians and between Sicilians of different cities. There seems not to have been much commercial exploitation; tax collecting was normally in the hands of the Greeks themselves, not of Roman *publicani;* and smallholdings continued to be the rule in many cities. The wealth of the cities, both free and tributary, may be seen from the speeches of Cicero in prosecution of Gaius Verres (*q.v.*), who in three years' governorship (73–71) had plundered widely, with especial attention to works of art. He also failed to defend the province against pirates.

Sicily was again a battlefield between 43 and 36 B.C., when Sextus Pompeius held Messana and cut off the corn supply of Rome. In the division of provinces between Augustus and the senate, Sicily fell to the latter. It had perhaps received Latin rights from Julius Caesar. Augustus planted colonies at Panormos, Syracuse, Tauromenium, Thermae, Tyndaris and Catania. But the island remained Greek; not only the old Greek cities but also the old Sicel towns, which had long been completely assimilated to the Greek, used Greek as their everyday language, though Latin was the official language. Christianity was early introduced to Syracuse, where the catacombs and early churches (belonging mainly to the Byzantine period) are second only to those of Rome. The island has little history in the Roman imperial period; its continued prosperity is shown by the Roman amphitheatres of Syracuse and Catania and by Roman remains in other cities large and small; a few cities declined and show now only Greek remains, but these were in general replaced by others. The tourist traffic in Sicily is as old as the Roman period. Cicero praises the climate of Syracuse, and the emperor Hadrian climbed Etna in A.D. 126 to see the sunrise. In the late 3rd century a rich Roman, perhaps a high court official, perhaps even an emperor (the name of Maximian has been suggested), retired and built a huge villa, with floor mosaics with lively pagan scenes, near the modern town of Piazza Armerina. In the reorganization of the late 3rd–early 4th century Sicily became one of the provinces of Italy and, together with Africa, Raetia and western Illyricum, was made part of the Italian prefecture; with Sardinia and Corsica, it was part of the diocese of Italy and the vicariate of Rome. (T. J. DN.; X.)

4. The Byzantine Period, to 827.—The earlier Germanic invaders of the Roman empire did not reach Sicily, though Alaric had designs upon it. The Vandals under Gaiseric, however, raided and invaded Sicily from Africa for about 40 years intermittently until in 476 they ceded the greater part of it to Odoacer, the barbarian king of Italy, against the payment of an annual tribute. The Ostrogothic successors of the latter paid no tribute to the Vandals after 491, but Theodoric gave Lilybaeum to the Vandal king Trasamund, to whom his daughter Amalafrida was married in 500. The Byzantine general Belisarius, having destroyed the Vandal kingdom in 533–534, occupied Sicily in 535, when his emperor began hostilities against the Ostrogoths in Italy; and though Totila recovered part of the island in 550, it was lost again before the final overthrow of the Ostrogothic kingdom at Taginae (552).

Sicily had thus come under Byzantine rule. When the emperor Heraclius and his successors divided the empire into themes (provinces), Sicily became one of them, placed between the exarchate of Ravenna in the north and that of Carthage in the south; it was administered by a *patricius* responsible to the government at Constantinople. Probably after the Italian revolt of 726 and certainly after the fall of the exarchate in the middle of the 8th century, the Byzantine dominions in southern Italy were incorporated in the theme of Sicily, an arrangement which lasted until the time of the Arab conquest, when the mainland dominions were formed into the themes of Calabria and Longobardia.

Ecclesiastically, Sicily remained at first under the papacy, which in addition to rights of jurisdiction had considerable interests in the island, arising from its vast Sicilian estates. Gregory I's letters admirably illustrate the importance attached to them by the pope. The Iconoclastic controversy and the ensuing revolt against Byzantine rule under papal leadership (726) led to the confiscation by the emperor of the papal estates in Sicily and southern Italy; soon after, the ecclesiastical jurisdiction of these regions passed to the patriarch of Constantinople.

These political and ecclesiastical changes corresponded to demographic and cultural developments of longer standing. In Sicily and southern Italy, the Greek element had been greatly strengthened since the end of the 6th century as a result of emigration from other Byzantine provinces after the Avar and Slav invasions of Greece and, later, of the Persian and Arab conquests. The hellenization of Sicily, which appears to have been all but complete in the 8th century, is revealed in the history of the Sicilian Church. The Greek rite, already used at the time of Gregory I, spread during the 7th century from the eastern coast over the whole of the island. By the time of its separation from Rome, the Sicilian Church was virtually Greek. Sicily was thus well on the way to becoming a fully integrated part of the Byzantine empire when the Arab conquest began in 827.

Ever since the 7th century, Arab expansion in North Africa had constituted an immediate threat to Sicily and to southern Italy, the occupation of which would moreover expose Greece, the exarchate of Ravenna and Dalmatia to Saracen attack. Sicily consequently became a vital link in the imperial defense against Islam. Constans II was the first Byzantine emperor fully to appreciate this; in 663–668 he tried to strengthen the Byzantine position in Sicily by his own presence. Constans, however, was assassinated during a rebellion at Syracuse in 668. Mizizes (Mezezius) who was then proclaimed emperor by the Byzantine militia, fell from power in 669.

This rebellion was followed by other risings in Sicily in 718 and in 781, which, although primarily of a military nature, doubtless also reflected separatist tendencies. A more dangerous revolt, that of the naval commander Euphemius about 826, was to be the immediate cause of the Arab conquest of Sicily.

5. The Arabs.—Following the conquest of North Africa, Arab descents from Africa began in 703. They were interrupted for the invasion of Spain (710), and the next serious attacks were probably not made before 740 and 752–753. Energetic Byzantine reaction on this last occasion was followed by a respite lasting more than 50 years. But in 827 Euphemius, who had been proclaimed emperor in Sicily, appealed for help to the Aghlabid amir of Africa. But once the Arabs had landed they began the conquest on their own account. Euphemius was murdered by Sicilians in 828; and Palermo fell in 831 and became the base for further conquests. When a stronger policy was adopted under the emperor Michael III, it was too late; after the Byzantine defeat near Butera in *c.* 845, the whole of the Val di Noto fell into Arab hands; its occupation was practically completed with the capture of Enna in 859. A Byzantine offensive under Basil I led to some temporary successes, but the fall of Syracuse in 878 all but sealed the fate of Sicily; when they captured Taormina in 902, the Arabs were practically masters of Sicily. Local resistance and insurrections, however, continued. Together with internal discords in the Arab camp they did much to slow up the Arab advance, despite the scant help from Constantinople. The Val Demone (*i.e.*, the region between Etna, Messina and Caronia) held out longest; and the last Byzantine stronghold, Rametta, was not lost until 965. The Byzantines, however, did not abandon hope of reconquering Sicily. Basil II was planning a Sicilian expedition at the time of his death (1025); it was actually dispatched in 1038 under the great Byzantine general, George Maniaces. The campaign was highly successful, and a large part of eastern Sicily, including Messina and Syracuse, was recaptured; but after Maniaces was suddenly recalled, the Byzantine position on the island collapsed.

The Arab conquest separated Sicily not only from Constantinople but also from the Italian mainland, where the Arabs did not succeed in establishing themselves permanently. The history of Arab Sicily, on the other hand, was marked by growing independence from Africa. Until 909, it was under the Aghlabids; after their fall, it passed under the Fatimids, who moved their capital to Cairo in 972. But from the middle of the 10th century the office of the governor (amir) became hereditary in the dynasty of the Kalbids, until the anarchy after Maniaces' conquests led

to the fall of that dynasty and to the division of Sicily into a number of principalities, while Palermo acquired self-government.

During and after the conquest, large Arab immigration from Africa took place; this, together with the conversions of Christians, contributed to make Sicily not only politically but also culturally part of the Arab world. However, the Greek Christian element remained predominant in the Val Demone; and even after the Arabs gained a majority in the rest of Sicily, Christian groups remained scattered over the island.

As in other Arab countries, the Christians, although placed in a position of legal inferiority, enjoyed religious toleration and a measure of self-government in return for paying taxes which may have been often less burdensome than those levied by the Byzantines. Relics of the Greek episcopate seem to have survived to the end of the Arab period; so did a number of Basilian monasteries; both provided the principal link with the Byzantine world outside (particularly with Calabria, where the Greek population had been strengthened by emigrants from Sicily). As far as the Sicilian clergy, secular and regular, was concerned, this emigration seems to have been less the result of persecutions than of the gradual spread of Islam over Sicily.

6. The Norman Conquest: Roger I.—When the Normans began to conquer Sicily in 1060, they were welcomed as liberators and their progress was doubtless assisted by the Christian population. They were no newcomers to the island. Norman mercenaries from the mainland, among them two sons of Tancred of Hauteville, had fought in Maniaces' army and taken part in the capture of Messina. In 1059 Pope Nicholas II invested another son of Tancred, Robert (*q.v.*) Guiscard, with his past and future conquests not only in Apulia and Calabria but also in Sicily. In his oath of allegiance Robert styled himself "by the grace of God and St. Peter duke of Apulia and Calabria and, with their help, hereafter of Sicily." (*See* NAPLES, KINGDOM OF.)

The conditions under which the Norman expedition took place were to leave a profound mark on the history of Sicily. The papal enfeoffment of Robert with Sicily may have been legally contestable—it was naturally not recognized by Constantinople—but it forged a link between the papacy and Sicily which long outlasted Norman rule.

It was Robert's brother, Count Roger I (*q.v.*), who as the former's vassal and with his assistance became the real conqueror of Sicily. Internal conflicts among the Arabs served him well. The amir of Syracuse and Catania, at war with his brother-in-law, the amir of Girgenti (Agrigento) and Castrogiovanni (as Enna came to be called), went so far as to offer Roger his help to conquer the island. The first landing (1060), near Messina, was followed by an equally inconclusive attack on that town in 1061; the third, made in greater strength and with the participation of Robert, succeeded (summer, 1061). The possession of Messina gave the Normans control of the straits and a military base for further advance; the capture of Palermo (1072) concluded the first phase of the conquest; and the capture of Noto (1091) completed it. Although Robert's assistance was at first all-important, Roger soon assumed the leading role. After the conquest Robert probably retained only Palermo—apart from the suzerainty over the whole island, which suzerainty became entirely nominal under Roger Borsa, Robert's weak successor as duke of Apulia (1085–1111), who in 1091 surrendered half of Palermo to his uncle.

If Roger was the real conqueror of Sicily, he was also the founder of the Norman Sicilian state. By distributing fiefs sparingly, he gave feudalism a less important place in Sicily than it had acquired on the mainland, where Robert Guiscard had established his ducal power by making the Normans accept him as their ruler. At the same time he accepted much of the existing law and institutions. In his treatment of the Arab majority, his policy of religious toleration resembles that previously practised by the Arabs. Their legal conditions varied considerably, ranging from the liberties that they enjoyed at Palermo, where they had their own quarter and mosques, to the serfdom of the mass of the country population. Roger made use of their military and administrative services; many leading Arabs, however, seem to have left the country after the conquest.

Roger showed much favour to the Greeks, as appears from his lavish patronage of Basilian monasticism. He founded or restored at least 14 Greek monasteries as against 4 Benedictine ones. The Norman conquest was followed by an increase of the Greek population, Basilian monks from Calabria forming only one element of the new immigration. The Sicilian Church, however, became Latin, according to the promise made by Robert Guiscard to Pope Nicholas II in 1059. But the papacy was left only little influence in it; the concession of the apostolic legation to Roger (1098) made the Sicilian Church practically independent of Rome by sanctioning an already existing state of monarchical control.

7. Roger II and the Foundation of the Kingdom.—The survival of Roger I's work during the difficult period after his death (1101) was a measure of his success. Roger's son and heir Simon died in 1105 and was succeeded by his brother Roger II (*q.v.*); but their mother Adelaide ruled as regent from 1101 until Roger attained his majority in 1112. Roger continued his father's efforts to take advantage of the difficulties of the duke of Apulia (Roger Borsa had been succeeded by his son William in 1111), not only eliminating the last relic of ducal authority on the island by obtaining the second half of Palermo in 1122 but also extending his lands and influence in the duchy of Apulia and in Calabria. After Duke William's death in 1127, Roger II crossed over to the mainland to assert his claims as his heir.

Roger's expedition opens a period of struggles which lasted until 1139. His claims were opposed not only by many lords and towns but also, until 1128, by Pope Honorius II. By supporting the antipope Anacletus II against Innocent II, Roger obtained the royal title for Sicily, Apulia and Calabria (1130). But this led to an alliance of Innocent II and the emperor Lothair against the ruler who by his support of the antipope was primarily responsible for the prolongation of the schism. Lothair's invasion of the kingdom in 1136 in alliance with Pisa and Venice and, perhaps, Byzantium brought Roger's fortunes to their lowest point. But after Lothair's withdrawal, Roger soon recovered lost ground; and after the death of Anacletus (1138), Innocent II confirmed him in the royal title. The treaty of Mignano (1139) meant the final acknowledgment by the papacy of the Norman kingdom, which included the mainland provinces of Apulia, Calabria and Capua as well as Sicily.

The political problems of the kingdom were, to no small extent, a legacy of the different territories of which it was formed. Thus the conflicts with the German and Byzantine emperors belong largely to the history of the South Italian provinces; while Roger's and his successors' African policy is rooted in the earlier history of Sicily. The first expeditions against the Zirid prince of Mahdia in North Africa (1118, 1123) were failures; but in 1134–35 internal discords in the Zirid state provided Roger with fresh opportunities for intervention, which was facilitated by the growth of the Sicilian navy; the expedition led to the occupation of the island of Djerba. The capture of Tripoli in 1146 initiated a series of conquests which culminated in that of Mahdia in 1148; by that year, Roger's African empire extended from Tripoli to Tunis, from the desert of Barca to Kairouan. That it was short-lived was primarily due to the failure to check the growing power of the Almohads; and the death of the grand admiral George of Antioch, the conqueror of Africa, in 1151–52 and that of Roger II in 1154 jeopardized its survival. The crisis was to begin in 1156 with an Arab rising in Africa; by 1160, the African empire was lost. The Norman kings, however, did not abandon their African ambitions: William II was to send a fleet to lay siege to Alexandria in 1174, and further expeditions were to be undertaken against the North African coast in the following years.

Expansion in Africa had sharpened the antagonism between Norman Sicily and Constantinople. At first, when the Byzantines had received support from the Western emperors, the alliance of the two empires constituted a formidable threat to the Sicilian kingdom; but the clash between Byzantine and Western imperial claims led, after Frederick I Barbarossa's Roman coronation (1155), to the end of the alliance, Frederick continuing an anti-Norman policy of his own. Roger took the initiative against the Byzantines in 1147 and 1149, seized Corfu and invaded Greece. The Byzantines, on the other hand, reconquered part of Apulia in

1155, after Roger's death; but they lost the conquered territories again after their crushing defeat at Brindisi (1156). Peace was concluded in 1158 and was not broken until 1185, when King William II invaded the Byzantine empire, took Durazzo (Durres) and Thessalonica (modern Salonika) and advanced toward Constantinople, with the ultimate aim of seizing the Eastern imperial crown. But the counteroffensive under the new emperor Isaac II Angelus put an end to this expedition. It was the last great enterprise of the Norman kings.

8. Internal Development of Norman Sicily.—Roger II's internal policy was based on that of his father and was continued by his son William I (*q.v.;* 1154–66) and by the latter's son William II (*q.v.;* 1166–89); but his own contribution to the building of the Sicilian state was very great. The combination of Norman, Arabic and Byzantine elements is perhaps the most striking but at the same time a very natural characteristic of the government and civilization of the kingdom. Roger's Assizes of Ariano (1140) derive largely from Justinian and later Byzantine law; very strong Byzantine influence can be found in the judicial and fiscal administration and in the central financial department, the *duana* (diwan), the personnel of which was originally Arabic. The Norman chancery issued documents in Latin, Greek and Arabic. On the other hand, the curia was primarily modeled on that of the northern European states; and not only were feudal institutions accepted but feudal barons also played an important part in the provincial and local administration. At the same time, some of the most influential ministers of the Norman period, the grand admirals such as George of Antioch were Greeks. Roger's capital was cosmopolitan Palermo.

But despite the role played by Greeks and Arabs in the bureaucracy, Sicily became progressively latinized. While the Muslim population was decreased by conversions, the Latin element was strengthened by the settlement of colonies of "Lombards" (*i.e.,* mainlanders), Greeks and Arabs being gradually reduced to small minorities. This process is also reflected in the large number of Latin monasteries founded by the kings, after the initial favour shown to the Basilian monks. But Sicilian civilization retained the composite character that it had possessed from the beginning; Sicilian scholars translated Greek classical texts into Latin; Idrisi, on the orders of Roger II, composed one of the outstanding works of Arab geography; and Sicilian architecture was the product of Roman, Byzantine, Arabic and Norman influences.

9. The Hohenstaufen Accession.—William II's death (1189) was followed by a struggle for the succession. William I and William II had long supported Pope Alexander III in his conflict with Frederick Barbarossa; but a truce had been concluded in 1177 and the subsequent rapprochement between king and emperor had resulted in William II's sanctioning the betrothal of Frederick's son Henry to Constance, daughter of Roger II and heiress apparent to the kingdom (1184; marriage 1186). This created a situation of great potential danger for the papacy, as it strengthened imperial claims on the kingdom of Sicily by the addition to them of Constance's rights. On William's death, however, there was strong Norman opposition to the prospect of German rule, and Tancred (*q.v.*), an illegitimate grandson of Roger II, was crowned king. Frederick's son, who had succeeded his father as Henry VI, failed to make good his claims in 1191, when he was forced to raise the siege of Naples; but the sudden death of Tancred early in 1194 proved a turning point. The reign of his young son, William III, under the regency of Queen Sibylla lasted only ten months; Henry VI finally occupied the kingdom and was crowned king at Palermo in 1194. A new period in the history of Sicily had begun.

Henry aimed at including Sicily permanently in the empire, which he unsuccessfully tried to make hereditary in his family, the house of Hohenstaufen; and the distribution of high offices and lands among his followers was to strengthen his control of the kingdom. His death in 1197 temporarily put an end to imperial domination. Constance, as ruler with her young son, the future emperor Frederick II (*q.v.*), returned to a strictly Norman policy. But she died in 1198, leaving Frederick under the guardianship of Pope Innocent III. The following years were marked by growing anarchy, due to the weakness of the monarchy, to the attempts of German lords to seize hold of the kingdom, to the Arabs' endeavours to improve their position and to the commercial expansion of Pisa and Genoa. After Frederick had reached his majority (1208), Innocent handed the government over to him.

10. The Emperor Frederick II (Frederick I of Sicily).—In 1211 Pope Innocent saw himself forced to support Frederick's renewed election as king of the Romans or as German king. Frederick promised Innocent, just before the latter's death in 1216, to renounce the Sicilian kingdom in favour of his son Henry after his imperial coronation; but the papacy's hope to prevent in this way the reunion of Sicily and the empire was not fulfilled, for in 1220 Frederick succeeded in having Henry elected German king. The eventual union of the German and Sicilian crowns, foreshadowed by this election, made papal insistence on Frederick's renunciation obsolete; and when Frederick was crowned emperor in the same year, Honorius III tacitly recognized him also as Sicilian king.

Frederick devoted the following years to the restoration of royal power in the kingdom; and in this formidable task he revealed himself the true successor of Roger II. In Sicily Frederick acted as Norman rather than German ruler: symptomatically, his edict on the resignation of privileges (1220) put the deadline at the death of William II, not of Henry VI. His preoccupation with the affairs of the kingdom was largely responsible for the postponements of his crusade, which led in 1227 to his excommunication. Papal troops invaded the kingdom during his absence in Palestine (1228–29); but Frederick's return meant his immediate victory, and the peace of San Germano (1230) was followed by further internal reforms. The Constitutions of Melfi (1231), a legal code inspired by Roger II's Assizes, give a remarkable insight into the organization of the kingdom and also into the political ideas of its ruler. Frederick carried the evolution of the Sicilian administration considerably beyond what had been achieved by the Normans; on the other hand, the use that he made of assemblies of estates, including municipal representatives (from 1232), was a landmark in the history of the Sicilian parliament. In his highly centralized government Frederick was the heir to his Norman predecessors, whose work he continued. The same continuity can also be seen in Sicilian culture, but at Frederick's brilliant cosmopolitan court Italian poetry provided a new formative element.

11. The End of Hohenstaufen Rule.—Frederick's work was jeopardized and his dynasty ruined by the union of Sicily with the empire. The conflicts with the papacy and with its allies, the Lombard communes, meant a serious strain on the Sicilian economy and started a chain of events which was finally to lead to the establishment of the house of Anjou in the kingdom. Innocent IV, having declared Frederick deposed from the imperial throne (1245), wanted to deprive him and his descendants of the Sicilian crown as well. Frederick's death in 1250 and that of his son Conrad IV in 1254 provided the papacy with new opportunities, but the success of Manfred (*q.v.*), one of Frederick's illegitimate sons, in establishing control over the kingdom even before Conrad's death complicated the situation and gave rise to lengthy and tortuous negotiations in which papal offers of the Sicilian crown to foreign princes alternated with rapprochements to Conrad and to Manfred. Manfred was crowned king at Palermo in 1258—a usurpation of the rights of Conradin, Conrad IV's young son and heir. In 1263 after negotiations with Manfred had broken down for the last time Pope Urban IV announced the choice of Charles of Anjou (see CHARLES I of Naples), brother of Saint Louis, as king of Sicily (Edmund Lancaster, son of Henry III of England, who had been invested by the papacy with the kingdom in 1255, had failed to substantiate his claims). In 1265 Charles was invested with the kingdom by Pope Clement IV in Rome; near Benevento, in 1266, he defeated Manfred, who was killed.

12. Charles of Anjou.—After Benevento there was no serious resistance to Charles in the kingdom. But his rule had still to stand its test when Conradin came to Italy in 1267 to seize his inheritance. A rising in his favour swept the kingdom and showed the fragility of Charles's position. In 1268, however, Conradin was defeated at Tagliacozzo and executed at Naples. There followed severe suppression of the revolt, especially in Sicily, where

it had been more widespread than on the mainland. Charles tried to put his power on a firm foundation by a large-scale distribution of fiefs among his French nobility. At the same time he preserved, in its main outlines, Frederick's system of government. The "French colonization," however, and the sense of grievance that it caused among the Sicilians was partly responsible for the great revolt of 1282, known as the Sicilian Vespers (*q.v.*), which severed Sicily once more from the mainland. (N. R.)

13. The Aragonese.—The revolt of the Vespers precipitated the arrival in Sicily of Peter III of Aragon, who had a claim to the crown through his marriage (1262) to the Hohenstaufen heiress Constance, Manfred's daughter. The Sicilians, however, were jealous of their independence: it had been only after the papacy had rebuffed their appeal for recognition of a communal regime of their own that they had addressed themselves to Peter. While Peter was accepted as king (Peter I), Constance governed the island for him till his death in 1285; but then, whereas their eldest son succeeded to Aragon as Alfonso III, the Sicilian crown was transmitted to their second son, who became James I of Sicily. The termination of the personal union between Aragon and Sicily was gratifying to Sicilian separatist feeling, but its effect was prejudiced when Alfonso, having already deserted the Sicilian side in the continuing War of the Sicilian Vespers, died without issue in 1291. In his will he had stipulated that if his brother James were to succeed him in Aragon as James II, he should renounce Sicily to their youngest brother Frederick, in repetition of the previous arrangement; but James, heir to Aragon in his own right, would not at first be bound by this condition and simply nominated Frederick as his lieutenant in Sicily. Subsequently, however, he found that from Aragon's point of view it was indeed expedient to renounce the Sicilian connection; and by the treaty of Anagni (1295) he agreed to surrender Sicily to the papacy.

The Sicilians reacted against this new betrayal by taking Frederick as their king. Though in fact he was Frederick II of Sicily, he took the style of Frederick III (*q.v.*).

James II of Aragon allied himself with Charles II of Naples to execute the treaty of Anagni against the Sicilians, but the allies' efforts came to nothing; and the treaty of Caltabellotta (1302) brought the 20 years' War of the Vespers to an end: Frederick was to have the kingdom for his lifetime, after which it should revert to the Neapolitan Angevins. Pope Boniface VIII, whose predecessors had steadfastly opposed the Aragonese over Sicily, could only insist that Frederick should style himself "king of Trinacria" (*i.e.*, "of the Three-cornered Isle") instead of "king of the Island of Sicily."

The Angevin-Aragonese peace did not last long. Having taken the emperor Henry VII's side against Robert of Naples and the Guelphs, Frederick resumed the title "king of Sicily" and, in 1314, induced the Sicilian parliament to declare that on his death the crown should pass to his son Peter. Successive Neapolitan attacks on the island failed to reduce him, and on his death his son succeeded him as Peter II (1337).

Frederick's position had obliged him to rely on the support of the Sicilian parliament. He organized it as an assembly of three *bracci*, or houses, representing the feudatories, the clergy and the towns of the royal domain, on the model of the three estates of the kingdom of Aragon; and he associated it with himself in the exercise of sovereignty. Himself a strong ruler, he was able to keep his barons in order. Under his successors, however, the barons began to assert themselves both in encroachment on the royal authority and in internal warfare of their own; *e.g.*, between the "Latin" faction (the older Sicilian nobility) and the "Catalan" (the Aragonese newcomers) during the reign of King Louis, who was only four years old when he succeeded his father Peter II in 1342.

The next king, Frederick III (properly so numbered), who succeeded his brother Louis in 1355, managed to withstand renewed attacks by the Neapolitan Angevins (who occupied Messina for a short time in 1356) and finally came to terms with Joan I of Naples in 1373: she agreed that Sicily, officially called Trinacria again, should be a separate kingdom in vassalage both to the Holy See and to her own kingdom of Naples-Sicily.

Frederick III died in 1377, leaving a daughter, Mary, as his heiress. There ensued a long period of disorder. Peter IV of Aragon, on the grounds that the testament of Frederick III (II) precluded female succession to the Sicilian crown, claimed it for himself as the nearest male heir (he was also the father of Mary's mother and the husband of her aunt); and Mary underwent a series of abductions. Peter, however, in the face of objections from the papacy and the Angevins, in 1380 ceded his pretensions to his second son, Martin, duque de Montblanch, whose son Martin was to marry Mary. Peter IV died in 1387, leaving Aragon to his elder son John I; the queen of Sicily was brought to Spain in 1388; and her marriage to the younger Martin took place in 1390. In 1392 the couple landed in Sicily with Martin of Montblanch and began to reign as queen and king-consort, despite strong local opposition. Mary died in 1401, leaving her widower to reign alone as Martin I of Sicily; but meanwhile the duque de Montblanch had become king of Aragon as Martin I in 1395 through the death of John I. When Martin I of Sicily died without legitimate issue in 1409, he left his kingdom, with his second wife Blanche of Navarre as regent, to his father, who thus became Martin II of Sicily.

Martin II, who had no surviving children of his own, intended that Sicily at least, if not Aragon too, should go to his grandson Fadrique (Frederick) de Luna, a bastard of Martin I of Sicily. On Martin II's death, however, in 1410, this succession was contested; and Ferdinand of Antequera, son of Peter IV's daughter Leonor, having been chosen king of Aragon as Ferdinand I in 1412, defeated Fadrique's partisans and re-established Blanche's authority as his regent in Sicily. Thenceforward the Aragonese (later the Spanish) and the Sicilian crowns were to remain united for nearly 300 years (till the War of the Spanish Succession).

Alfonso V of Aragon (I of Sicily), having succeeded his father Ferdinand I in 1416, used Sicily as his base for his expeditions to Naples during Joan II's precarious tenure of that kingdom; and the Aragonese conquest of Naples after Joan's death temporarily restored the personal union of the Neapolitan and Sicilian crowns. Alfonso, however, kept the two kingdoms theoretically separate, arranging as early as 1443 that his bastard son Ferdinand (Don Ferrante) was to inherit Naples, whereas Sicily passed at his death in 1458 to his brother John II of Aragon (John I of Sicily). John's son Ferdinand II (*q.v.*) finally recovered Naples for the legitimate line of Aragon (1503).

The dynastic troubles of the period 1337–1412 had further helped the Sicilian feudatories and the towns to extort privileges from the kings as the price of support. The royal lieutenants or viceroys who ruled for the Aragonese kings from 1415 had extensive powers, but still had always to reckon with the Sicilian parliament. By the middle of the 15th century it had become an accepted principle of the constitution of the kingdom that no new taxes should be levied without the parliament's consent; and this power of veto over subsidies to the king made the *bracci* a formidable obstacle to royal or viceregal absolutism, as well as to any attempt to reduce feudal, ecclesiastical or municipal privilege.

14. The Spanish Habsburgs.—Ferdinand II, dying in 1516, left Sicily, with Naples and all his Spanish inheritance, to his grandson, the Austrian Habsburg Charles I of Spain, who three years later became Holy Roman emperor as Charles V. The reign began with a rising (1516) of the privileged orders against the viceroy, Hugo de Moncada, whom Charles recalled without canceling the edicts that had provoked the trouble. A rising against the new viceroy, Ettore Pignatelli, duque de Monteleone, developed however into widespread outrages against the nobility, which reacted by co-operating with the viceroy. Thereafter the privileged orders in general professed themselves loyal to the distant monarchy, which stood as the guarantor of their privileges, while the viceroys had to bear the burden and the odium of actual government, raising the necessary taxes and trying to keep order in an island still prone to baronial vendetta and to jealous rivalries between town and town. The arrival of the Ottoman Turks in the western Mediterranean exposed Sicily to danger from a new quarter, and it became a base for the emperor's counterattacks on the North African coast.

In 1555 Charles V abdicated Naples to his son, the future Philip

II of Spain. Pope Paul IV then promoted a scheme whereby the French should conquer Naples for a prince of their own house and seize Sicily for the Venetians; but Philip, to whom Charles V had abdicated Spain and Sicily in 1556, forced the pope to recognize him in 1557 and secured peace from the French in 1559. The battle of Lepanto (1571) checked the gravest menace from Turkey.

Contributions from Sicily were required for Philip II's enterprises in western Europe and for the defense of Spanish interests in the west and in Italy during Philip III's reign (1598–1621). With the growth of the French challenge to Spain during Philip IV's reign (1621–65; *see* THIRTY YEARS' WAR) the viceroys had constantly to raise more troops and more money, and the measures to which they turned were resented as oppressive or contrary to the traditional privileges of Sicily. In May 1647, when the viceroy Pedro Fajardo, marqués de Los Vélez, had put a tax on grain and at the same time tried to keep the old price of loaves by reducing their weight, the people of Palermo rose in revolt and forced him to repeal the tax. Then, in July, the Palermitans, led by Giuseppe Alessio, rose again to demand the repeal of all taxes imposed since Charles V's death and the reservation of the viceroyalty and other public offices to Sicilians: the viceroy withdrew from Palermo and, fearing lest the Sicilian barons and the other towns might join the rebellion, began negotiations with Alessio. The latter, however, lost control over his riotous followers and was assassinated; Messina, always the envious rival of Palermo, remained loyal; the French took no advantage of the situation; and in September the viceroy and his Spaniards returned to the capital, where he died a few weeks later. His successor, faced with another revolt, had to concede a general amnesty. The events of the early summer had largely inspired the parallel revolt of Masaniello (*q.v.*) in Naples.

The revolt of Messina in 1674, during the reign of Charles II of Spain, was no less characteristic of Sicily under foreign rule than that of Palermo and was more serious because the French, at war with Spain again (*see* DUTCH WARS), were able to exploit it. The governor of Messina, Luis del Hoyo, wishing to break the stranglehold of the municipal oligarchy over the senate of Messina, tried to exploit popular discontent during a food shortage in order to introduce a popular element into the senate. The oligarchy stirred up its own popular reaction; Del Hoyo was replaced by Diego de Soria as governor; but conflict between the senate's and the governor's factions went on till, in Aug. 1674, Soria was besieged in his palace and forced to capitulate. The Messinese then appealed to Louis XIV of France, a French fleet arrived in September with supplies of food, and in April 1675 the duc de Vivonne (Louis Victor de Rochechouart) was sworn in as Louis XIV's viceroy of Sicily. Though Palermo declared itself for Spain against its traditional rival, the whole island might have fallen to the French if the Dutch had not sent a fleet, under the great M. A. de Ruyter, to support their Spanish ally. The naval war round the island was indecisive and Messina itself held out against the Spaniards till the French, making the peace of Nijmegen (1678), deserted the Sicilian cause. A viceregal amnesty to the rebels was revoked, and Messina's senate and privileges were abolished.

15. The Savoyard and Austrian Habsburg Interval.—Charles II's death (1700) was followed by the War of the Spanish Succession (*see* SPANISH SUCCESSION, WAR OF THE), during which Sicily's eventual fate was continually in discussion between the belligerent powers. The Austrians overran the kingdom of Naples (1707); and the Bourbon Philip V of Spain, under pressure from the English, ceded Sicily to Victor Amadeus II of Savoy at the peace of Utrecht (1713). The Franco-Austrian peace of Rastatt (1714) left the Austrians in possession of Naples and also gave Sardinia to them, but Spain was not a party to this peace.

The Spaniards, having seized Sardinia in 1717, invaded Sicily in July 1718. Victor Amadeus was unable to defend his kingdom without Austrian help. Austria, however, wanted to exchange Sardinia with him for Sicily; and this exchange had been included in the plan on which the Quadruple alliance of Great Britain, France, the United Provinces of the Netherlands and Austria (Aug. 1718) was formed for the settlement of the Austro-Spanish dispute.

The British admiral George Byng (later Viscount Torrington, destroyed the Spanish fleet off Cape Passero; the Austrians crossed the straits into Sicily; Victor Amadeus agreed to the plan for the exchange; and declarations of war against Spain were issued from Great Britain in Dec. 1718 and from France in Jan. 1719. By the treaty of The Hague (Feb. 1720) Spain accepted the terms of the alliance, and Sicily passed to Austria.

16. The Bourbons.—Extremely disliked in Sicily, the Austrian rule there came to an end during the War of the Polish Succession (*q.v.*), when the Spanish Bourbon infante Don Carlos, having first conquered Naples and the mainland for himself, was crowned king at Palermo in July 1735. Having initiated his great program of enlightened reform for his two Italian kingdoms (*see* NAPLES, KINGDOM OF), he abdicated Naples and Sicily to his third son, Ferdinand (III of Sicily), when he himself became king of Spain as Charles III in 1759. The antifeudal policy of the viceroy Domenico Caracciolo in the 1780s was resented by the Sicilian baronage, but Ferdinand was able to take refuge in Sicily when he was expelled from Naples by the French Revolutionary army in 1799. On his second expulsion from Naples, during the Napoleonic period, Ferdinand established himself in Sicily again (1806), to remain there, under British protection, for a far longer time. The British occupation, with Lord William Bentinck as ambassador in practical control of affairs from 1811, was chiefly important to Sicily for the constitution of 1812: the old parliament, ever jealous of its right to bargain with the crown, had been obstructing the war effort and so was superseded by a bicameral organ on British lines. Though this "English" constitution went some way to satisfying Sicilian particularism, it offended not only the king but also the more reactionary elements in Sicily insofar as it curtailed their privileges, notwithstanding its aristocratic bias.

The restoration of the Bourbons to Naples (1815) put an end to the constitution; and in 1816 the unitary kingdom of the Two Sicilies was proclaimed, subjecting Sicily to centralized government from Naples, with the king now styled Ferdinand I (*q.v.*). The Sicilians hated this; and, in 1820, Palermo rose in revolt for their constitution of 1812. The contemporary Neapolitan revolution, on the other hand, was for the democratic Spanish constitution of 1812, and troops from Naples suppressed the Sicilian autonomist movement before Naples in turn was put down for Ferdinand by the Austrians.

Ferdinand I was succeeded by Francis I (1825–30), the latter by Ferdinand II (*q.v.*; 1830–59). When revolution broke out in Sicily in Jan. 1848, the latter promised a constitution; but the Sicilians, acting again in disunion with the mainland, soon declared the deposition of the Bourbon dynasty and offered the crown to a prince of the house of Savoy. After the defeat of the Neapolitan revolution in 1849, Carlo Filangieri (*q.v.*) reduced Sicily to obedience again in April–May.

Francis II was the last Bourbon king of Sicily. Within a year of his accession, Giuseppe Garibaldi (*q.v.*) landed at Marsala with his "Thousand" on May 11, 1860 (*see* also ITALIAN INDEPENDENCE WARS OF; ITALY: *History*). A provisional government was formed at Palermo in June, a constitution was proclaimed in August; but after Garibaldi's conquest of Naples and his junction with the Sardinian-Piedmontese army, a Sicilian plebiscite in October decided by an overwhelming majority under universal suffrage for annexation to Sardinia. Messina, which held out for Francis, fell to the Piedmontese in March 1861. Sicily thus became part of the new kingdom of Italy.

17. The Italian Regione.—Despite the plebiscite, the house of Savoy found Sicily no more docile than earlier foreign dynasties had found it. Garibaldi's abortive *pronunciamiento* of 1862 was launched from Sicily and supported by the Sicilians, whom the royal government had to reduce by force. Brigandage and conspiracy continued into the 20th century, feeding on separatist tradition, on Catholic resentment at the government's treatment of the papacy before the Lateran treaty and on general discontent at the north's neglect of the backward south. Symptomatic of this state of affairs was the persistence of the Mafia (*q.v.*).

After World War II, during which Sicily had become a battleground (*see* WORLD WAR II: *The Sicilian Campaign*, May–August

1943), the Italian government gave more sympathetic attention to the island's problem. For the royal statute of autonomy (1946) and the republic's constitutional statute (1948) *see* below, *Regional Government*. (X.)

IV. POPULATION AND ADMINISTRATION

1. Population.—The population at the 1961 census was 4,631,382. The average density was 466.6 per sq.mi. compared with the national density of 429 per sq.mi. The chief characteristic of the distribution of population is the universal tendency to live in close-packed towns and villages, both in the coastal towns and in the agricultural towns of the interior which contain from 5,000 to 10,000 inhabitants or more. The birth rate rose from 21.5 in 1958 to 22.1 in 1961, while the death rate also rose from 8.6 in 1958 to 8.8 in 1961 so that the natural increase was 13.3 per 1,000. The rate of infant mortality dropped steadily from 55.4 per 1,000 live births in 1957 to 47.0 in 1961, but this was still high compared with northern Italy. Moreover the average figure for the whole island ignores the much higher figures reached in the mining province of Caltanissetta or in some communes of Agrigento province. Sicily's most serious problem is overpopulation. Before 1914 about 80,000 Sicilians emigrated overseas every year. Since 1951 there has been some emigration abroad, but the main movement of population has been to north Italy and within Sicily from the hill villages to the coastal towns. Industrial development associated with the oil production of the Ragusa-Gela district is beginning to alleviate Sicily's unemployment problem (see *The Economy*, below).

2. Regional Government.—Sicily is one of the five regions of Italy which obtained a special form of local autonomy under the constitution of 1948 (*see* ITALY). Sicilian regional status was conferred by an Italian constitutional law (1948). The regional capital is Palermo and there are nine provinces (Agrigento, Caltanissetta, Catania, Enna, Messina, Palermo, Ragusa, Syracuse [Siracusa] and Trapani). The background of regional government in Sicily is totally different from that of the other regions enjoying local autonomy. The history of Sicily, occupied by many invaders and for long periods separated politically from the mainland, had produced a people differing greatly in character and habits from the inhabitants of the mainland. Sicily accepted incorporation in Italy in 1860, but the political institutions of Piedmont were not well suited to it, and the centralized control imposed by the government in Rome, afraid of any threat to Italy's recently won unity, checked the possibility of local autonomy. There was no political separatist movement before 1940, but there was a general feeling of discontent and widespread belief that the interests of Sicily were sacrificed to those of the industrial north.

In 1943 it gradually became clear that very few Sicilians wanted separation from Italy but that a large number wanted a considerable measure of local autonomy. The Italian government installed a high commissioner with plenary executive powers in Jan. 1944; and a royal decree law of May 15, 1946, granted Sicily a statute of autonomy which was converted unchanged into the constitutional statute of Feb. 26, 1948.

The statute gave Sicily a very considerable degree of self-government. The regional assembly (*consiglio regionale*; the local parliament) can be dissolved only for persistent violation of the statute and then only after a debate in the Italian parliament. The assembly's legislative powers are more extensive than those of any other region. Elementary education is in its hands, and legislation on higher education (including university education) is limited only by the need to conform to the general principles of national education. Without prejudice to the national parliament's plans for agrarian and industrial reform the regional assembly has exclusive power to legislate on all questions concerning agriculture, land reclamation, industry and commerce.

The president of the region is elected by the assembly from among its members. He not only represents the region, promulgates regional laws and, together with the *giunta* (the executive committee of the assembly), constitutes the government, but also has the right to attend meetings of the Italian council of ministers with the rank of minister and to speak when Sicilian questions are under discussion (though he has not the right to vote). The president and *giunta* are invested by the statute with the supremely important executive function of maintaining public order by means of the national police force, although at their request the national government can act jointly with them. The president also represents the national government in Sicily.

The financial provisions of the statute are peculiar to Sicily. The island is expected to provide for its own expenditure by raising its own taxes, but the Italian state has agreed to hand over to the region a share of the taxation raised from firms which have factories or offices in Sicily though their central offices are not in the island. The state has also to pay over a lump sum annually for public works in Sicily.

3. Social Conditions.—Sicily is the richest of the southern regions in natural resources, but the standard of living of the majority of Sicilians is as low as that of south Italy. In 1960, 39.4% of the working population was engaged in agriculture compared with Italy's average of 30.8%. The contrast between the highly productive areas round Palermo and Catania, on the lower slopes of Etna and along parts of the coast and those in the mountainous interior still given over to extensive arable farming remains great. Life in many of the agricultural towns of the interior is primitive. Many such towns are without adequate sanitation, some still without a water supply, and serious overcrowding is common. Disease is consequently rampant. There is serious agricultural unemployment and above all underemployment. In the southeast corner of Sicily there has been a great change since 1957. Industry is growing rapidly associated with the development of oil. Since 1960 the number of industrial workers has increased steadily.

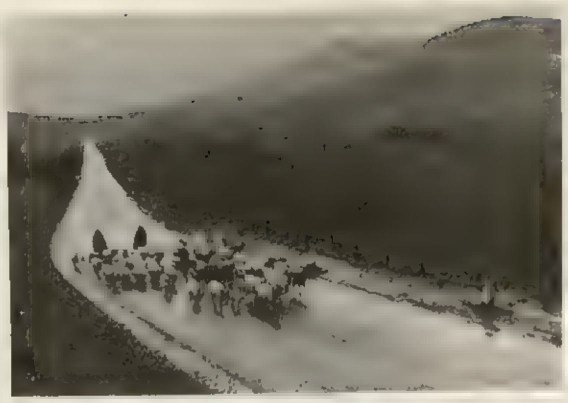
Education follows the general lines of Italian national education, although the regional government has complete control of primary education. The rate of illiteracy is high, double that of Italy, and there is a great shortage of schools. There are three universities, at Palermo, Messina and Catania, with a total of about 23,000 students. Provision is slowly beginning to be made for technical education.

Social services as a whole are organized like those on the mainland but the standard is poor. Housing conditions in the towns as well as the villages are not good and there is a great shortage of medical services. The hand of the Mafia (*q.v.*) undoubtedly lies heavily on the whole social life of Sicily and has put great obstacles in the way of social improvement. In 1962 the Italian government appointed a parliamentary commission to enquire into the causes of the Mafia and to propose measures necessary for its repression. (M. M. C.)

V. THE ECONOMY

1. Agriculture.—In 1936 more than half the employed population of Sicily was engaged in agriculture. World War II did not materially alter the economic structure of the island and in 1951 the proportion was 51%. The outstanding feature of Sicilian land use is the very high proportion of the total area given over to agriculture and forest, more than 95%.

Sicilian agriculture is dominated by the lack of water, which limits production and restricts the use of modern technical methods. Both intensive and extensive methods of cultivation exist. On Aug. 10, 1950, a law was passed by the Italian government to initiate a fifteen-year economic development plan (1950–65) for the southern regions including Sicily. Under the direction of the *Cassa per il Mezzogiorno* public works such as land reclamation, flood and erosion control projects, construction of aqueducts and sewers, provincial roads and railways were being undertaken. Among the many completed schemes are the irrigation of the Gela district (including the Disueri dam) and of the Catania and Lentini plains, and the reafforestation, flood and erosion control projects of the upper Simeto and Salso basins, where the Pozzillo dam has been constructed to irrigate the Catania plain. Areas where irrigation has long been part of the agricultural system are on the slopes of Etna and of the Iblei mountains, around Marsala and in the Conca d'oro behind Palermo. Citrus fruits and vegetables grow in abundance and provide the main source of the island's


(ABOVE AND TOP RIGHT) BRUCE DAVIDSON—MAGNUM; (RIGHT) SERGIO LARRAIN—MAGNUM

(ABOVE) DISPLAY OF THE DAY'S CATCH AT MONDELLO, A FISHING VILLAGE ON THE TYRRHENIAN SEA; (TOP RIGHT) GOATHERDS AND FLOCK IN THE EREI MOUNTAINS; (RIGHT) RELIGIOUS PROCESSION IN A SOUTHERN SICILIAN VILLAGE

agricultural exports. In these fertile areas small holdings are important and the majority of the farms are less than seven acres. Extensive farming prevails all over the interior where the practically treeless land produces the wheat for which Sicily has been famous for more than 2,000 years.

The characteristic estates of the interior were the *latifondi*, large, extensively cultivated estates frequently owned by absentee landlords and run on capitalist lines by middlemen who employed a few workers engaged by the year (*salariati*) and large numbers of badly paid day labourers. Before 1939 farms of more than 125 ac., which constituted less than 1% of the total number of farms in Sicily, occupied 36% of the land. Large areas were also sublet in fragmentary holdings to peasants who farmed the land extensively and thus created what is known as the *latifondo contadino* (the peasant's *latifondo*).

Subject to general conformity with the Italian parliament's plans for land reform, the regional government of Sicily has exclusive power to legislate on all agrarian questions. To change the system of extensive *latifondo* farming, however, was a difficult proposition. The shortage of water throughout the interior meant that any plan for abolishing the great estates and for establishing peasant proprietors would be meaningless unless it made provision for a regular and sufficient water supply, which alone would make it possible to grow other crops than wheat; and supplies of water and better local roads were also required if the peasants were to be persuaded to live in small villages near their fields. Yet the Sicilian regional government made a beginning. On Dec. 27, 1950, a law was issued on the expropriation of land applicable to the whole island: 69,500 ac. were listed for expropriation (this amount was expected to rise to 250,000 ac.). Intensively cultivated land was exempt. The work of expropriation and redevelopment was entrusted to the Sicilian Land Reform agency in Palermo under the general supervision of the Italian ministry of agriculture.

Sicily's most important crop is wheat and in 1961 nearly 645,000 ha. of the arable land were sown to wheat; production amounted to 700,640 metric tons, more than 8% of the total Italian wheat production. The yield, however, was low: Sicily produced only 10.9 quintals per hectare, as compared with the average of 19.6 for the whole of Italy and 35.0 for Lombardy. The second most important crop consists of citrus fruits, particularly lemons, of which Sicily produced 443,830 metric tons in 1961, about 90% of Italy's total lemon crop. The island also produces large quantities of fresh vegetables, particularly tomatoes, artichokes, peas and beans, and their early season makes them a profitable export crop. Much wine is produced, mostly for local consumption; Marsala, however, is a valuable export wine.

Stockraising.—The estimated number of cattle in Sicily rose from 168,000 in 1930 to 277,000 in 1961; although the latter figure compared favourably with those for southern Italy it was admittedly low. The supply of fodder is poor, there are few permanent meadows, and much of the pasture is suitable only for sheep and goats, which numbered fewer than 900,000. Donkeys and mules are used everywhere by peasants, and with their bells and packs are a characteristic feature of Sicilian life.

2. Fisheries.—In 1959, 24% of the Italian fishing fleet was registered in Sicily, and fishing was an important source of employment. The most valuable fish caught is tunny (tuna); some is eaten fresh in Sicily but most is canned for export. The total Sicilian catch amounted to a considerable proportion of the national catch and the tunny catch exceeded three-quarters.

3. Industry.—The small canning industry of fish, fruit and vegetables was greatly expanded after 1945. The sulfur deposits of the central-southern part of Sicily are mined where the sulfur content reaches 16%–20%. Output is declining because of competition from the U.S., high production costs and the need to modernize the mines. Oil discovered around Ragusa, Gela and Fontanarossa is rapidly developing into one of the largest European oil fields. In the early 1960s production amounted to about 2,000,000 metric tons. Associated industries have developed and industrial growth generally has been greatly stimulated by financial aid supplied by the *Cassa per il Mezzogiorno*. A plant for the fabrication of polythene and ethylic alcohol is in operation at Ragusa. Oil refineries are at Ragusa, Gela and Augusta. Chemical and pharmaceutical plants are at Catania and Gela, and large fertilizer plants at Porto Empedocle and Licata. The cement industry is growing in importance and new factories have been constructed at Palermo and Augusta.

4. Communications.—Sicily's three chief ports are Palermo, Messina and Catania. A regular shipping service connects Palermo with Naples, and Messina is linked to the mainland by the train-ferry across the straits. The three chief ports are connected by three main railway lines, all single-track: one along the north coast from Palermo to Messina; one along the east coast from Messina to Catania and Syracuse; and one running inland from Catania to Palermo via Enna. The remarkable narrow-gauge circum-Etna railway, which climbs to 3,196 ft., is not a mere tourist line but carries heavy loads of agricultural produce. There are other secondary lines, but these are subject to severe gradients and curves.

Main roads between important towns are good, but there is a dearth of good secondary roads, and some small hill towns or villages can be reached only by steep mule tracks. The main airports are near Palermo and Catania.

See also references under "Sicily" in the Index. (G. KH.)

BIBLIOGRAPHY.—*General: Legge constituzionale 26 Febraio 1948*, No. 2; Nunzio Prestianni, *L'Economia agraria della Sicilia* (1947);

Touring Club Italiano, *Guida d'Italia, Sicilia e isole minori;* L. Franchetti and Sidney Sonnino, *La Sicilia,* 2nd ed. (1925), the report of their inquiry in 1876. (X.)

Archaeology: Preclassical Archaeology: see the original descriptions in the various writings of Paolo Orsi; reports of subsequent investigations in official Italian publications; relevant articles in Max Ebert, *Reallexikon der Vorgeschichte,* 15 vol. (1924–32); fresh surveys with bibliographies in B. Pace, *Arte e civiltà della Sicilia antica,* 4 vol. (1935–49); L. Bernabò Brea, *Sicily Before the Greeks* (1957), *Musei e monumenti in Sicilia* (1958).

Classical Archaeology: see articles on individual towns in Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft* (1893 *et seq.*); G. E. Rizzo, *Monete greche della Sicilia* (1946); B. Pace, "Arti ed artisti della Sicilia antica," *Memorie della Reale Accademia dei Lincei,* 5th series, xv, 471–628 (1917); E. Douglas Van Buren, "Archaic Terra-cotta *Agalmata* in Italy and Sicily," *Journal of Hellenic Studies,* xli, 203–216 (1921), *Archaic Fictile Revetments in Sicily and Magna Graecia* (1924); B. Ashmole, *Late Archaic and Early Classical Greek Sculpture in Sicily and South Italy* (1936); L. Quarles van Ufford, *Les Terrescuites siciliennes* (1941); W. J. Anderson, rev. W. B. Dinsmoor, *The Architecture of Ancient Greece* (1950); Lucy T. Shoe, *Profiles of Western Greek Mouldings* (1952); G. V. Gentili, *La Villa romana di Piazza Armerina* (1951); and current reports in *Fasti archaeologici* (annually). (A. W. V. B.)

History: For a general introduction: G. Libertini and G. Paladino, *Storia della Sicilia* (1933); L. Natoli, *Storia di Sicilia* (1935); F. de Stefano, *Storia della Sicilia dal secolo XI al XIX* (1948). For ancient Greek Sicily: E. A. Freeman, *The History of Sicily,* 4 vol. (1891–94); J. Bérard, *La Colonisation grecque de l'Italie méridionale et de la Sicile* (1941) and *Bibliographie des principales cités grecques de l'Italie . . .* (1941), supplemented by T. J. Dunbabin in *Papers of the British School at Rome,* vol. xviii, pp. 104–116 (1950); T. J. Dunbabin, *The Western Greeks,* for history to 480 B.C. (1948); A. G. Woodhead, *The Greeks in the West* (1962). For the Roman period: V. M. Scramuzza, "Roman Sicily," in *An Economic Survey of Ancient Rome,* ed by Tenney Frank *et al.,* vol. iii (1937). For the Byzantines: P. Orsi, *Sicilia bizantina* (1942); F. Giunta, *Bizantini e bizantinismo nella Sicilia normanna* (1950). For the Muslim period: M. Amari, *Storia dei Musulmani in Sicilia,* 2nd ed. (1930–39). For the Normans: F. Chalandon, *Histoire de la domination normande en Italie et en Sicile* (1904). For the Hohenstaufen period: M. Schipa, *Sicilia ed Italia sotto Federico II* (1928); various authors, *Atti del convegno internazionale di studi federiciani* (Palermo, 1952). For the Angevin and early Aragonese period: S. Runciman, *The Sicilian Vespers* (1958), with bibliography. For later history: L. Genuardi (ed.), *Parlamento siciliano* (1924), being section 1 of the *Atti delle assemblee costituzionali italiane,* 1st series; monographs on the persons mentioned in the text above; publications of the Società Siciliana di Storia Patria; and works indicated in the bibliography to the article NAPLES, KINGDOM OF; also J. Rosselli, *Lord William Bentinck and the British Occupation of Sicily* (1956). (X.)

SICKERT, WALTER RICHARD (1860–1942), influential British painter, was born at Munich on May 31, 1860, the son of Oswald Adalbert Sickert, a Danish-born German draftsman who settled in England in 1868. After several years on the stage, Sickert went in 1881 to the Slade school of art in London; he retained, however, something of the make-up of a late Victorian actor-manager. He studied under Alphonse Legros but through his father knew the works of Thomas Couture and Gustave Courbet. In 1882 he became a pupil of Whistler and in 1883 met Degas in Paris; these artists much influenced his work and personality. His first pictures of London music-hall interiors appeared in 1886, and, at the New English Art club, founded in that year, Sickert exhibited until 1917.

Like Whistler, Sickert used tonal methods to gain his effects, and his colours up to about 1904 are sombre. To Degas he was indebted for the ability to establish a situation merely by the attitudes of the figures—an innovation in England, where narrative painting was then governed by literary principles. Sickert coupled this with a refreshing vein of satire.

Between 1885 and 1905 Sickert spent most summers at Dieppe and worked in Venice. Returning to London in 1905 he became the focus of a group of painters which included Lucien Pissarro, Spencer Gore, Harold Gilman and Augustus John. Through Gore's and Pissarro's influence, Sickert's work began to show the influence of Neoimpressionism, though he was unmoved by the more advanced French styles then beginning to be known in London. He was a founder of the Camden Town and London groups (1911 and 1913) and was made a royal academician in 1934, resigning in 1935.

Sickert painted at Brighton and Bath in the 1920s and 1930s and wrote occasional criticism, supporting the principles of drawing observed by Degas. Such principles, and Sickert's own style, have been vital to English art and teaching. He died at Bath Somerset, on Jan. 22, 1942. His grandfather, Johann Jürgen Sickert (1803–64), and his brother Bernhard Sickert (1862–1932) were also painters.

BIBLIOGRAPHY.—R. Emmons, *The Life and Opinions of Walter Richard Sickert* (1941); Lillian Browse (ed.) and R. H. Wilenski, *Sickert* (1944). *See* also W. R. Sickert, *A Free House!,* ed. by Sir Osbert Sitwell (1947). (D. C. T. T.)

SICKINGEN, FRANZ VON (1481–1523), an enigmatic figure of the early years of the Reformation in Germany, was born at Ebernburg near Worms on March 2, 1481. A member of the *Reichsritterschaft,* or class of free knights, he became a sort of German *condottiere* and acquired considerable wealth and estates in the Rhineland as the result of immensely profitable campaigns against private individuals and against cities such as Worms (1513) and Metz (1518). In 1518 he led the army of the Swabian league against Ulrich I of Württemberg. After the death of the Holy Roman emperor Maximilian I in 1519, Sickingen, while accepting bribes from Francis I of France, used his influence to support the election of Charles V as emperor. The presence of his troops around Frankfurt, where the election was held, certainly helped Charles. For this service he was made imperial chamberlain and councilor.

Sickingen supported both the nationalist schemes of Ulrich von Hutten and the religious reforms of Martin Luther. He protected Johann Reuchlin against the Dominicans in 1519 and helped guarantee Luther's safe conduct to Worms in 1521. He harboured many humanists and reformers in his castles, which became, in Hutten's words, "a refuge for righteousness." When the knights' war broke out in 1522, as a rising of his class in defense of its ancient status, Sickingen, who the previous year had led an abortive and inglorious expedition to France, placed himself at the head of it. He declared war against his old enemy Richard of Greiffenclau, archbishop of Trier and a friend of the French. He sadly underestimated the opposition. Trier remained loyal to the archbishop, and princes such as Philip of Hesse rallied to his support; Sickingen was repulsed, his support fell off, and he was declared an outlaw. He was forced on the defensive; the Swabian league declared against him; his castles fell one by one; and on May 6, 1523, he capitulated in his last stronghold at Landstuhl. He died the next day and was buried there. In 1889 a splendid monument was raised at Ebernburg to his memory and to that of Hutten.

Sickingen's life in the end was a tragic waste in spite of all his outstanding qualities. He championed a dying cause. The class of knights was being overtaken by the growing influence both of the towns and of the princes and was an anachronism in the new modern world that was evolving, the main occupation of its members being private warfare. Had they been properly organized and led, the knights might well have harnessed the wave of nationalist feeling in Germany, and particularly in the Rhineland, to productive ends. Sickingen, for all his power and influence, was no more able to do this than was Hutten. On the one hand a champion of the poorer classes, a sincere Lutheran and genuine patriot, he was on the other hand a self-interested opportunist who probably aimed at high office. In the end he accomplished little.

BIBLIOGRAPHY.—H. Ulmann, *Franz von Sickingen* (1872); H. Prutz, "Franz von Sickingen," in *Der neue Plutarch* (1880); the "Flersheimer Chronik" in Hutten's *Deutsche Schriften,* ed. by O. Waltz and S. Szamatolski (1891); K. Rendenbach, *Die Fehde Franz von Sickingens gegen Trier* (1933), no. 224 of *Historische Studien.* (D. G. D.)

SICULI (English SICELS; Greek SIKELOI), an ancient Sicilian tribe. In historical times they occupied the eastern part of the island, to which they gave their name. There was an ancient tradition that the Siculi once lived in central Italy east and even north of Rome, but were driven out and finally crossed to Sicily, leaving remnants behind; *e.g.,* at Locri (*q.v.*). They were distinct from the Sicani (*q.v.*) in the west of the island. They are hard to identify archaeologically ("Siculan," as used in archaeology, being purely a conventional term; *see* SICILY: *Archaeology*); some words of their Indo-European language are known.

The Siculi lived in independent towns; they were easily dis-

placed by the Greek colonies (*see* SICILY: *History*) and did not react *en masse* till the 450s B.C. under Ducetius (*q.v.*). The most important Sicel gods were the Palici, protectors of agriculture and sailors, who had a lake and temple near the Symaethus (Simeto) River; Adranus, sometimes said to be the father of the Palici, a god akin to Hephaestus, whose temple and everlasting fire were guarded by hounds; and the goddess Hybla or Hyblaea, whose sanctuary was at Hybla Geleatis. The chief Sicel towns were Agyrium (modern Agira), Centuripa or Centuripae (Centuripe), Henna (Enna), and three towns called Hybla (*q.v.*).

BIBLIOGRAPHY.—J. Whatmough, *The Foundations of Roman Italy,* ch. 15–16 (1937); T. J. Dunbabin, *The Western Greeks,* esp. ch. 1 and 5 (1948); L. Bernabò Brea, *Sicily Before the Greeks,* ch. 6 (1957).

SICYON (or SECYON, the local form; modern Greek SIKION), an ancient Greek city in the northern Peloponnese, about 11 mi. (18 km.) NW of Corinth and 2½ mi. (4 km.) SW of modern Sikionia. Its site is a low triangular plateau about 2 mi. (3 km.) from the Corinthian Gulf, at the confluence of the Asopus (Asopos) and Helisson (Elisson) rivers. Between city and port lay a fertile plain.

Its original name was said to have been Aegialeia ("beach town"), and it was inhabited in Mycenaean times. In the *Iliad* it is mentioned as subject to Agamemnon and previously ruled by Adrastus, king of Argos, who was still honoured as a hero at Sicyon in historical times. After the Dorian invasion it had the three Dorian tribes (*see* DORIANS) with a non-Dorian tribe called Aigialeis and a class of serfs, the "club-bearers" (*korunephoroi*) or "smock-wearers" (*katonakophoroi*). For several centuries Sicyon was subject to Argos. Independence was established in the 7th century by non-Dorian tyrants, called Orthagorids after their founder, Orthagoras; their mild rule lasted longer than any other Greek tyranny (*c.* 655–*c.* 555 B.C.). The most famous of the Orthagorid tyrants, Cleisthenes (*q.v.*), had connections with many commercial centres of Greece and south Italy and held an international competition before finally marrying his daughter to Megacles of Athens. Besides reforming the city's constitution and replacing Dorian cults with that of Dionysus, he took a chief part in the Sacred War (*c.* 590) in the interest of Delphi. Cleisthenes' successor, Aeschines, was expelled by the Spartans, but his institutions survived till the end of the 6th century, when Dorian supremacy was re-established and the city joined the Peloponnesian League. Henceforth Sicyonian policy was usually determined by Sparta or Corinth, though Argos retained some ascendancy (perhaps religious) in the 490s.

Sicyon supplied ships and 3,000 infantry to fight the Persians in 480–479. Its 6th-century bronze sculptors included Canachus (*q.v.*). In the 5th century it suffered like Corinth from the commercial rivalry of Athens and was repeatedly harassed by Athenian ships. In the Peloponnesian War (431–404) Sicyon followed Sparta and Corinth, and in the Corinthian War (394–386) it sided with Sparta. In 369 or 368 it was captured and garrisoned by the Thebans. Soon after this Euphron became tyrant with popular support, but his position was insecure and exiled Sicyonian oligarchs murdered him on an embassy to Thebes (*c.* 365). The oligarchy was restored, but new tyrants soon arose with the help of Philip II of Macedonia. During this period Sicyon reached its zenith as a centre of art: Apelles studied painting there, and Sicyonian sculpture culminated in Lysippus and his pupils. After participating in the Lamian War under a second Euphron (killed 322), and in the struggles between the successors of Alexander the Great, Sicyon was captured (303) by Demetrius Poliorcetes, who transplanted the inhabitants to the acropolis and renamed the site Demetrias. In the 3rd century it passed from tyrant to tyrant, until in 251 it was liberated by Aratus (*q.v.*) and enrolled in the Achaean League (*q.v.*). The destruction of Corinth (146) brought Sicyon additional territory and presidency over the Isthmian Games; but in 58 B.C. the Sicyonians had to sell their works of art to meet the city's debts. Under the Roman empire it was quite obscured by the restored Corinth and Patrae; Pausanias (*c.* A.D. 150) found it almost deserted after an earthquake. In Byzantine times it had a bishop, and there was an Early Christian basilica at Sikionia. Sicyon's later name, "Hellas," may mean that it was a refuge for Greeks fleeing the Slav invasions.

Ancient fortifications are still visible, and remains of a theatre, a gymnasium, a stadium, and other buildings, including a Roman bath, now the museum.

See C. H. Skalet, *Ancient Sicyon* (1928; with bibliography); A. Andrewes, *The Greek Tyrants,* ch. 5 (1956).

SIDAMA, derived from the tribal name Sidamo, is a term applied to Hamites (Cushites) of southwestern Ethiopia (*q.v.*), consisting of (i) the Sidamo, Darasa, Hadya, Kambatta, Tambaro, and Alaba, in the area between the Omo River and lakes Zeway, Shala, and Awasa; (ii) the Wolamo, Kucha, Gamo, Gofa, Konta, Kullo, Zala, and Kuera or Koyra, between the Omo and Lake Abaya (Margherita); (iii) the Chara, Basketo, Zaysse, Doko, and other minor tribes southwest of group ii. Linguistically, groups ii and iii are known as Ometo, "people of the Omo." Population figures for this area were very uncertain in the 1960s, the total being estimated at somewhere between 250,000 and 1,500,000. (*See* also GURAGE; JANJERO.)

All groups are agricultural, though many of them keep cattle. The principal food is ensete (Abyssinian banana); chicken or fish is not eaten. Until they were conquered late in the 19th century by the Ethiopians, all the Sidama tribes were small independent kingdoms. There is an age-grade system (*gada; see* AGE SET) derived from the Galla (*q.v.*) among the Sidamo, and in many tribes there are submerged classes; *e.g.,* smiths, weavers, potters, tanners. There is a widespread cult of a sky-god (called Wa'a in group i, Tosa or Tsosa in group ii) associated with a cult of nonancestral spirits. Some of the peoples are Monophysite Christians (*see* MONOPHYSITES) or Muslims, but paganism seems to predominate. *See* also AFRICA: *Ethnography* (*Anthropology*): *Northeast Africa;* CUSHITIC PEOPLES.

BIBLIOGRAPHY.—Enrico Cerulli, *Studi Etiopici,* vol. ii, iii (1938); Ernesta Cerulli, *Peoples of South-West Ethiopia and Its Borderland* (1957); C. F. Beckingham and G. W. B. Huntingford, *Some Records of Ethiopia, 1593–1646* (1954); G. A. Lipsky *et al., Ethiopia: Its People, Its Society, Its Culture* (1962). (G. W. B. H.)

SIDDONS, SARAH (née KEMBLE) (1755–1831), one of the greatest English tragic actresses, was born at the Shoulder of Mutton Inn, Brecon, Wales, on July 5, 1755, the eldest of the 12 children born to Roger Kemble and his wife, Sarah (née Ward; *see* also KEMBLE). As a child she acted with her father's strolling company, gaining what education she could as they traveled about, and soon proved an excellent actress. At the age of 18 she married William Siddons, a handsome but somewhat insipid actor who had been a member of her father's company. Having broken away from the family and joined another company, Sarah made her first success at Cheltenham. When David Garrick sent a representative to see her, she was playing Rosalind in *As You Like It* in a barn in Worcestershire and was also expecting a baby, but Garrick nevertheless offered her an engagement. This she could not accept until after the child was born. When she appeared at Drury Lane, London, with Garrick in 1775 she was a failure. Consequently she went back on tour in the country, and there she earned great favour, becoming the tragedy queen of the English stage and drawing vast crowds everywhere she appeared.

In 1782, at the request of Richard Brinsley Sheridan (who had succeeded Garrick), she consented, although with reluctance, to appear at Drury Lane again. There was no mistake this time: she chose the part of Isabella in *Isabella, or The Fatal Marriage,* and her success was phenomenal. From that time onward she reigned as queen at Drury Lane until, in 1803, she and her brother John Philip Kemble went to Covent Garden. In 1783 she was appointed to teach elocution to the royal children. She retired from the regular stage on June 29, 1812, with a farewell performance that was memorable: the audience would not allow the play, *Macbeth,* to proceed beyond the sleepwalking scene, in which she gave an exhibition of perfection, for it was as Lady Macbeth that Mrs. Siddons excelled.

She played all the great roles of tragedy, wisely leaving comedy alone (she nearly failed as Rosalind in *As You Like It*), but as a tragic actress she was one of the greatest the English stage—and possibly that of the world—had seen. Her power over her audi-

ences was complete; she held them under her sway. Lady Macbeth, Isabella, Belvidera in *Venice Preserv'd*, Jane Shore, Katharine in *Henry VIII*, Constance in *King John*, Zara in *The Mourning Bride*, and Volumnia in *Coriolanus* were her great parts. Just over middle height, she had a superb figure, classical features, large and eloquent eyes, a rich, resonant voice, and perfect diction. Her success was due to her complete concentration upon the character she played: she identified herself with it and seemed possessed by it, oblivious of all else around her, and when she uttered a shriek she seemed able to pierce the very soul of her audience. The mother of five children, she died in London on June 8, 1831.

See Yvonne Ffrench, *Mrs. Siddons: Tragic Actress* (1936).

(W. J. M.-P.)

SIDE (modern SELIMIYE, Turkey), an ancient city on the Pamphylian (central southern) coast of Asia Minor about 5 mi. (8 km.) W of the mouth of the Melas (Manavgat) River. It was founded by Greeks from Cyme in Aeolis (though a peculiar non-Greek language was also spoken there). It was the most important place in Pamphylia, having a good harbour of two small moles and two artificial harbours for larger vessels at the apex of its triangular site. Alexander the Great occupied it (333 B.C.), and there the Rhodian fleet defeated that of the Seleucid king Antiochus III (190 B.C.). In the 1st century B.C. the Cilician pirates made it their chief slave market. The great ruins cover a large promontory, fenced from the mainland by a ditch and wall, which survives with Byzantine modifications. Within this is a maze of structures out of which rises the colossal ruin of the theatre, built upon arches; it is the finest in Asia Minor.

See A. M. Mansel, *Die Ruinen von Side* (1963).

SIDERITE (CHALYBITE), a mineral consisting principally of ferrous carbonate, but frequently containing substantial amounts of manganese and magnesium and lesser amounts of calcium substituted for iron. Also known as spathic iron ore, its principal use is as a source of iron for the manufacture of iron and steel, especially in Great Britain and Austria (*see* IRON: *Compounds of Iron;* IRON AND STEEL INDUSTRY: *Iron Ore*). Ferrous carbonate, not necessarily produced directly from siderite, is used in the treatment of anemia and in the preparation of other iron compounds. The composition is expressed by $FeCO_3$. The hardness is 3.5–4, the specific gravity is 3.96 (for pure $FeCO_3$) and the colour is usually a shade of brown because of partial oxidation.

Siderite is found in small amounts in cavities in basaltic igneous rocks and as a subordinate gangue mineral in hydrothermal metallic veins and in some pegmatites. It is widely distributed as a fine-grained concretion (*q.v.*) and in thin beds associated with shales and coal seams. Organic matter preserved in the latter rocks has been taken to indicate that they were formed in reducing environments, and indeed siderite has been observed in some swamps forming under such conditions. Siderite is also found in sedimentary rocks of Jurassic age in England and western Europe admixed with varying amounts of calcitic shells and oölites of the iron silicate chamoisite. It is a constituent of the banded iron ores of the Lake Superior district. (D. L. G.)

SIDEROSTAT, an instrument which, like the coelostat and heliostat (*qq.v.*) reflects a portion of the sky in a fixed direction notwithstanding the diurnal motion of the heavens. The name is applied especially to the polar siderostat, a form of telescope in which the observer looks down the polar axis onto a mirror; by adjusting the mirror he can bring any part of the sky into the field of view without changing his own position.

SIDGWICK, HENRY (1838–1900), English philosopher, best known for his theories concerning ethics, was born at Skipton in Yorkshire on May 31, 1838. He was educated at Rugby and at Trinity college, Cambridge. In 1859 he was elected fellow of Trinity and then lecturer in classics. In 1869 he exchanged this lectureship for one in moral philosophy, but resigned his fellowship on religious grounds.

He was appointed praelector in 1875 and elected to an honorary fellowship in 1881. In 1883 he was appointed Knightsbridge professor of moral philosophy; and in 1885, the religious test having been removed, Trinity again elected him to a fellowship on the foundation. He died on Aug. 29, 1900.

Sidgwick took an active part in university business and especially in promoting the higher education of women through the foundation of Newnham college. He had married in 1876 Eleanor Mildred Balfour (sister of A. J. Balfour) who succeeded Miss A. J. Clough as principal of Newnham in 1892. He was active in social and philanthropic work and deeply interested in psychical phenomena, being a founder and first president of the Society for Psychical Research. But his energies were mainly directed to the study of religion and philosophy.

Sidgwick's major work was *Methods of Ethics* (London, Cambridge, 1874; 6th ed., London, 1901; reissue of 7th ed., 1930). His three "methods" are egoism, utilitarianism and intuitionism. Egoism is the theory that justifies an action by its contribution to the greatest happiness of the agent, utilitarianism by its contribution to the greatest happiness of all those affected by it. Sidgwick uses "intuitionism" to designate all theories which recognize ultimate ends other than happiness or ultimate rules other than the rule that the maximum happiness ought to be promoted. In support of intuitionism, Sidgwick admits that ordinary men regard many rules as binding on them independently of the happiness produced and also that they often regard ends other than happiness (ends such as knowledge, beauty and virtue) as intrinsically desirable. But he argues that the rules of common sense are vague and indefinite; they allow exceptions and conflict with each other. Thus they cannot be regarded as providing an adequate rational basis for conduct. Most of them can, however, be justified as tending to promote general happiness, and the principle of utility can always be invoked to render their application more precise, to explain their exceptions and to resolve their conflicts.

While rejecting psychological hedonism, Sidgwick points out that any action can be rationally justified by demonstrating its contribution to the happiness of the agent. This egoistic principle, indeed, if taken as the sole guide to life, seems repellent to the moral sentiment; and when a man's private good conflicts with the general good most men agree that the former should yield. Yet even such lofty moralists as Joseph Butler and Samuel Clarke are found suggesting that enlightened self-love is the supreme principle of rational conduct. Utilitarianism appears to Sidgwick to be the only principle capable of giving rational unity to moral consciousness. But he finds it difficult to reconcile its moral claims with the rational supremacy of self-love. There is no empirical evidence that a man's achievement of the greatest general happiness will in fact entail his own greatest happiness. . Some thinkers, such as Plato and Kant, relied here on a belief in the divine ordering of the world to guarantee this coincidence. Sidgwick concludes that only such a belief (whether based on theology or imported as an *ad hoc* postulate) will serve to bring unity to the world of rational conduct.

Sidgwick's claim to distinction does not rest on the originality or conclusiveness of the doctrines summarized above. His work is marked by caution and thoroughness, a masterly exposure of confusions and ambiguities, an acute and balanced presentation of complex argument and an open-minded candour and honesty.

Sidgwick's chief works are *Principles of Political Economy* (1883; 3rd ed., 1901); *The Scope and Method of Economic Science* (1885); *Outlines of the History of Ethics* (1886; 6th ed., 1931), enlarged from his article "Ethics" in the *Encyclopædia Britannica*, 9th ed., vol. viii (1879); *Elements of Politics* (1891; 2nd ed., 1897; reissue of 4th ed., 1929). The following were published posthumously: *Philosophy: Its Scope and Relations* (1902); *Lectures on the Ethics of T. H. Green, Mr. Herbert Spencer and J. Martineau* (1902); *The Development of European Polity* (1903); *Miscellaneous Essays and Addresses* (1904); *Lectures on the Philosophy of Kant* (1905). (J. D. M.)

SIDGWICK, NEVIL VINCENT (1873–1952), English chemist whose conception and elaboration of electronic bonding was a valuable interpretation of chemical union, was born on May 8, 1873, at Oxford. Sidgwick received his M.A. at Oxford university and his Ph.D. at Tübingen. He became fellow and tutor at Lincoln college, Oxford, in 1901, reader in chemistry at Oxford in 1926 and professor in 1935, a position which he held until 1945. His books are: *Organic Chemistry of Nitrogen* (1910), *Electronic Theory of Valency* (1927), *The Covalent Link in Chemistry* (1933) and *The*

Chemical Elements (1950) in two monumental volumes. Sidgwick was Baker lecturer at Cornell (N.Y.) in 1931, vice-president of the Royal society in 1935–37, president of the Faraday society in 1932–34 and president of the Chemical society in 1935–37. He received the Royal medal of the Royal society in 1937 and the Longstaff medal in 1945. He died March 15, 1952, at Oxford.
(V. Bw.)

SIDHI, a town in Madhya Pradesh, India, and the headquarters of the district of the same name. Pop. (1961) 5,021. It is situated on the northern outcrops of the Vindhya Mountains at a height of 706 ft. (215 m.) above sea level overlooking the Son River. Sidhi has no railway communications; one metaled road, 46 mi. (74 km.) long, connects it with Rewa (*q.v.*) to the northwest.

Sidhi District is the most northeasterly district of Madhya Pradesh. It covers an area of 4,060 sq.mi. (10,515 sq.km.) and has a population (1961) of 580,129. The southern part of the district is hilly and undulating. The northern part is flat country and is drained by the Son and its tributary toward the northeast. The climate is hot and moist, the average annual rainfall being 45 in. (1,143 mm.). The soil is mostly laterite and crops include rice, wheat, barley, and gram. Nearly nine-tenths of the population is engaged in agriculture. The hilly tracts of the south are inhabited by tribal peoples. (M. N. K.)

SIDI BEL ABBÈS, the chief town of an *arrondissement* in the Oran (Ouahran) *département*, Algeria, lies at an altitude of 1,552 ft. on the Mékerra River, 48 mi. (77 km.) S of Oran by rail. Pop. (1960) 105,357 (commune), about 65,000 Muslim. The town derives its name from the tomb of the marabout (saint) Sidi Bel Abbès, and was founded by French troops in 1843, becoming a *commune* in 1856. Sidi Bel Abbès was the depot and recruiting station of the Foreign Legion. The Legion's barracks contained the Legion Museum, comprising the Salle d'Honneur, with the colours of the Legion's regiments and their principal trophies, and the Musée du Souvenir. The town was evacuated by the Legion and all French troops after Algerian independence. It was formerly surrounded by walls and bastions, with four gates, but these were demolished and replaced by boulevards and squares, beyond which are the spreading suburbs. Other buildings include that of Les Coopératives Agricoles (Union of Farmers), the town hall, a mosque, churches, and a theatre; there are a *lycée* and other schools. The town is on the Oran-Tlemcen-Oujda railway and an airport is 3 mi. distant. The surrounding area, once swampy, now produces cereals (mainly wheat and barley) and vines.
(A. Am.)

SIDMOUTH, HENRY ADDINGTON, 1st Viscount (1757–1844), English statesman, prime minister from 1801 to 1804, and home secretary from 1812 to 1822, was the son of Anthony Addington, a physician, and was born in London on May 30, 1757. Educated at Winchester College and at Brasenose College, Oxford, he became a member of Parliament for Devizes, in Wiltshire, in 1784. He was a personal friend and strong supporter of William Pitt, and in 1789, with Pitt's help, became speaker of the House of Commons. Rejecting offers of a position in the Cabinet, he held this post until 1801. In that year, after Pitt had quarreled with George III on the question of Catholic emancipation, the more pliable and very Protestant Addington became prime minister, a position he held until April 1804. He was responsible for the Treaty of Amiens in March 1802, a settlement which at first increased his popularity in the country, but when it proved only temporary Addington's fortunes became less favourable. War had been renewed in May 1803, and it became increasingly clear that Addington lacked the necessary gifts of a war leader, and Pitt, who began openly to oppose the government, took his place.

Addington's continued importance in politics depended not on his personality and his abilities so much as on a little group of about 50 followers in the House of Commons. In January 1805 Addington and his group supported Pitt, and Addington himself, who was now created Viscount Sidmouth, reentered the Cabinet as lord president of the council. In July 1805 he broke away again but entered the opposition ministry of Charles James Fox and George Grenville in February 1806 as lord privy seal. Remaining true to his Protestant, Church of England allegiance, he left the new ministry in 1807 when the government proposed to throw open commissions in the Army and Navy to Roman Catholics and dissenters. As a spokesman of the Church of England party he attempted in 1811 to bring in a private bill requiring all dissenting ministers to be licensed, and restraining unlicensed preachers. The bill was unsuccessful, but a year later he returned to the Cabinet as lord president of the council under Spencer Perceval. After Perceval's assassination (May 1812) and the formation of the earl of Liverpool's ministry, Sidmouth became home secretary in June.

He had to deal with a discontented population disposed to demonstration and riot. The first year of his office was one of high prices, commercial difficulties, and large-scale unemployment; to control the situation, in face of demonstrations both by manufacturers and by Luddites, he increased the power of the magistrates. After 1815 he carried this policy of firmness much further. In 1817, for instance, he issued a circular to the lords lieutenant declaring that magistrates might apprehend and hold to bail persons accused on oath of seditious libels, and was supported by Parliament even after a publisher, William Hone, had been acquitted by a jury on charges which followed Sidmouth's instructions. He was mainly responsible for the later policy embodied in the Six Acts of 1819, which, among other provisions, limited the rights of public meeting and the circulation of political literature. His instructions to magistrates and his use of police spies made him detestable to all Radicals and Liberals. Exhausted by his efforts, he resigned office in January 1822, but, with the strong support of George IV, he remained in the Cabinet until November 1824, when he resigned because of his disapproval of the recognition of the South American republics. Thereafter he took little part in public affairs, although he remained a strong supporter of church and king, in 1829 speaking against Catholic emancipation and in 1832 voting against the Reform Bill. He died at the White Lodge, Richmond Park, on Feb. 15, 1844. Sidmouth was a statesman of limited imagination and no outstanding ability, but his character and integrity could not be questioned: his inability to assess the forces of change in English politics prevented him from leaving any permanent mark on English history.

Bibliography.—G. Pellew, *The Life and Correspondence of . . . Viscount Sidmouth*, 3 vol. (1847); Sir G. C. Lewis, *Essays on the Administrations of Great Britain from 1783 to 1830* (1864); Sir S. Walpole, *History of England from . . . 1815*, 5 vol. (1878–86); W. R. Brock, *Lord Liverpool and Liberal Toryism, 1820–1827* (1941). (A. Bri.)

SIDMOUTH, a seaside and resort town and urban district in the Honiton parliamentary division of Devon, Eng., 15 mi. (24 km.) ESE of Exeter by road. Pop. (1961) 10,890. Area 17.9 sq.mi. (46.4 sq.km.). It lies in the valley formed by the Sid in the red sandstone cliffs edging the western end of Lyme bay. Traces of a Bronze Age camp have been found on High Peak; on Salcombe Hill is the Norman Lockyer Observatory, now belonging to the University of Exeter. The beach is pebbly, with sand at low tide, and most of the houses were built during the Regency period when Sidmouth became popular. Agriculture is a main industry.

SIDNEY (Sydney), **ALGERNON** (1622–1683), English politician executed for treason, exercised a lasting posthumous influence in Great Britain and America as a champion of constitutional liberties. The second surviving son of Robert, 2nd earl of Leicester, Sidney was born at Penshurst Place, Kent. In October 1641 he went to Ireland with his brother Philip, and helped to suppress the rebellion there. The brothers returned to England in August 1643 and, suspected of supporting the royalist cause in the Civil War, were immediately detained by the parliamentarians. By this time, however, Sidney's attachment to the cause of constitutional liberties was already formed and he willingly joined the parliamentary side.

Made a captain of horse in the earl of Manchester's army on May 10, 1644, he soon became a lieutenant colonel and charged at the head of his regiment at Marston Moor (July 2), receiving serious wounds. In April 1645 he was given the command of a cavalry regiment in Oliver Cromwell's division of Sir Thomas Fairfax's army, and became governor of Chichester in Sussex in May.

In 1647 he again went to Ireland, where his brother was lord lieutenant, and was made lieutenant general of horse and governor of Dublin. But these appointments aroused jealousies in Ireland and Algernon was immediately recalled and placed at Dover as governor. In 1648, nominated against his will as a commissioner for the trial of Charles I, he refused to take part. He was deprived of his post at Dover in 1651, retired to Holland for some months, and, although appointed to the Council of State (1652), consistently withheld his support from Cromwell after the latter became lord protector (December 1653). He caused more offense in 1656 by taking the role of Brutus in performances of *Julius Caesar* at Penshurst.

On the restoration of the Rump of the Long Parliament (May 1659) he was again placed on the Council of State. When Charles II was restored to the throne, Sidney, then in Denmark, was not disposed to make his peace with the Stuarts and instead made his way to Rome (November 1660). Welcomed by Roman society, he enjoyed the tranquil pleasures of study, but in 1663 the attraction of the world of affairs proved too strong and he went to Brussels and to The Hague. His movements were closely watched by the English government and on the outbreak of the Second Dutch War (1665–67) attempts were made to seize or assassinate him. From The Hague he urged the invasion of England and after France had allied with Holland he offered at Paris to raise rebellion in England for 100,000 crowns. Nothing came of these schemes and in 1666 he retired to the south of France.

However, by 1677 the declining health of his father, who had continued to pay him a small maintenance, gave him compassionate grounds to return to England. The earl of Leicester died in November, but a protracted chancery suit over the will detained Sidney from returning to France. He now found himself the centre of liberal expectations, and was almost instantly drawn into the politics of opposition to Charles II. In the excitement of the "Popish Plot" (*see* ENGLISH HISTORY) Sidney, a welcome recruit to the Country Party, made a series of unsuccessful attempts to get into the House of Commons, aided by his friend William Penn. Sidney's opinions were similar to those of the earl of Shaftesbury, but they did not associate closely and it was as an individual, not as member of a faction, that Sidney received bribes in 1679 and 1680 from the French ambassador Paul Barillon to assist in undermining Charles II's government. Meanwhile Sidney worked on his *Discourses Concerning Government* (published in 1698 after his death), inspired, like John Locke's *Two Treatises,* by the move to exclude James, duke of York, from the succession, and by the menacing implications of Sir Robert Filmer's absolutist work *Patriarcha,* published for the first time in 1680.

Although relatively ineffectual as a politician, Sidney's reputation as a leader of popular opposition made it certain that the government would try to entrap him. His alleged but improbable implication in the Rye House Plot of 1683 to assassinate the king and the duke of York was therefore not unexpected. Denounced by informers, Sidney was arrested on June 26, 1683, and tried before Judge George Jeffreys in November. As in the trial of his fellow victim, Lord William Russell, the rules of evidence and procedure were heavily weighted against him. Passages from the manuscript of the *Discourses* were used to supplement Lord Howard of Escrick's testimony that Sidney believed in the lawfulness of resistance and had conspired for the death of the king. He was found guilty, and beheaded on Tower Hill, London, on Dec. 7, 1683. His remains were buried at Penshurst.

The example of Sidney's lifelong resistance to royal absolutism proved enduring, and his *Discourses* were a potent vehicle for liberal ideals. They became a popular "textbook of revolution" in the American colonies and were well-known in 18th century France and Germany. Although Sidney declared "I am persuaded to believe that God had left nations to the liberty of setting up such governments as best pleased themselves" it is misleading to call him a republican. His *Discourses* prescribed a limited monarchy, regulated by the concept of a voluntary civil contract, ruling in the interests of the free community of citizens in whom resided ultimate authority. Like the political philosopher James Harrington, Sidney recognized social and economic changes under the political tendencies of his time, and he lamented the extinction of the old nobility in the rising tide of middle class wealth. Sidney's liberalism was essentially aristocratic and a patrician hauteur was a commonly observed feature of his behaviour. It accounts for some of the vicissitudes of his life, as well as the nobility of his bearing on the scaffold.

See A. C. Ewald, *The Life and Times of the Hon. Algernon Sidney, 1622–1683,* 2 vol. (1873). (H. G. Ro.)

SIDNEY, SIR HENRY (1529–1586), English lord deputy of Ireland from 1565 to 1571 and from 1575 to 1578, was born on July 20, 1529, probably at Baynard's Castle, London. The eldest son of the soldier and courtier Sir William Sidney, he was brought up at court and enjoyed the favour of successive sovereigns. A gentleman of the privy chamber to Edward VI and knighted in 1550, Sidney's importance was so great that John Dudley, earl of Warwick (later duke of Northumberland), secured him as husband for his daughter Mary. Sidney signed Edward VI's will leaving his throne to Lady Jane Grey, but rapidly went over to Mary I and thus retained his position at court.

Sidney's sister Frances married (1555) Thomas Radcliffe, Lord Fitzwalter (afterward earl of Sussex), and when Fitzwalter went to Ireland as lord deputy in April 1556, Sidney accompanied him as vice-treasurer. He took part in the military expeditions (1556 and 1557) against the Scots in Antrim and acted as lord justice during Fitzwalter's absences. In 1559 he returned to England and was appointed lord president of the Welsh Marches. He resided much at court, until returning to Ireland as lord deputy (1565). Like Sussex, he played a direct, if equally inglorious, part in negotiating with Shane O'Neill (*see* IRELAND; *History*). Sidney encouraged O'Neill to believe that Elizabeth I would accept him as ruler of Tyrone, at the same time urging on the queen a harsh policy. But Sidney's military moves against O'Neill were unsuccessful and he was reduced to intriguing against him with the O'Donnells and with Sorley Boy MacDonnell. After the assassination (1567) of O'Neill by the MacDonnells, the lord deputy's claim to have conquered the Ulster leader was brazenly paraded in the preamble to the act (1569) of forfeiture which declared that most of the Northern Province belonged to the queen. But Sidney's weakness is shown in the fact that while parliament decreed the abolition of the title O'Neill, Turlough Luineach was proclaimed in succession to Shane without suffering adverse consequences.

At the Parliament of 1569–71, Sidney showed statesmanship in the care he took to avoid religious dissensions. With the Butlers as well as the Geraldines in alliance against the Dublin administration, an insistence on conformity to Protestantism would have made the war in the south more general and more difficult to terminate. Sidney contented himself with securing a temporary amendment of Poynings' Law, which enabled him to negotiate for the enactment of other parts of the government's program.

In his later periods of office in Ireland, Sidney made efforts to enlighten the London government about Irish conditions, to systematize Irish administrative arrangements, and to extend English rule in Ireland. He undertook long journeys in order to enforce his ideas and tried both to moderate fanatical English officials and to force recalcitrant Irish chiefs to observe the English laws. Yet his insistence on arbitrary taxation and impositions for the viceroy's forces caused much discontent and ultimately led to his recall. Thereafter he served the queen only as president of the Council of Wales and of the Marches, apparently giving general satisfaction. He died at Ludlow, Shropshire, on May 5, 1586.

Sidney's letters of state suggest he had statesmanlike qualities. He gave admirable advice to London on securing impartial administrations in Ireland and he expressed his dissatisfaction when the queen favoured Ormonde in his quarrel with Desmond. But as time went on he grew too sophisticated and complaisant. No judge of character, he was too ready to assist profit-making projects by needy and unscrupulous English adventurers. He was also rather vain of honours, and patronized antiquaries who gave him a fictitious pedigree. He resented the queen's failure to reward his Irish services, but showed little anxiety for a peerage,

which might have involved him in ruinous financial burdens.
(R. D. Es.)

SIDNEY, SIR PHILIP (1554–1586), the finest English example of the Renaissance ideal of the perfect gentleman. He was virile and cultivated: a soldier, statesman, courtier, poet and patron of scholars and poets. The eldest son of Sir Henry Sidney and Mary Dudley, he was born at Penshurst, Kent, on Nov. 30, 1554. On Oct. 17, 1564, he and his lifelong friend Fulke Greville were entered at Shrewsbury school. Sidney's training was based on the assumption that he would be heir to his uncle, the earl of Leicester. He visited Kenilworth, Leicester's home, in Aug. 1566 and with Leicester took part in Queen Elizabeth I's progress to Oxford. In 1568 he went to Christ Church, Oxford. As lord deputy Sir Henry was much in Ireland and Sidney spent some vacations with the Cecils: a match was proposed in 1569 with Sir William Cecil's daughter Ann, but came to nothing.

DETAIL FROM A PORTRAIT BY AN UNKNOWN ARTIST; PHOTO BY COURTESY OF THE EARL OF WARWICK

SIR PHILIP SIDNEY AT THE AGE OF 22

On May 25, 1572, Queen Elizabeth granted Sidney a licence to travel for two years to learn foreign languages. He left England in the earl of Lincoln's suite and reached Paris, where he lodged with the English ambassador, Sir Francis Walsingham, on June 8. The massacre of St. Bartholomew's day caused his departure for Frankfurt, where he lodged with the scholar-printer, Andrew Wechel, and met Hubert Languet, an ardent Huguenot who, like other mature scholars and men of affairs, conceived warm affection and admiration for Sidney. In the early summer, 1573, he went on to Heidelberg, Strasbourg and Vienna and in October left for Italy, spending some time in Venice where his portrait was painted by Paolo Veronese. In Oct. 1574 he returned to Vienna, and after a second visit to Prague in Feb. 1575, to England. He spoke French and Italian well, and his letter to Lord Burghley (Dec. 1574) shows that he had also a firm grasp of European politics.

Insufficient use was made of his training and exceptional ability during the next ten years. His father had once more been sent to Ireland, and this brought Sidney into touch with the earl of Essex, earl marshal for Ireland, who desired a marriage between Philip and his daughter, Penelope—the "Stella" of his sonnets. But nothing came of this, and Essex died in 1576. In 1577 Sidney was sent on an embassy to the emperor Rudolph II. He met Don John of Austria at Louvain, proposed a Protestant league to the emperor at Prague and on the way back was warmly received by William of Orange. On his return he defended his father's government of Ireland, particularly from the attacks of Lord Ormond, writing the *Discourse on Irish Affairs* for the queen. In 1580 he was bold enough to oppose the queen's proposed marriage to the duc d'Alençon in a *Letter to the Queen*. He was knighted in 1583 to enable him to receive the Garter as Prince John Casimir's proxy.

Sidney divided his time between his uncle's town residence, Leicester house, and Wilton, where his sister Mary, now countess of Pembroke, lived. He took part in the court festivities and composed *The Lady of May* for Leicester's entertainment of the queen at Wanstead in 1578. In this year he met Spenser at Leicester house and discussed with him and Edward Dyer the use of classical metres in English verse. Spenser's *The Shepheardes Calender* (1579) was dedicated to Sidney. During these years he himself wrote the *Arcadia, An Apologie for Poetrie* and *Astrophel and Stella*—some of the sonnets in this cycle would seem to have been written after Penelope Devereux's enforced marriage to Lord Rich toward the end of 1581. Sidney's literary output was the result of his lack of employment in serious affairs, and it is unlikely that his untimely death prevented further imaginative works.

In the autumn of 1583 he married Frances Walsingham; Elizabeth stood godmother to the daughter born in 1585. Sidney still desired active service and took an interest in the colonizing enterprises of Martin Frobisher, Richard Hakluyt and Sir Walter Ralegh. He advocated a direct attack on Spain, and was himself preparing to sail with Drake in 1585 when the queen recalled him and appointed him governor of Flushing. Leicester, in command of the English forces in the Netherlands, leaned heavily on his advice during the next difficult months.

In July 1586 Sidney made a successful raid on Axel, near Flushing; on Sept. 22 he was with a small force under Sir John Norris which Leicester ordered to intercept a convoy of provisions near Zutphen. According to Fulke Greville, Sidney cast off his cuisses because Sir William Pelham was without his, and so received the fatal bullet wound in his thigh. Greville is also responsible for the story that he refused a cup of water in favour of a wounded soldier with the words, "Thy necessity is yet greater than mine." Greville was not present, but both stories are in keeping with Sidney's character. He died at Arnhem on Oct. 17, 1586.

Sidney's death was an occasion for universal mourning. The most famous of some 200 elegies is Spenser's *Astrophel*, published, with others by the countess of Pembroke, Fulke Greville, Ralegh, Matthew Roydon and Lodowick Bryskett, in his *Colin Clouts Come home againe* (1595).

Writings.—None of Sidney's works was printed during his lifetime. Between 1577 and 1580 Sidney wrote the first version of his prose romance *Arcadia*, in five books or acts, for the entertainment of his sister and her friends. This version remained in manuscript until A. Feuillerat printed it from the Clifford manuscript in 1926. Between 1580 and 1584 Sidney embarked on a radical revision, using nearly all the old material, but enlarging books i and ii with many additional episodes and writing a new book iii, which breaks off in the middle of a sentence before he had reached the point where the old book iii began. This version was published as *The Countesse of Pembrokes Arcadia* (1590), probably under the supervision of Fulke Greville, who claimed that it truly represented Sidney's serious and moral intentions. In 1593 Hugh Sanford prepared a folio edition which included the 1590 version and books iii–v of the old *Arcadia* with some changes which may represent Sidney's intentions, but were probably made by his sister. This was the *Arcadia* known to countless later writers and readers. Sidney's early pastoralism seems to have been mainly influenced by Jacopo Sannazzaro's *Arcadia* (poems with short prose links) and his later epic narrative by Heliodorus' *Aethiopica* (a Greek novel with a complicated "Chinese box" structure of stories within the main story). The golden world of innocent shepherds and shepherdesses is combined by Sidney with a sterner world of violent action and crime. The main narrative concerns the adventures of two princes, Musidorus and Pyrocles, who fall in love with Pamela and Philoclea, the daughters of Basilius and Gynecia. The characters of the two girls are well differentiated and in her struggle with illicit passion Gynecia has been likened to Phèdre. There are some good comic and dramatic episodes, and many serious debates on such topics as justice, atheism and suicide. The sentences are long and laden with rhetorical devices, for example:

> There were hills which garnished their proud heights with stately trees; humble valleys whose base estate seemed comforted with refreshing of silver rivers; meadows enameled with all sorts of eye-pleasing flowers; thickets, which being lined with most pleasant shade, were witnessed so to by the cheerful disposition of many well-tuned birds; each pasture stored with sheep feeding with sober security, while the pretty lambs with bleating oratory craved the dams' comfort; here a shepherd's boy piping, as though he should never be old; there a young shepherdess knitting and withal singing, and it seemed that her voice comforted her hands to work, and her hands kept time to her voice's music.

Sidney's use of this style was deliberate. In all his work he observed the principle of decorum, suiting style to subject. That he could be direct is shown in this letter to his father's secretary:

> I assure you before God that if I know you do so much as read any letter I write to my Father, without his commandment, or my consent, I will thrust my dagger into you. And trust to it, for I speak it in earnest.

Many poems in which Sidney experimented in metrical forms are introduced into the *Arcadia;* they include a few sonnets. These, however, do not reach the high standard of those in his *Astrophel and Stella.* Sir Thomas Wyat and the earl of Surrey had introduced the sonnet into English poetry, but *Astrophel and Stella* was the first English *canzoniere*—a series of songs and sonnets telling a love story. Sidney's sonnets were partly autobiographical, but they were also indebted to Petrarch, Du Bellay, Ronsard, Desportes and others. His metre is extremely regular. The rhyme scheme of the octave is that adopted in Europe; Sidney usually followed Surrey in ending the sestet with a couplet. He often adopted Wyat's colloquial tone; for example,

> "Fool," said my Muse to me, "look in thy heart and write!"

The fondness for word play and the use of personification apparent in the *Arcadia* reappear here. More striking is the freshness he managed to infuse into the stale Cupid imagery.

> With how sad steps, o Moon, thou climb'st the skies,
> How silently, and with how wan a face!
> What, may it be that even in heav'nly place
> That busy archer his sharp arrows tries?

After Shakespeare's sonnets, *Astrophel and Stella* is the finest Elizabethan sonnet cycle. Thomas Newman published a faulty quarto edition, with preface by Thomas Nashe and some sonnets by other writers, in 1591. He followed this with a better text in the same year. The 1598 *Arcadia* folio offered a still better text and included *Certain Sonnets* and *The Lady of May.* To the 1613 folio was added *A Dialogue Between Two Shepherds. Two Pastorals* appeared in Francis Davison's *Poetical Rhapsody* (1602). (On the authenticity of these last two items, *see* W. Ringler, "Poems Attributed to Sir Philip Sidney," *Studies in Philology,* xlvii, pp. 126–151; 1950.)

Sidney seems to have had Stephen Gosson's *The Schoole of Abuse* (1579) in mind when he sprang to the defense of poetry, though the exact date of his treatise is not known. It was published in 1595, by W. Ponsonby as *The Defence of Poesie* and by H. Olney as *An Apologie for Poetrie,* and Ponsonby included it in the 1598 *Arcadia.* It is the first work to introduce the critical ideas of Renaissance theorists into England. Sidney's arguments for the lofty nature of poetry and against its detractors have been repeated countless times—partly by reason of the easy and persuasive style. He is urbane and courteous where Gosson was shrill and abusive. He wrote just before the great Elizabethan age of poetry and drama, and without the example of Shakespeare, he cannot be blamed for missing the English genius for tragicomedy. He praised the best English works available.

A Woorke Concerning the Trewnesse of the Christian Religion, translated from Duplessis-Mornay, was completed and published by Arthur Golding in 1587. Sidney was responsible for the first 43 psalms in the metrical versions completed by his sister. His translations of the first *Semaine* of G. Du Bartas and of the first two books of Aristotle's *Rhetoric* are lost. Sidney's letters, especially the Latin correspondence with Languet (trans. and ed. by S. A. Pears, 1845), and his three political treatises, not published in his lifetime, the *Discourse on Irish Affairs,* the *Defence of Leicester* and the *Letter to the Queen,* are of historical interest.

Sidney's *Complete Works,* including both versions of the *Arcadia,* were ed. by A. Feuillerat, 4 vol. (1912–26). The texts are not altogether satisfactory, and for his poems see *The Poems of Sir Philip Sidney,* ed. by W. A. Ringler (1961). A modernized text of *The Countess of Pembroke's Arcadia* with the linking passage in bk iii supplied by W. Alexander in 1621, and the sixth book, added by R. Beling in 1627, was ed. by E. A. Baker (1907).

Bibliography.—Fulke Greville's *Life of the Renowned Sir Philip Sidney* (1625) was ed. by N. Smith (1907). The best biography is M. W. Wallace, *The Life of Sir Philip Sidney* (1915); the works are more fully treated by M. Wilson, *Sir Philip Sidney,* 2nd ed. (1950). There is a full bibliography by M. Poirier, *Sidney: le chevalier poète élisabéthain* (1948). *See* also J. Buxton, *Sidney and the English Renaissance* (1954); J. W. Lever, *The Elizabethan Love Sonnet* (1956); R. W. Zandvoort, *Sidney's Arcadia: a Comparison between the Two Versions* (1929); Jean Robertson, "Sir Philip Sidney and his Poetry," *Elizabethan Poetry,* pp. 110–29 (1960). (J. Rn.)

SIDON (modern Saïda), once the chief city of Phoenicia, is the fourth largest city in Lebanon. Situated on the Mediterranean about 25 mi. (40 km.) SSW of Beirut, it had an estimated population of 32,200 in 1961. It is an agricultural trading centre as well as a trading and fishing port. It has a large Christian community as well as an old Jewish community and quarter. Since 1950 the city has been the Mediterranean terminus of the pipeline from the Saudi Arabian oil fields.

History.—Sidon, one of the oldest Phoenician cities, was founded in the 3rd millennium B.C. and became prosperous in the 2nd. It was burned by the Philistines in the 12th century B.C. but rapidly regained its strength and scattered its colonies in the eastern Mediterranean (including Cyprus). It is frequently mentioned in Homer and in the Old Testament. It was ruled in turn by Assyria, Egypt, Persia, Alexander the Great, the Seleucids of Syria, the Ptolemies of Egypt, and the Romans. At this time Sidon was famous for its purple dyes and glassware. Herod the Great embellished the city, and Jesus visited its neighbourhood.

During the period of the Crusades Sidon changed hands several times, and more than once it was destroyed and rebuilt. It was destroyed again by the Mongols in 1260. Under Turkish rule almost continuously for 400 years from 1517, it blossomed into vigorous existence in the 17th century under Fakhr al-Din II, a semi-independent Druze amir, who encouraged and protected its commerce. In 1791, however, the governor of Lebanon, Ahmad al-Jezzar, drove the French merchants from its gates, thereby largely killing its trade. In 1837 it was ravaged by an earthquake but rebuilt. The later history of the city is inseparable from that of the rest of Lebanon (*q.v.*).

Archaeology.—Because much of the ancient city lies under the modern town, the archaeological history of Sidon is obscure. A large necropolis discovered southeast of the town in 1855 yielded numerous sarcophagi including those of two Sidonian kings, Eshmunazar and Tabnith, both with valuable Phoenician inscriptions. In 1887, 17 more sarcophagi were discovered including the famous "Alexander" sarcophagus, depicting battle and hunting scenes, now at Istanbul. The Beirut Museum possesses 31 more sarcophagi and a rich collection of jewelry of the 5th–4th century B.C. found in some of them during further excavations carried out in the necropolis in 1963–64. Since 1961 the Lebanese General Directorate of Antiquities has undertaken systematic excavations at the site of the Persian palace and in the temple of the god Eshmun, discovered in the 19th century 2.8 mi. (4.5 km.) N of the town.

The ancient city's sea mole and wall can still be located, and it still has the twin harbours used in antiquity and two castles of the crusader period. *See* also Phoenicia.

Bibliography.—F. C. Eiselen, *Sidon* (1907); O. Hamdy Bey and T. Reinach, *Une Nécropole royale à Sidon* (1892); E. Kukahn, *Anthropoide Sarkophage in Beyrouth* (1955). (R. D. B.)

SIEBENGEBIRGE ("Seven Hills"), actually a group of about 40 wooded hills of volcanic origin in Germany between Königswinter on the right bank of the Rhine and the Cologne-Frankfurt am Main *Autobahn.* The district is a popular tourist resort and a nature reserve. The hills form the northwestern part of the Westerwald region. The seven principal hills seen from Bonn (whence the name) are: Drachenfels (1,053 ft. [321 m.]), reached by rack railway from Königswinter and surmounted by a ruined castle; Wolkenburg (1,063 ft.); Petersberg (1,086 ft.), with a motor road to the summit hotel which until 1952 was the seat of the tripartite Allied High Commission; and, to the south, Ölberg (1,522 ft. [464 m.]), the highest of the group; Löwenburg (1,493 ft.); Lohr-berg (1,427 ft.); and Nonnenstrom-berg (1,101 ft.). Quarries yield basalt for building (*e.g.,* Cologne and Limburg cathedrals) and paving, on the lower slopes behind Königswinter are some of the northernmost vineyards in Europe.

(R. E. Di.)

SIEBOLD, KARL THEODOR ERNST VON (1804–1885), German zoologist who investigated the morphology and life history of lower animals, was born in Würzburg on Feb. 16, 1804. After practising medicine briefly, Siebold was professor of zoology at Erlangen, Freiburg, Breslau and Munich. In 1848 he founded the *Zeitschrift für wissenschaftliche Zoologie.* He died at Munich on April 7, 1885.

Siebold was probably the first to discover that the Protozoa

consist of a single independent cell. This important finding emphasized the cell as a basic unit of living matter, and suggested that evolution from one- to many-celled organisms, and also embryonic development, involved a progressive specialization and division of labour among cell units. Siebold also clearly showed that eggs of some insects were normally parthenogenic, developing without fertilization. His studies on the life histories of tapeworms demonstrated to medical men of his day the bearing of basic zoological research on certain clinical problems. His work on the anatomy of invertebrates was one of the first important texts on the comparative anatomy of lower animals. His major works include *Lehrbuch der vergleichenden Anatomie der Wirbellosen Thiere* (1845–48; Eng. trans., W. I. Burnett, *Anatomy of the Invertebrata*, 1854); *Überdie Band-und Blasenwürmer* (1854); *Wahre Parthenogenesis bei Schmetterlingen und Bienen* (1856; Eng. trans., W. S. Dallas, *On a True Parthenogenesis in Moths and Bees*, 1857). (H. H. S.)

SIEDLCE, a town in Warsaw *wojewodztwo* (province), Poland, a district capital, and the seat of a Roman Catholic bishopric, lies 55 mi. (89 km.) E of Warsaw. Pop. (1960) 32,300. As a settlement situated on the Podlasian plain at the meeting place of trade routes, Siedlce obtained town rights *c.* 1557. Its most flourishing period was during the 18th century, when its owners, the Oginski family, transferred there their residence (the present bishop's palace) and church. After the third partition of Poland (1795) Siedlce was occupied by Austria. From 1809 to 1815 it was included in the duchy of Warsaw, and then in the Polish kingdom, dependent on Russia. It was returned to Poland in 1918. During World War II five Nazi concentration camps were situated there. Siedlce is on the Warsaw–Moscow road and railway and is an economic centre of the southeastern part of Warsaw *wojewodztwo*. Industry, only poorly developed, is represented by food and metalworking factories. (T. K. W.)

SIEGBAHN, KARL MANNE GEORG (1886–), Swedish physicist, awarded the 1924 Nobel Prize in Physics for his discoveries and investigations in X-ray spectroscopy, was born in Örobro, on Dec. 3, 1886. He obtained his doctor's degree in 1911 from the University of Lund. He served as assistant to J. R. Rydberg in the physics institute of the university from 1907 to 1915 and in the latter year was appointed deputy professor of physics. On Rydberg's death in 1920 he succeeded to the chair. In 1923 he became professor of physics at Uppsala and in 1937 he accepted the corresponding chair at Stockholm and became director of the Nobel Institute of Experimental Physics of the Swedish Royal Academy of Sciences, Stockholm. His early researches were mainly concerned with problems in electricity and magnetism, but after 1914 he was engaged in the systematic investigation of the X-ray spectra of the different elements. In 1916 he discovered a new spectral series, the long-wave M-lines. With the aid of high-vacuum spectrographs and other apparatus of his own invention he carried the measurement of wavelengths to an extraordinary level of accuracy. In 1924 he and his colleagues succeeded in furnishing proof of the diffraction of X rays in prisms; he later investigated the immediate field between X rays and ultraviolet rays. In the 1960s he was an honorary member of the International Committee on Weights and Measures after more than 25 years service on that body.

See S. Lindroth, *Swedish Men of Science, 1650–1950* (1952). (W. J. Br.)

SIEGBURG, a town of West Germany in the *Land* (state) of North Rhine-Westphalia, Federal Republic of Germany, lies on the Sieg River 7 mi. (11 km.) NE of Bonn by road. Pop. (1961) 33,974. Siegburg grew around the abbey founded in 1064 by St. Anno, Archbishop of Cologne, on the summit of the Michaelsberg, a volcanic hill. The Church of St. Servatius, dating from about 1150, contains an important treasury of Romanesque art. Siegburg, which has a large factory producing synthetic fibres, is skirted eastward by the Cologne–Frankfurt am Main *Autobahn*. (K. Ku.)

SIEGEN (Sechten), **LUDWIG VAN** (1609–*c.* 1680), Dutch engraver, the inventor of mezzotint (*q.v.*), was born in Utrecht. He spent most of his early life in the services of the landgravine Amelia Elizabeth and the landgrave William of Hesse-Kassel. He lived in Amsterdam from 1641 to about 1644 and here it is supposed he was influenced by Rembrandt. Later he served Johann Philipp von Schönborn, elector of Mainz, and the duke of Brunswick while in Wolfenbüttel, where he is mentioned for the last time in 1676. His first mezzotint was a portrait of Amelia Elizabeth, in the dedication of which he claims the invention of the process as one not of lines, but of dots. There are seven known rouletted mezzotint plates by him.

See A. M. Hind, *History of Engraving and Etching, from the 15th Century to the Year 1914* (1923). (H. Es.)

SIEGEN, a town of West Germany in the *Land* (state) of North Rhine-Westphalia, Federal Republic of Germany, lies on the Sieg River 70 mi. (113 km.) ESE of Cologne by road. Pop. (1961) 49,404. The town's two castles were formerly residences of branches of the House of Nassau and the lower (Orange-Nassau) castle has an interesting royal crypt. The late Romanesque Nikolaikirche (13th century) is hexagonal in shape and has a magnificent font. Siegen is the centre of the Siegerland iron ore mining district, and has metalworking and machine-building industries, and colleges of mining and technology. The Flemish painter Rubens was born there.

SIEGFRIED (Sigurd), a figure from the heroic literature of the ancient Teutons, known to us as Sigurd in Old Norse and as Siegfried in German literature, although these two branches of the tradition do not always agree. He plays a part in the story of Brunhild (*q.v.*), in which he meets his death, but in other stories he is the leading character and triumphs. A feature common to all is his outstanding strength and courage. One story tells of his fight with a dragon, and another of how he acquired a treasure from two brothers who quarreled over their inheritance. These two stories are combined into one in the *Edda* (*q.v.*) and told in detail, whereas in German, where they are kept entirely separate, the information is scant and largely contained in allusions. Siegfried plays a major part in the *Nibelungenlied* (*q.v.*), where this old material is used but much overlaid with more recent additions. *Das Lied vom hürnen Seyfrid,* not attested before about 1500, still retains it in identifiable form, although the poem's central theme is the release of a maiden from a dragon; and an *Edda* poem tells how Sigurd awakened a maiden, a valkyrie, from a charmed sleep. Here, too, many critics have tried to establish a connection between German and Norse; not only, however, are there many important differences, but there is great doubt, on internal evidence, about the antiquity of both poems.

In the original stories Siegfried was presented as a boy who, although of noble lineage, grew up bereft of parental care; this shows through clearly, although in the full accounts in both Norse and German it is overlaid with elaborate accounts of his courtly upbringing. As with Brunhild, it is still disputed whether the figure is of mythical or historical (Merovingian) origin.

See *Das Lied vom hürnen Seyfrid,* edited by K. C. King (1958). (K. C. K.)

SIEMENS, a family best known as engineers, inventors and manufacturers in the electrical industry.

Werner von Siemens (1816–1892), the chief founder of the electrical firm of Siemens and Halske, was born on Dec. 13, 1816, at Lenthe, Hanover. Between 1838 and 1848 he held a commission in the artillery, was entrusted with many specialized works and in particular became acquainted with the recently developed electric telegraph. In 1847 he founded, together with a skilled mechanic, J. G. Halske, the firm of Siemens and Halske for the manufacture of telegraphic apparatus. This firm under Siemens' guidance became one of the most important electrical undertakings in the world, with branches in different countries of which those in England and Russia were particularly important. It carried out many large telegraphic projects and later expanded into other electrical fields, as new applications of electricity were developed.

Many of Werner von Siemens' inventions relate to telegraphic apparatus. He used gutta-percha as an insulant for telegraphic cable in 1847; this form of insulation was later widely used for electric light cables. The Siemens armature, which he invented

in 1856 for use in telegraphy, afterward found further application in larger generators for electroplating and lighting and, following the invention of the ring armature, was modified by F. von Hefner-Alteneck, Siemens' designer, into the basic form of the modern armature. One of the most important of Siemens' discoveries was that of the dynamo-electric principle, the principle governing the self-excitation of the dynamo. This idea was also put forward by C. A. Varley and by Sir Charles Wheatstone about the same time, but Siemens appears to have more fully appreciated the possibilities of the invention and was certainly more responsible for their development. He died at Charlottenburg, Berlin, on Dec. 6, 1892.

Sir William Siemens (Karl Wilhelm; 1823–1883), brother of Werner, is known for his work in electricity and in the application of heat. In both fields he combined the functions of innovator, manufacturer and successful man of business. He was born at Lenthe, Hanover, on April 4, 1823. After attending the University of Göttingen he entered, as a pupil, the manufacturing concern of Count Stolberg at Magdeburg. At the age of 19 he first visited England in the hope of introducing an electroplating process invented by himself and Werner, which he succeeded in selling to Messrs. Elkington of Birmingham. He returned to Germany, but in 1844 was again in England, this time with another invention, the "chronometric," or differential, governor. Finding that British patent law afforded the inventor a protection then lacking in Germany, he henceforth made England his home.

The next few years were spent in trying to develop his inventions, of which at this time his water meter was commercially the most successful. His activities made him a respected figure in scientific circles: his paper "On the Conservation of Heat Into Mechanical Effect," read to the Institution of Civil Engineers in 1853, gained him the Telford medal, and in 1862 he was elected a member of the Royal society. William's chief work in the field of heat was concerned with regenerative heating and consequent improvements in steelmaking processes. He invented the regenerative condenser in 1848 and, together with his brother Friedrich, first tried to apply it to the steam engine, using the heat from the regenerator to preheat the boiler feed water. When this did not meet with success, other applications were sought and the idea occurred of applying the principle to furnaces, using the heat regained from the flue gases to heat the air supply to the furnace. This was patented in Friedrich Siemens' name in 1856 and met with great success for use both in glassmaking and in steel manufacture. Later the use of gas instead of solid fuel greatly extended the use of the regenerative furnace.

At that time the quality and reliability of steel was much inferior to that of iron and Siemens did much to develop steelmaking processes, using a large-scale trial plant equipped with regenerative furnaces. This work led to the Siemens-Martin process, shown at the Paris exhibition of 1867, and to the founding of the firm of Landore-Siemens, which was established for the application of the process.

In the field of electricity William, though making no great inventions, became an acknowledged authority and leader. From 1848 onward he represented the firm of Siemens and Halske in London, and when the separate firm of Siemens Brothers was established in 1865 he became a partner and director. At first the chief business was the erection of overland telegraph lines and the laying of submarine telegraph cables. William was, however, in constant close liaison with all the ideas and projects of his brother Werner in Berlin and, when the latter discovered the dynamo-electric principle, William introduced it to England by reading to the Royal society, in Feb. 1867, a paper entitled "On the Conversion of Dynamical Electric Force Without the Aid of Permanent Magnetism." Gradually in the late 1870s and 1880s the electric-light side of the business grew. One of the last projects with which William was associated was the Portrush electric railway in the north of Ireland, opened in 1883, which utilized water turbines driving a Siemens dynamo. The power so produced was transmitted to another machine acting as a motor on the tramcar. William Siemens was knighted in 1883 and died in London on Nov. 19 of the same year.

Alexander Siemens (1847–1928), nephew of William, was born in Hanover on Jan. 22, 1847. In 1867 he went to England, where he worked first in the workshops of Siemens Brothers at Woolwich, and consequently in the erection of the Indo-European telegraph line in Persia (1868) and in the laying of the Black sea cable (1868). In 1878 he became a naturalized British subject. The following year he took over the management of the electric-light department of Siemens Brothers, and was responsible for the installation of electric light at Godalming, Surrey, the first English town to be so lighted. Like many other members of the family, Alexander patented several inventions. After the death of Sir William he became a director of the company, a position he retained until 1918. He took an active part in public activities associated with his profession, was a member of several important committees and was twice president of the Institution of Electrical Engineers. He died at Milford-on-Sea, Hampshire, on Feb. 16, 1928. (M. K. W.)

SIENA, a city and archiepiscopal see of Tuscany, Italy, and capital of the province of Siena, lies 59 mi. (95 km.) S of Florence by rail. Pop. (1961) 64,383. Siena was the site of Etruscan and Roman settlements, and remains essentially a medieval town. Standing on three small hills, its medieval walls and gates enclose brick and stone buildings and narrow climbing streets which twist to reveal the green Tuscan country beyond. The oldest examples of Sienese architecture, including parts of the cathedral which stands on the southwesterly of the three hills, are Romanesque in style; a number of early campaniles show Lombard forms. The predominant type of civic building in this period was the fortified inhabited tower (*casatorre*), such as the Torre dei Forteguerri. To a later, Gothic era belong the enlarged cathedral and its baptistery, a number of churches, and many palaces; much of this building is in a mixture of brick and stone. Probably the oldest of these palaces is the Palazzo Tolomei, built before 1205; others are those of the Buonsignori, Sansedoni, Salimbeni, and, finest of all, the Palazzo Pubblico (1297–1310), the seat of civil government. The lowest part of the Palazzo Pubblico's facade is in stone while the upper part, in brick, contains windows grouped in threes and set in pointed arches, the whole being topped with ornamental battlements; the interior is decorated by the great masters of Sienese painting (*see* below). To one side rises the slender 334 ft. (98.75 m.) bell tower, the Torre del Mangia (1338–48); at the foot of the Palazzo is a chapel (1352–76) begun as a public thank offering after the terrible plague of 1348. The Palazzo and its tower stand along one curved side of the shell-shaped Piazza del Campo, the centre of Siena's civic life set in a hollow among the city's hills.

The cathedral, begun in the 12th century in the Romanesque style, was later transformed into one of the finest examples of the Italian Gothic. Its nave (293 ft. long) was begun in the 13th century and the greater part of the church, built inside and out of black and white marble, was finished, together with the baptistery and the fine campanile, in the following century. It was then decided to turn the nave into the transept of an enormous new cathedral, but that scheme was abandoned in 1355 after the crisis which followed the Black Death, and only a few ruined walls bear witness to it. The splendid west front (1376 onward; with modern restoration) resembles one begun considerably earlier at Orvieto, and is decorated with a multitude of columns and statues. Inside, Niccolò Pisano's octagonal pulpit (1265–68) and the marble floors which include Old Testament scenes designed by Domenico Beccafumi are outstanding. The frescoes in the chapel of St. John Baptist were painted (1504) by Pinturicchio, as were those (1503–08) of the life of the Sienese Pope Pius II in the Piccolomini Library adjoining the cathedral; the library was founded in honour of Pius II by his nephew, Cardinal Francesco Piccolomini, later Pius III. Beneath the cathedral, serving as its crypt, is the Church of S. Giovanni which has a beautiful but incomplete 14th-century facade and a fine outer flight of marble steps (1451); inside is a magnificent font with bas-reliefs by Jacopo della Quercia, Donatello, Lorenzo Ghiberti, and other 15th-century sculptors.

Siena is famed for its fountains: the Fonte Branda, already

BY COURTESY OF THE ITALIAN STATE TOURIST OFFICE—CHICAGO

(LEFT) THE MANGIA TOWER OF THE PALAZZO PUBBLICO AND (CENTRE) THE CAMPANILE OF THE CATHEDRAL RISING ABOVE THE CITY OF SIENA

existing in the 11th century and immortalized by Dante (*Inferno*, xxx, 78), the Fonte Gaia, and the Fonte Nuova. Siena's churches are especially notable for their paintings: for example, the 13th-century brick Church of S. Domenico, with a contemporary portrait of St. Catherine of Siena by Andrea Vanni (*c.* 1332–1414); Sta. Maria dei Servi, with Matteo di Giovanni's "Massacre of the Innocents" (1491); and S. Agostino with works by Simone Martini, Matteo di Giovanni, and others. That Siena's 15th-century architecture was mainly inspired by Florence is shown by the Palazzo Piccolomini, the Palazzo Spannocchi, the Loggia del Papa, and other early Renaissance buildings. Later monuments include the Church of Sta. Maria di Provenzano designed by Schifardini (1594); that of S. Agostino, rebuilt in 1755 by L. Vanvitelli; the theatre built from Bibiena's designs in 1753 (restored 1951); and the Teatro Rozzi erected in 1816 by the Congrega dei Rozzi, a company which dated from the 16th century and had an important place in Italian theatrical history.

Since building activity was largely suspended in the 16th century and because the modern quarters have developed outside the walls, Siena's medieval character remains largely unspoiled and the town now flourishes on the resultant tourism. Visitors come especially to see the Palio delle Contrade, the famous horse races of medieval origin which are held yearly on July 2 and Aug. 16 in the appropriate setting of the Piazza del Campo, amid colourful festivities. Siena is connected by road, rail, and air with Rome, and by road and rail with Florence, Empoli, Orvieto, Grosseto, and Arezzo. Many of the greatest masterpieces of Sienese painting and sculpture are in the Pinacoteca Nazionale, the Palazzo Pubblico, and the Museo dell'Opera del Duomo, which contains Duccio di Buoninsegna's "Maestà." The Museo Etrusco Senese houses archaeological treasures; St. Catherine's rooms are in the Santuario Cateriniano; and the Palazzo Chigi is famous for its collection of musical instruments and its summer school of music, the Accademia Musicale Chigiana. The university dates from the 13th century and maintains faculties of law, medicine, and pharmacy. The city is an administrative and agricultural centre, and manufactures chemicals, fertilizers, and machinery. Its wines and sweet cakes (*panforte*) are famous. Siena escaped serious damage in World War II and was liberated from the Germans in 1944. It has survived as a charming provincial town of great beauty.

Art.—After defeating the Florentines in 1260 the Sienese enjoyed nearly a century of prosperity during which they lavished great wealth upon the creation of a graceful city. The Gothic cathedral was adorned with two masterpieces, the octagonal pulpit carved (1265–68) by Niccolò Pisano and the "Maestà" painted (1308–11) by Duccio to celebrate the 50th anniversary of the victory of 1260, for the Virgin was the protector of the Sienese whose coins were from 1279 inscribed *Sena vetus civitas Virginis.* Niccolò Pisano and his son Giovanni Pisano (*qq.v.*), who also worked in Siena (1284–99), may be considered among the originators of Italian Gothic sculpture. Among their Sienese followers were Camaino di Crescentino, creator of Siena's Fonte Nuova (1298); his son, Tino di Camaino, sculptor of the tomb of the emperor Henry VII in Pisa (1315); Lorenzo Maitani (1275–1330), architect of Orvieto cathedral; and Lando di Pietro, who was entrusted in 1339 with the ambitious but abortive scheme to enlarge Siena's cathedral. Like this cathedral project, the grandiose civic centre was intended to outdo Siena's rivals, especially Florence. The civic centre was placed distinctly apart from the ecclesiastical; the Piazza del Campo and the Palazzo Pubblico were the product of communal pride and patronage, as was their decoration. The theme of Duccio's "Maestà" was repeated in the frescoes in the great council chamber of the Palazzo Pubblico, where the Virgin (1315) by Simone Martini (*q.v.*) faces his equestrian portrait of the conquering Sienese captain, Guidoriccio da Foligno (1328). Frescoes of Good and Bad Government (*c.* 1339) by Ambrogio Lorenzetti (*q.v.*) are in the Sala della Pace of the same building. Later, Jacopo della Quercia carved one of the finest fountains of his time, the Fonte Gaia (1419) the remains of which are in the Palazzo Pubblico; a reproduction (1868) stands in the Piazza del Campo.

Until the time of Duccio di Buoninsegna (*q.v.*) Sienese painting was a flourishing but provincial school working in the Romanesque tradition. Duccio produced beautiful and tender Byzantinesque Madonnas, often with the child pulling playfully at his mother's headdress. He began the transformation of the Sienese school from a hierarchic Byzantine stiffness to a looser, flowing, linear style. Simone Martini painted in a more vivacious, poetic, and original manner, his hands and draperies more freely drawn. Simone, like his brilliant contemporaries the brothers Pietro and Ambrogio Lorenzetti, painted what came to be the typical Sienese Madonnas. Siena played an important part in the development of Renaissance painting, but the death of Simone Martini and the Lorenzetti brothers coincided with Siena's economic decline and the plague of 1348. Thereafter Sienese painting took the path not of the rationalism and scientific inquiry connected with the Renaissance in Florence but of the mysticism of Siena's two great saints, Catherine and Bernardino (*qq.v.*). Siena's painters maintained their somewhat conservative Gothic style and their colourful linear rhythms, achieving an elegant, spiritual beauty. Perhaps only the lyrical designs, reminiscent of Ambrogio Lorenzetti, produced by Stefano di Giovanni, known as Il Sassetta, and the strange angular landscapes of Giovanni di Paolo compare with earlier Sienese masterpieces. Taddeo di Bartolo (*c.* 1393–1422), who executed new frescoes for the Palazzo Pubblico, Domenico di Bartolo (*c.* 1400–47), Sano di Pietro (1406–81), and others produced work of lasting beauty, but the school was in decline during the 15th century.

Sienese sculpture was carried to its height by Jacopo della Quercia. Among those who followed him were three painter-sculptors, all to some extent receptive to Florentine influences: his pupil Lorenzo Vecchietta (*c.* 1412–80); Neroccio di Landi (1447–1500); and Francesco di Giorgio (*q.v.*). Sienese painting required new life and inspiration, and these were supplied to some degree by Domenico Beccafumi (*q.v.*) and Giovanni Antonio de' Bazzi, known as Sodoma, who arrived in Siena in 1501 and worked mainly there until his death in 1549. For the rest, the 16th century saw the disappearance not only of Siena's political independence but also of its artistic originality.

SIENA PROVINCE (area 1,475 sq.mi. [3,820 sq.km.]; pop. [1961]

272,111) lies in the central Tuscan hills with the Chianti Hills to the east. The province is drained mainly by the Ombrone River and its tributaries. The beautiful countryside is fertile and well wooded in places, with considerable areas of rather desolate landscape, bleaker than other parts of Tuscany. Though there has been a small development of chemical and light industries, the province is primarily an agricultural area centred on the market town of Siena. About two-thirds of the population is engaged in agricultural activities: cattle raising, the cultivation of grain and olives, and above all the production of wine, Chianti being perhaps the best-known of Italian wines. The population, always sparse, has recently been drifting away from the depressed areas of the countryside. There are no other real towns, and only a few of the 36 communes have over 10,000 inhabitants; the most notable centres are Poggibonsi, Montepulciano, and San Gimignano (*q.v.*), famous for its many towers.

History.—An Etruscan site and a Roman colony, Siena lay at the point where the road from Rome to the north, the Via Francigena, met a route running eastward toward Arezzo and the Adriatic ports. The Roman settlement, known as Sena Julia, was largely destroyed, and Siena grew again, perhaps as a defensible point of refuge from invading barbarians during the disintegration of the Roman Empire. By the 7th century Siena had a bishop, and under the Lombard kings was ruled by royal representatives, *gastaldi*, whose title changed in Carolingian times to that of count. The Lombard ruling class apparently fused with those whom it ruled, and the counts and other feudal lords achieved independence of the emperor. This local aristocracy seized power in the surrounding countryside (*contado*) but had to share its predominance with the bishops who, during the 11th century, made themselves independent, at least within the city.

Siena's fertile *contado* produced oil, wine, and grain. The *mezzadria* system, by which the landowner provided the peasant with tools and seed, receiving half the produce in return, was being practised near Siena as early as 821. This was one way of financing the process of land reclamation which, especially through drainage schemes, increased Tuscany's agricultural productivity and helped to make possible the growth of urban life. The townsfolk increased their prosperity and power, securing the highways on which their trade depended by subduing the lords and castles of the *contado*. Early in the 12th century they replaced the rule of the counts and bishops with that of their own free commune and consuls (*see* COMMUNE [MEDIEVAL]). This amounted to rebellion against the emperor. In 1186 the emperor Frederick I's son, the future emperor Henry VI (1191–97), successfully besieged Siena; he then confirmed Siena's position as an imperial vassal with rights of self-government through its consuls, but freed the smaller towns and nobles of the *contado* from dependence upon Siena. When Henry VI was succeeded by an infant son in 1197, Tuscany revolted and Siena reasserted control over the *contado*, so becoming involved once more in the long-standing and primarily economic conflict with Florence. Florence was anti-imperial, or Guelph; Siena became the centre of Tuscan Ghibellinism, but imperial support did not bring victory against Florence.

In the 13th century Sienese merchant companies, favourably placed on international routes, were trading in London and Bruges and at the fairs of Champagne, where their agents acquired Flemish cloth. Siena became an important centre for banking and was especially active in transferring ecclesiastical incomes for the popes, who were often resident at nearby Viterbo. High interest rates and profits brought great fortunes; the Magna Tabula of the Buonsignori was the greatest merchant-banker company of the time. During the 13th century Siena fought to control Grosseto in the unfounded hope that a port could be opened up nearby at Talamone, and in the 14th century the city declined in the face of competition from Florence which was in a better position geographically and demographically to develop its industry and trade.

Commercial prosperity brought social and political unrest and the emergence of an urban opposition to the ruling oligarchy. From 1199 the consuls were replaced by a *podestà*, usually from another town, who governed for one year, while an oligarchic consular class of nobles, the *magnati* or *grandi*, who lived and traded in the town and possessed feudal estates outside it, continued to exercise real power. A revolt of the people (*il popolo*) followed a Sienese defeat by the Florentines in 1235, and in 1240 the oligarchy of the *grandi* gave way to a council of 24 (12 *grandi* and 12 *popolani*) which ruled for about 30 years. The death of the emperor Frederick II in 1250 was followed, in 1253, by an uprising which increased the strength of the *popolo* through a modification of the powers of the *podestà* and the creation of a *capitano del popolo*. Siena reached the height of its political success when on Sept. 4, 1260, an army, which included forces provided by Frederick II's illegitimate son Manfred, crushed the Florentines at Montaperti.

Siena's triumph was short-lived. Commercial prosperity suffered when the pope ordered the sequestration of the goods of the church's Ghibelline enemies, and hard-hit Sienese merchants, who thus lost their goods and credits, began to go into exile; some even financed the Guelph leader, Charles of Anjou, who defeated the last imperialists, Manfred and Conradin, at Benevento (1266) and Tagliacozzo (1268) respectively. Now Florence was triumphant, and in Siena there was party strife as *magnati* and merchants gradually abandoned the Ghibelline cause. Siena became Guelph after the Ghibelline leader, Provenzano Salvani, was defeated in 1269. In 1277 the *magnati*, such as the Tolomei, Salimbeni, and Piccolomini, turbulent and still in part Ghibelline, were excluded from government office. For nearly 80 years a merchant oligarchy, Guelph and closely allied to Florence, maintained Siena's prosperity and embellished the city. The republic was ruled by a council with a membership fixed in 1287 at nine, the *nove*. Those eligible for office, the *noveschi*, formed a genuine oligarchy; not only nobles, but lawyers, doctors, artisans, and others were excluded. The *noveschi*, many of them risen out of the *popolani*, were often middle-class or "new" men, others were rich or socially well-established; some owned both town and country property, and some were connected by marriage or interest to the *magnati*. Distinctions were seldom clear-cut, for *magnati* had always participated in the merchant-banking companies and grown rich in trade, just as many merchants used their wealth to purchase country estates.

The oligarchic rule of the *nove* provided experience and continuity in government but involved endless conspiracy on the part of those excluded. Affairs largely beyond the commune's control produced discontent: late in the 13th century Florence defeated the Pisans and its economy boomed; early in the 14th century Italy experienced an overall economic decline which was reflected in Siena; Tuscany was visited by wars, famines, and in 1348 by the Black Death. In 1355 the *popolo* allied with the *magnati* and overthrew the *noveschi* regime, establishing a popular government of 12, the *dodici popolani*, representing a class rather inferior to the *noveschi*, who were now excluded from government. Siena entered a disturbed era of faction and revolt. In 1368 the *magnati* party, the *monte dei gentiluomini*, ejected the *dodici popolani* but were themselves overthrown and replaced by successive coalitions of the *dodicini*, the *noveschi*, and a new party, the *monte dei riformatori*. The *riformatori*, representing lower middle-class or artisan but not proletarian interests, governed through a further period of violence until upset in 1385 by a party representing even lower social strata, the *monte del popolo*. Henceforth, Sienese politics consisted of successive predominantly middle-class coalitions; for decades there was discord but no major revolution. During the period after 1387, Siena survived a new war with Florence only by accepting the despotic rule of Gian Galeazzo Visconti, Lord of Milan. But from 1409, when King Ladislas of Naples advanced into Tuscany, Siena and Florence formed a republican alliance which excluded both the Milanese and Neapolitan despots from Tuscany, and lasted into the 16th century. War and civil strife continued with brief interludes, for Siena was never sufficiently industrialized to prevent the steady enfeeblement of its commerce and banking, or to destroy the power of the nobles who brought turbulence to city politics and wrought havoc in the *contado*, where they reasserted their independence. Losing its place in the socioeconomic pattern of northern Italy, Siena reverted to a predominantly agricultural economy characteristic of southern Italy, and seems to have turned to religion; this was

the period of Siena's mystic saints, Catherine and Bernardino.

Politically, the urban ruling class was slowly transformed into a new aristocracy based on political privilege and increasing investments in land. After 1385, the *gentiluomini* and *dodicini* were mainly excluded from government, while entrance to the *monti* and thus to the full citizenship of the governing class was increasingly restricted. From this sterile situation, in which social mobility was reduced to a minimum and the parties ceased to reflect economic differences and became like hereditary castes, there finally emerged a despot. In 1487 an exiled member of the *noveschi* party, Pandolfo Petrucci, succeeded in driving out the government. Power in Siena was now vested in a *balìa*, or committee, in which the Petrucci were predominant, a constitutional arrangement cloaking Pandolfo's tyranny, a rule secured by a policy of lies, banishments, and assassinations. From 1494 Italy was a battlefield for French and Spanish armies; in a difficult situation Pandolfo managed to preserve his control until his death in 1512 and his regime was continued by his family until 1524.

Siena was now unable to preserve its independence. In 1526 an army sent by the Florentine pope, Clement VII, was defeated outside the walls but Siena had to seek protection from Clement's enemy, the emperor Charles V. In 1530 Charles sent a Spanish garrison to defend Siena, but protection was transformed into domination and, after a number of insurrections, the Sienese succeeded, with French help, in expelling the Spaniards in 1552. Siena was of strategic importance to both French and Spaniards, and in 1554 the Spaniards, supported by the Florentines, began a long siege which lasted until April 1555, when Siena's French garrison surrendered; the town's population had been reduced from about 40,000 to 8,000. Many surviving Sienese went into exile, and in 1557 Philip II of Spain ceded Siena to Cosimo I de' Medici, ruler of Florence. For two centuries Siena retained a separate administration but its history increasingly merged with that of Florence, Tuscany, and finally Italy.

See also references under "Siena" in the Index.

BIBLIOGRAPHY.—L. Douglas, *A History of Siena* (1902); W. Heywood and L. Olcott, *Guide to Siena* (1903); F. Schevill, *Siena: the Story of a Mediaeval Commune* (new ed. 1964); E. Sandberg Vavalà, *Sienese Studies: the Development of the School of Painting of Siena* (1953); E. Hutton, *Siena and Southern Tuscany* (1955); I. Origo, *The World of San Bernardino* (1962). (A. T. L.)

SIENKIEWICZ, HENRYK (pseudonym LITWOS) (1846–1916), Polish author of novels which achieved great popularity, and winner of the Nobel Prize for Literature in 1905, was born on May 5, 1846, at Wola Okrzejska. He was at school in Warsaw, and studied literature, history, and philology at the university there, but left, without taking a degree, in 1871. He had begun to publish critical articles showing "positivist" influence in 1869. His first novel, *Na marne* (trans. as *In Vain*, 1889), appeared in 1872, and his first short story, *Stary sługa* (*An Old Retainer*), in 1875. Much of his early work treats social problems in the positivist manner. He traveled in the United States (1876–78) as special correspondent of the *Gazeta polska*, and after his return to Poland, via Italy and France, published successful short stories, among them *Janko Muzykant* (1879; *Yanko the Musician*), *Latarnik* (*The Lighthouse-Keeper*), and *Bartek Zwycięzca* (*Bartek the Conqueror;* both 1882). From 1882 to 1887 he was co-editor of the daily *Słowo*. In 1891 he made a hunting trip to East Africa, described in *Listy z Afriki* (1892).

In 1900, to celebrate the 30th year of his career as a writer, he was presented by the Polish people with the small estate of Oblegorek, near Kielce, where he lived until 1914. Revered in Poland, especially after the award of the Nobel Prize "for his outstanding contribution to the prose epic," and well known abroad, he assumed the role of unofficial spokesman for his countrymen on political issues. In World War I he promoted the cause of Polish independence and organized relief for Polish war victims. He died at Vevey, Switzerland, on Nov. 15, 1916.

Sienkiewicz' great historical trilogy, which began to appear in *Słowo* in 1883, was written "to fortify the hearts of his countrymen." It comprises *Ogniem i mieczem* (four volumes, 1884; *With Fire and Sword*, 1885); *Potop* (six volumes, 1886; *The Deluge*, two volumes, 1895); and *Pan Wołodyjowski* (three volumes, 1887–88; *Pan Michael*, 1895). Set in the later 17th century, it describes Poland's struggles against Cossacks, Tatars, Swedes, and Turks, stressing Polish heroism in a vivid style of epic clarity and simplicity. Although historians deplore its lack of historical accuracy, literary critics praise its narrative power, even when they condemn its theatricality. Stylistically it is of great interest for its artistic re-creation of 17th-century language.

The most famous of Sienkiewicz' other novels is the widely translated *Quo Vadis* (three volumes, 1896; English translation 1898), a historical novel of Rome under Nero. *Bez dogmatu* (three volumes, 1891; *Without Dogma*, 1893) is a penetrating, though overlong, psychological novel; *Rodzina Połanieckich* (1893–94; three volumes, 1895; *Children of the Soil*, 1895) is a didactic *roman de moeurs*. Other works include the medieval historical novel *Krzyżacy* (four volumes, 1900; *The Teutonic Knights*, 1943) and *W pustyni i w puszczy* (1911; *In Desert and Wilderness*, 1912; reprinted 1945), a story for children.

BIBLIOGRAPHY.—The definitive edition of Sienkiewicz' works is *Pisma*, by J. Krzyżanowski, with biographical material and bibliography, 60 vol. (1948–60). English translations include *Yanko the Musician and Other Stories* (1893); *Tales from H. Sienkiewicz* (1931); *Portrait of America: Letters of H. Sienkiewicz*, ed. and trans. by C. Morley (1959). *See* also K. Czachowski, *H. Sienkiewicz, Obraz twórczości* (1931); J. Birkenmajer, "H. Sienkiewicz," in *Thought*, XIV (1939); W. Lednicki, *Henryk Sienkiewicz* (1960). (L. R. LR.)

SIERRA LEONE, an independent sovereign state in West Africa and a member of the Commonwealth of Nations, is bounded north and east by the Republic of Guinea, south by Liberia, and west by the Atlantic Ocean. Area 27,925 sq.mi. (72,326 sq.km.). Before achieving independence in 1961, Sierra Leone was divided into two parts, a British colony and protectorate. The area of the former colony is wholly coastal, consisting of the Sierra Leone Peninsula, at the northern end of which stands the capital and chief port, Freetown; Sherbro and smaller islands, with the town of Bonthe; and Turner's Peninsula. The former protectorate included the headwaters of the Niger River, which flows northeast, away from the Atlantic, eventually reaching the Gulf of Guinea in Nigeria.

PHYSICAL GEOGRAPHY

Relief and Drainage.—The mountains of the Sierra Leone Peninsula, consisting of igneous (norite) rocks, run parallel to the sea for about 25 mi. (40 km.) with a maximum width of 12 mi. (19 km.) and reach 2,912 ft. (888 m.) in Picket Hill. They are well watered, thickly wooded, and much dissected by deeply cut ravines. The name Sierra Leone derives from that given to the range by the Portuguese explorer Pedro da Cintra in 1462. Along the Atlantic coast there extends a flat, low-lying, and frequently flooded coastal plain, 20–40 mi. (32–64 km.) wide and composed of sands and clays. The numerous estuaries are fringed below high-tide mark by extensive mangrove swamps. In the south the plain gives way inland to rolling wooded country with isolated hills rising to more than 1,000 ft. (300 m.). In the north the plain is backed by an upland plateau reaching to 1,500 ft. or more. The plateau is principally composed of granite, but to the west it is bounded by a narrow outcrop of metamorphic rocks, the Kambui schists, in which valuable metal ores, including gold, iron, and chromite occur. Above the plateau surface rise several mountain ranges, reaching 6,390 ft. (1,948 m.) in Bintimane (Loma Mountains) and 6,080 ft. (1,853 m.) in Sankan Biriwa (Tingi Hills).

Numerous rivers rise in the well-watered Fouta Djallon Plateau and flow down its steep slopes southwest to the Atlantic Ocean. Their middle courses are interrupted by rapids, but they are navigable in parts and those in the south are important for the commercial movement of goods. From north to south the principal rivers are: the Great Scarcies (or Kolenté), Sierra Leone's northwestern boundary with the Republic of Guinea; Little Scarcies (Kaba); Rokel (Seli), known in its lower course as the Sierra Leone River, with the large, sheltered harbour of Freetown at its mouth; Jong (Teye); Sewa (Bum); Waanje; Moa; and Mano (Moro), forming the country's boundary with Liberia.

Climate.—Conditions are generally hot, with average tempera-

tures of 24°–29° C (75°–85° F) on the coast and between 20.5° and 35° C (69° and 95° F) in the interior, and, during the rains (May–October), very wet. During the dry season (November–April) the harmattan wind from the northeast frequently blows, bringing with it fine Saharan dust, poor visibility, and very low relative humidity. The rainy season is introduced by a series of squalls with thunder and lightning.

Relative humidity during the rains may be as high as 90% for considerable periods, particularly during the wettest months, July to September. Annual average rainfall totals range from more than 150 in. (3,810 mm.) in the peninsula mountains to 80 in. (2,032 mm.) in the north. Freetown (128 in. [3,251 mm.]) recorded 13.40 in. (340 mm.) in one day in July 1948 and 5.91 in. (152 mm.) in an hour in September 1944.

Sierra Leone was formerly known as "the white man's grave," but during the 20th century living and health conditions greatly improved.

PRINCIPAL PHYSICAL FEATURES, TOWNS, AND RAILWAYS OF SIERRA LEONE

Vegetation.—Forest probably covered most of Sierra Leone in the past, but has been greatly reduced in area, particularly through felling for cultivation. The wetter areas are still forest-covered, especially where forest reserves have been constituted, as on the hills of the peninsula and in the Gola Forest near the Liberian border, but savanna woodland (sometimes known as orchard bush) and grassland are increasingly common. Valuable timber trees exploited include species of *Khaya* (African mahogany) and African teak (*Oldfieldia africana*).

In the interior there is much secondary forest, where trees have at some past time been felled to permit cultivation; in these areas the African oil palm (*Elaeis guineensis*) is often common and has great economic value as a source of palm oil and palm kernels. Where the annual rainfall decreases, savanna grassland or woodland with fire-tolerant trees, such as lophira, become increasingly common. Grassland has developed throughout the country where the soils are too thin or the slopes too steep for tree growth and is particularly characteristic on lateritic patches. Swamps, which flood seasonally, abound along the beds of streams, and mangroves flourish in the saline tidal areas of river estuaries, though they are being drained and cleared in places for the cultivation of swamp rice. The swamps of the southern coastlands are important producers of the fibre piassava, used in the manufacture of strong brooms and obtained from the swamp palms *Raphia vinifera* and *R. gaertneri*.

Animal Life.—The relatively few large game animals are rarely seen because of the density of the vegetation; they include elephant, leopard, bush cow or dwarf buffalo (*Syncerus nanus*), and antelope (including duiker, the harnessed antelope, and, in the Gola Forest, the bongo). Tiger cats, civets, chimpanzees, various species of monkey, porcupines, bush pigs, and cane rats (also called "cutting grass") are common. Some of these are responsible for much damage to growing crops. During the dry season lions occasionally come in from Guinea in search of game. Birds and insects abound, the latter including mosquitoes, tsetse flies, termites, locusts, and sand flies. Hippopotamuses, both normal and pygmy, crocodiles, and manatees are common in the rivers, and the estuaries are often shark-infested. Fish, caught both in the rivers and at sea, include tarpon, barracuda, mullet, and "bonga" (a species of shad). The coastal waters off Sierra Leone constitute an excellent fishing ground, which is as yet not fully exploited. (R. W. SL.)

THE PEOPLE

The Mende (*q.v.*), Loko, Kono, Vai (*q.v.*), Koranko, Susu (*q.v.*), and Yalunka peoples belong to the Mande linguistic group (*see* MANDINGO), and the Temne (*q.v.*), Limba, and Bulom to the semi-Bantu group. The tribes have many cultural features in common, from long association with the Muslim Fulani and Mandingo, the dominating influence of the two largest tribes (Mende and Temne), and contact with Europeans. The Mende in the centre and south, and the Temne in the north, divide into 60 and 44 chiefdoms respectively, each ruled by a paramount chief and a council of section chiefs. The main economic activity is agriculture; in most areas rice is the staple crop, but the Bulom grow cassava and fishing is important among them. Descent is normally patrilineal, although there are close links with certain maternal relatives. The Temne divide into 25 patriclans. The Poro and Sande secret societies (for men and women respectively) sanction behaviour in most aspects of life; the Temne have the Ragbenle Society in addition. The population in the former colony area also includes Creoles, the descendants of repatriated slaves mainly from Nova Scotia and the West Indies.

The only lingua franca is a form of pidgin English which is

fairly widespread, though by no means universal. There are diverse animist beliefs, rites, and practices among tribes and families. Islam is followed in parts of Sierra Leone, and Christianity is well established. Freetown is the seat of the Roman Catholic bishop of Freetown and Bo, and of the Anglican archbishop of West Africa. (ME. F.)

Population and Towns.—According to the census of April 1963 the population was 2,180,355. Freetown (*q.v.*) is the capital and chief port, but iron ore is shipped from Pepel, a short distance up the Sierra Leone River. The other chief towns are Bo (the former capital of the protectorate) in Southern Province, the diamond-mining centre of Kenema in Eastern Province, and the trading centre of Makeni in Northern Province. Lunsar, near the Marampa iron mines, grew rapidly in the 1950s. Other important settlements are Waterloo, a Creole town on the peninsula; Kabala in Northern Province; and Kambia in the Scarcies River rice district. (R. L.)

HISTORY

Sierra Leone was originally divided into many small independent kingdoms or chiefdoms; each had its own ruler whose power was checked by his council of subchiefs. In many areas there were also secret societies, of which the Poro Society is the best known, which maintained law and order as well as instructing initiates in the traditions and customs of the country (*see* SECRET SOCIETIES, PRIMITIVE). The Bulom people have been settled immemorially on the coast. The Temne, by tradition migrants from the north, were well established on the coast by the 15th century. The Mende reached it only by slow migration in the 19th century.

Portuguese voyagers gave the name Serra Lyoa ("Lion Mountain"), later corrupted to Sierra Leone, to the mountainous peninsula at the mouth of the Rokel River. From the late 15th century European ships of all nationalities put in regularly, near the site where Freetown now stands, to take on water and firewood and to trade manufactured goods for slaves and ivory. Though English trading posts were built on Bunce and York islands in the 17th century, no European power exercised jurisdiction in Sierra Leone. Traders settled there under the protection of the African rulers, who welcomed them for the goods they brought. In the early 18th century the Fulani and Mandingo Muslim peoples in Fouta Djallon (Futa Jalon), north of Sierra Leone (later in Guinea), started a holy war of conversion. From Fouta Djallon, Islam spread gradually to the coast. By the end of the 19th century it was firmly established in northern Sierra Leone and in the 20th century began to spread among the Mende.

A group of freed slaves of African birth or origin arrived in Sierra Leone from England to form a settlement in 1787. Its sponsor, the English slave abolitionist Granville Sharp, called it "the Province of Freedom" and hoped it would become a base against the slave trade. King Tom, a Temne subchief, gave the colonists a strip of land, but his successor, King Jimmy, drove them away in 1789. The settlement was revived in 1791 by the Sierra Leone Company, a trading company sponsored by opponents of the slave trade, with headquarters in London. The town was rebuilt and named Freetown. The company brought from Nova Scotia, as settlers, some former slaves who had gained their freedom by serving the British in the American Revolutionary War. They were joined in 1800 by "Maroons," free Negroes from the mountains of Jamaica, who had been deported to Nova Scotia for insurrection. These settlers, African in origin, were English-speaking and many were literate and Christian.

After the British Parliament made the slave trade illegal in 1807, the British government took over the settlement (Jan. 1, 1808) as a naval base against the slave trade and as a centre to which slaves, captured in transit across the Atlantic, could be brought and freed. Between 1807 and 1864, when the last slave ship case was adjudicated in the Freetown courts, the British Navy brought in more than 50,000 "recaptives." Drawn from all over West Africa, these heterogeneous people lacked any common language or culture. Inspired by Sir Charles MacCarthy, governor from 1814 to 1824, the government undertook a deliberate policy of turning them into a homogeneous, Christian community. Missionaries of the (Anglican) Church Missionary Society (CMS) and the Methodist Missionary Society, and the pastors of the Freetown settler churches, worked among them with such success that within a generation the policy was virtually fulfilled. The CMS, as well as opening boys' and girls' secondary schools, founded an institution at Fourah Bay, near Freetown, to train teachers and missionaries (see *Education, Welfare, and Defense* below).

The recaptives and their children (known as Creoles) prospered as traders, opening stores or bartering imported European goods in the neighbourhood for exportable palm produce. Many left the colony to trade along the coast or to work there as clerks, teachers, or missionaries. At their suggestion, English missions were started in the Yoruba country (later part of Nigeria), the homeland of many recaptives. Thus they formed an educated West African elite, bringing to their homeland the new ways they had learned.

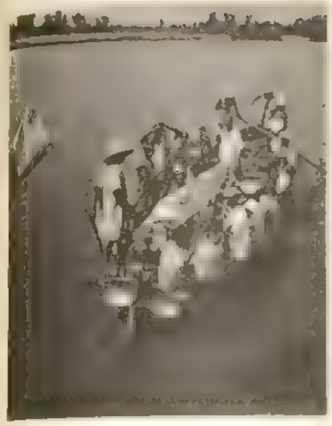
The most famous recaptive was Samuel Adjai Crowther, who became an Anglican priest in 1843 and bishop in the Niger territories in 1864. Among distinguished Creoles were Africanus Horton and William Davies, who qualified in Britain in 1859 as doctors and served as officers in the British Army; and Samuel Lewis, a barrister, who served many years on the colony's Legislative Council and was knighted in 1896.

Colony and Protectorate.—The colony made treaties of friendship with most of the neighbouring chiefs and gradually acquired jurisdiction over the adjoining coastline. The Creoles wanted to extend the colony inland, but the British government was unwilling to accept new West African responsibilities. By 1890, however, it was realized in London that the French, rapidly advancing inland, would soon hem the colony into a tiny enclave and destroy Freetown's value as a naval base. A more expansive policy was sanctioned, frontiers were delimited with the French and Liberian governments, and a British protectorate was proclaimed in 1896 over the area within the frontier lines. The British government made no contribution toward governing the new protectorate so Sir Frederic Cardew, governor from 1894 to 1900, introduced a hut tax to raise extra revenue to pay for the enlarged administration. The chiefs, who had not been consulted about the protectorate, objected. A revolt broke out in the north in 1898 under an experienced war chief, Bai Bureh. It spread among the Mende but was suppressed by the end of the year. There was no further armed rising against the British.

The protectorate was governed on the principles of indirect rule, afterward introduced by Frederick (later Lord) Lugard into Nigeria. The chiefs retained much of their power, under the supervision of British district commissioners. Traditional ways were encouraged, and for the first decades of the 20th century little was done to extend education in the protectorate. During the 19th century many Creoles held senior official posts and looked forward to governing themselves ultimately. But after the protectorate was assumed they were gradually removed from office and both colony and protectorate were ruled by British administrators. A new constitution in 1924 allowed a few Creoles to be elected, and protectorate chiefs nominated, to the Legislative Council, but this did not satisfy nationalist aspirations.

Independence.—After World War II British policy changed in Sierra Leone, as elsewhere in West Africa, it was agreed to constitute democratic institutions through which the dependent territories could evolve into independent states. The small educated minority of Creoles hoped to entrench themselves politically, lest under a fully democratic constitution they should be overwhelmed by the protectorate peoples. But the 1951 constitution gave power to the majority. The government elected under it was led by Milton (later Sir Milton) Margai, a physician and leader of the Sierra Leone People's Party (SLPP), a predominantly protectorate party.

During the 1950s parliamentary institutions on the British pattern were introduced in successive stages. In 1956 the Legislative Council was renamed the House of Representatives; in 1957 elections were held with direct voting, women being eligible to vote; in 1958 the ex officio members left the legislature, and in 1960 the governor gave up his reserved powers. The last stage was reached on April 27, 1961, when Sierra Leone became an indepen-



(Above) Rice harvesters of the Susu tribe on their way to work on the Great Scarcies River; (top right) the University College of Sierra Leone near Freetown; (right) workers wash gravel for diamonds near Kenema

(TOP LEFT) CAMERA PRESS; (TOP RIGHT) PAUL CONKLIN—PIX; (BOTTOM RIGHT) A. F. KERSTING

Hospital beds in establishments conducted by the government, missions, and mining companies exceeded 1,500, and there were in addition about 90 government dispensaries and health centres, as well as treatment centres attached to the Endemic Diseases Control Unit.

Wages and conditions of employment are regulated by joint industrial councils and wages boards. By the mid-1960s there were a number of registered labour unions, including mineworkers, railway workers, and maritime and waterfront workers.

The Royal Sierra Leone Military Force comprises one battalion, supporting services, and a volunteer naval force.

dent state within the Commonwealth. The governor, Sir Maurice Dorman, elevated to the position of governor general, was succeeded in 1962 by a Sierra Leonean, Sir Henry Lightfoot Boston.

Though the SLPP formed an alliance in 1960 with some of the rival political groups, a vigorous opposition party, the All People's Congress, emerged, and during the early years of independence Sierra Leone remained a two-party state. In 1964 Sir Milton Margai died and was succeeded as prime minister by his brother, formerly the finance minister, Sir Albert Margai. (C. Fy.)

ADMINISTRATION AND SOCIAL CONDITIONS

Administration.—The constitution provides for a governor general appointed by the British sovereign on the advice of the Sierra Leone prime minister; for a House of Representatives of not fewer than 60 members and a normal life of five years; for universal adult suffrage; and for executive responsibility by a cabinet of ministers appointed by the prime minister and presided over by him. There are a Supreme Court and a Court of Appeal, with further appeal to the Judicial Committee of the British Privy Council in certain cases. The chief justice is appointed by the governor general on the advice of the prime minister; other judges are appointed on the advice of the Judicial Service Commission. The constitution includes the customary provisions governing citizenship and safeguarding individual rights and freedoms.

Outside the Freetown Peninsula (or Western Area) the country is divided for purposes of local administration into the Southern, Eastern, and Northern provinces, which are subdivided into districts. Sierra Leone's 146 chiefdoms form the basic unit of local government; in each chiefdom the tribal authorities elect a paramount chief who sits in the District Council with the elected members of the council.

Education, Welfare, and Defense.—The chief institution of higher education is the University College of Sierra Leone, a reconstitution (1960) of Fourah Bay College, founded at Leicester near Freetown by the Church Missionary Society in 1827. In 1876 it was affiliated to Durham University, in England, and in the 1950s moved to Mt. Aureol overlooking Freetown. It has about 550 students. Another university college, specializing in agricultural education, was opened at Njala in 1965. Illiteracy is still widespread, but by the mid-1960s there were more than 800 primary schools with about 135,000 pupils; more than 50 secondary schools (with about 13,500 pupils); 6 teacher-training colleges, and 2 technical institutes (Freetown and Kenema).

THE ECONOMY

Although since the 1930s minerals have been of increasing importance, Sierra Leone still has a subsistence economy based on rice growing on the upland slopes or in the swamps and estuaries.

Agriculture.—Agriculture occupies about 80% of the population. Besides rice, the main food crops are palm oil and cassava. The chief export crops are palm kernels (about 60,000 tons annually), cocoa (4,000 tons), coffee of the species *Coffea robusta* (4,500 tons), and piassava brush fibre (5,000 tons). Cultivation, transport, and marketing are assisted by cooperatives and by marketing and credit societies. Market gardening is carried on around Freetown and in Northern Province around Kabala in the Koinadugu District. Koinadugu District is also the centre of the livestock industry.

Mining.—Iron ore (about 2,000,000 tons annually) is extracted from strip mines at Marampa; deposits of rutile and bauxite exist near the coast in Southern Province. Diamonds, the country's most valuable product, were first found in river gravels of the Kono District, about 140 mi. (225 km.) E of Freetown. Subsequent prospecting revealed diamond-bearing deposits over much of the Bafi-Sewa River system. In 1935 a British company obtained a 99-year concession covering the whole country. Following World War II, local poverty combined with boom prices for diamonds encouraged illicit mining, which was carried on not only by the inhabitants but also by thousands of immigrants (largely from French Guinea) and organized by the Lebanese community. The company's monopoly was rescinded in 1955 except for 450 sq.mi. (1,165 sq.km.) around Yengema in Eastern Province. The government thereafter opened up the surrendered area to licensed Sierra Leonean diggers and in 1959 established its own buying office. By these measures, illegal exports of diamonds were materially reduced. The Sierra Leone deposits lie in gravel layers up to four feet thick beneath a shallow alluvial covering and half the diamonds they yield are gem stones. The concessionary company produces between one-quarter and one-third of the total output; during the dry season about 3,300 licensed diggers are at work. The value of diamond exports in 1964 was nearly £20,000,000.

Industry and Power.—The main industries are concerned with processing agricultural and forest products (oil mills, rice mills, sawmills). Furniture, cigarettes, soft drinks, nails, and oxygen and other compressed gases are produced locally. Further industrial development depends chiefly on power supplies. By the mid-1960s the generating capacity in Freetown was about 10,000 kw., and there were about 20 small power stations in the provinces. Fishing is important, and village industries include fish curing and smoking, hand-expressing of palm oil, and cracking palm kernels.

Trade and Finance.—Sierra Leonean overseas trade expanded substantially in the 1950s. More than 50% of imports come from other Commonwealth countries (the bulk from the U.K.); the main imports are foodstuffs, cotton fabrics, clothing and footwear, machinery, motor vehicles, petroleum products, and tobacco. Minerals account for about 70% of total exports by value, agricultural products for about 16%; most of the exports go to the U.K.

The unit of currency is the leone (1 leone = 10*s*. sterling = $1.40 U.S.), divided into 100 cents, which is replacing the West African shilling.

Transport and Communications.—A state-owned railway of 2 ft. 6 in. gauge runs east from Freetown to Pendembu (227.5 mi. [366 km.]) near the Liberian frontier. A branch line from Bauya runs northeast to Makeni. The railway has many difficult gradients and limited carrying capacity and has never covered its capital costs. An iron-ore railway of 3 ft. 6 in. gauge joins Marampa with Pepel. Inland waterways total about 500 mi. (800 km.) and there is a considerable volume of river traffic, carried by launches, including coastwise routes from Freetown southward to Bonthe and northward on the Great and Little Scarcies rivers. Freetown, the finest natural harbour in Africa, is a regular port of call for numerous shipping lines and can accommodate large ships alongside the Queen Elizabeth II Wharf, opened in 1954. The international airport of Lungi is on the north bank of the Sierra Leone River opposite Freetown, but domestic air services use the smaller Hastings Airfield at Freetown. The principal towns are accessible by road and the majority of villages can be reached by truck; in the mid-1960s the system comprised about 2,175 mi. (3,500 km.) of main roads and 2,200 mi. (3,540 km.) of secondary roads.

The Sierra Leone Broadcasting Service (founded 1934) transmits in English and a number of indigenous languages.

See also references under "Sierra Leone" in the Index.

(R. L.; X.)

BIBLIOGRAPHY.—H. C. Luke, *A Bibliography of Sierra Leone* (1925); Merran McCulloch, *Peoples of Sierra Leone Protectorate* (1950); H. R. Jarrett, *A Geography of Sierra Leone and Gambia* (1954); *Colonial Reports* (HMSO, annually until 1958); N. A. Cox-George, *Finance and Development in West Africa* (1961); C. Fyfe, *A History of Sierra Leone* (1962), (ed.), *Sierra Leone Inheritance* (1964).

Current history and statistics are summarized annually in the *Britannica Book of the Year*.

SIERRA MORENA, Spain, forms the mountainous southern edge of the Meseta (plateau), stretching for about 200 mi. (320 km.) from the Sierra de Alcaraz (5,899 ft. [1,798 m.]) in the east to the Portuguese border in the west. It includes many minor ranges which run transversely: *e.g.*, the Sierra Madrona, Sierra de Almadén, and Sierra de Aracena. It forms the main watershed between the rivers Guadiana and Guadalquivir, and its southern edge, overlooking the plains of Andalusia, is marked by a fault line; streams flowing south to join the Guadalquivir have eroded deep valleys in the scarp face, which descends from more than 1,500 ft. (450 m.) to land below 300 ft. (90 m.). Most of the Sierra Morena is wild desolate country thickly covered with dense evergreen bushes and shrubs. Its great breadth has long made it a formidable barrier in the social and military history of Spain. The chief communications route is via the Puerto de Despeñaperros which links Andalusia with Castile. The geological structure is very complex, exhibiting faulting and associated metamorphism resulting in the formation of minerals. Silver, lead, and copper are mined in the region of Linares; mercury at Almadén; copper at Ríotinto and at Tharsis in the Sierra de Aracena.

(J. M. Ho.)

SIERRA NEVADA, a range of mountains in southeast Spain. It is the highest range in the Baetic Cordillera (*see* PENIBÉTICO, SISTEMA), sometimes known as the Sierra Nevada system. The Nevada itself is clearly defined by the faulted troughs of the Vega of Granada to the northwest, the Guadix tableland to the northeast, and the Alpujarras (*q.v.*) depression to the south. The Nevada is a domed mountain elongated for about 26 mi. (41 km.) from east to west. The central dome, composed of hard slates (*lastra*) and softer micaceous schists presenting a smooth skyline, has been faulted so that the northern slopes rise precipitously, while the south and southeast flanks are gentler. On the edges of the centrally breached anticline are situated the main peaks: the Cerro de Mulhacén (11,411 ft. [3,478 m.]), the highest point of the Iberian Peninsula, and the Picacho de Veleta (11,128 ft. [3,392 m.]). Several other summits rise above the snowline, which lies at 10,000 ft. (3,050 m.), and have given significance to its name "the snowy range." Partly glaciated during the Würm period, leaving a trail of moraines and cirques, there is still a minute glacier in the Corral de Veleta. The range lies only a few miles inland from the Mediterranean coast and rises to above 10,000 ft.; the combination of Mediterranean climate and height has produced a sequence of flora from subtropical to alpine species.

BIBLIOGRAPHY.—J. Carandell, "La morfologia de la Sierra Nevada," *Rev. Acad. Madr.* vol. 19 (1921); J. Sermet, "Sierra Nevada," *Estudio. Geogr.* vol. 3 (1943); B. Messerli, "Sierra Nevada," *ibid.*, vol. 23 (1962)

(J. M. Ho.)

SIERRA NEVADA (Spanish, "Snowy Mountains"), of California, one of the major ranges of the United States, extends from Tehachapi Pass north-northwestward 400 mi. to Lake Almanor. Its breadth varies from 40 to 80 mi. The highest peak is Mt. Whitney, 14,495 ft.

The range consists chiefly of granite, flanked in places by belts of metamorphosed sediments, and overlain elsewhere by Tertiary volcanic materials. The range was created by block faulting: this great section of the earth's crust was broken apart from its surroundings, raised and tilted into a new position. The eastern edge became a massive escarpment, towering nearly three miles above Owens Valley. Streams cut deep valleys into the gentle western slope and deposited the resultant detritus onto the depressed western portion, creating the vast alluvial plains of the San Joaquin-Sacramento Valley ("Great Valley"). During the Pleistocene Epoch alpine glaciers reshaped some valleys, creating the spectacular scenery of Yosemite Valley and producing peaks, pinnacles, and rock-rimmed lakes along the eastern crest. In a few high sheltered valleys, small glaciers still remain.

This massive, unbroken barrier lifts the prevailing winds moving eastward off the Pacific Ocean, producing heavy precipitation, especially in winter, on the western slope (30–40 ft. of snowfall at 6,000 ft. in an average winter) and creating a desert eastward into Nevada. The range is a vast reservoir for water, supplying irrigation requirements for the Great Valley and municipal requirements for several cities, notably San Francisco and Los Angeles.

The great differences in altitude produce marked contrasts in vegetation. On the west, the grasslands of the Great Valley are succeeded upslope by a belt of oak woodland, replaced in turn by chaparral (drought-resistant dense brush), yellow pine forest, white fir–red fir forest, lodgepole pine forest, subalpine forest, and culminating finally in alpine meadows and bare rock slopes above the timberline. On certain unglaciated interfluves amidst the yellow pine forests are groves of Sequoias (*Sequoia gigantea* "big tree"), world renowned for their great age and size. On the steep eastern slope the transition from forest to desert is abrupt. Some lumbering is carried on in the west-slope forests. The upper zones are grazed by sheep in summer and the foothills support cattle through the year.

Gold was discovered in 1848 in Sacramento Valley gravels derived from the Sierra Nevada. Soon, thereafter, the bedrock source or "mother lode" was discovered on the western slope. After a wild period of mining booms and a longer period of steady production, mining activity gradually declined and the old mining towns became more important as tourist attractions than as mineral producers.

Tourists also visit the Sierra Nevada mountain recreation centres in great numbers in both summer and midwinter. Among the chief attractions are the three national parks—Yosemite, Sequoia and Kings Canyon (*qq.v.*)—and several ski centres in the national parks, the Donner Pass-Tahoe-Reno region, and the Mammoth area on the eastern slope.

Donner Pass (7,017 ft.), used by the Southern Pacific Railroad and highway U.S. 40, is the most important transmontane route, connecting San Francisco with Reno, Nev. Farther south, the mountains constitute an almost impenetrable barrier to land trans-

portation: from Tioga Pass to Walker Pass, a distance of 165 mi., no road crosses the range.

Aside from scattered recreation areas, lumber camps, and small towns serving the transmontane transportation routes, the Sierra Nevada range is virtually devoid of permanent inhabitants. Much of it is included in the national parks and several national forests.

BIBLIOGRAPHY.—Clarence King, *Mountaineering in the Sierra Nevada* (1872, 1935); John Muir, *The Mountains of California* (1894); Roderick Peattie, *The Sierra Nevada: the Range of Light* (1947); F. E. Matthes, *The Incomparable Valley* (1950), *Sequoia National Park* (1950); Norman E. A. Hinds, "Evolution of the California Landscape," *Bull. 158*, California Division of Mines (1952). (R. F. LN.)

SIEYÈS, EMMANUEL JOSEPH (1748–1836), French abbé and statesman who was one of the chief theorists of the French Revolution and an architect of the coup that replaced the Directory by the Consulate. He was born at Fréjus, in Provence, on May 3, 1748. Sent to the Sorbonne in Paris to be educated for an ecclesiastical career, he took less interest in theology than in the teachings of Locke, Condillac, and other political philosophers but nevertheless entered the church. He rose to be vicar-general (1780) and chancellor (1788) of the diocese of Chartres.

In January 1789, after the summoning of the Estates-General for the following May (*see* FRANCE: *History*), Sieyès published his pamphlet *Qu'est-ce que le Tiers État?* ("What Is the Third Estate?"). He began his answer thus: "Everything. What has it been hitherto in the political order? Nothing. What does it desire? To be something." The pamphlet had a great vogue, and its author was elected as the last of the 20 Paris deputies of the Third Estate. He strongly advised that the Estates-General meet in one chamber as the National Assembly, but he objected to the abolition of tithes and the confiscation of church lands without compensatory payment. He supported individual rights and civil equality, but not absolute political equality. Wishing to keep real power in the hands of the enlightened *bourgeoisie*, he resisted the idea of referendum and "imperative mandate": on the contrary, he thought that representatives elected for a specified period should be free to vote as they might see fit. In the committee on the constitution he argued against allowing the king to have a right of veto. He had a considerable part in the framing of the system of *départements*, but after the spring of 1790 he was eclipsed by men of more determined character and more revolutionary aims. Only once (June 1790) was he elected to the post of fortnightly president of the Constituent Assembly.

Ineligible for the Legislative Assembly of 1791 (as was every member of the Constituent Assembly), Sieyès was elected to the National Convention by three *départements* (Gironde, Orne, Sarthe) in September 1792. In the Convention he sat with the centre. He voted for the death sentence on Louis XVI. Soon afterward he effaced himself step by step, partly from disgust, partly from timidity. He abjured the Christian faith when the cult of Reason was instituted. It is said that he later summarized his conduct during the Terror in the ironical phrase *"J'ai vécu"* ("I remained alive").

After the fall of Robespierre (July 1794), Sieyès became more active. In 1795 he was on the Committee of Public Safety for six months, concerning himself with foreign policy; and his mission to The Hague was followed by a peace treaty with the Dutch and by France's alliance with the new Batavian Republic. He tried in vain to ensure that the new French Constitution of the Year III (1795) should establish a "constitutionary jury" with power to annul unconstitutional laws. Elected to the Council of the Five Hundred in October 1795, he was then voted a member of the Directory by the Chambers but declined to serve in this latter post. In May 1798 he was sent to Berlin to induce Prussia to make common cause with France; but though he negotiated adroitly he failed in his main object.

Sieyès was elected a director in place of J. F. Reubell in May 1799. He was already contemplating the overthrow of the Constitution, which was supposed to remain in force for six years from 1795 before it could be revised. To promote his plans he relied on the moderate republicans, took action against the Jacobins, and made overtures to Gen. B. C. Joubert for a *coup d'état*. Joubert's death and Napoleon Bonaparte's return from Egypt blocked his schemes; he then managed to come to an understanding with Bonaparte. On 19 Brumaire (Nov. 10, 1799) he became a provisional consul with Bonaparte and P. R. Ducos. His proposals for the Constitution of the Year VIII (Dec. 25, 1799) were characterized by his desire to maintain an oligarchy of men who had emerged in the original period of the Revolution: he wanted the political assemblies to be composed of members co-opted rather than directly elected by the citizens; and he believed that a chamber consisting of life members should have power to annul unconstitutional laws. Bonaparte accepted these two points but secured the rejection of the complex system proposed by Sieyès for balancing powers and for preventing the supremacy of any one man. Sieyès declined the title of second consul; he became the first member of the Senate and played a leading part in the first recruitment of senators, legislators, and tribunes.

From 1800 to 1802 Sieyès may have given secret encouragement to opposition groups. Thereafter he played a progressively diminishing role. He was nominated grand officer of the Legion of Honour in 1804 and a count of the empire (together with all the other senators) in 1808. Though he was one of the senators who voted against Napoleon in April 1814, he accepted a seat in the Chamber of Peers during the Hundred Days (1815). Banished at the Second Restoration, he went to Brussels. After the revolution of 1830 he returned to France and died in Paris on June 20, 1836.

During the Revolutionary period Sieyès enjoyed great fame as a political theorist, but his pride soon made him enemies, and his feeble voice and lack of oratorical power reduced his direct influence on large audiences. His prudence moreover led him to prefer to act indirectly save in 1799. In political conflict he was keen-sighted, resilient, and practical; but with regard to institutions he was dogmatic and uncompromising. Thus he had more influence on political doctrine than on any actual constitutional system.

BIBLIOGRAPHY.—A. Neton, *Sieyès* (1900); P. Bastid, *Sieyès et sa pensée* (1939); A. Vandal, *L'Avènement de Bonaparte* (1902); J. Bourdon, *La Constitution de l'an VIII* (1942). (C. E. DU.)

SIGEBERT, the name of three Frankish kings of the Merovingian dynasty.

SIGEBERT I (d. 575), king of the region afterward known as Austrasia, was the third son of Clotaire I and Queen Ingund. In the division made at his father's death (561), he received the northeastern part of the kingdom. To this was added, at the death (567) of his eldest brother Charibert, lands south of the Garonne and a large area in the west, including Touraine and Poitou. Menaced by the Avars (established in Germany since 558), he moved his capital from Reims to Metz and had twice to repel their attacks (562 and *c.* 568). He married (567) Brunhilda (*q.v.*), daughter of the Visigothic king Athanagild, whose other daughter Galswintha married Sigebert's half-brother Chilperic I. When Chilperic had Galswintha murdered (567 or 568) in order to marry his mistress Fredegund, Sigebert in revenge exacted from Chilperic his sister-in-law's dowry (Bordeaux, Limoges, Quercy, Béarn, and Bigorre). Prolonged warfare followed, during which Sigebert twice sought help from barbarian tribes beyond the Rhine. They ravaged the Paris area in 575, and Chilperic fled to Tournai. Sigebert, now controlling all the territory between the Loire and the Seine, pursued him. At his moment of triumph, when he was on the point of being acclaimed king at Vitry by Chilperic's subjects, he was stabbed by two assassins sent by Fredegund (early December 575).

SIGEBERT II (601–613), eldest of the four sons of Theuderic II, "reigned" only a few months and was never acclaimed king. On his father's death (613), his great-grandmother Brunhilda proposed his succession, but the Austrasian aristocracy rebelled and sought help from Clotaire II of Neustria. Brunhilda's army refused to fight that of Clotaire, and Theuderic's children fell into Clotaire's hands. Sigebert and one of his brothers were killed, a third died later at Clotaire's court, and the fourth disappeared.

SIGEBERT III (II) (630/631–656), was king of Austrasia under his father Dagobert I (*q.v.*) from 633 or 634, and sole king from

639. To conciliate the Austrasians, resentful that Dagobert had moved his residence from Metz to Paris, the three-year-old Sigebert was established at Metz (633 or 634) with a household ruled by Chunibert, bishop of Cologne, and the duke Adalgisile. On Dagobert's death (January 639), Pepin the Old (d. *c.* 640; *see* CAROLINGIANS) regained his influence in Austrasia, where he had been mayor of the palace before Dagobert's accession. By 643 Pepin's son Grimoald was mayor of the palace and Sigebert was thereafter completely under his tutelage. Before the birth of Sigebert's son, the future Dagobert II, Sigebert had adopted Grimoald's son Childebert, whom Grimoald later established as king for a few months (660 or 661). Sigebert died on Feb. 1, 656, and was succeeded by Dagobert II.

See F. Lot *et al.*, *Les Destinées de l'empire en occident de 395 à 888* (1928), vol. i of *Histoire du moyen âge* in G. Glotz (ed.), *Histoire générale;* L. Dupraz, *Le Royaume des Francs . . . 656–680* (1948).

SIGER OF BRABANT (d. between 1281 and 1284), professor of philosophy at Paris and leader of heterodox Aristotelianism, was a native of the duchy of Brabant. From 1266 (when his name first appears) to 1276 he was prominent in the dissensions that troubled the faculty of arts in Paris. From his earliest years as a teacher he professed a disquieting Aristotelianism, without regard for orthodox Christian doctrine. His teaching was attacked by Bonaventura, the minister general of the Friars Minor, and by Thomas Aquinas, the head of the Dominican school. On Dec. 10, 1270, Étienne Tempier, bishop of Paris, condemned 13 errors taken from the teaching of Siger and his partisans. On Nov. 23, 1276, the inquisitor of France summoned Siger, Goswin of the Chapel and Bernier of Nivelles before his tribunal; they fled to Italy and probably appealed to the tribunal of the papal curia. On March 7, 1277, Tempier pronounced the condemnation of 219 propositions, aiming especially at the teaching of Siger and Boëtius of Dacia. Siger was probably condemned to stay at the curia in the company of a cleric; he was stabbed at Orvieto, by his cleric, who had gone mad, and died under the pontificate of Martin IV, before Nov. 10, 1284. Dante (*Paradise*, x, 133–138) puts Siger in the Heaven of Light in the brilliant company of 12 illustrious souls who are glorified because of their fidelity to their own providential mission: Siger fought all his life for the autonomy of philosophy.

By the middle of the 20th century 14 certainly authentic works of Siger's and at least 6 probably authentic commentaries on Aristotle were known. The most important are: *Quaestiones in Metaphysicam; Impossibilia* (six exercises in sophistry); *Quaestiones de necessitate et contingentia causarum; Quaestiones in tertium de Anima;* and *Tractatus de anima intellectiva.*

Siger is an important representative of that school of radical Aristotelianism which arose in the faculty of arts at Paris when Latin translations of Greek and Arabic works had led to the discovery of philosophy by the masters there. Some of these masters, after 1260, inaugurated a purely rational teaching, without any concern with the exigencies of Christian faith; and Siger appears as the leader of this group. His capital source is Aristotle; secondary sources are Proclus and Avicenna (chiefly in metaphysics), Averroes (chiefly in psychology), Albert the Great and Thomas Aquinas. Siger's most typical doctrines are: the First Being is the immediate cause of a single creature, the first intelligence; all other creatures derive indirectly from God, by way of a progressive emanation; there is no real distinction between essence and existence in creatures; the created world is necessary and eternal, and every species of being (mankind, for instance) is eternal; there is only one intellectual soul for mankind and consequently one will; this unique soul is eternal, but the human individuals are not immortal (on the question of the soul, however, Siger's teaching was not consistent); human will is a passive potency moved by the intellect. Siger never accepted the theory of a double truth (one of reason and one of faith). To characterize his philosophy as "Latin Averroism" is inaccurate: his system must be called a radical or heterodox Aristotelianism.

BIBLIOGRAPHY.—P. Mandonnet, *Siger de Brabant et l'averroïsme latin au XIIIe siècle*, 2nd ed. (1908–11); F. van Steenberghen, *Siger de Brabant d'après ses oeuvres inédites* (1931–42) and *Aristotle in the West* (1955); E. Voegelin, "Siger de Brabant," *Philosophical and Phenomenological Research*, iv, no. 4 (1943–44); B. Nardi, *Sigieri di Brabante nel pensiero del Rinascimento italiano* (1945); C. A. Graiff (ed.), *Siger de Brabant: Questions sur la Métaphysique* (1948); J. J. Duin, *La Doctrine de la providence dans les écrits de Siger de Brabant* (1954). (F. V. SN.)

SIGERIST, HENRY ERNEST (1891–1957), Swiss medical historian who emphasized the social aspects of medicine, was born on April 7, 1891, in Paris and educated in France, Switzerland, England, and Germany. Combining interest in the humanities with medicine (M.D., Zurich, 1917), in 1925 he became professor of the history of medicine at Leipzig and in 1932 professor and director of the Johns Hopkins University Institute of the History of Medicine, the first American medico-historical research centre. A brilliant teacher and strong proponent of socialized medicine, Sigerist attracted disciples interested in medical history, sociology, and public health. His 27 books and 454 papers range from studies in medieval philology (*Studien und Texte zur frühmittelalterlichen Rezeptliteratur*, 1923) to analyses of American and Soviet medicine (*American Medicine*, 1934; *Socialized Medicine in the Soviet Union*, 1937). Upon retirement in 1947 he undertook a comprehensive *History of Medicine* of which the sections on primitive, archaic, and Greek medicine (vol. i, 1951; vol. ii, 1961) were completed before his death at Pura, Switz., on March 17, 1957. (GE. MI.)

SIGHT, SENSE OF: *see* VISION.

SIGHTS, GUN, are mechanical or optical devices that help the gunner aim his weapon at a target. They range from simple iron sights for small arms weapons to complex "fire control systems" for long-range artillery. To understand their nature and use it is essential first to know something of the behaviour of projectiles during flight.

When a projectile is fired from a gun, its flight path, or trajectory, forms a curve due to the action of the force of gravity and the resistance of the air. The force of gravity causes the trajectory to bend downward. If the projectile is rotating about its longitudinal axis, air resistance will cause it to drift to the right or left, depending upon the direction of rotation. These and other factors affecting the trajectory of a projectile are discussed quantitatively in the article BALLISTICS.

In view of these characteristics of the trajectory, it is obvious that the axis of a gun must be pointed above and to one side of the target to make the projectile hit that target. The process

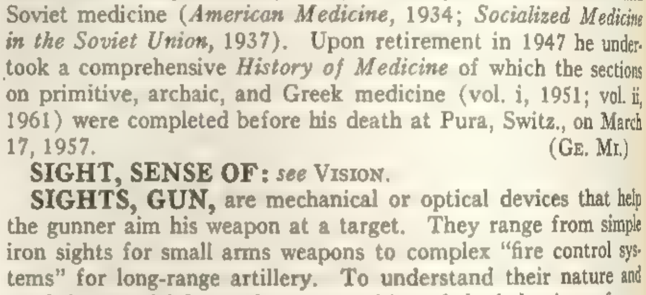


FIG. 1.—SIGHTING TERMS

of pointing a gun in the proper direction and at the desired angle of elevation is known as "laying" the gun. The devices employed to determine the proper direction and elevation for laying the gun constitute the sight or fire control system.

Terminology.—In discussing the principles of sighting, the following terms need to be clearly understood (*see* fig. 1):

Line of Sight.—A straight line joining gun and target.

Angle of Sight.—The angle between the line of sight and the horizontal plane through the gun.

Elevation.—The degree to which the axis of the gun is elevated above some line of reference. Quadrant elevation is the angle between the axis of the gun and the horizontal plane; tangent elevation, or superelevation, is the angle between the axis of the gun and the line of sight.

Rifles and Machine Guns.—The sights used for these types of weapons differ very little throughout the world. They are based on the principle that two points in fixed relation to each other may be brought into line with a third point. A typical rifle sight con-

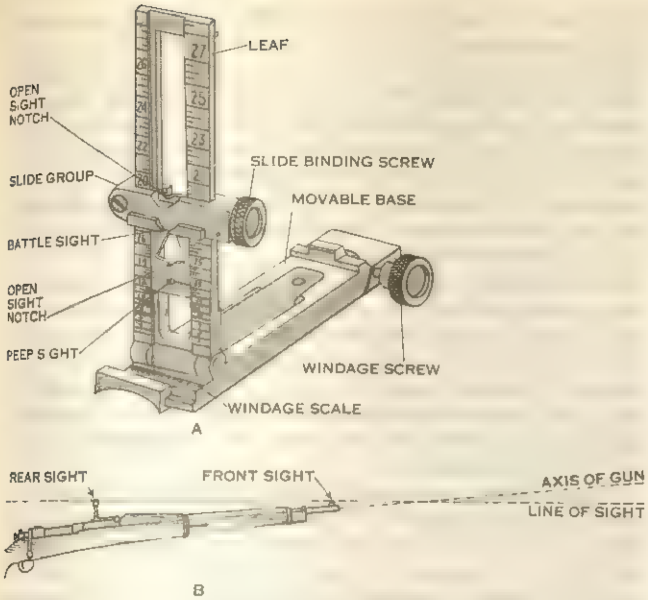


FIG. 2.—RIFLE AND MACHINE-GUN SIGHTS: (A) CLOSE-UP VIEW OF LEAF SIGHT; (B) ARRANGEMENT OF RIFLE SIGHTS

sists of a fixed foresight near the muzzle and a rearsight that is movable in a vertical plane, as shown in fig. 2. Vertical movement is calibrated in range (yards or meters) and is constrained to follow a curve that compensates for lateral drift of the bullet. A lateral adjustment is also provided to permit correction for the effect of wind.

Field Artillery, Mortars and Free Rockets.—Before discussing the sighting systems of these weapons, it is necessary to understand the arrangement of the basic parts of a gun mount. The top carriage is that portion of a gun mount that rotates (traverses) about a vertical axis and forms the support for the tipping parts of the gun. Any component of the sighting system attached to the top carriage will move with the gun in traverse, but will not move with it in elevation. The gun proper is usually mounted in

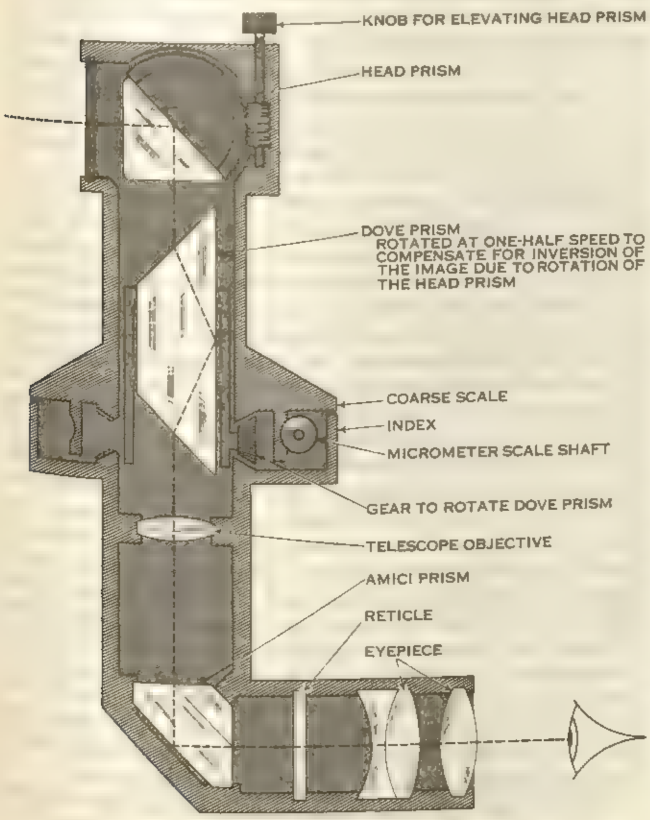


FIG. 3.—PANORAMIC TELESCOPE

a cradle attached to the top carriage by means of trunnions that form the horizontal axis about which the cradle and gun (tipping parts) rotate in elevation. The gun recoils in the cradle along the axis of the bore. Components of the sighting system attached to the cradle will move with the gun in both elevation and traverse, but will not move with the gun in recoil.

Since the elevating and traversing mechanisms are independent of each other, the laying of the gun can be simplified by having one member of the crew lay the gun in direction, and another lay it in elevation. It is, therefore, common practice to have the devices used for laying the gun in direction attached to the top carriage on one side (usually the left side) and the devices for laying in elevation attached to the cradle on the other side of the gun.

Mobile weapons are usually out of level (canted) due to irregularities in the ground and the axes of rotation are therefore not truly horizontal and vertical. When the gun is traversed under this condition, quadrant elevation will change and, conversely, when the gun is elevated the direction of fire will change. Sighting devices are so constructed that reference planes can be adjusted to the true vertical, hence laying accuracy is not impaired.

For protection from hostile fire, these weapons are usually emplaced in a position not visible to the enemy. Since the target can not be seen from the gun, laying in direction is accomplished

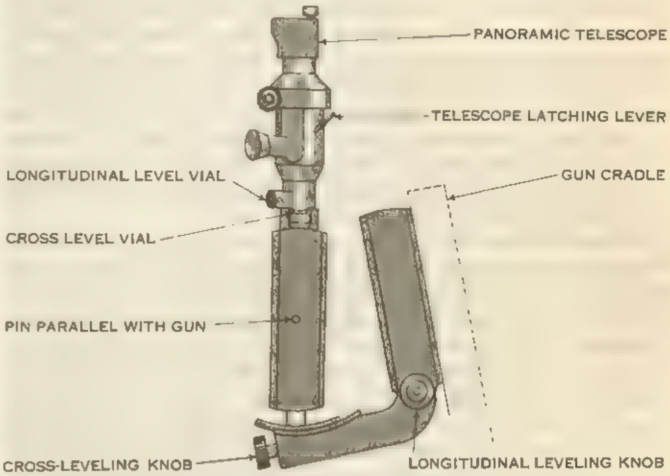


FIG. 4.—CANT COMPENSATING TELESCOPE MOUNT

by sighting on stakes set in the ground at a known direction from the gun. The horizontal angle between the gun axis and the gun-aiming stake line required to hit the target is measured from accurate maps and corrected by ballistic computations. A panoramic telescope which permits sighting in any direction from the same eyepiece position is used to measure this horizontal angle. (*See* fig. 3.) The panoramic telescope is mounted on the cradle and the mount compensates for the out-of-level condition by leveling the telescope about a bar parallel to the gun axis. (*See* fig. 4.) This determines the vertical plane containing the gun and permits the panoramic telescope to accurately measure the direction of fire.

Laying in elevation is accomplished by a device known as an elevation quadrant mounted on the cradle. The quadrant is also cross-leveled by rotation about a bar parallel to the gun axis so that elevation will be measured in the vertical plane. The precomputed quadrant elevation is set into the quadrant, offsetting a fore-and-aft spirit level. The gun is then elevated until the spirit level indicates that the gun has been elevated the desired amount. The elevation quadrant is sometimes incorporated into the compensating mount of the panoramic sight. Sights for mortars are usually so designed since laying of the mortar is performed by one man.

Antitank Guns and Recoilless Rifles.—As these weapons are employed at greater range than rifles or machine guns, they employ telescopic sights to increase the accuracy of fire. The telescopic sight contains a reticle with range markings (yards or metres) calibrated to the ammunition and offset at the longer

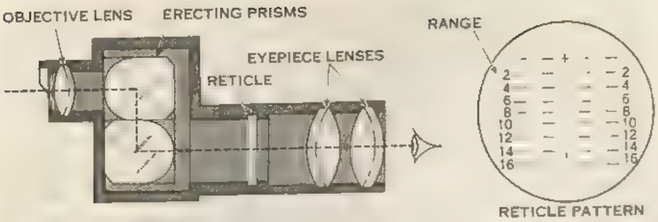


FIG. 5.—DIRECT FIRE TELESCOPIC SIGHT

ranges to compensate for drift. (*See* fig. 5.) The sight is mounted on the cradle (or directly on the recoilless rifle, which has no cradle) and in some cases provision is made for leveling the sight about an axis parallel to the gun to compensate for the error due to the gun being canted. In most cases the range of the target is estimated, but some guns are equipped with small-calibre spotting or ranging rifles to help determine the range. The spotting rifle is ballistically matched to the main gun. Both range and windage corrections are determined by trial fire with the small gun until a target hit is obtained before firing the main gun.

Tank Gunnery.—High-power telescopic sights are always used for tank gunnery because direct hits are required on small targets at long range. The physical configuration of the vehicle often makes it impractical to use a simple telescope. An articulated telescope, one with a rotating joint parallel with the gun trunnion, is sometimes used to bring the eyepiece to the gunner and eliminate movement as the gun is elevated. More frequently this problem is resolved by mounting a periscope in the turret roof and mechanically linking it to the gun to follow it in elevation. In some modern tanks, optical range finders (*see* RANGE FINDERS) are used to measure target range. Mechanical or electrical devices accurately compute the vertical and lateral corrections for each type of ammunition fired by the tank gun.

Naval Gunnery.—Modern naval guns are traversed and elevated by electronic and hydraulic servomechanisms. Direct laying by the gun crew is used only in emergencies. In normal practice, guns are sighted and the range of the target is determined either by radar or by optical means from director stations located high on the ship. The sighting data is electrically transmitted to a plotting room in the hull where mechanical or electronic computers generate the gun orders for electrical transmission to the turrets. A stable vertical gyroscope in the plotting room supplies data to the computer, usually called a range keeper, for conversion of director data into true horizontal and vertical tracking angles thus eliminating errors due to pitch and roll of the ship. The firing direction and quadrant elevation required to hit the target are computed by the range keeper with allowance for speed of both the firing ship and target. The laying angles are then converted, by data from the stable vertical gyroscope, into angles of turret elevation above deck plane and turret traverse in deck plane. The servomechanisms then keep the guns at the desired quadrant elevation and direction.

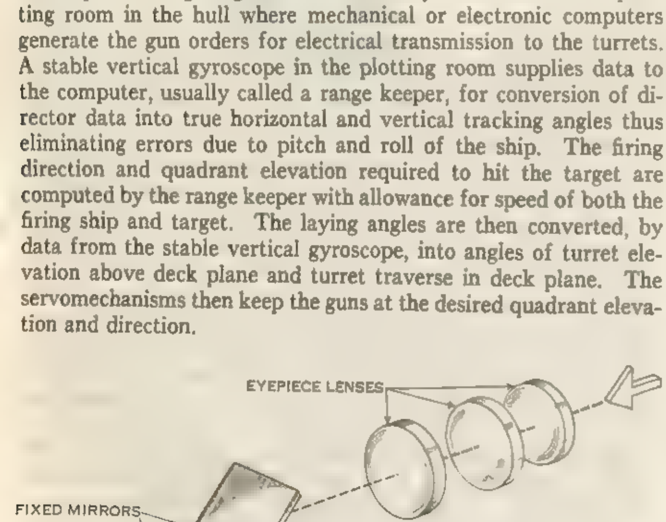


FIG. 6.—LEAD COMPUTING SIGHT

The guns are fired at any time, for ship motion does not disturb the lay of the guns.

See GUNNERY, NAVAL.

Antiaircraft Artillery.—Low-flying aircraft can be engaged successfully by antiaircraft artillery because the flight time of the shell is short. The high speed of jet aircraft makes it almost impossible to compute accurately the movement of the target during the time the projectile fired by an antiaircraft gun is in flight. The only effective weapons for use against high-flying aircraft are guided missiles or fighter planes.

Radar is used with antiaircraft artillery and missiles to measure the range to the target, but telescopic tracking in elevation and traverse usually gives greater accuracy than radar tracking, which is used only at long range or when visibility is poor. Electronic or mechanical computers are used to calculate the angle the gun must "lead" or be aimed ahead of the target. The gun is moved by electronic and hydraulic servomechanisms since the tracking rates are high.

Gyroscopes are used as the basic element in one type of lead-computing sight. Gyroscopic precession (*see* GYROSCOPE) is induced by the rotation of the sight while tracking the target in elevation and traverse. Precession is controlled by springs and weights designed to match the ballistics of the gun and adjusted as the target range changes. The amount the gyros precess against the restraining springs and weights is a measure of the required lead. This movement of the gyro is used to tilt mirrors in the sight (*see* fig. 6) which causes the target to appear ahead and above its true position; hence the gunner will track ahead by the appropriate amount.

Aircraft Gunnery.—Fighter aircraft carry rockets, machine guns or small-calibre cannon to engage both aerial and ground targets. Rigidly mounted guns are aimed by pointing the aircraft in the proper direction while approaching the target. Radar is used to measure the range to the target. A gyro lead-computing sight similar to that described above is used to compute the lead required to compensate for the motion of the firing aircraft and the target. (G. M. T.; W. T. AL.)

SIGIRI (SIGIRIYA), the Lion Rock, the ruin of a remarkable stronghold 45 mi. NNE of Kandy in the Central province of Ceylon. There a solitary pillar of granite rises to a height of 1,144 ft. from the plain, and the top actually overhangs the sides. On the summit of this pencil of rock there are five or six acres of ground; and on them, in A.D. 477, Kasyapa the Parricide, hoping to find an inaccessible refuge from his enemies, built his palace.

For a description of the Lion Rock and for the story of Kasyapa's crime *see* T. W. Rhys Davids, "Sigiri the Lion Rock" in *Journal of the Royal Asiatic Society*, vol. vii, pp. 191–220 (1875).

SIGISMUND (1368–1437), Holy Roman emperor from 1433, also king of Hungary by right of marriage from 1385 and by coronation from 1387, German king from 1411, king of Bohemia by inherited title from 1419 and by treaty from 1436, and Lombard king from 1431. A member of the House of Luxembourg, he was born on Feb. 15, 1368, a son of the emperor Charles IV (*q.v.*) by his last consort, Elizabeth of Pomorze. From his father he received Brandenburg (except the Neumark). Charles IV had also arranged the betrothal of Sigismund to Mary, daughter of King Louis I of Hungary and Poland; and after Charles IV's death (1378) Sigismund's half brother, the German king Wenceslas, who was also king of Bohemia, sent the boy to the Hungarian court. When Louis I died (1382), Sigismund's hopes sustained a setback: the Poles took Mary's sister Jadwiga instead of her as their queen; Mary's mother, Elizabeth of Bosnia, tried to prevent Sigismund's marriage; and a strong faction in Hungary offered the crown to Charles III (*q.v.*) of Naples. Sigismund however consummated his marriage to Mary in 1385, Charles of Naples was murdered in 1386, and finally Sigismund was crowned on March 31, 1387. Even so, his rule was still challenged by partisans of the Neapolitan king Ladislas, Charles III's son; to raise funds, he pawned Brandenburg to his cousin Jobst of Moravia in 1388; and Mary's death (1395) weakened his position.

Sigismund meanwhile kept his eye on the Bohemian succession: sometimes in collaboration and sometimes in rivalry with Jobst,

he intervened in the movements against Wenceslas. In 1396 he came to terms with Wenceslas. Then he turned his attention to the Turks, who had taken advantage of the troubles in Hungary to overrun Serbia and Bulgaria. Pope Boniface IX called for a crusade against them, and Sigismund advanced down the Danube Valley with a great army of the chivalry of Europe. He captured Vidin and laid siege to Nicopolis (Nikopol); but in September 1396 Sultan Bayazid I inflicted a crushing defeat on the crusaders. Escaping via Constantinople, Sigismund returned to Hungary in 1397.

In August 1400 Wenceslas was deposed from the German kingship, which was transferred to Rupert of the Palatinate. Sigismund sought to exploit this situation to get Bohemia from Wenceslas. In April 1401 he was arrested by conspirators in Buda: held prisoner at Siklos, he was released in the autumn. Returning to Bohemia, he had Wenceslas arrested in March 1402 but was soon called away to confront an invasion of Hungarian Dalmatia by Ladislas of Naples. During this Dalmatian War, Wenceslas was released and restored to his Bohemian throne. Ladislas, forced to abandon his attempt on the Hungarian crown, sold his Dalmatian conquests to the Venetians, against whom Sigismund became involved in intermittent war for years. In 1408 Sigismund married Barbara of Cilli, who bore him a daughter, Elizabeth, in 1409.

When King Rupert died, three of the German electors chose Sigismund as king on Sept. 20, 1410, but others chose Jobst in October. Jobst died in January; and on July 21, 1411, at Frankfurt, a new election assured the crown to Sigismund.

Having gone to Italy to reduce the Venetians, with whom he made a five-year armistice in April 1413, Sigismund had an interview with Pope John XXIII (*q.v.*), whom Ladislas of Naples had expelled from Rome. It was Sigismund who induced John to announce the summoning of the Council of Constance (*q.v.*) for 1414. Then at last Sigismund entered his German kingdom. Having been crowned at Aachen on Nov. 8, 1414, he arrived at Constance, where he made himself a leading spirit in the council. The degree of his responsibility for the arrest and burning of John Huss (*q.v.*) is a matter of controversy. Sigismund was concerned, however, not only to heal the great schism of the papacy but also to establish peace in Christendom so that a new crusade could be launched against the Turks. After going to Languedoc to persuade Ferdinand I of Aragon to disavow the antipope Benedict XIII (1415), he went to Paris, then to London in the hope of reconciling Charles VI of France and Henry V of England. Eventually, however, he allied himself with Henry; and this change of policy compromised his role of mediator at Constance.

The death of Wenceslas (1419) made Sigismund heir to Bohemia, but the Hussite Wars there cost him years of effort, with one military humiliation after another, before he could get his title recognized (*see* HUSSITES). His preoccupation with central Europe led him to neglect German affairs, but three of his acts were of great consequence: (1) his granting of Brandenburg by successive stages, from 1411, to Frederick VI of Hohenzollern (*see* HOHENZOLLERN); (2) the marriage of his daughter, in 1422, to Albert V of Austria, the future German king Albert II (*see* HABSBURG); and (3) his grant of the Saxon electorate, in 1423, to Frederick IV of Meissen (*see* WETTIN). After the Imperial Diet at Nürnberg in 1422, when he appealed in vain for effective help against the Hussites, he did not return to Germany till 1430. The Union of Bingen (1424), between the German electors, was a step toward strengthening the princes against his royal authority.

A campaign against the Turks resulted in a defeat for Sigismund in 1428, but peace was made next year. In 1431 he went to Italy again, to receive the Lombard crown in Milan on Nov. 25. The wars of the Italian states and the conflict between Pope Eugenius IV and the Council of Basel (*q.v.*) delayed Sigismund's coronation as Holy Roman emperor till May 31, 1433.

The *Compactata* of Prague (*see* HUSSITES) paved the way for peace in Bohemia. In July 1436 Sigismund approved them, and in August he was finally received as king in Prague. He died at Znojmo, in Moravia, on Dec. 9, 1437.

See also references under "Sigismund" in the Index.

BIBLIOGRAPHY.—There is a German translation of the life of Sigismund by his contemporary Eberhard Windecke, *Das Leben König Sigmunds* (1886). The fullest biography is J. Aschbach, *Geschichte Kaiser Sigmunds*, 4 vol. (1838–45). *See* also *Die Urkunden Kaiser Sigmunds*, ed. by W. Altmann (1896–1900); H. Horváth, *Zsigmond király es kora*, with summary in German (1937).

SIGISMUND I THE OLD (Polish ZYGMUNT STARY) (1467–1548), king of Poland from 1506, was born on Jan. 1, 1467, the fifth son of Casimir IV Jagiello and Elizabeth of Austria. He learned the art of government as prince of Glogów, which he transformed into a model state, and subsequently as governor of Silesia and margrave of Lusatia under his brother Vladislav of Bohemia.

Sigismund was elected grand prince of Lithuania on Sept. 13, 1506, and king of Poland on Dec. 8, 1506, in succession to his brother Alexander. In Poland his first step was to recover control of the mint and to put it in the hands of capable merchants and bankers (Jan Boner, the Betman family, and others), who reformed the currency and developed commerce. Apart from them, he relied on the support of the magnates for internal affairs and came into conflict with the middle *szlachta*, or gentry, which at that time was beginning to assert itself. Popular movements against the patrician governments of the Prussian cities were suppressed with severity. On the other hand, though he was a devout Catholic himself, Sigismund would not tolerate the persecution of non-Catholics; he upheld the rights of his Orthodox subjects against the Catholic minority of magnates which dominated the Lithuanian senate; and he protected the Jews.

In foreign policy Sigismund's principal advisers were Jan and Jaroslaw Laski and Jan Tarnowski. For the first 20 years of his reign Poland was almost continuously at war with Muscovy; and in order to break the coalition between Muscovy and the Habsburgs he had to resign to the latter his claims to the Hungarian crown (Treaty of Vienna, July 22, 1515). His southeastern borderland, meanwhile, was so frequently ravaged by the Tatars that a small permanent army had eventually to be established there. Quarrels with the grand masters of the Teutonic Order (*q.v.*), who wanted to release Prussia from Polish suzerainty, were ended in 1525, when the grand master Albert (*q.v.*), having professed Lutheranism, did public homage as duke of Prussia to Sigismund at Cracow.

On Feb. 8, 1512, Sigismund had married Barbara, daughter of the Hungarian prince Stephen Zápolya, but she died three years later, leaving only daughters. The king's second marriage (April 18, 1518) was to the beautiful Bona Sforza (1494–1557), daughter of Gian Galeazzo Sforza, duke of Milan. By her he had a son, who was to succeed him as Sigismund II (*q.v.*), and four daughters: Isabella (1519–59), married in 1539 to the Hungarian king John Zápolya; Sophia (1522–75), married in 1556 to Henry II of Brunswick-Wolfenbüttel; Anna (1523–96), married in 1576 to Stephen (*q.v.*) Báthory; and Catherine (1526–83), married in 1562 to John III of Sweden, from whom the Vasa kings of Poland descended. Bona Sforza, however, used her financial talents mainly for her own benefit and became generally detested for corruption.

A lover of the fine arts and, especially, of architecture and sculpture, Sigismund brought Italian artists to Cracow and promoted the development of the Polish variety of Italian Renaissance style. From his 60th year onward, however, his energy and capacity declined, though his herculean physique retained the appearance of vigour. Sigismund died in Cracow on April 1, 1548.

See Z. Wojciechowski, *Zygmunt Stary* (1946); W. Pociecha, *Królowa Bona*, 4 vol. (1949–58). (ST. HE.)

SIGISMUND II AUGUSTUS (Polish ZYGMUNT AUGUST) (1520–1572), king of Poland as sole ruler from 1548, was born in Cracow on Aug. 1, 1520, the son of Sigismund I and Bona Sforza. He was crowned in 1530, during his father's lifetime, and ruled Lithuania from 1544. In 1543 he married Elizabeth of Austria, daughter of the German king and future Holy Roman emperor Ferdinand I; she died childless in 1545 and Sigismund in 1547 married the beautiful Barbara Radziwill, who belonged to a family of Lithuanian magnates (*see* RADZIWILL).

On his father's death (April 1, 1548), Sigismund's actual reign in Poland began. At his first *Sejm* (Diet), in October 1548, he came into conflict with the rising power of the *szlachta*, or gentry. With secret support from Austria and from the queen mother, the *szlachta* threatened to renounce its allegiance to Sigismund unless he repudiated his wife and shook off the influence of the Radziwills. His firm refusal produced a reaction, and at the second *Sejm* (1550) the *szlachta* was less turbulent. Barbara was crowned queen of Poland on Dec. 7, 1550, but died childless in 1551. In 1553 Sigismund married Catherine, his first wife's sister; this union also was childless.

Sigismund's reign was a period of internal turmoil and external expansion (*see* POLAND: *History*). The doctrines of the Reformation, which had been received in Poland clandestinely during Sigismund I's reign, spread very widely throughout the *szlachta*, and the democratic upheaval placed all political power in the latter's hands. The *szlachta* forced the magnates to restore the crown lands which had come into their possession, and demanded a centralized administration of the state.

The collapse of the Livonian branch of the Teutonic Order led to Poland's acquisition of Livonia in 1559, but Sigismund's most striking and personal achievement was the real union of Poland and Lithuania, brought about during the *Sejm* at Lublin in 1569, despite opposition by the separatist Lithuanian magnates. The last male of the Jagiellos in the direct line, Sigismund died at Knyszyn on July 7, 1572.

See O. Halecki, *Dzieje unii jagiellońskiej*, vol. ii (1920). (ST. HE.)

SIGISMUND III (1566–1632), king of Poland from 1587 and king of Sweden from 1592 to 1599, was born at Gripsholm on June 20, 1566, the elder son of John III of Sweden and his consort Catherine, sister of Sigismund II of Poland. He belonged to the Vasa dynasty through his father, and to the Jagiello dynasty through his mother, who educated him as a Catholic. After the death of Stephen (*q.v.*) Báthory, he was elected king of Poland on Aug. 19, 1587, thanks to his aunt, Anne (Stephen's widow), and to the Polish chancellor, Jan Zamoyski (*q.v.*). He was crowned at Cracow on Dec. 28, 1581.

From the first Sigismund was antipathetic to the majority of his subjects. The *szlachta*, or gentry, who could not understand his refinement, regarded his calm and reserved manner as haughtiness. His political views, moreover, were diametrically opposed to the omnipotent Zamoyski's. In foreign policy, Sigismund aimed at alliance between Poland and the Austrian Habsburgs with the double object of forming a solid Catholic front against the Turks and of exerting greater influence on Sweden. No benefit, however, could be expected from any alliance unless the Polish constitution were reformed so as to increase the royal power.

On May 31, 1592, Sigismund married Anne, daughter of the archduke Charles of Inner Austria; and at the *Sejm* or Diet in the autumn Zamoyski accused him of wanting to cede his throne to the Habsburgs. Finally, however, king and chancellor patched up their quarrel; and after the death of Sigismund's father (Nov. 17, 1592) Zamoyski obtained the Diet's permission for Sigismund to go to Sweden to secure the succession there (summer 1593).

Having consented to the maintenance of Lutheranism in Sweden, Sigismund was crowned at Uppsala on Feb. 19, 1594. On July 19 he departed for Poland, leaving his uncle, Charles (*see* CHARLES IX), as regent. The regent, however, soon began to subvert Sigismund's authority. In 1598 Sigismund landed in Sweden with an army, but was defeated at Stångebro (Sept. 25) and had to accept humiliating terms before leaving the country. The next year Charles had him formally deposed by the Swedish Diet. Sigismund refused to renounce his rights and embarked on hostilities against the Swedes (1600).

In Poland, meanwhile, Sigismund had favoured the ecclesiastical Union of Brześć (1596), which brought numbers of his Orthodox subjects into the Roman Catholic communion. The oppressed Cossacks of the Ukraine rebelled in the name of Orthodoxy, and the Polish Protestants made common cause with them. To reinforce the pro-Austrian faction at his court, Sigismund, a widower from 1598, soon made plans to marry his first wife's sister Constantia, thereby antagonizing Zamoyski again. At the *Sejm* of 1605 the king tried to introduce decision by a majority of votes instead of unanimity, but Zamoyski frustrated this reform. Zamoyski died in June, and on Dec. 11, 1605, Sigismund and Constantia were married.

Zamoyski's death removed all restraint from his extremist partisans, and from 1606 to 1608 Poland was in a state of civil war and anarchy. There followed a long war against Muscovy (1609–19; *see* RUSSIAN HISTORY); and in 1617 the Polish-Swedish War, which had been interrupted by a truce in 1611, broke out again. Since Sigismund at the same time brought Poland into conflict with Turkey (1617–21), the Swedish king Gustavus II Adolphus was able to occupy Livonia, capturing Riga in 1621. After another truce (1622–25), Gustavus in 1626 attacked Polish Pomorze also. Finally the six-year Truce of Altmark was signed in 1629.

Sigismund died in Warsaw on April 30, 1632. His son by Anne succeeded him as Wladyslaw IV. By Constantia he left the future king John Casimir and four other children. (ST. HE.)

SIGNAC, PAUL (1863–1935), French painter, developer, with Georges Seurat, of the technique called Neoimpressionism, was born in Paris, Nov. 11, 1863. After youthful attempts to study architecture, he turned, at the age of 18, to painting. Through Armand Guillaumin, Signac became a convert to the tone division of Impressionism, and he was represented by several landscapes at the exhibition of the Maison Dorée in 1886.

In 1884 Signac had helped found the Salon des Indépendants. There he met Seurat, whom he initiated into the broken-colour technique of Impressionism. The two went on to the method they called Pointillism or Neoimpressionism. They continued to divide their canvases into minute dabs of pure pigment, as had the Impressionists, but they adopted an exact, geometrical system of applying the dots instead of the somewhat intuitive application of the earlier masters. This technique was not unlike the mechanical process of printing photographic illustrations in colour by means of halftones. Some critics denounced Signac's landscapes as monotonous and mathematical rather than artistic, but others found that the optical mixture of detached colours separated by pure white created an admirable luminosity. In water colours Signac used the principle in a much freer manner. After 1886 Signac took part regularly in the annual Salon des Indépendants, to which he sent landscapes, seascapes, and decorative panels. He traveled widely over Europe with Seurat and other painters, and in his later years painted many scenes of Paris, Viviers, and other French cities.

Signac produced much critical writing and was the author of *De Delacroix au néo-impressionnisme* (1899) and *Jongkind* (1927). The former book is an exposition of Pointillism or Neoimpressionism as developed by Seurat and himself. Signac died in Paris on Aug. 15, 1935. Paintings by Signac may be seen at the Museum of Modern Art and in the Frick collection in New York city; at the Musée National d'Art Moderne in Paris; and in many other public collections.

SIGNAL, a word common in slightly different forms to nearly all European languages, derived from Lat. *signum*, "a mark" or "sign," a means of transmitting information, according to some prearranged system or code, in cases where a direct verbal or written statement is unnecessary, undesirable, or impracticable. For military signaling *see* SIGNAL COMMUNICATION, MILITARY; *see* also RAILWAY: *Communications and Signaling: Signals*.

SIGNAL COMMUNICATION, MILITARY. Signal communication, or signaling, has long played an important role in warfare. It serves to provide the means for transmitting information from reconnaissance and other units in contact with the enemy, and the means for exercising command by transmitting the orders and instructions of commanders to their subordinates. It comprises all means of transmitting messages, orders, and reports, both in the field and between army headquarters and field installations. In U.S. Army usage, signal communication includes automatic transmission of data over military wire and radio circuits; between the electronic computers and control systems employed in gun laying, air navigation, and guided-missile control; and in the various logistic and clerical activities of the military arms and services. Signal communication in some armies also

includes photography and television.

The history of the science of communication can be divided into four phases: (1) the early period of origin and growth up to the invention of the electric telegraph and the Morse code of dots and dashes in the mid-19th century; (2) the advent of the telegraph, the telephone, and wireless radio and their development up to the outbreak of World War I in 1914; (3) development during World War I and up to the beginning of World War II; and finally (4) the great expansion during and after World War II.

Early Development.—Messengers have been employed in war since ancient times and still constitute a valuable means of communication. Alexander, Hannibal, and Caesar each developed an elaborate system of relays by which messages were carried from one messenger post to another by mounted messengers traveling at top speed. They were thus able to maintain contact with their homelands during their far-flung campaigns and to transmit messages with surprising speed. Genghis Khan at the close of the 12th century not only emulated his military predecessors by establishing an extensive system of messenger posts from Europe to his Mongol capital but also utilized homing pigeons as messengers. As he advanced upon his conquests he established pigeon relay posts across Asia and most of Europe. He was thus able to utilize these messengers to transmit instructions to his capital for the governing of his distant dominions. Before the end of the 18th century European armies used the visual telegraph system devised by Claude Chappe, employing semaphore towers or poles with movable arms. The Prussian Army in 1833 assigned such visual telegraph duties to engineer troops.

At the same time that these elementary methods of signal communication were being evolved on land, a comparable development was going on at sea. Early signaling between naval vessels was by prearranged messages transmitted by flags, lights, or the movement of a sail. Codes were developed in the 16th century that were based upon the number and position of signal flags or lights or on the number of cannon shots. In the 17th century, Adm. Sir William Penn (1621–70) and others developed regular codes for naval communication, and toward the close of the 18th century, Adm. Richard Kempenfelt developed a plan of flag signaling similar to that now in use. Later Sir Home Popham increased the effectiveness of ship-to-ship communication by improved methods of flag signaling.

Advent of Electrical Signaling.—Despite the early pioneering efforts on land and sea the real development of signal communication in war did not come until after invention of the electric telegraph by Samuel F. B. Morse. In his successful demonstration of electric communication between Washington, D.C., and Baltimore, Md., in 1844, he provided a completely new means of rapid signal communication. The development of the Morse code of dots and dashes used with key and sounder was soon used to augment the various means of visual signaling. Vice-Adm. Philip Colomb's flash signaling, adopted in the British Navy in 1867, was an adaptation of the Morse code to lights. The first application of the telegraph in time of war was made by the British in the Crimea in 1854, but its capabilities were not well understood and it was not widely used. Three years later, in the Indian Mutiny, the newly established telegraph, which was controlled by the British, was a deciding factor.

In the U.S. Civil War (1861–65) wide use was made of the electric telegraph. In addition to its employment in spanning long distances under the civilian-manned military telegraph organization, mobile field service was provided in the Union Army by wagon trains equipped with insulated wire and lightweight poles for the rapid laying of telegraph lines. Immediately before and during the Civil War visual signaling also received added impetus through the efforts of two British officers, Capt. Frank Bolton of the Army and Capt. Colomb of the Royal Navy, and Maj. Albert Myer, signal officer of the U.S. Army (*see* SIGNAL CORPS). The two British officers, working together, and Myer, working independently, arrived at the same end. They developed a system, applying the Morse code of dots and dashes, for spelling out messages by flags by day and lights or torches by night. Another development for light signaling placed a movable shutter, controlled by a key, in front of a strong light. An operator, opening and closing the shutter, could produce short and long flashes in accordance with the Morse code and this could spell out messages.

Simultaneously with U.S. and British developments in military telegraph, the Prussian and French armies also organized mobile telegraph trains. During the short, decisive Prussian campaign against Austria in 1866, field telegraph enabled Count Helmuth von Moltke, the Prussian commander, to exercise command over his distant armies. Soon afterward the British organized their first field telegraph trains in the Royal Engineers. In the British Abyssinian campaign, 1867–68, Sir Robert Napier's forces made extensive use of Bolton and Colomb's new method of visual signaling and of the field telegraph. Success was immediate and remarkable.

Another instrument was added to the techniques for visual signaling through the development of the heliograph. It employed two adjustable mirrors so arranged that a beam of light from the sun could be reflected in any direction. The beam was interrupted by a key-operated shutter that permitted the formation of the dots and dashes of the Morse code. Where climatic conditions were favourable this instrument found much use, notably by the British Army in India and the United States Army in the southwestern U.S. Because consistency and regularity of sunshine were important, the heliograph was never widely adopted throughout the armies of continental Europe.

The invention of the telephone in 1876 was not followed immediately by its adoption and adaptation for military use. This was probably due to the fact that the compelling stimulation of war was not present and to the fact that the development of long-distance telephone communication was not achieved for many years. The telephone was used by the United States Army in the Spanish-American War, by the British in the South African War and by the Japanese in the Russo-Japanese War. This military use was not extensive and it made little material contribution to the development of voice telephony. Between the start of the 20th century and the outbreak of World War I, military adaptation of the telephone did take place, but its period of major growth had not yet arrived.

Near the close of the 19th century, a new means of military signal communication made its appearance—the wireless telegraph, or, as named in the United States and Canada—the radio. In 1894 Sir Oliver Lodge successfully demonstrated that the electromagnetic waves first discovered by Heinrich Hertz, and later studied by James C. Maxwell, could be guided and used for signaling. He was followed by Marconi, who developed but did not discover the use of these waves for signaling. The major powers throughout the world were quick to see the wonderful possibilities of this new agency for military and naval signaling. Its development was rapid and continuous and, by 1914, it was adopted and in extensive use by all the armies and navies of the world. It soon became apparent that wireless telegraphy was not an unmixed blessing to armies and navies, because it lacked secrecy and messages could be heard by the enemy as well as by friendly forces. This led to the development of extensive and complicated codes and ciphers as necessary adjuncts to military signaling. The struggle between the cryptographer and the cryptanalyst expanded greatly with the adoption of radio and continued to be a major factor affecting its military use. (*See* CRYPTOLOGY.)

From World War I to 1940.—The onset of World War I found the opposing armies equipped to a varying degree with modern means of signal communication but with little appreciation of the enormous load that signal systems must carry to maintain control of the huge forces that were set in motion. The organization and efficiency of the armies varied greatly. At one end of the scale was Great Britain, with a small but highly developed signal service, and at the other end stood Russia, with a signal service inferior to that of the American Union Army at the close of the Civil War. The fact that commanders could not control, coordinate, and direct huge modern armies without efficient signal communication quickly became apparent to both the Allies and the Central Powers. The Germans, despite years of

concentration on the Schlieffen plan, failed to provide adequately for communication between higher headquarters and the rapidly marching armies of the right wing driving through Belgium and northern France. This resulted in a lack of coordination between these armies which caused a miscarriage of the plan, a forced halt in the German advance, and the subsequent withdrawal north of the Marne. On the Allied side, the debacle of the Russian forces in East Prussia—a crushing defeat at the hands of Gen. Paul von Hindenburg in the Battle of Tannenberg—was largely due to an almost total lack of signal communication.

As the war progressed there was a growing appreciation of the need for improved electrical communications of much greater capacity for the larger units and of the need within regiments for electrical communications, which had heretofore been regarded as unessential and impractical. Field telephones and switchboards were soon developed and those already in existence were improved. An intricate system of telephone lines involving thousands of miles of wire soon appeared on each side. Pole lines with many crossarms and circuits came into being in the rear of the opposing armies and buried cables and wires laid in trenches were installed in the elaborate trench systems leading to the forwardmost outposts. The main arteries running from the rear to the forward trenches were crossed by lateral cable routes roughly parallel to the front. Thus there grew an immense gridwork of deep buried cables, particularly on the German side and in the British sectors of the Allied side with underground junction boxes and test points every few hundred yards. The French used deep buried cable to some extent but generally preferred to string their telephone lines on wooden supports set against the walls of deep open trenches. Thus electrical communication in the form of the telephone and direct-current telegraph gradually extended to the smaller units until front-line platoons were frequently kept in touch with their company headquarters through these mediums.

At the beginning of the war most of the armies were using military telegraph instruments employing an alternating current at frequencies from 500 to 700 cycles per second. These instruments could transmit well through the long and generally defective wire lines used in open warfare. However, when used in the trench systems in close proximity to the enemy it was found that their signals could be readily intercepted by using amplifiers associated with earth induction pickup devices. A new type of telegraph instrument was soon developed which transmitted an interrupted direct current that could not be picked up by the listening devices then in use.

Despite efforts to protect the wire lines, they were frequently cut at critical times as the result of the intense artillery fires. This led all the belligerents to develop and use radio (wireless) as an alternate means of communication. Prewar radio sets were too heavy and bulky to be taken into the trenches and they also required large and highly visible aerials. Radio engineers of the belligerent nations soon developed smaller and more portable sets powered by storage batteries and using low, inconspicuous aerials. Although radio equipment came to be issued to the headquarters of all units, including battalions, the ease of enemy interception, the requirements for cryptographing or encoding messages, and the inherent unreliability of these early systems caused them to be regarded as strictly auxiliary to the wire system and reserved for emergency use when the wire lines were cut. Visual signaling returned to the battlefield in World War I with the use of electric signal lamps. Pyrotechnics, rockets, Very pistols, and flares had a wide use for transmitting prearranged signals. Messenger service came to be highly developed and motorcycle, bicycle, and automotive messenger service was employed. Homing pigeons were extensively used as one-way messengers from front to rear and acquitted themselves extremely well. Dogs also were used as messengers and, in the German Army, reached a high degree of efficiency.

A new element in warfare, the airplane, introduced in World War I, immediately posed a problem in communication. During most of the war, communication between ground and air was difficult and elementary. To make his reports the pilot had to land or drop messages, and he received instructions while in the air from strips of white and black cloth called "panels" laid out in an open field according to prearranged designs. Extensive efforts were made to use radiotelegraph and radiotelephone between the airplanes and ground headquarters. The closing stages of the war saw many planes equipped with radio but the service was never satisfactory or reliable and had little influence on military operations.

During World War I, wireless telegraph communication was extensively employed by the navies of the world and had a major influence on the character of naval warfare. High-powered shore and ship stations made possible wireless communication over long distances.

World War I left a host of lessons to be studied by the army, naval, and air force personnel of all nations. Important among these lessons was the need to improve mobility, firepower, communication, and the effectiveness of air power. Advances in military communications after the war were greatly aided by the large increase in electronic and communication development for civilian purposes.

The growth of radiobroadcasting following World War I brought into existence in most nations a greatly increased interest in communication engineering and the expansion of manufacturing facilities to exploit the new art. Carrier telephony on wire systems also made great advances in providing for several unconflicting conversations over the same wires. There was a rapid growth in long-distance telephone techniques and plant. The design of amplifiers for telephone repeaters made rapid progress.

One of the war lessons learned in most of the major nations was the compelling need for scientific research and development of equipment and techniques for military purposes. Although the amount of funds devoted to military development during the period from World War I to World War II was relatively small, the modest expenditures served to establish a bond between industry, science, and the armed forces of the major nations.

Of great importance in radio communication was the pioneering by the amateurs and by industry and science in the use of very high frequencies. These developments opened up to the armed services the possibilities of portable short-range equipment for mobile and portable tactical use by armies, navies, and air forces. Military work in these fields was carried out actively in Germany, Great Britain, and the United States. As early as 1938 Germany had completed the design and manufacture of a complete line of portable and mobile radio equipment for its Army and Air Force.

The increased interest in scientific research and development on the part of the armed services brought about the birth of radar (*q.v.*), which was destined to become one of the major technological advances of World War II. Signifying radio detecting and ranging, radar was developed for long-range location and tracking of enemy aircraft and for the direction of antiaircraft fire. It used sharply beamed short pulses of radio energy that were reflected back from the surfaces of aircraft in flight to give a visual signal indication on the face of a receiver oscilloscope. Using radar, it became possible to detect and locate enemy aircraft at distances of 100 to 150 mi., which enabled the defending air force then to take off and intercept enemy bombers. In Britain, radar became one of the major factors that enabled the British to meet and neutralize mass German bombing attacks during the Battle of Britain. Although, strictly speaking, radar is not used for communication, it formed part of the complex communication systems required for antiaircraft warning and fire-control nets. The employment of radar in IFF systems (identification, friend or foe) to distinguish friendly aircraft from hostile on the oscilloscopes of search sets did in fact serve a communication function.

Between World Wars I and II the printing telegraph, commonly known as the teleprinter or teletypewriter machine, came into civilian use and was incorporated in military wire-communication systems but military networks were not extensive. Before World War II military radioteleprinter circuits were nonexistent.

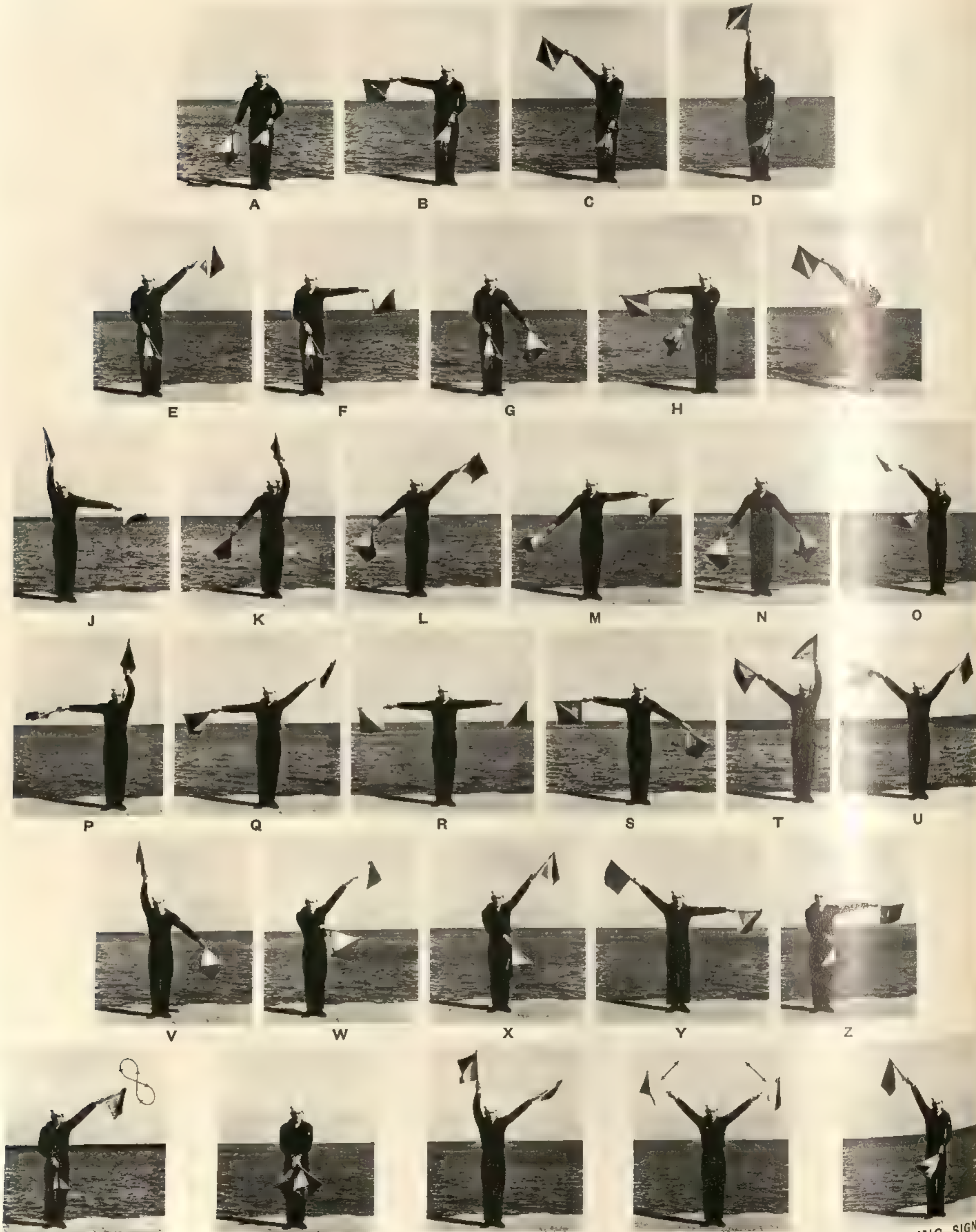
Another major communication advance that had its origin and early growth during the period between World Wars I and II was frequency-modulated (FM) radio. Developed during the late 1920s and early 1930s by Edwin H. Armstrong, inventor and a major in the United States Army Signal Corps during World War I,

BY COURTESY OF THE NATIONAL GEOGRAPHIC SOCIETY. PAINTED FOR ENCYCLOPÆDIA BRITANNICA, INC., BY SIDNEY DICKENS

INTERNATIONAL CODE FLAGS AND PENNANTS

The two-letter signal R Y flying from the masthead signifies "Crew have mutinied." The other two-letter groups on the halyards hoisted to the yardarm have the following meanings: N C—"I am in distress and require immediate assistance"; K A—"My vessel is very seriously damaged"; K T—"Have you a line-throwing apparatus?"; I X—"I have received serious damage in collision"; C G—"You should alight as near to me as possible"; A D—"I must abandon my vessel."

SIGNAL COMMUNICATION, MILITARY



SEMAPHORE FLAG SIGNALS

Numerals from 1–9 are the same as the first nine letters of the alphabet, sent after the "numeral" signal; zero is the same as the letter J

this new method of modulation offered heretofore unattainable reduction of the effect of ignition and other noises encountered in vehicular radio applications. It was first adapted for military use by the United States Army, which, prior to World War II, had under development tank, vehicular, and man-pack frequency-modulated radio transmitters and receivers.

On the eve of World War II, all nations employed generally similar methods for military signaling. The messenger systems included foot, mounted, motorcycle, automobile, airplane, homing pigeon, and the messenger dog. Visual agencies included flags, lights, panels for signaling airplanes, and pyrotechnics. The electrical agencies embraced wire systems providing telephone and telegraph service, including the printing telegraph. Both radiotelephony and radiotelegraphy were in wide use, but radiotelephony had not as yet proved reliable and satisfactory for tactical military communication. The navies of the world entered World War II with highly developed radio-communication systems, both telegraph and telephone, and with development under way of radar and sonar (sound navigation and ranging) and many other electronic navigational aids. Blinker-light signaling was still employed. The use of telephone systems and loud-speaking voice amplifiers on naval vessels had also come into common use. Air forces employed wire and radio communication to link up their bases and landing fields and had developed airborne long-range, medium-range, and short-range radio equipment for air-to-ground and air-to-air communication.

World War II and After.—In the communication and electronics field, World War II was in one sense similar to World War I—the most extravagant prewar estimates of military, naval, and air communication requirements soon proved to represent only a fraction of the actual demand. Requirements for all kinds of communication equipment and for improved quality and quantity of communications pyramided beyond the immediate capabilities of industry. Expansion of manufacturing facilities became vital, and the expansion of research and development activities in the communications-electronics field was unprecedented. The early German blitzkrieg, with tank and armoured formations, placed a new order of importance on reliable radio communication. The development of the air, infantry, artillery, and armoured team created new requirements for split-second communication by radio between all members. Portable radio sets were provided as far down in the military echelons as the platoon. In every tank there was installed at least one radio and in some command tanks as many as three. In the field of wire communications multiconductor cables were provided; they could be rapidly reeled out and on them as many as four conversations could take place simultaneously through the use of carrier telephony. The Germans were the first to bring forth this type of military long-range cable and their example was promptly followed by both the British and the U.S. forces. High-powered mobile radio sets became common at division and regimental level. With these sets telegraph communication could be conducted at distances of more than 100 mi. with vehicles in normal motion on the road. Major telephone switchboards of much greater capacity were needed. These were developed, manufactured, and issued for use at all tactical headquarters to satisfy the need for the greatly increased number of telephone channels required to coordinate the movements of field units whose mobility had been expanded many times.

Radio relay, born of the necessity for mobility, became the outstanding communication development of World War II. Sets employing frequency modulation and carrier techniques were developed and used, as were also radio relay sets that used radar pulse transmission and reception techniques and multiplex time-division methods for obtaining many voice-channels from one radio carrier. Radio relay telephone and teletypewriter circuits spanned the English Channel for the Normandy landing and later furnished important communication service for Gen. George S. Patton, after his breakout from the Normandy beachhead.

The need for communication between the homelands and many far-flung theatres of war gave rise to the need for improved long-range overseas communication systems. A system of radioteletypewriter relaying was devised, by which a radioteletypewriter operator in Washington, London, or other capitals could transmit directly by teleprinter to the commander in any theatre of war. In addition, a system of torn-tape relay centres was established so that tributaries could forward messages through the major centres and retransmission could be effected in a minimum of time by transferring the perforated tape message from the receiving to the transmitting positions. In addition a system of holding teletypewriter conferences was developed. These conferences were called "telecon" and enabled a commander or his staff at each end to view on a screen the incoming teletypewriter messages as fast as the characters were received. Questions and answers could thus be passed rapidly back and forth over the thousands of miles separating the Pentagon in Washington, D.C., for example, from SHAEF headquarters in Europe or Gen. Douglas MacArthur's headquarters in the Far East.

During the latter years of the war, new and improved communication and electronic devices came forth from research and development in ever-increasing numbers. A new long-range electronic navigation device, known as loran, used for both naval vessels and aircraft, appeared, as did short-range navigational systems, called shoran. Combinations of radar and communications for the landing of aircraft in zero visibility were perfected. One such system was the GCA or ground-controlled approach system. Combinations of radio direction-finding, radar, and communications systems were developed and used for ground control of intercept aircraft—the system called GCI (ground-controlled intercept). Radio-controlled guidance of falling bombs was brought forth to enable the operator in the bomber to direct the bomb to the target. Electronic countermeasures made their appearance in the form of jamming transmitters to jam radio-communication channels and radar, navigation, and other radio aids to military operation.

At the end of World War II the military forces of the Allied nations were called upon to occupy the territories of their former enemies. The Allied signal communications staffs were usually charged with responsibility for the supervision of military and civilian communication within the occupied areas.

The military services learned well from their wartime experiences the importance of scientific research and development in all fields, including communications electronics. Advances were made toward increasing the communication capacity of wire and radio relay systems and toward improving electronic aids for navigation and for the detection of enemy forces. Measures to provide more comprehensive and more reliable communication and electronic equipment continued to be stressed in the armies, navies, and air forces of the major powers. Further impetus to these developments came from the conflict in Korea, which reaffirmed the vital importance of signal communication in the coordinated and combined operations of air, ground, and naval forces.

After mid-century, accordingly, military efforts in all the many facets of signal communication continued to intensify almost as extravagantly as during World War II. Two major additions in the U.S. Army were television and "electronic brain" equipment. The latter, in many forms of digital and analogue computers and of such data-processing devices as punch-card machines, were applied increasingly to personnel record handling and to depot and supply operations interconnected over wide areas by signal-communication networks.

Television systems proved valuable as training aids in military schools where mass instruction, especially in manual functions, was needed and where instructors were few. Such systems enabled a single instructor to teach many small classes simultaneously, each grouped before a receiver where they could watch demonstrations closely. Two-way communication systems permitted the instructor to call and question any student in any classroom and, inversely, enabled any student to put questions to the instructor. Portable television equipment in the field proved valuable for sending back to headquarters, by antenna radiation or coaxial cable, a picture of any scene of operations such as a river crossing. Equally valuable was a television camera in the hands of a forward scout or in a reconnaissance aircraft, whether piloted or remotely controlled, to scan enemy territory.

Thus signal communication, combining in itself the powers of photography, television, radar, and other instruments using the electromagnetic radiation spectrum, moved into such new areas of military communications electronics as battle area surveillance and electronic warfare devices to interfere with, or jam, enemy transmitters. In the U.S. Army, battle area surveillance activity radically augmented conventional reconnaissance methods. An electronically controlled target acquisition system, to discover enemy troops or transport on the ground or in the air, was being developed by the Signal Corps using optical, sonic, photographic, infrared, and radar equipment. The aggregate of information gathered by these devices over a wide enemy front can be electronically assembled and displayed in the headquarters where the combat commander can quickly estimate the situation and make tactical decisions.

Military signal communications in the 1960s began to employ a new technology that enabled it to attain very great ranges, coupled with high-quality service. This was communications by satellite relay. The military application by the U.S. Department of Defense followed initial experiments by the U.S. Army Signal Corps, which produced Score and Courier in 1958 and 1960 (both launched into orbit by the Air Force). An approach to an effective system supplementing the global U.S. military networks was accomplished by the two Syncom satellites developed and launched by the National Aeronautics and Space Administration in 1963 and 1964. The U.S. Department of Defense took control of the Syncom system in 1965 for communications relay through Army ground stations and Navy shipboard terminals.

See also references under "Signal Communication, Military" in the Index.

Bibliography.—Hanns Grosser, *Die Führungstruppe* (1944); T. B. Gravely, *The Second World War 1939–45: Signal Communications* (1950); Dulany Terrett, *The Signal Corps: the Emergency* (1956); G. R. Thompson *et al.*, *The Signal Corps: the Test* (1957); G. R. Thompson and D. Harris, *The Signal Corps: the Outcome* (1966).

(G. I. Bk.; G. R. T.)

SIGNAL CORPS. Organizations of this or an equivalent title have existed in the armies of many nations since the creation of the first signal corps in the Union and Confederate armies of the American Civil War.

United States.—The Signal Corps, U.S. Army, dates from June 21, 1860, when Congress authorized the post of signal officer, filled on June 27 by Maj. Albert J. Myer, assistant surgeon, who had won War Department acceptance of a simple but effective method of flag signaling (wigwag). Obtaining officers on detail, he trained a number, some of whom later served with the Confederate Army and helped establish the Confederate Signal Corps in 1862.

The provision of signal communication and related tasks of observation brought the Corps a variety of responsibilities, including weather reporting over Army wire lines from 1870 until that responsibility was transferred to the Weather Bureau in 1891, observation balloons, early Army aircraft, and Army photography. The Signal Corps made still pictures and, later, motion pictures, for many purposes, including military analysis and tactical reconnaissance; it produced and distributed all Army training films. Responsibility for electrical communications (lost during the Civil War to the civilian military telegraph organization but recovered soon after) led to wire-line construction and operations to serve Army posts and camps throughout the U.S. territorial west, and later to the provision of all communications in Alaska. The Alaska communication system, assigned to the Signal Corps in 1900, became the principal means by which civilians and government agencies in Alaska maintained rapid communication with the United States. Installed and operated by the Signal Corps, the system utilized submarine cable, overland pole lines, and radio.

Because the Signal Corps had responsibility for Army communications it was also made responsible for the use of radar and television. Army radar equipment was developed by the Signal Corps prior to World War II. Many types were used during the war, especially in aircraft detection and in gun laying. In 1946, using high-power radar designed at its laboratories at Fort Monmouth, N.J., the Signal Corps succeeded in receiving radar reflections from the moon. Television closed-circuit systems for use in Army schools and field television for use in battle area surveillance were developed (*see* Signal Communication, Military).

Signal Corps research and development laboratories worked upon a wide array of electronic aids, such as individual helmet radios for use in short-range communication with each soldier in combat, and theatre-wide area communication systems for the interconnection of headquarters by multiple routes. Such developments as these were necessary to provide the flexibility and universality needed for the command control of widely dispersed mobile forces.

To meet its responsibility for the Army strategic communications network, the Signal Corps installed and operated automatically relayed high-capacity radio and wire circuits the world over. Experimental work was done toward global communications by satellite relay, leading to the first successful satellite for communication purposes, the relatively simple Score, in December 1958, and a more sophisticated type, Courier, in October 1960.

The combat support mission of the Signal Corps included the installation, operation, and maintenance of communications within all commands down to and including the brigade. Communications within regiments and smaller units were handled by the personnel of the several arms. Field units of the Signal Corps were organized in signal operations, construction, and service battalions and in operations, construction, photographic, repair, depot, and service companies. Service companies and battalions were composed of teams, organized and equipped to perform one of the technical functions of the Signal Corps.

Unlike the signal organizations of the armies of other nations, the U.S. Army Signal Corps exercised for over a century full responsibility for development, production, and distribution of military communications equipment. But following a reorganization of the Army in 1962 these functions were removed to the newly created Army Materiel Command.

Another reorganization of March 1, 1964, eliminated the remaining Signal Corps operating organization when Signal Corps schools were transferred to the Continental Army Command, and the strategic communications function to the newly created U.S. Army Strategic Communications Command. The chief signal officer, redesignated chief of communications–electronics, retained advisory and supervisory functions only in the U.S. Army General Staff.

Other Countries.—In European armies the signal function, dating generally from the first military applications of the electrical telegraph in the mid-19th century, developed under engineer organizations. But the rapid multiplication of the means of signal communication, especially with the advent of radio early in the 20th century, took military signals from the engineers and established the function as a separate and independent organization. The commander usually serves on the general staff while directing also the many-sided operations of the signal communication system of a modern army. In Great Britain the director of signals in the War Office became head of the British Royal Corps of Signals. This Corps, which took form in 1920 from the signal service of the Royal Engineers, provides communication facilities down to brigades and forward to infantry battalions and artillery battery headquarters. In France the signal service of the army was lodged, after World War I, with the engineers until midcentury when it separated as the *Corps des Transmissions*. The commander, directly responsible to the minister of national defense, is charged with the organization and operations of the corps, the training of signal personnel and the provision of technical advice on research and development and on supply requirements. In Germany signal troops served as a separate organization, the *Nachrichtentruppe*, after 1899.

Signal communications in the U.S.S.R. operated as a general staff function directly under the Soviet war minister. Soviet signal troops serve under the ground forces along with the infantry, cavalry, and engineers. In Japan the signal section, one of the technical services of the national safety forces of the country, is represented on the general staff by the chief of the signal section. A signal company serves each army division, performing functions similar to those in the U.S. Army.

See R. F. H. Nalder, *The History of British Army Signals in the Second World War* (1953); Dulany Terrett, *The Signal Corps: the*

Emergency (1956); G. R. Thompson *et al.*, *The Signal Corps: the Test* (1957); G. R. Thompson and D. Harris, *The Signal Corps: the Outcome* (1966). (G. R. T.)

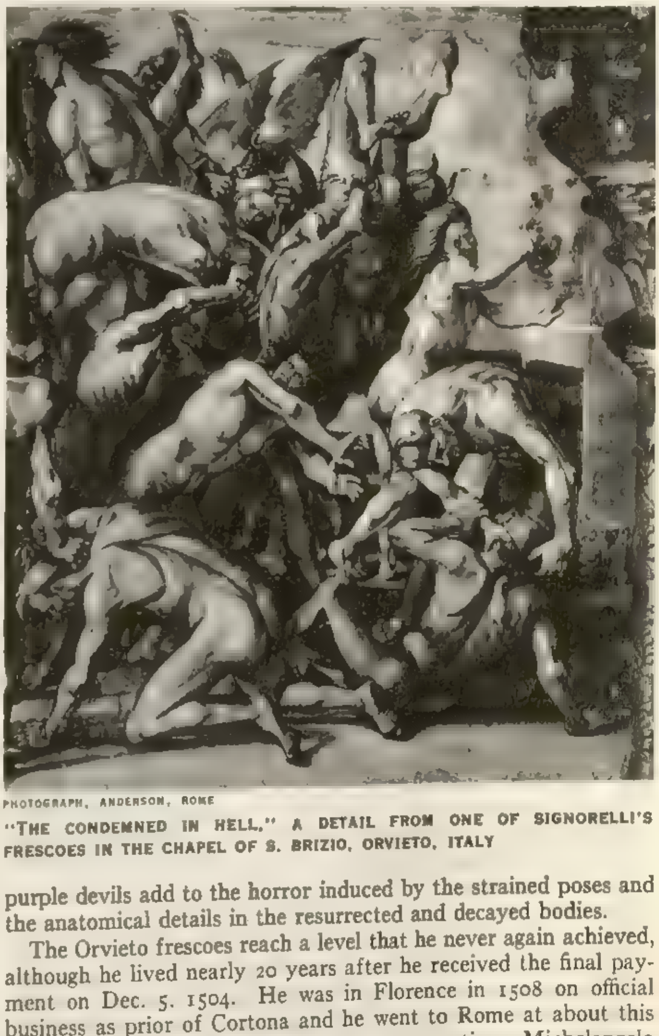
SIGN LANGUAGE, a means of communication through bodily movement. Organized sign language is found in many parts of the world as a kind of lingua franca (*q.v.*). Among the Plains Indians (*q.v.*) of North America, tribes with mutually unintelligible speech communicated through a system of hand movements that was understood throughout the Great Plains. Possibilities for communication were considerable; for example, the Kiowa Apache communicated with the Kiowa mainly by gestures. This sign language was comparable to that of deaf-mutes, but not identical (*see* DEAF AND HARD OF HEARING, TRAINING AND WELFARE OF). Among the Angami Nāga tribe of Assam, a remarkably flexible sign language permits real inventiveness and ingenuity on the part of the communicator, but the meaning of the signs is generally clear. Gestures formed the basis for an extended system of signs among such Australian aborigines as the Arunta.

Gestures, facial expressions, and other bodily movements facilitate communication among all peoples, kinesically supplementing and emphasizing verbal behaviour (*see* LANGUAGE: *Language in Culture: Communication*), providing traditional forms of greeting (*see* SALUTATIONS; SALUTE), and allowing for emotional expression and release of tension (*see* EMOTION). For ritualistic and theatrical aspects of sign language *see* DANCE; DANCE, PRIMITIVE; MIME; PANTOMIME.

BIBLIOGRAPHY.—G. Devereux, "Some Mohave Gestures," *American Anthropologist*, vol. 51 (April 1949); L. J. Davidson, "Some Current Folk Gestures and Sign Languages," *American Speech*, vol. 25 (Feb. 1950); W. D. Brewer, "Patterns of Gesture Among the Levantine Arabs," *American Anthropologist*, vol. 53 (April 1951); A. Lyall, "Italian Sign Language," *Twentieth Century*, vol. 159 (June 1956); W. Tomkins, *Universal Sign Language of the Plains Indians of North America*, 13th ed. (1958); W. C. Stokoe, Jr., *Sign Language Structure . . .* (1960); G. J. Brault, "Kinesics and the Classroom: Some Typical French Gestures," *The French Review*, vol. 36 (Feb. 1963); R. M. Hughes, *The Gesture Language of the Hindu Dance* (1964). (F. R. E.)

SIGNORELLI, LUCA (LUCA DA CORTONA, originally LUCA D'EGIDIO DI VENTURA DE' SIGNORELLI) (*c.* 1445/50–1523), Italian painter, the most powerful exponent of the nude before Michelangelo, was born in Cortona, Umbria, probably between 1445 and 1450. He was first recorded as a painter in 1470, but it is very likely that he was a pupil of Piero della Francesca, along with Perugino, in the 1460s. The first certain surviving work by him is a fragmentary fresco, now in the museum at Città di Castello, which he finished in Nov. 1474. This shows a strong influence from Piero della Francesca, and the style is responsible for the attribution to Signorelli at this date of several other pictures, including the beautiful Madonna in Christ Church library, Oxford. His first signed work was a processional banner with a Madonna on one side and a Flagellation on the other; these are now in the Brera, Milan, as separate pictures. They still show links with the style of Piero, but the dominant influence is that of Florence, and especially that of the Pollaiuoli, which suggests that Signorelli visited Florence in the 1470s. The scientific naturalism represented by the Pollaiuoli now becomes the principal feature of Signorelli's style. In 1479 he was elected to the council of 18 in his native Cortona and for the rest of his life he was active in politics.

About 1483 he went to Rome, for the "Testament of Moses" fresco in the Sistine chapel is unanimously attributed to him; yet his name does not occur in the only two documents—of 1481 and 1482—known to us. It is likely that his friend Perugino got him the commission to help complete the cycle. By this date his style was fixed and his interest in dramatic action and the expression of great muscular effort mark him as an essentially Florentine naturalist working outside Florence. The altarpiece for Perugia cathedral, of 1484, shows the same qualities. Between 1497 and 1498 he was at work on a fresco cycle of scenes from the life of St. Benedict in the monastery at Monteoliveto Maggiore, near Siena, but these are hardly more than a prelude to his masterpiece, the frescoes of the "End of the World" and the "Last Judgment" in the chapel of S. Brizio in Orvieto cathedral. A small part of the ceiling had been painted by Fra Angelico in the 1440s, and he had determined the subject by painting the "Christ in Judgment" over the altar. On April 5, 1499, Signorelli was commissioned to complete the ceiling and it was not until April 27, 1500, that he received the more important commission to paint the walls. At this moment Italy was in turmoil, partly because of the incursions of the French armies and partly because of the preaching and death of Savonarola, who seemed to have predicted the French invasion as a judgment. The French had actually by-passed Orvieto on their way south, and the completion of the frescoes was therefore in the nature of a thanksgiving. It also explains the grim and terrible treatment of a subject well suited to Signorelli's genius. The large scenes on the walls represent the last days of the world, with the portents of doom followed by the appearance of Antichrist and the false miracles worked by him before his final destruction. In the frescoes nearer to the altar we see the next stages, the "General Resurrection" and the "Last Judgment," "Heaven" and "Hell." The "Judgment" itself is represented only by the small figure of Christ over the altar, painted by Angelico. Behind the altar are a Dantesque "Hell" (with the classical figure of Charon and the River Styx) and a part of the "Blessed in Heaven." Signorelli had little sense of colour, but here his greenish and



PHOTOGRAPH, ANDERSON, ROME

"THE CONDEMNED IN HELL," A DETAIL FROM ONE OF SIGNORELLI'S FRESCOES IN THE CHAPEL OF S. BRIZIO, ORVIETO, ITALY

purple devils add to the horror induced by the strained poses and the anatomical details in the resurrected and decayed bodies.

The Orvieto frescoes reach a level that he never again achieved, although he lived nearly 20 years after he received the final payment on Dec. 5, 1504. He was in Florence in 1508 on official business as prior of Cortona and he went to Rome at about this time, and again in 1513, but the new generation—Michelangelo and Raphael—had now won all the commissions and Signorelli returned to his less sophisticated Umbrian clientele. Most of his later works betray the hands of his numerous assistants, but among his better works are the altarpiece in the National gallery, London (1515), the "Deposition" (1515/16) in Sta. Croce, Umbertide, and the "Assumption" and "Immaculate Conception" begun in 1519 and 1521 (both in the Museo Diocesano, Cortona). One of his finest works, the "Pan" formerly in Berlin, was destroyed in World War II. He died in Cortona on Oct. 16, 1523.

See G. Vasari, *Lives . . .*, Eng. ed. by C. du G. de Vere, vol. iv (1913); L. Dussler, *Signorelli* (1927); M. Salmi, *Luca Signorelli* (1953). (P. J. My.)

SIGURD, the name of two kings of Norway.

SIGURD I (*c.* 1090–1130), nicknamed Jorsalafar (Jerusalem-farer), king from 1103 to 1130, an illegitimate son of Magnus II Barfot, succeeded to the throne, together with his elder and younger brothers Eystein and Olaf, on Magnus' death. About 1107 Sigurd sailed for the Holy Land, first visiting England, France, Spain, and Sicily. He bathed in the Jordan and was lavishly entertained by Baldwin I, king of Jerusalem. Leaving his ships at Constantinople, as a gift to the emperor Alexius I, he reached Norway (1110) overland. After the deaths of Olaf (1115), who never ruled in fact, and of Eystein (1122), Sigurd was sole ruler until his own death in 1130. He strengthened Norway's defenses and placed the Church on a secure footing by introducing tithes. In his latter years he became mentally unbalanced.

SIGURD II (1134–1155), nicknamed Munnr (Mouth), was an illegitimate son of Harald IV Gille, on whose death (1136) he and Harald's legitimate son Ingi, both minors, were proclaimed joint kings. Their supporters withstood other pretenders to the throne until, in 1142, they were joined by Eystein, who claimed also to be a son of Harald and was given a third of the kingdom. Sigurd was killed in 1155 by the supporters of Ingi. (G. T.-P.)

SIGURD, a figure in Germanic heroic literature: *see* SIEGFRIED.

SIGURDSSON, JÓN (1811–1879), Icelandic statesman and scholar who, as leader of the Patriotic Party, worked tirelessly to obtain some degree of self-government for his country. He was born in the west of Iceland on June 17, 1811. After studying classical philology and ancient history, and later political theory and economics at the University of Copenhagen (1833–35), he worked (1835–48) on the Arnamagnean collection of Icelandic manuscripts and was thereafter, until his death, secretary of the Arnamagnean Commission. Throughout his life he edited, either alone or in cooperation with others, a series of ancient texts, including a number of sagas (in *Islendinga Sögur*, vol. i and ii, 1843–47), the *Edda Snorra Sturlusonar* (1848–52), *Lovsamling for Island*, 17 vol. (1853–77), *i.e.*, a collection of Icelandic law, and the first volume (1857) of the series *Diplomatarium Islandicum*.

One of the greatest of Icelandic scholars, Sigurdsson achieved equal distinction as a statesman. Working in Copenhagen, he became involved in the discussions which preceded the authorization (1843) by the Danish king Christian VIII of the revival, as an advisory body, of the *Althing*, the ancient legislative assembly of Iceland. Elected a representative for its first meeting (1845), Sigurdsson rapidly became one of the most influential members of the *Althing* and was for many years speaker of the lower house. On the establishment (March 1848) of representative government in Denmark, Sigurdsson immediately demanded virtual self-government for Iceland. But at a national assembly held at Reykjavík (1851) it was made clear that the Danes would grant to Iceland only the limited autonomy allowed to the various districts of Denmark. Acceptance of such terms would have constituted Iceland a mere province of Denmark, and Sigurdsson, supported by the majority of delegates, categorically rejected them.

For the following 20 years Sigurdsson and his party continued to reiterate their demands. He also sought to improve his country's economic situation, and partly as a result of his agitation the Icelandic trade was finally freed from Danish monopoly and opened to all nations in 1854. He also encouraged the development of modern methods of agriculture and fisheries. The crown of Sigurdsson's achievements came in 1874, the one-thousandth anniversary of the settlement of Iceland, when King Christian IX granted Iceland a constitution. The *Althing* was now to control Icelandic finance, and to share legislative power with the crown. Complete home rule was not won until the 20th century, but this preliminary step was described by Sigurdsson as "a good beginning." Sigurdsson died in Copenhagen on Dec. 7, 1879, and was buried at Reykjavík.

See T. Johnson, *Jon Sigurdsson, the Icelandic Patriot* (1887); P. E. Ólason, *Jón Sigurtsson* (1940). (T. J.)

SIIRT, capital town of an *il* (province) of the same name in southeastern Anatolia, Turkey, lies on the Buhtan River in the southeastern foothills of the Taurus Mountains, and 95 mi. (153 km.) E of Diyarbakir. Pop. (1960) 22,944. Before the disintegration of the Ottoman Empire Siirt was an important trade centre of a large region which included northern parts of Syria and Iran. Its economic importance is now local. The chief industrial activity is the manufacture of blankets, which are well known throughout Turkey. The town is linked by road with Diyarbakir, Cizre, and Bitlis.

SIIRT IL extends from the southeastern foothills of the Taurus in the northeast to the Tigris Valley in the southwest. Area 4,447 sq.mi. (11,518 sq.km.). Pop. (1960) 232,243. It is drained by several tributaries of the Tigris, the chief being the Buhtan and the Garzan. The climate is severe with heavy snowfall in winter. The mountains are well forested, whereas the foothills and plains in the south have steppe vegetation. The cultivation of cereals and livestock produce are the chief sources of income. In the southwest is the Raman oil field, and oil is refined at Batman nearby. (N. Tu.; S. Er.; E. Tu.)

SIKANG (HSI-K'ANG SHENG), a former Chinese province made up of parts of western Szechwan and eastern Tibet in 1928 and dissolved in 1955. Area 204,247 sq.mi. (529,000 sq.km.); pop. (1953) 3,381,064. Chinese claims to controlling all Tibet vary both in age and degree of control, but at least as far back as 1751 the Ch'ien-lung emperor established his Chinese commissioners in Lhasa with full political power over the Dalai Lama. In the late 19th century, the British authorities in India contested this greatly weakened control, and in 1913 supported a Tibetan independence movement, urging that China divide the area into an Inner and Outer Tibet with a consequent division between Chinese and British influence. Although China did not accept these changes, in 1928 eastern or Inner Tibet ("Kham" in Tibetan) was split into Ch'ing-hai and Sikang provinces, while Outer Tibet was declared a special territory within China. Under the Chinese People's Republic the eastern half of Sikang, from Ya-an, the former capital, near the Ta-tu River west to Pa-tang on the Chin-sha or upper Yangtze, was incorporated into Szechwan province, July 1955, while the western half (Chamdo) was placed back in Tibet or Hsi-ts'ang province.

Sikang averages 13,000 ft. above sea level, a vast plateau marked by jagged snow-covered ranges, canyons, coniferous forests, streams, and lakes. Through it run the upper courses of the Yangtze with several of its tributaries, and the Mekong, Salween, and Brahmaputra rivers. Its highest known peak, Minya Konka, rises 24,891 ft. above sea level. The dominant economies are subsistence farming of barley, dry rice, potatoes, and a few vegetables on the narrow valley floors and stream terraces, and herding on the plateaus. Horses, wool, furs, musk, medicinal plants, gold, and asbestos have been the main products exchanged for brick tea, salt, tobacco, cloth, and other consumer goods; coal and rich iron ore deposits are found in the eastern parts. The Chinese built two heavy motor roads into the area, largely along ancient caravan trails.

See also CH'ENG-TU. (TE. H.)

SIKAR, a town in Rajasthan, India, and the administrative headquarters of the district of the same name, lies about 65 mi. (105 km.) NW of Jaipur city on the Western Railway. Pop. (1961) 50,636. About seven miles to the southeast of the town is a ruined temple, which is reputed to be many centuries old.

SIKAR DISTRICT has an area of 3,027 sq.mi. (7,840 sq.km.) and a population (1961) of 820,286. The soil is mainly sandy; in the eastern part of the district sandy loam soil is also found. The average annual rainfall is about 15 in. (381 mm.). The chief crops are jowar, bajra, barley, and gram. Other important towns are Fatehpur, Lachmangarh Sikar, Ramgarh, Neem-Ka-Thana and Sri Madhopur.

There is a cotton ginning factory at Neem-Ka-Thana and *khadi* (homespun cotton cloth) is made at Reengus, which is also an important junction on the Western Railway. (S. M. T. R.)

SIKELIANOS, ANGELOS (1884–1951), one of the leading modern Greek lyrical poets, was born on the island of Leukas.

His first important work, the *Alafroiskiotos* was published in 1909, and revealed his lyrical powers. It was followed by a group of outstanding lyrics. His next period was introduced by the philosophic poem *Prologos tes Zoes* ("Prologue to Life"; 1917) and includes the long works *Meter Theou* ("Mother of God") and *Pascha ton Hellenon* ("The Greek Easter"), culminating in the *Delphikos Logos* ("Delphic Utterance"; 1927). The Greek tradition and the national historic and religious symbols are here given a mystic turn and a universal significance. In the 1930s and 1940s there appeared a second group of lyrics, which display the full power of Sikelianos' art. They express in rich and incisive language and with forceful imagery the poet's belief in the beauty and harmony of the world. The tragedies of Sikelianos (*Sibylla, Daedalus in Crete, Christ in Rome, The Death of Digenis* and *Asklepius*, which are introduced by the long dramatic poem *The Dithyramb of the Rose*) are more notable for their lyric than their dramatic qualities.

In his mature works Sikelianos is fulfilling in poetry the aspirations of the "demotic movement" of the 1880s, which sought to combine Greek tradition with western thought, and to introduce as a consciously literary language the idiom of the people. Although occasionally the power of his inspiration drives him to grandiloquence which blunts the poetic effect of his work, some of his finer lyrics are among the best in 20th-century western literature.

Sikelianos died at Athens, June 19, 1951.

See P. S. Sherrard, *The Marble Threshing Floor; a Study in Modern Greek Poetry* (1956). (Ce. A. T.)

SIKHISM, an Indian religion combining Islamic and Hindu beliefs, founded in the late 15th century A.D. by Nanak, the first guru (teacher). The word "Sikh" is derived from the Sanskrit *shishya* ("disciple"). Sikhs are disciples of their ten gurus, some of whose writings are compiled in the Granth Sahib, the Sikh sacred book (*see* GRANTH). An act of the Indian legislature defines a Sikh as one who "believes in the ten gurus and the Granth Sahib." The Sikhs, approximately 8,000,000 in number, are mainly concentrated in East Punjab (*see* PUNJAB) where in the 19th century they created a powerful state under Ranjit Singh before the British took over the country. The majority of them belong to agricultural tribes of Indo-Scythian stock. Small communities of Sikhs live also in the United Kingdom, Canada, the United States, East Africa, Burma, Malaya, and Hong Kong.

The Ten Gurus.—Muslims began invading India in the 8th century A.D. By the 15th century they had conquered the whole of northern India and converted to Islam large numbers of Hindus, particularly in the Punjab and Bengal. This impact of Islam on Hinduism generated movements aiming at a compromise between the two faiths. Nanak (1469–1539), the founder of Sikhism, was the son of a Hindu revenue official of a village called Talwandi now in West Pakistan. From an early age he began to seek the company of wandering hermits. Although he took occasional employment, married and reared a family, he remained preoccupied with spiritual matters. At the age of 30 he gave up other pursuits and became a teacher. The main point of his teaching was to emphasize the need of a rapprochement between Hinduism and Islam. He was influenced by the teachings of the exponents of *bhakti* (the devotional movement), notably Kabir (*q.v.*), and by the Muslim Sufis (*see* SUFISM). He rejected many of the cherished beliefs and practices of Hinduism and accepted instead those of Islam. He was a monotheist, opposed to the worship of idols, and a strong critic of the caste system. He advocated the repetition of *nam* (name of the Lord) as a means to salvation. Nanak traveled extensively all over India and through the Arab countries of the Middle East.

Nanak chose as his successor a disciple named Angad (1504–52), who traditionally invented the Gurmukhi script (*see* PANJABI LANGUAGE) and built *gurdwaras*, as Sikh temples are called, from which the teachings of the founder could be propagated. On his death, Angad chose a disciple, Amardas (1479–1574), to be the third guru. Amardas nominated his son-in-law, Ramdas, to be the fourth, and thereafter the gurus came from one family. Ramdas (1534–1581) had a large tank dug, around which his son Arjan (1563–1606) built the city of Amritsar (*q.v.*), making it a place of pilgrimage for the Sikhs. Arjan compiled the Granth by collecting the writings of his predecessors and of some Hindu and Muslim saints whose teachings agreed with those of the gurus, and adding to them some hymns of his own. He also built many temples, and for the purpose received offerings from his followers. The Mughal emperor Jahangir, perturbed by Arjan's influence, had him arraigned on a charge of nonpayment of taxes and of vilifying Islam. He was tortured and died on May 30, 1606, at Lahore.

After the execution of Arjan the Sikhs began to change from a pacifist to a militant sect. Arjan's son, Hargobind (1595–1644), began to drill his followers and equip them with arms, as a result of which he was arrested and kept in Gwalior fort for some time. On his release he resumed his activities and fought a few minor engagements with the Mughals. Hargobind's grandson Har Rai (1630–1662), followed by his young son Harikrishen (1656–1664), made no significant contribution to the development of the Sikh religion. The ninth guru, Hargobind's son Teg Bahadur (1621–1675), who organized Hindu resistance against the bigotry of the Mughal emperor Aurangzeb, was executed in Delhi on Nov. 11, 1675.

Teg Bahadur's son, Gobind Rai (1666–1708), took the final steps in militarizing the Sikhs. On the Hindu New Year in 1699 he assembled his followers at Anandpur, founded by his father, in the foothills of the Himalayas, and initiated five of them as members of a fraternity which he named the Khalsa ("pure"): they drank *amrit* (nectar) out of the same bowl (though they all came from different castes), received new names with the suffix Singh ("lion"), and swore to keep the five K's: to wear long hair (*kesh*), a comb (*kangha*) in the hair, soldier's shorts (*kachkha*), a steel bangle (*kara*) on the right wrist, and a sabre (*kirpan*). After initiating the five, the guru was in his turn initiated by them and renamed Gobind Singh. At the end of the initiation the guru hailed his new converts with the words *Wah Guru ji ka Khalsa, Wah Guru ji ki Fateh*—"The Khalsa are the chosen of God, victory be to God."

In the days following, some 80,000 persons were initiated into the Khalsa fraternity. Many Sikhs, however, known as Sahajdharis (those who take time to adopt the new faith), preferred to remain clean-shaven. The division between the Khalsa and the Sahajdhari continues to this day.

Gobind Singh lost all his four sons in his lifetime and declared the succession of gurus at an end, investing the Granth, of which he produced the final edition, as the immortal guru of the Sikhs. His own hymns were compiled separately in a volume called the *Dasam Granth*, which is not given the sanctity accorded to the Granth Sahib. Gobind Singh was assassinated on Oct. 7, 1708, while he was in Nanded in Hyderabad state.

Sikh Rise to Power.—Before his death Gobind Singh charged a Hindu ascetic, Lachhman Das (1630–1716), to avenge the tyranny perpetrated by the Mughals on the Sikhs. Lachhman Das, who took the name Banda Singh Bahadur at his initiation, roused the non-Muslim peasantry and in a series of whirlwind campaigns destroyed the Muslim landed class in southern Punjab. After eight years of defiance Banda was starved to surrender, and along with 700 of his followers he was executed in Delhi on June 19, 1716.

For many years following Banda's execution the Sikhs were savagely persecuted by the Muslim governors of the Punjab. But the situation eased with the invasion of the Persians under Nadir Shah in 1738–39, followed by the series of incursions of the Afghans led by Ahmad Shah Durrani, who in 1761 defeated the Marathas at Panipat. The Sikhs, who had kept out of the way of the invader, filled the vacuum he left under their remarkable leaders, chiefly Kapur Singh and Jassa Singh Ahluwalia. They divided the Sikh forces into 12 *misls* or militias with their own spheres of operations. The strength of the *misls* varied from 20,000 horsemen of the most powerful, the Bhangis of Gujrat, to a few hundred of the smaller ones. The *misls* built chains of forts and levied *rakhi* (protection money) from towns and villages. In 1761 Jassa Singh Ahluwalia occupied Lahore for a short while and struck coins in the name of the *misl* confederacy. In 1765, when the

Bhangis took Lahore, the power of the Sikh *misls* extended from the Indus in the west to the Ganges in the east, and from the Himalayas in the north to the desert wastes of Sindh in the south.

The *misls* did not retain their cooperative character for long. As soon as they had dispossessed the Mughals, they began to fight among themselves, until Ranjit Singh of the Sukarchakia *misl* absorbed the others and created a unified Sikh State.

The Sikh Empire.—Ranjit Singh (*q.v.*) became chief of the Sukarchakias at the age of 12. With the help of his mother-in-law, Sada Kaur, widow of a chieftain of the Kanahya *misl*, he captured Lahore in the summer of 1799 and proclaimed himself maharaja of Punjab. Sada Kaur also helped him to take Amritsar and destroy the *misls* north of the Sutlej as well as the petty Muslim and Hindu kingdoms of the Punjab plains and hills. Ranjit Singh's ambition to extend his power over the rest of the Punjab was curbed by the English, who in 1809 forced him to sign the Treaty of Amritsar and to accept the Sutlej River as his eastern boundary. He captured Multan in 1818 and Kashmir in 1819. In 1822 he began to modernize his army by employing officers of Napoleon's disbanded army to train his troops. In 1823 he defeated the Afghans and planted his flag on the ramparts of Peshawar; Sikh rule was firmly established there in 1834. His Dogra underlord, Gulab Singh, extended the Punjab's frontiers beyond Ladakh into little Tibet (Baltistan). In 1638 Ranjit Singh signed a tripartite treaty with the British and the exiled Afghan Shah Shuja for a joint invasion of Afghanistan.

Dissolution of the Sikh Empire and the Anglo-Sikh Wars.—After the death of Ranjit Singh, the Sikh kingdom rapidly disintegrated into anarchy. His eldest son and successor, Kharak Singh, died of drink, and Kharak Singh's son, Naonihal Singh, was killed in an accident on the same day (Nov. 5, 1840). Naonihal Singh's wife, Maharani Chand Kaur, ruled for a few months till she was ousted (and later murdered) by Ranjit Singh's second son, Sher Singh. Sher Singh ruled for two years till his assassination on Sept. 15, 1843. Ranjit Singh's youngest son, Dalip Singh, was proclaimed maharaja with his mother, Maharani Jindan, as queen regent, but power in fact passed into the hands of the army, which began to dictate to the maharaja and his ministers.

The English, apprehending hostilities, began to move their forces to the Punjab frontier, and there followed what are known in English history as the Sikh Wars of 1845–46 and 1848–49. For these, *see* SIKH WARS. After the Sikh defeat, the Punjab was annexed to British India. Maharaja Dalip Singh, granted a pension, left India to reside in England, and died on Oct. 22, 1893.

Sikhs Under British Rule.—After three years, the three-man Board of Administration set up to rule the Punjab was abolished and Sir John Lawrence became chief commissioner. Having seen the fighting quality of the Sikhs, he threw open recruitment to them, a policy which paid dividends as the Sikhs helped the British to quell the Indian Mutiny of 1857. After the Mutiny, the proportion of Sikhs in the British Army in India was increased. New regulations requiring Sikh soldiers to observe the outward symbols of the Khalsa faith (*i.e.*, keeping their hair and beards unshorn) were largely responsible for retaining the separate identity of the Sikhs.

After the British annexation several movements, religious, social, and political, occurred among the Sikhs. The Nirankaris aimed at reforming religious ritual. The Namdharis or Kukas under Ram Singh sought to reestablish Spartan traditions of the early Khalsa and restore Sikh rule. In January 1872 the Namdharis clashed with the police and 66 of them were apprehended and blown off from cannons at Maler Kotla; Ram Singh was exiled to Rangoon where he died in 1885. The Singh Sabha movement which came into prominence at the turn of the century was essentially educational and literary. It enjoyed the patronage of English administrators, with whose collaboration Sikh schools and colleges were founded in different parts of the Punjab. The movement stimulated interest of European scholars in Sikhism, the most notable of them being M. A. Macauliffe who produced a six-volume study of Sikh religion (1909). Among the Sikh scholars associated with the movement was the poet-philosopher Vir Singh (1872–1957).

Anglo-Sikh collaboration was closest during World War I. Recruitment for active service was considerably higher among the Sikhs than among any other Indian community, and Sikhs fought with distinction in France, Africa, and the Middle East. Differences arose during the war, however, between the Sikhs and the British government. Sikh settlers in British Columbia and California who were subjected to discrimination formed a revolutionary organization known as the Ghadr ("Mutiny"). Many returned to India and for several months of 1915 carried on terrorist activities in central Punjab. The Ghadrites were crushed. Immediately after the war Sikhs clashed with the government over the control of their *gurdwaras*. The government supported the hereditary Mahants who were in possession. A group of radicals formed the party known as Akali ("immortals") to take over the *gurdwaras*. More than 30,000 men were imprisoned in the Akali movement, which lasted from 1922 to 1925. In that year the Sikhs Gurdwaras Act was passed, handing over all historic shrines to a 160-man elected body known as the Shiromani Gurdwara Prabandhak Committee. During World War II the Sikhs again cooperated with the British, but with considerably less zeal than they had done before. An influential group belonged to the Indian National Congress and a substantial proportion of the Japanese-sponsored Indian National Army was composed of Sikhs.

Sikhs in Independent India.—The relinquishment of power by Britain was accompanied by the partition of the country. The western part of Punjab became Pakistani territory, the eastern part Indian. The Sikhs left in Pakistan (about 2,500,000) moved to India and in turn ejected Muslims living in East Punjab. This mass exchange of populations was accompanied by extreme violence resulting in heavy loss of life. Most of the refugee Sikhs were rehabilitated in East Punjab. In many districts bordering on West Pakistan the Sikhs came to form a majority of the population. They also spread into the Terai region of Uttar Pradesh, Rajasthan, and Madhya Pradesh. The scale of compensation for refugees was so low as to cause severe hardship. A sense of grievance against the administration was generated, and the demand for an autonomous Sikh state gathered momentum. In March 1966 the Congress Party agreed to create in Punjab a separate Punjabi-speaking state in which Sikhs would constitute about 55% of the population. (*See* PUNJAB [INDIA].)

Religion and Way of Life.—The Sikh religion is an amalgam of the Muslim faith and Hinduism, simply and clearly expressed in the Panjabi language. There is one God, who is not represented by idols or images. Man should serve him by leading a good life in obedience to his commands, and by prayer, in particular by repeating the name of God, until, after his soul has passed through various existences by transmigration, he ultimately becomes one with God. Religious rites and customs which are meaningless formalities, such as bathing in "holy" rivers and making offerings to the dead, are to be rejected. Practices which injure health, such as drinking and smoking, are forbidden. There is no professional priesthood, all adults of either sex being eligible to carry out religious ceremonial, although in big Sikh communities there are professional singers and readers of the Granth. The Sikh religion is not ascetic, and the ideal is married life within normal human society. Although Sikhism officially deprecates the caste system and pronounces the equality of all men before God, it has been difficult in practice to escape from caste divisions, especially where marriage is concerned. The relationship between Sikhs and Hindus is in fact close, and Hindus in the Punjab honour the gurus.

The Sikh place of worship is the *gurdwara*, in which, in the most prominent position, is a copy of the Granth that is treated with great honour. The chief *gurdwara* is the Havimandir (known as the Golden Temple) at Amritsar, founded by Guru Arjan, several times destroyed, and finally rebuilt by Ranjit Singh. Sikh worship consists largely of reading from the Granth. Sikhs keep the Hindu festivals by performing ceremonies in their *gurdwaras*, and they also celebrate festivals on their own holy days, such as the birthdays of some of the gurus, by taking the Granth in procession through the streets. Sikh boys and girls undergo the initiation ceremony of the five K's at the age of puberty and boys take the additional name of Singh (though not all persons named Singh are Sikhs). The corresponding name for Sikh women is Kaur.

Sikhs make excellent farmers, soldiers, and mechanics. The proportion of literacy is higher among them than among any of the other major communities of India. Apart from the Khalsa College at Amritsar there are a number of medical, engineering, and technical institutes affiliated to the universities of the Punjab, Delhi, and Bombay.

BIBLIOGRAPHY.—J. D. Cunningham, *History of the Sikhs*, 2nd ed. (1853; rev. ed. by H. L. O. Garrett. 1919); L. Rama-Krishna, *Les Sikhs* (1933); Khushwant Singh, *A History of the Sikhs 1469–1839*, vol. i (1963); Hari Ram Gupta, *History of the Sikhs 1739–1799*, 3 vol. (1939–44); Ganda Singh, *Banda Singh Bahadur* (1935); M. A. Macauliffe, *The Sikh Religion*, 6 vol. (1909); Sher Singh, *Philosophy of Sikhism* (1944); *Selections from the Sacred Writings of the Sikhs*, Eng. trans. by Trilochan Singh *et al.* (1960). (K. SI.)

SIKHOTE-ALIN, a mountain complex of the U.S.S.R., is located in the Soviet Far East, along the coasts of the Sea of Japan and the Tatar Strait. The chain lies in the Primorskiy (Primorski; Maritime) Kray and Khabarovskiy (Khabarovsk) Kray of the Russian Soviet Federated Socialist Republic. It is orientated northeast–southwest for about 800 mi. (1,300 km.) from the lower Amur to Vladivostok. It is bounded by the sea to the southeast and the tectonic trench of the Ussuri Valley to the northwest. The Sikhote-Alin was formed in the Mesozoic folding period, but was much affected and uplifted in the Alpine period. The area is bounded by major fault lines and there is a general tilt to the west. The relief is complex, with eight main ranges rising to a maximum height of 6,818 ft. (2,078 m.) in Mt. Tardoki-Yani. The general westerly tilt means that the watershed lies fairly close to the coast and the main rivers, Khungari, Anyuy, Khor, Bikin, and Iman drain to the Ussuri-Amur. Although the highest, rounded summits are bare, most of the mountains are forested with pine, larch, fir, and birch. The lower slopes toward the Ussuri are densely covered with "jungle" of eastern deciduous species. Although a wide range of minerals has been found in the Sikhote-Alin, few are exploited, except coal in the south. The whole area is thinly populated. (R. A. F.)

SIKH WARS, two campaigns fought between the Sikhs and the British, which resulted in the conquest and annexation by the British of the Punjab in northwestern India.

The Sikh state in the Punjab (*q.v.*) was built up into a formidable power by Maharaja Ranjit Singh (*q.v.*; ruled 1792–1839). It extended from the Sutlej River to the Afghan hills and from Multan to the Kashmir mountains and Tibet. Its strength rested on its Western-trained army with a formidable train of artillery. Within six years of Ranjit's death, however, the state government broke down in a series of palace revolutions and assassinations. There were four rulers, the last a boy. (*See* SIKHISM.) The army got out of hand and organized itself in *punches* (committees). Finally the regent, Maharani Jindan, and her minister, Lal Singh, determined to invade British India under the pretext of forestalling a British attack.

The First Sikh War (1845–46).—Such were the conditions in the Punjab when the first Sikh War broke out in 1845. Relations with the British had already been strained by the refusal of the Sikhs to allow the passage of British troops through their territory during the first Afghan War (1838–42), and Sikh suspicions were heightened by Lord Ellenborough's defeat of the Gwalior Army in 1843. British troops were concentrated at Ferozepur, Ludhiana, and Ambala, so that neither side was taken by surprise. British forces amounted to 40,000 men with 94 guns as compared with the Sikh regular army of about 60,000 men with several hundred guns (some estimates are higher). With due allowances, the two forces were not unequally matched. The British had far greater reserves, but an early Sikh success might have roused all north and central India. The British Army possessed the advantage of unified leadership and confident belief in itself, whereas the Sikhs were divided, hesitant, and depressed.

The Sikh Army crossed the Sutlej on Dec. 11, 1845, and threatened Ferozepur. British preparations, however, were complete, and their main army marched 16 mi. from Ambala on the 12th and reached the neighbourhood of Ferozepur by the 18th. In the two battles at Mudki and Ferozeshah, the Sikhs under Lal Singh were thrown back across the Sutlej. A second Sikh army under Tej Singh failed to seize the opportunity of falling on the British when the troops were exhausted on the morning after the Ferozeshah battle. The losses were heavy on both sides, but the result was not decisive. In January 1846 the Sikhs made a dash for Ludhiana but were repulsed at Aliwal on Jan. 28. On Feb. 10 Sir Hugh (later Field Marshal Viscount) Gough, his heavy artillery having arrived, attacked and routed the Sikhs at Sobraon. The collapse of a bridge of boats caused the breakup of the Sikh Army and the Treaty of Lahore was concluded March 9.

By the treaty the Jullundur region and Jammu and Kashmir were ceded to the British and an indemnity of £500,000 was exacted. The Sikh Army was reduced to 20,000 infantry and 12,000 cavalry, and a British resident was installed at Lahore. Annexation was rejected on account of the martial nature of the people. For nearly two years the resident, Sir Henry Lawrence, successfully controlled the government, with his powers enhanced by another treaty signed at Bhairowal in December 1846. But the Sikhs remained restless and unconvinced that they had been finally defeated.

The Second Sikh War (1848–49).—Within months of Lawrence's departure on leave, Dewan Mulraj, the Sikh governor of Multan, rose against the Lahore government (April 1848). The governor general, Lord Dalhousie, was advised to wait until the cold weather before suppressing him; as a result the Sikh Army joined the insurgents, thus turning a local rising into a national war. Gough, again in command, crossed the Ravi on Nov. 16. The Sikhs were led by Sardar Sher Singh Attariwala. Two bloody but inconclusive battles at Ramnagar and Chillianwalla led to the decisive Battle of Gujrat on Feb. 21, 1849, when Gough, who had been much criticized for his preference for frontal attack to cannon fire, at last gave his artillery full scope. The Sikh leaders surrendered on March 12 and the Punjab was then annexed. The losses at Chillianwalla were so great (2,446 casualties, 4 guns, 3 colours) that Sir Charles Napier was appointed to supersede Gough, but Gujrat was won before Napier could arrive.

See J. D. Cunningham, *History of the Sikhs*, 2nd ed. (1853; rev. ed. by H. L. O. Garrett, 1919); Sir C. Gough and A. D. Innes, *The Sikhs and the Sikh Wars* (1897). (T. G. P. S.)

SI KIANG: *see* HSI CHIANG.

SIKKIM, a country in the eastern Himalayas and a protectorate of the government of India, is bounded on the west by Nepal, on the north by Tibet, on the east by Bhutan, and on the south by India. It has an area of 2,828 sq.mi. (7,325 sq.km.). The capital is Gangtok.

Physical Geography.—Sikkim is a mountainous country consisting of part of the main chain of the Himalayas and of ranges projecting southward and gradually lessening in height. Kangchenjunga (*q.v.*) in the north rises to more than 28,000 ft. (8,500 m.), and in the south the outer hills are about 700 ft. (215 m.) in height. Between these ranges the country is split up into mountain ridges and deep valleys where every type of climate from arctic to subtropical can be found. Accordingly the vegetation is very diverse. Below the line of eternal snows (17,000 ft. [5,100 m.]) to 12,000 ft. (3,600 m.) is grassland for cattle; down to 9,000 ft. (2,700 m.) are forests. On the slopes between 4,000 and 7,000 ft. crops are grown. In the valley bottoms there is luxuriant tropical growth. The chief river is the Tista, which is formed by the confluence of the Lachen and the Lachung in the north of Sikkim. Rainfall is heavy, the annual average varying from about 130 in. (3,300 mm.) at Gangtok to about 30 in. (760 mm.) in the more sheltered areas in the north. The abundant vegetation includes many types of flowering plants, orchids, and rhododendron. There are also a large number of species of moths and butterflies.

History.—The Lepchas are thought to be the earliest inhabitants of Sikkim. Their numbers began to diminish after the Tibetans started to colonize Sikkim at the end of the 16th century. Definite control by the Tibetans was acquired during the first half of the 17th century when several lamas fled from the Chinese in Tibet. They made Buddhist lamaism the state religion of Sikkim and established a government under a Tibetan king (1642), from whom the succeeding line of maharajas claimed descent.

Sikkim remained a Tibetan dependency until the end of the 18th

century, when British expansion in India changed the situation. Under a British-Nepalese treaty (1816) Nepal was obliged to return the Terai or submontane portion of Sikkim to the Sikkimese. During the war between Nepal and Sikkim (1834–35) Britain intervened again and in 1839 obtained the site of Darjeeling from Sikkim. Ten years later, after much friction between the British and the Sikkimese, Britain seized the Terai and some of the outer hills. The trouble resulting from this measure was settled by a treaty which Britain forced upon Sikkim in 1861. This treaty established British suzerainty over Sikkim, at least in the British view, though this was not shared by the Tibetans, who proceeded to engage in official activities in Sikkim. In 1888 a British military expedition drove the Tibetan army out of Sikkim and pursued it into the Chumbi Valley in Tibet. The Chinese—suzerains of Tibet—ordered their resident in Lhasa to make peace, but disagreement regarding suzerainty over Sikkim delayed a treaty until 1890, when Sikkim's borders were finally defined and China recognized British control over Sikkim's internal administration and foreign relations. For this purpose, a British political officer was appointed to assist the maharaja. In 1893 a supplementary treaty between China and Britain, relating to trade, communications, and pasturage, gave the British special rights in Sikkim and in Tibet but was not recognized by the Tibetans.

When India obtained independence in 1947, it succeeded to British rights in Sikkim under a temporary standstill agreement. By this time the maharaja had introduced minor social reforms and political liberties. Essentially, however, his rule and that of the big landowners in their areas were autocratic. The peasants were helpless and exploited. Dissatisfaction with the system caused such civil disorders between 1947 and 1949 that the maharaja had to call for Indian assistance, and in June 1949 Sikkim had again become a protectorate, this time of India. An agreement of 1950 entitled India to handle Sikkim's defense, communications, and foreign affairs. With the approval of the Indian government several political parties continued to agitate for reforms, which were gradually being granted. In the meantime, the government was proceeding with the modernization of the cultural and economic life of the people, in particular the construction of motorable roads and hydroelectric power plants.

Population.—At the 1961 census the population of Sikkim was 161,080, compared with 137,725 at the 1951 census. Most of the increase was accounted for by the Nepalese, who form about three-quarters of the population, the minority groups—Bhotias and Lepchas—showing virtually no increase at all. The communities do not intermarry since the Nepalese are mostly Hindus and the Bhotias and Lepchas Buddhist. The Nepalese live as cultivators in the middle altitudes; the Bhotias are mainly graziers in the high altitudes; the Lepchas usually live in the remotest valleys. There are no towns and the villages are widely scattered. In 1961 Gangtok, which lies on the main trade route between India and Tibet, had a population of 6,848.

Government and Social Conditions.—An Indian-appointed dewan (quasi prime minister) shares power with the maharaja who remains supreme ruler. The maharaja is assisted by an executive council and by a state council whose 20 members are partly elected and partly appointed by the maharaja. The subjects which the members of the state council are permitted to discuss are restricted, and the maharaja retains a veto over all their decisions. In the general elections held in 1960 the National Party (which has the maharaja's support) won seven of the 14 elective seats, the Nepalese-supported National Congress four seats, and the State Congress three seats. An election scheduled for the end of 1962 was postponed as a result of the emergency caused by China's invasion of India. Instead a People's Consultative Committee was formed, in which all population groups and political parties were represented.

In Sikkim there were in the early 1960s almost 100 government schools and about 50 government-aided schools, with approximately 10,000 pupils. Medical services included 4 hospitals and 20 dispensaries.

The Economy.—The economy of Sikkim is based on its agriculture, and the chief crops are rice, maize (corn), millets, barley, oranges, apples, pineapples, cardamom, and potatoes. Woolen cloth is also produced and handloom cotton weaving is an important domestic industry. The main exports are cardamom, oranges, apples, and potatoes. Most trade is with India, which supplies Sikkim's essential imports, including textiles and foodstuffs. Sikkim has no railways or air communications but, with Indian help, motor roads of strategic importance are being built, and by the early 1960s about 300 mi. of such roads had been constructed. In the mid-1960s Sikkim was engaged in an intensive development program, financed entirely by India. Emphasis in the second Five-Year Plan was on the improvement of agricultural products and methods, communications, and on the training of Sikkimese youth. (Wr. L.)

Bibliography.—J. D. Hooker, *Himalayan Journals* (1854); L. A. Waddell, *Among the Himalayas*, 2nd ed. (1900); J. C. White, *Sikkim and Bhutan* (1909); J. Morris, *Living with Lepchas* (1938); G. Gorer, *Himalayan Village* (1938); W. Levi, "Bhutan and Sikkim: Two Buffer States," *World Today*, 15:492–500 (Dec. 1959); P. P. Karan and W. M. Jenkins, Jr., *The Himalayan Kingdoms: Bhutan, Sikkim, and Nepal* (1963); Government of Sikkim, *Second Five-Year Plan, 1961–66* (1961); Desmond Doig, "Sikkim," *National Geographic*, 123:398–429 (March 1963).

SIKORSKI, WLADYSLAW (1881–1943), Polish military commander and prime minister of the Polish government-in-exile during World War II. He was born on May 20, 1881, at Tuszów Narodowy in Galicia. The son of a teacher, he graduated in engineering at Lwow (Lvov). A reserve officer of the Austrian Army, he was, with Jozef Pilsudski (*q.v.*), one of the founders, in 1908, of the Polish secret military organization, and from August 1914 to September 1916, during World War I, he headed the military department of the Supreme National Committee, organizing the Polish Legions to fight on the Austrian side against Russia (Pilsudski objected to his pro-Austrian policy as shortsighted). Having distinguished himself in the Russo-Polish War (*q.v.*) in command of an army operating in Polesie and then, from August 1920, of one fighting north of Warsaw, Sikorski was appointed chief of the Polish general staff on April 1, 1921.

After the assassination of Pres. G. Narutowicz, Sikorski became prime minister in December 1922 but resigned in May 1923. As minister of military affairs from February 1924 to May 1925, he contributed to the modernization of the armed forces. Commander of the Lwow army corps from December 1925, he remained neutral during Pilsudski's *coup d'état* of May 1926. Even so, he was dismissed from his post in March 1928. He then began to write on military and political matters and joined the anti-Pilsudski group headed by I. J. Paderewski, Wincenty Witos, and Wojciech Korfanty.

On the outbreak of World War II, Sikorski went to Paris to become prime minister of the Polish government-in-exile and commander in chief of the Polish forces. After the collapse of France in 1940 he moved his government to London. He won the esteem of Winston Churchill and on three visits to the U.S. (1941–42) discussed problems of grand strategy, as well as revision of Poland's western frontier, with Pres. Franklin D. Roosevelt.

Sikorski concluded an agreement on political and military cooperation with the U.S.S.R. (July 30, 1941) and visited Stalin in Moscow, where a joint declaration of mutual assistance was signed (Dec. 4, 1941). In April 1943, however, his government asked the International Red Cross to investigate the murder, at Katyn, of thousands of Polish officers who had been prisoners of war in Soviet hands. Stalin thereupon severed Soviet-Polish diplomatic relations. On July 4, 1943, on a return flight from the Middle East to London, Sikorski perished in an air crash at Gibraltar.

See Gen. M. Kukiel *et al.*, *General Sikorski* (1954). (K. Sm.)

SIKORSKY, IGOR IVAN (1889–), pioneer airplane designer and manufacturer. He was born in Kiev, Russia, May 25, 1889, and educated in Kiev, Petrograd and Paris. At an early age he turned his attention to aeronautics and in France and Germany studied the work of Blériot (*q.v.*) and Zeppelin. In 1909 he designed and built his first flying machine, a helicopter but it was not successful and he turned to conventional fixed-wing airplanes. By 1911 he had produced five successful machines and become an accomplished pilot.

During the period 1913–17 Sikorsky designed, built and flew planes that were remarkable for their size and flying range. They were powered by four engines and were the precursors of the modern airliners and large bombers.

In 1919 Sikorsky moved to the United States and became a citizen in 1928. He established an aircraft firm on Long Island that later became a division of United Aircraft corporation. His organization turned out a long series of land planes, seaplanes and amphibians that were used by the military services and commercial airlines.

In 1939, Sikorsky returned to his original interest in the helicopter, and produced the first successful direct-lift machine in the United States. He continued active development of this type until his retirement from United Aircraft corporation in May 1957. (S. P. J.)

SILAGE: *see* ENSILAGE.

SILAS, SAINT (fl. A.D. 50), early Christian prophet and missionary, was the companion of St. Paul (*q.v.*) on the second journey, when he took the place formerly held by Barnabas. The tour included southern Galatia, Troas, Philippi (where he was imprisoned), Thessalonike (Salonika), and Beroea, where Silas was left with Timothy, though he afterward rejoined Paul at Corinth.

He is in all probability the Silvanus (the abbreviation is similar to that of Prisca for Priscilla and Sopater for Sosipater) who is associated with Paul in the letters to the Thessalonians, mentioned again in II Cor. 1:19, and the bearer and amanuensis of I Pet. (5:12). It is possible, indeed, that he has an even closer connection with this letter (*see* PETER, EPISTLES OF SAINT: *I Peter: Authorship and Date*), and some scholars are inclined to give him a prominent place among the writers of the New Testament. He was of Jewish birth and probably also a Roman citizen. St. Silas' feast day is July 13 in the West, July 30 in the East.

SILCHESTER, a parish and village in Hampshire, England, 10 mi. SW of Reading. The parish contains the site of the Romano-British town of Calleva Atrebatum, excavated in 1864–78, 1890–1909, and 1938–39 (results published in *Archaeologia*). Soon after the Roman conquest Calleva became the cantonal capital of the Atrebates and the centre of a road system. The outer earthworks enclosed 230 ac., the inner (about A.D. 160–170) 100 ac. The street plan divided Calleva into 37 *insulae* and there were a forum, basilica, public baths, several temples, a Christian church, and an amphitheatre. (*See* BRITAIN: *Roman Britain*.) The main collection of antiquities from the site is in the Reading Museum; the Calleva Museum (opened 1951) illustrates the town's life. Pop. (1961) 511.

See G. C. Boon, *A New Guide to the Roman City Calleva Atrebatum at Silchester, Hampshire* (1952).

SILENCER: *see* MUFFLER.

SILENUS: *see* SATYRS AND SILENI.

SILESIA (Polish SLASK, Czech SLEZSKO, German SCHLESIEN), an old Polish province which became a possession of the Bohemian crown in 1335, passed with that crown to the Austrian Habsburgs in 1526, was taken by Prussia in 1742, and returned to Poland in 1945. Bounded by the Sudeten Mountains to the southwest, by the Beskid Range to the south, and by the Cracow-Wielun plateau to the northeast, it covers the basin of the upper and middle Oder (Odra) River.

The Name.—The German historian Ignaz Imsieg, writing in 1830, derived the name Silesia from that of the Silingi, a Vandal tribe, but that tribe stayed in Silesia only for a relatively short period and migrated southward about the end of the 3rd century. Polish historians maintain that the name derived from the Sleza River (German Lohe), a left-bank, northward-flowing tributary of the Oder, to the east of the mountain of the same name. Frantisek Palacky, the Czech historian, believed that *Slezane* (the Czech name for the Silesians) was a Slavonic transcription of Silingi. Wladyslaw Semkowicz, a Polish historian, demonstrated in 1933 that the name *Silingi* itself described the people living along the Sleza. By the 9th century the population of Silesia was exclusively Slavonic: the Dziadoszanie and the Bobrzanie occupied the north of the country, the Slezanie the

THE REGION OF SILESIA

middle, and the Opolanie and the Golensicowie the south. Because of their central position, the Slezanie gave their name to the province, which the old Latin chronicles of the Poles called *Slansco* (1132) or *Zlesia* (1163).

The First Polish Period.—By the 10th century Silesia was a disputed territory between the Czech dynasty of the Premyslids and the Polish dynasty of the Piasts, but between 989 and 992 Mieszko I, prince of Poland, acquired Silesia from Boleslav II, prince of Bohemia. Boleslaw I the Brave, king of Poland, consolidated this possession by founding a bishopric in 1000, with its seat first at Smogorzow (German Schmograu), when it was later transferred to Wroclaw (Breslau). This diocese, covering the territory of historic Silesia, was subject to the archbishop of Gniezno, whose seat was at that time the capital of Poland.

When Boleslaw III the Wry-Mouthed, prince of Poland, died in 1138, he left a statute dividing his lands between his four sons. Wladyslaw, the eldest, received Silesia, with overlordship over all the other Polish territories; in 1146 he was evicted by his brothers, against whom he had allied himself with the German king Conrad III. In 1163, however, under pressure from Conrad's successor, the Holy Roman emperor Frederick I Barbarossa, Silesia was restored to Wladyslaw's two elder sons, Boleslaw I the Tall (d. 1201) and Mieszko the Knock-Kneed (d. 1211): Boleslaw became prince of Lower Silesia, with Wroclaw as his capital, Mieszko of Upper Silesia, with Opole (Oppeln). In 1203 a third principality, with Glogow (Glogau) as capital, was created for Konrad, Wladyslaw's third son.

The princes of Wroclaw constituted the senior branch of the Silesian Piasts. Henry I the Bearded, Boleslaw's son, who reigned from 1201 to 1238, tried to reunite the Polish kingdom. His son, Henry II the Pious, pursued the same policy, but was killed in the great Battle of Legnica (Liegnitz) on April 9, 1241, where at the head of an army of Polish and German knights he checked the Tatar invasion which had devastated the country. All the Silesian Piasts encouraged the immigration of German peasants, artisans, and merchants to promote economic growth; forests were cleared, swamps drained, new towns founded, and the mining and textile industries developed—at the cost, inevitably, of a progressive germanization of the country.

As the Silesian Piasts were prolific, the three existing principalities were subdivided again, and by the end of the 15th century there were no less than 16 Silesian principalities. They were (from

the northwest to the southeast): Glogow, Zagan (Sagan), Wolow (Wohlau), Legnica, Jawor (Jauer), Olesnica (Oels), Wroclaw, Swidnica (Schweidnitz), Brzeg (Brieg), Grodkow-Nysa (Grottkau-Neisse), Ziebice (Münsterberg), Opole, Raciborz (Ratibor), Krnov (Jägerndorf), Opava (Troppau), and Cieszyn (Teschen, or Tesin). All of these acknowledged a certain unity, which was expressed in periodical meetings of the ruling princes, under the presidency of the senior of Piast of the Silesian line.

The Bohemian Period.—The petty Silesian princes often sought the help of the powerful kings of Bohemia against their own brothers and cousins, thus enabling those monarchs to revive their old claims to overlordship. In 1335, at Visegrád, an arbitral award rendered by Charles I of Hungary (Charles Robert of Anjou-Naples) assigned all Silesia to the Bohemian crown in return for the renunciation by the Bohemian king of his further claims to the Polish crown; this settlement was accepted by the Polish king Casimir III the Great. The change was, however, only one of overlordship: the Piast princes remained hereditary rulers of their own principalities so long as their lines survived; and only on the extinction of any line could the king of Bohemia take possession of the principality and assign it to a new prince. Though Bohemia was part of the Holy Roman Empire, the Piasts of Silesia were allowed to retain their status as Polish princes.

Silesia belonged to the lands of the Bohemian crown, but not to Bohemia. On various occasions on which the will of its Estates had to be consulted (chiefly on questions of succession) Silesia took a line different from Bohemia, though in practice usually identical with that of Moravia. In the 15th century Silesia took the side of the emperor Sigismund against the Bohemian Hussites (*q.v.*) and was in consequence heavily ravaged in the Hussite Wars of 1425–35, the German element suffering severely. Feeling in the country itself was divided, and when George of Podebrady ascended the throne of Bohemia in 1457, most of the Silesian princes recognized his suzerainty. The towns, however, led by Wroclaw, resisted, and, because of this, the Hungarian king Matthias I (*q.v.*) Corvinus was able to rule Silesia (with Moravia) from 1469 to 1490. On his death, Silesia reverted to the Bohemian crown, from whose weak holder, Vladislav II, the Estates exacted concessions which made them virtually independent.

The Habsburg Period.—On Dec. 5, 1526, the Silesian Estates accepted without demur the succession of the Austrian archduke Ferdinand, of the Habsburg dynasty (*see* FERDINAND I, Holy Roman emperor) to the Bohemian crown, though the surviving Piast princes claimed that this could not affect their own rights, and one of them, Prince Ferdinand II of Legnica, concluded a *pactum mutuae successionis*, or reciprocal treaty of inheritance, of his own with the elector Joachim II of Brandenburg (1537). This had no immediate effect, and Silesia remained entirely under the Habsburgs, who reduced its internal liberties, while tending to emphasize its independence of Bohemia. When the archduke Rudolf (the future emperor Rudolf II) was elected king of Bohemia in 1575, the Silesian princes became princes of the Empire.

The Reformation turned Silesia almost entirely Protestant. Consequently, on the outbreak of the Thirty Years' War, the Silesians joined the Bohemians in their struggle against the Habsburgs. Though intervention by Lutheran Saxony, from 1621, obliged the Habsburgs to acknowledge some formal measure of religious tolerance in Silesia, so that the Counter-Reformation could not be fully enforced there, the country was repeatedly overrun by contending armies and pillaged by mercenary bands. Under the Peace of Westphalia (1648), thanks largely to pressure from Sweden, a few towns were allowed to have Protestant churches, besides the freedom of worship which the princes enjoyed; and relations with Protestant communities abroad were permitted. In 1707, during the great Northern War (*q.v.*), Charles XII of Sweden obtained the Convention of Altranstädt from the emperor Joseph I, whereby the latter undertook to restore 128 churches to the Protestants, with permission to build more. Silesia was again the most Protestant part of the Austrian Habsburgs' dominions. Meanwhile it had been making an economic recovery: the Austrian mercantilist reformers of the late 17th and early 18th centuries made the development of its mining and textile industries the cornerstone of their plans, and before long Silesia counted as the richest of all the Austrian provinces, while Wroclaw was now one of the largest and richest cities of the empire.

The Prussian Period.—It was chiefly its wealth that tempted Frederick II of Prussia to wrest Silesia from the Habsburg heiress Maria Theresa in the War of the Austrian Succession (*q.v.*), which he began in 1740. His excuse was the old treaty of 1537, although his own great-grandfather had in 1686 renounced his claims under that treaty in return for the cession of the district of Swiebodzin (Schwiebus). Under the Treaty of Berlin of July 28, 1742, Maria Theresa was obliged to cede to Frederick all of Lower Silesia and Upper Silesia except the districts of Krnov, Opava, and Cieszyn. The Austrian attempt to recover all Silesia in the Seven Years' War (*q.v.;* 1756–63) came to nothing. What was left of Austrian Silesia was united to Moravia till 1849, when it was made a separate crownland.

Frederick's action in seizing Silesia had not been at all unwelcome to the German and Protestant Silesians, and it brought them many benefits. Frederick devoted much attention to his new acquisition, which was at first placed under a special *Landesminister*. The old Estates were abolished, and a more efficient administration was introduced. Great attention was paid to economic development. In 1815 Prussian Silesia was enlarged by the incorporation of Saxon Lusatia. In 1818 the diocese of Wroclaw ceased to be subject to the Polish metropolitan of Gniezno and passed under direct administration of the Holy See.

With the increasing importance of coal, in which Upper Silesia proved very rich, economic development proceeded even more rapidly than before, until it became an industrial district second in Germany only to the Ruhr area. Lead and zinc, as well as coal and iron, were mined and utilized in factories on the spot.

Lower Silesia was by this time purely German. In Upper Silesia, except for its western portions, the population was mixed. Most of the towns were almost entirely German, but many of the agricultural districts were Polish, and a fair proportion of the miners and unskilled industrial workers were Poles.

The Settlement of 1919–21.—In 1919, after World War I, both Poles and Czechs laid claim to parts of Prussian Silesia, while conversely the Germans of Austrian Silesia asked to be incorporated in Germany. Further, while the Czechs claimed the whole of Austrian Silesia, partly on grounds of historic right, Poland claimed Upper Silesia and part of Cieszyn Silesia as Polish by majority. On July 28, 1920, the principal Allied powers laid down a frontier which divided the Cieszyn area and the town itself between Poland and Czechoslovakia. (*See* TESCHEN.)

There remained the largest question of all: that between Germany and Poland. When the draft of the Treaty of Versailles (*q.v.*) was handed to the German delegation on May 7, 1919, it comprised the transfer to Poland of the greater part of Upper Silesia according to the ethnic boundary between the two nationalities, easy to establish on the basis of the Prussian census of 1910. However, in response to Germany's vigorous protests, contained in the *Remarques de la délégation allemande sur les conditions de la paix* (May 29), this decision was modified under the pressure of David Lloyd George, the British prime minister. The final treaty provided that the population should declare by plebiscite, by communes, whether it wished to belong to Germany or Poland.

In the poll (March 20, 1921), 706,820 votes were cast for Germany (including 182,288 emigrant votes brought from Germany to the plebiscite area) and 479,414 for Poland. In all, 682 communes voted for Poland and 792 for Germany, but the Poles secured preponderance in the southeastern plebiscite area, which economically was the most important; and in May 1921, the Polish plebiscite commissioner, Wojciech Korfanty (*q.v.*), ordered an armed rising. The dramatic struggle for Upper Silesia ended seven months later when, on Oct. 20, 1921, the principal Allied powers endorsed the award of the Council of the League of Nations of Oct. 12, whereby three-quarters of the coal-producing area and nearly two-thirds of the steelworks went to Poland. Lower Silesia, meanwhile, was left to Germany. (*See* POLAND: *The*

Treaty of Versailles and Frontier Problems.)

The Settlement of 1945.—Germany overthrew the settlement of 1921 by attacking Poland on Sept. 1, 1939, thus starting World War II. Polish Upper Silesia was annexed to Germany, which on Dec. 12, 1940, divided its newly reunited Silesian province in two *Gaue:* Lower Silesia, with Breslau (Wroclaw) as its capital; and Upper Silesia, with Kattowitz (Katowice). Polish patriots, especially those who took part in the Upper Silesian rising of 1921, were arrested and shot without trial, and educated Poles were deported from Silesia to the German-administered *General-Gouvernement* of Poland. Silesia was subjected to a policy of forcible germanization and settled by Germans evacuated from the territories which, according to the Ribbentrop-Molotov agreement of Sept. 28, 1939, were temporarily described as the Soviet sphere of interest. (*See* POLAND: *Poland During World War II.*)

In 1942 the Polish government-in-exile in London, supported by the resistance leaders in Poland itself, claimed a radical revision of Poland's western frontier, arguing that this was necessary for political, economic, and moral reasons. Politically, it was important to ensure Poland's security by giving it the shortest possible frontier with Germany, namely the Oder-Lusatian Neisse line; economically, it was imperative to compensate Poland for the loss of eastern provinces (to the U.S.S.R.) so that it could transfer westward its refugees from the east as well as the surplus of its population in central provinces; and morally, the new territorial settlement should show to the German people that Nazi war crimes could not go unpunished. It was in this spirit that the representatives of the United Kingdom, the United States, and the U.S.S.R. decided at Potsdam, on Aug. 2, 1945, to assign nearly all German Silesia to Poland and to transfer its German population to the Allied-administered Germany.

Silesia as a German *Provinz* covered until 1939 an area of 36,696 sq.km. (14,168 sq.mi.) with a population of 4,815,800 (census of May 17, 1939), including more than 500,000 Poles. It had a density of population of 131 inhabitants per sq.km. (340 per sq.mi.).

The pre-1939 Polish province (*województwo*) of Silesia composed the part of Upper Silesia attributed to Poland in 1921 (3,205 sq.km.) and the Cieszyn area (1,011 sq.km.). The total area of Polish Silesia was 4,216 sq.km. (1,628 sq.mi.) with a population of 1,312,000 (estimate of Sept. 1, 1939).

After World War II the major part of German Silesia became Polish, with the exception of the districts west of the Lusatian Neisse (Hoyerswerda, Rothenburg, and the western half of Görlitz) attributed to the German Democratic Republic. Five of the northern districts of formerly German Silesia (Zielona Gora [Grünberg], Kozuchow [Freystadt], Glogow, Wschowa [Fraustadt], and Szprotawa [Sprottau]) became parts of the new Polish province of Zielona Gora. The remainder was united with Polish Silesia, to which had been added five districts of the former province of Kielce (Bedzin, Zawiercie, Myszkow, Czestochowa, and Klobuck). In 1950 Polish Silesia was divided into three provinces with capitals in Wroclaw, Opole, and Katowice. Their total area was 38,076 sq.km. (14,701 sq.mi.), with an almost entirely Polish population of 6,440,377 (1960). The density of population was 182 inhabitants per sq.km. (471 per sq.mi.).

See also references under "Silesia" in the Index.

BIBLIOGRAPHY.—German viewpoint: C. Grünhagen, *Geschichte Schlesiens*, 2 vol. (1884–86); J. Partsch, *Schlesien*, 2 vol. (1896–1911); P. Weber, *Die Polen in Oberschlesien* (1914); E. Birke, *Schlesiens Schicksal* (1951).

Polish viewpoint: S. Kutrzeba, W. Semkowicz *et al.*, *Historia Śląska*, 3 vol. (1933–36); J. Weinstein, *Haute-Silésie, pays de contrastes* (1932); K. M. Smogorzewski (ed.), *La Silésie polonaise* (1932). (K. SM.)

SILHOUETTE. The word silhouette especially means a profile portrait cut out of black paper; but it is also used for any outline or sharp shadow of an object.

The term was taken from the name of Étienne de Silhouette, the parsimonious finance minister of France in the middle of the 18th century whose hobby was the cutting out of paper portraits (the phrase "*à la Silhouette*" originally meant "on the cheap"). Silhouette's hobby was also that of later famous persons, including Johann von Goethe. Collectors today look for both the painted and the cut silhouettes of the 18th and early 19th centuries, notably items by such artists as Francis Torond, A. Charles, John Miers, C. Rosenburg, Mrs. Brown, Auguste Edouart, T. Hamlet and Mrs. Beetham (nee Isabella Robinson).

The art of the paleolithic peoples, of several different eras, that made mural paintings and drawings in the limestone caves of France and Spain, is distinguished by a brilliantly realistic representation that appears to have been first achieved by drawing the outline of the object's shadow, which was generally filled in with a flat colour. Representation by profile-drawing persisted long after pictures and sculptures became conceptual rather than realistic (*e.g.*, the tomb paintings and relief sculpture of Egypt and the finely executed decorations on Greek and Etruscan pottery). The ancient Greeks knew how to draw the outline of a person's shadow cast by sunlight. Profile portraiture executed by using a candle or a lamp to throw the shadow on a wall or screen had become fairly common in Europe by the 17th century. Professional practitioners began to paint shadow portraits on any suitable material—plaster, wax, glass, vellum, paper—and to mount them, sometimes quite elaborately; and various mechanical devices were adopted to facilitate correct drawing. When paper became generally available the outline of the shadow was often cut out, and the cleverer practitioners learned to cut the portraits directly from life and to make more elaborately composed pictures. In the 18th century the painted "shade" and the paper-cut both became fashionable. After the coming of the daguerreotype and photography a few genuine artists, such as Phil May, sometimes made use of the painted shadow style in caricature; otherwise it can be said to have died out, except that the cut silhouette persisted.

By the second half of the 20th century there were still a few good silhouette-cutters (for example the Englishmen Hubert Leslie and H. L. Oakley), but the underlying principles of the art lived most vigorously in the cartoon film, in the work, for instance, of Walt Disney and of Lotte Reiniger. (R. L. M.)

SILICA is a compound of the two most abundant elements of the earth's crust, oxygen and silicon (*q.v.*). It is the main constituent of more than 95% of the earth's rocks. Its resistance to change gives it special significance in geology and important applications in many chemical processes requiring a refractory (heat-resistant) substance. It occurs in a great variety of forms; the principal industrial sources are sand (quartzose types), quartzite and sandstone (*qq.v.*).

Silica is silicon dioxide, SiO_2. By weight, it contains 46.75% silicon and 53.25% oxygen. It constitutes 59% of the mass of the solid crust of the earth, occurring both as free silica and in combination in the silicate minerals (*see* SILICON: *The Silicates*). The free silica found in rocks is almost universally crystalline quartz. Sea sand is quartz in the almost pure state.

Uses.—Buildings and roads consume the largest tonnage of silica, for sand is the ever-present constituent of portland cement concrete and lime mortar in construction work, and sandstone is one of the most permanent of building stones. The hardness of quartz becomes useful in the application of sand to the cutting, grinding and polishing of glass and stone, and in the consumption of a large tonnage of engine sand to prevent the slipping of driving wheels of a locomotive on a steel rail. The high melting point and low expansion coefficient of cristobalite make silica brick, usually manufactured from quartzite, one of the most useful of refractories, particularly in the steel industry. Another refractory use, consuming a large tonnage of sand, is in foundry molds. Low thermal conductivity adds to high melting point in making diatomaceous silica a popular refractory insulator for furnaces. Unusual insolubility in water and acids, together with negligible thermal expansion, gives vitreous silica a preferred place in chemical apparatus, both in the laboratory and in the plant.

As a chemical raw material, silica sand and rock enter into the manufacture of various products in which the silica, as such, largely disappears, to reappear in other combinations; these include glass, soluble silicates (water glass), ceramic products of various kinds, silicon carbide, ferrosilicon and silicones.

Pure silica is a chemical product as well as a raw material.

When fused with an alkali, dissolved, precipitated, washed and partially dried, it is obtained in an amorphous, very finely divided form which has unusual absorptive properties for water vapour, gases and the minute impurities in certain liquids.

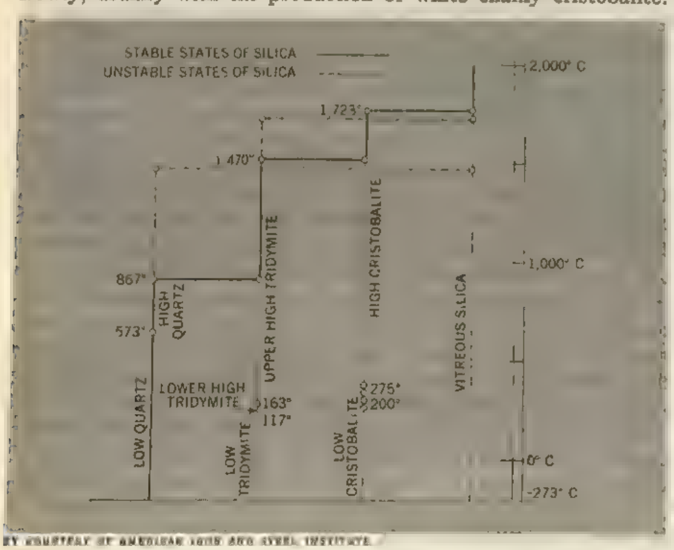
Principal Crystalline Forms.—Of the many forms of silica, the three principal crystalline forms are: quartz, tridymite and cristobalite. Quartz is the form that is stable, under atmospheric pressure, at all temperatures up to 867° C. Tridymite is the stable form from 867° C. to 1,470° C. Cristobalite is stable from 1,470° C. to its melting point at 1,723° C. (International Temperature Scale of 1948). The transformation of silica from one of these forms to another is sluggish, hence the two high-temperature forms can easily be preserved indefinitely at ordinary temperatures, while, conversely, quartz can be heated above 867° C. for hours or even for days and still remain either unchanged or only incompletely transformed into tridymite or cristobalite.

Three new crystalline forms of silica have been discovered since 1953: (1) coesite, which is stable only under high pressure and is the densest form of silica; (2) keatite, which is formed hydrothermally under high pressure and is possibly related to cristobalite in structure; and (3) a fibrous form discovered by A. and A. Weiss and sometimes designated silica W, which is produced by the oxidation of silicon monoxide, SiO.

Each of the three principal forms exists in more than one polymorphous modification, depending upon the temperature. In contrast with the sluggish inversions mentioned above, these high- and low-temperature modifications are produced almost instantaneously when the inversion point is attained; only under very unusual circumstances is it possible to hold the high-temperature modification at a lower temperature than its normal range. These quick-changing modifications were formerly designated by Greek letters, such as alpha and beta, but because of confusion arising from differences in the convention as to which modification should be called alpha, it is clearer to refer to them by the words high and low, or upper and lower. They are so designated in the figure.

The high-low inversion of quartz is at 573° C. Tridymite has two such inversions, at 117° and 163° C. Cristobalite is peculiar among polymorphous substances in that its high-low inversion point changes through a range of temperature (200°–275° C.) according to the previous history of the cristobalite.

Imitating the sluggishness of the inversion between the principal crystalline forms, molten liquid silica when cooled below 1,723° C., crystallizes sluggishly and can therefore be easily cooled to an amorphous glass, called vitreous silica, fused quartz, or silica glass. (This property persists in the silicate glasses.) Vitreous silica is often incorrectly called quartz, but quartz is a crystalline substance of completely different properties. Upon long-continued heating at temperatures above 1,100° C. vitreous silica devitrifies slowly, usually with the production of white chalky cristobalite.



BY COURTESY OF AMERICAN IRON AND STEEL INSTITUTE

STABLE AND UNSTABLE STATES OF SILICA

As the figure indicates, both quartz and tridymite can be melted at temperatures lower than the stable melting point of 1,723° C., but unless further heated to a temperature above 1,723° the fused silica so produced would gradually change over into cristobalite.

Silica is measurably volatile at temperatures above 1,600° C., but its vapour pressure is extremely small. The fluffy porous white deposits of silica often seen around electric furnaces usually are not due to direct distillation, but to reduction of SiO_2 by carbon monoxide and hydrogen to form volatile Si or SiO and gaseous SiH_4, which are subsequently reoxidized.

The Silica Minerals.—All of the three polymorphous crystalline forms mentioned above occur as natural minerals, and even the vitreous (glassy) form, though rare, has been found as a natural product. In combination with water in varying amounts up to about 12%, furthermore, silica occurs as the amorphous mineral opal. Finally, mixtures of hydrated and microcrystalline silica, over the whole range of possible proportions, occur in nature and can be grouped together as chalcedonic silica.

Cristobalite, the crystalline form stable at the highest temperature, always occurs as the low-temperature modification (*see* figure) but its external crystal form shows it to have grown at a temperature above the inversion range of 200°–275° C. Tridymite, likewise, occurs as the low modification, but its habit of growth in twinned hexagonal plates shows that it formed above 163° C. Both of these minerals are instances of Ostwald's rule, which states that a form of low or intermediate stability has a better chance of growing than the form that is most stable under the prevailing conditions.

Natural vitreous silica is called lechatelierite. It is the material constituting the peculiar tubes often found in sand, known as fulgurites, and produced by the intense local heating due to a stroke of lightning. It occurs in the meteor crater of Arizona, again a product of sudden and intense heating. An unexplained occurrence of lechatelierite is the scattered fragments found over a considerable area in the Libyan desert.

The foregoing (cristobalite, tridymite, lechatelierite) are relatively rare minerals. In extreme contrast is quartz, the world's second most abundant mineral (feldspar being first). It has received many names, differentiated according to external form, colour and visual effects caused by inclusions. Clear, colourless, well-crystallized quartz is often called rock crystal, and it received local names like "Herkimer diamond." It exhibits several types of twinning (*see* QUARTZ) and is frequently found as a pseudomorph of other minerals. The clear varieties which are coloured uniformly by various impurities include amethyst (purple, violet); sapphire quartz (indigo, blue); blue quartz; citrine and false topaz (yellow); rose quartz (pink and rose-red); smoky quartz and cairngorm stone (yellow to brown); and morion (deep brown). Inclusions of gas bubbles give rise to milky quartz; of red hematite, to ferruginous quartz; of needles of rutile (TiO_2) or tourmaline, to sagenitic quartz; of fine fibres and streaks of various composition, to cat's-eye and tigereye.

Like euhedral and massive quartz, the chalcedonic silica minerals have received a variety of names based upon visual characteristics. Because of their microfibrous structure, they all show a waxy lustre on a natural or a fractured surface. If visually homogeneous, nearly colourless and translucent, the mineral is chalcedony. Translucent green makes it chrysoprase, or prase if leek-green. Bright green due to chlorite inclusions produces plasma.

Other translucent varieties are bloodstone and heliotrope (green with red or yellow spots), carnelian (red or yellowish-red), and sard (brown). When banded instead of visually homogeneous, the mineral is agate, of which onyx and sardonyx are subvarieties characterized by banding in planes rather than curving surfaces. The less-prized, because opaque, varieties of chalcedonic silica are jasper, which is usually red, yellow or brown, and may even be green and bluish gray; flint, which is gray to black with conchoidal fracture; and chert, likewise gray to black with splintery conchoidal fracture.

Hydrated, apparently amorphous silica is usually given the name opal, of which the varieties used for gem cutting are characterized

by iridescent colours. Hyalite is the colourless and transparent form of opal. Although opal shows no evidence of crystallinity under the microscope, the X-ray reveals it to have, at least in part, the internal structure of cristobalite. An interesting structural form of opaline silica is the material called diatomite, which consists of microscopic skeletons of fresh-water diatoms. Sponge spicules are also essentially opaline silica. (*See* DIATOMACEOUS EARTH.)

Physical Properties.—The density of the principal forms at 0° C. is: quartz 2.651, tridymite 2.26, cristobalite 2.32, vitreous silica 2.203.

The thermal expansion of the crystalline forms is greatly influenced by the high-low inversions, in that the expansion coefficient becomes steadily greater as a polymorphous inversion point, like that of quartz at 573° C., is approached. At the inversion point there is a sudden increase of volume in going over to the high-temperature modification, after which there is relatively little change with further rise in temperature. High quartz actually contracts a little with rising temperature. Vitreous silica has an expansion coefficient which is near zero at all temperatures. This property gives it a special industrial value, for it can be suddenly heated, or chilled from a bright red heat (as by plunging in water) without breaking.

The specific heat of the forms at 0° C. is: quartz 0.1664, cristobalite 0.170, vitreous 0.1657. The curves of specific heat against temperature have the usual form, rising from zero at the absolute zero of temperature, interrupted by a discontinuity at the high-low inversions, and approaching a value of about 0.28 at the melting point.

The thermal conductivity of vitreous silica is uniformly low at all temperatures, about 0.003 cal. per square centimetre per second, per degree centigrade per centimetre at 0° C. Quartz, in contrast, while in a class with vitreous silica at ordinary temperature, rapidly increases in conductivity as the temperature falls and at the temperature of liquid hydrogen conducts as well as some metals.

The forms of silica are both hard and strong. Quartz is standard substance no. 7 in the Mohs' scale of hardness. A tensile strength as high as 160,000 lb. per square inch has been observed in drawn fibres of vitreous silica, and fibres that will carry 100,000 lb. per square inch are easily made. The vitreous form is singularly free from elastic anomalies and is therefore prized for various physical instruments.

The forms of silica are diamagnetic, with a mass susceptibility of about -0.5 per gram. They are poor conductors of electricity; the resistivity of vitreous silica at 20° C. is about 10^{19} ohms per square centimetre per centimetre, which makes it popular as an electrical insulator. It is especially valuable in such service because, unlike common glasses and many other substances, it does not condense a conducting layer of moisture upon its surface. (For other electrical properties, *see* QUARTZ.)

Both quartz and vitreous silica are colourless and transparent in visible light, and are unusually transparent to ultraviolet light. This property makes the vitreous form valuable in windows for transmission of whole sunlight, and in lamps for transmitting ultraviolet from the mercury arc. At wave lengths longer than visible light, on the other hand, though more transparent than other glasses to most of the infrared spectrum, silica reflects like a metal in the vicinity of wave length 9 μ, and has considerable reflecting power also near 13 and 23 μ.

The useful optical properties peculiar to crystalline quartz, particularly rotatory power, are discussed in the article QUARTZ.

Chemical Behaviour.—In the laboratory, crystalline and vitreous silica are virtually insoluble in pure water and dissolve very slightly in water solutions of the common mineral acids. Silica is attacked (etched), however, by dilute alkaline solutions and by concentrated phosphoric acid, and is rapidly dissolved by a solution of hydrofluoric acid (HF). The only halogen that attacks it is fluorine.

A homogeneous liquid containing only water and a few per cent of silica is easily obtained by acidifying the water solution of an alkaline silicate and dialyzing out the electrolytes. This liquid proves to be a hydrosol, containing silica units of large molecular weight, and not a true solution. Much experimental work was done on these sols in the hope of proving them to be silicic acids analogous to sulfuric and phosphoric acids, but without success. The sol can be concentrated to a stage where it becomes a rigid and elastic gel, corresponding to the natural mineral opal. The percentage of water in such gels is controlled by the temperature and by the water vapour pressure in contact with the gel, as well as by its previous chemical and physical history.

As the temperature rises, silica becomes measurably soluble in water, and continues to increase in solubility, if the pressure is increased, even above the critical temperature of water, 374° C., where surface tension disappears and gas and liquid become continuous in properties. At the critical end point of the solution, which is less than one degree higher than the critical temperature of water, and at the pressure corresponding to coexistence of quartz, liquid and vapour, quartz is soluble to the extent of 0.023% by weight. Its maximum solubility under these equilibrium conditions is 0.075%, at 332° C. Above the critical point, at higher pressures, a concentration of more than 0.3% has been attained. Vitreous silica attains an apparent equilibrium with water at about twice the concentration of quartz.

Silica in all forms reacts at temperatures above 1,000° C. with all metallic oxides, also with many salts when in the presence of air and moisture, to form crystalline or liquid silicates.

See also references under "Silica" in the Index.

BIBLIOGRAPHY.—W. Eitel, *The Physical Chemistry of the Silicates* (1954), *Silicate Science*, vol. i, *Silicate Structures* (1964), vol. ii, *Glasses, Enamels, Slags* (1965); C. Frondel, *Silica Minerals*, vol. iii of 7th ed., *Dana's System of Mineralogy* (1962); G. W. Morey, *Data of Geochemistry*, 6th ed., ch. 50, "Phase-Equilibrium Relations of the Common Rock-Forming Oxides Except Water," *Prof. Pap. U.S. Geol. Surv.* 440-L, (1964); R. B. Sosman, *The Properties of Silica* (1927) and *The Phases of Silica* (1965); J. G. Vail and J. H. Wills, *Soluble Silicates, Their Properties and Uses*, 2 vol. (1952); H. B. Weiser, *The Hydrous Oxides* (1926); R. Fricke and G. F. Hüttig, *Hydroxide und Oxydhydrate* (1937). (R. B. SN.)

SILICON is the second most abundant of the elements in the earth's crust, surpassed in quantity only by oxygen. It is found in measurable amount in practically every rock, in all natural waters, in the atmosphere (as siliceous dust), in many plants and in the skeletons, tissues and body fluids of some animals. Its presence in the sun and other stars is indicated by the spectroscope. It is never found in the free state, however, but always in combination—either with oxygen as silica (SiO_2), or with oxygen and aluminum, magnesium, calcium, sodium, potassium, iron and other elements in the numerous silicate materials which comprise rocks, soils and clays.

Prior to 1787 chemists had not differentiated silicon from its compound, silica, which they considered an element. In that year A. Lavoisier offered the speculation that silica was an oxide of an undiscovered element. In 1817 J. J. Berzelius found that the substance, free from oxygen, was present in cast iron; and six years later he established its existence as an element, although he was able to prepare it only in somewhat impure form. It was first obtained in reproducible crystalline form in 1854, when H. Sainte-Claire Deville crystallized it out of the alloy produced by electrolyzing a melt of mixed chlorides. The name silicon was derived from *silic-*, basal syllables of *silex*, Latin for "flint."

Silicon is chemical element no. 14 in the periodic system; symbol, Si. It is nonmetallic in character and has a chemical atomic weight of 28.086. Its isotopic composition, which has been found constant both in the earth's crust and in meteorites, is: Si^{28}, 92.28%; Si^{29}, 4.67%; Si^{30}, 3.05%. External electron arrangement: $1s^2\ 2s^2,\ 2p^6\ 3s^2,\ 3p^2$.

Forms.—Pure silicon crystallizes with isometric symmetry, and has the same structure as diamond. The distance between adjacent atoms is 2.35 Å.

Amorphous silicon has been frequently described, especially by the earlier workers. Some of it was undoubtedly impure as a result of the method of preparation, and was mixed with silicides and silica. Even when fairly pure, the supposed amorphous element has been found to give the same X-ray pattern as the recognizably crystalline form.

Preparation and Physical Properties.—Silicon is best prepared by reaction of metals such as magnesium, aluminum and potassium with the halogen compounds of silicon. It can be purified by recrystallization from metallic solution, as in aluminum, the excess of solvent metal being removed with acid. Because of the difficulty of preparing the pure element, the recorded physical properties are not as reliable as those of more easily purified substances.

The physical properties of silicon are as follows:

Density (grams per cubic centimetre)	2.42
Coefficient of thermal expansion at 0° C.	7×10^{-6}
Melting point	1,420° C.
Boiling point	about 2,600° C.
Specific heat at 0° C., calories	0.174
Refractive index, $\lambda = 589$ mμ	4.24

Chemical Behaviour and Uses.—Silicon is attacked by oxygen and water vapour at a red heat and higher, but the coating of SiO_2 formed by the reaction protects the element. It is vigorously attacked by gaseous fluorine, chlorine and hydrofluoric acid. It is not acted upon by sulfuric acid, hydrochloric acid, nitric acid or aqueous ammonia at ordinary temperature. Many oxides are reduced by silicon at high temperatures with the formation of metal or silicide, and silica.

Silicon is used in industry principally as the alloy of iron and silicon called ferrosilicon. This may contain from 14% to 94% Si. It is used for deoxidizing common steels and for making special silicon steels.

Compounds.—Silicon stands next to carbon in the fourth group of the periodic system. By simple analogy, this fact leads to the expectation that silicon can form chain and ring compounds with hydrogen and the halides, as well as a complex system of hydrogen-oxygen compounds. This expectation is to some extent realized in the simpler chain formulas of silicon compounds, and even the names, suitably modified, of the carbon-hydrogen-oxygen compounds have been carried over into the silicon series. New compounds are constantly being added to the silicon series.

Silicon has four electrons in the outer or valence shell. These are capable of pairing with the valence electrons of other elements to give four covalent bonds. Silicon is therefore commonly classed as tetravalent. The electrons can also be transferred to the outer shell of a strongly electronegative element such as fluorine to form bonds of ionic type. The covalent bonds of silicon to oxygen are estimated to be about half ionic in character. In the compounds of carbon, on the other hand, ionic bonds are rare. Further, carbon never produces more than four covalent bonds, and in some compounds the number descends to three and two.

Another important factor in the formation of compounds besides the electron arrangement is the size of the atom, controlling the space available for close approach of other atoms. Silicon has a larger atom than carbon. The co-ordination number of the silicon atom in its compounds is in nearly all cases four. With some atoms, however, such as fluorine and (occasionally) oxygen, it can be six. It is never less than four in any known compound.

These differences in external atomic structure find expression in the following striking differences between the carbon and silicon compounds of similar formula: (1) The carbon compounds form self-contained independent molecules which maintain their separate existence even though vaporized, liquefied, dissolved and modified by substitution of atom for atom; the silicon compounds polymerize and pass over into solid crystalline or amorphous aggregates, or structures extending infinitely in one, two or three dimensions. (2) In a world where water is always present as vapour or liquid, the carbon-based molecules survive untouched, the silicon-based molecules hydrolyze and end as hydrated silica or as water-bearing micas and clay minerals. (3) The carbon series build up their complexities by conjunction of carbon atom to carbon atom, while the silicon series build up their complexities mainly through the intermediary of oxygen atoms. It is easy to see that in a newly created world with prevailing temperatures of −50° to +100° C., carbon had within itself the possibility of an evolving life, while silicon was the possible parent mainly of cold rocks.

Hydrides.—Like carbon, silicon forms with hydrogen a series of compounds containing a progressively smaller proportion of hydrogen. The first member is silane, SiH_4, and the series has been prepared as far as Si_6H_{14}. All can be made by action of an acid on a silicide. All are gases at atmospheric temperatures. There are no unsaturated hydrides corresponding to the ethylene, acetylene and benzene types of hydrocarbons.

Oxides.—There is evidence for the existence of a monoxide, SiO, but little information is available on its properties.

The dioxide, SiO_2, usually called silica, is one of the most extensively investigated and best-known of chemical substances. (*See* SILICA.)

Compounds With Hydrogen and Oxygen.—A number of such compounds, called by the generic name siloxanes, had been prepared and studied by mid-20th century, but they lack the stability and the variety of behaviour characteristic of the analogous organic acids, esters, etc.

Halogen Compounds.—Silicon tetrafluoride, SiF_4, is a gas formed by reaction of hydrofluoric acid, HF, on silica. It combines with excess HF to form hydrogen silicofluoride, also called hydrofluosilicic acid or fluosilicic acid, H_2SiF_6, which is stable in water solution, giving the ion SiF_6^{--}. K_2SiF_6 is one of the few potassium salts that is difficultly soluble in water, making this acid useful in chemical analysis.

Silicon tetrachloride, $SiCl_4$, is a colourless liquid of boiling point 57° C. It reacts with water, either liquid or vapour, to form HCl and hydrated SiO_2. This reaction has been utilized, by including ammonia with the vapour, to produce a military protective dense white smoke consisting of particles of SiO_2 and NH_4Cl. As might be expected from the analogy with carbon, there are chlorides Si_2Cl_6, Si_3Cl_8 and others. Silicon tetrabromide is a colourless liquid boiling at 153° C. The tetraiodide is crystalline, melting at 121° C. and boiling at 290° C.

Numerous halogen derivatives of the silicon hydrides are possible, and many have been prepared and described. An example is silico-chloroform, $SiHCl_3$.

Carbide.—Silicon carbide, SiC, forms at electric furnace temperatures from reaction of carbon with silica. It exists in at least eight crystalline forms. It has become an important industrial abrasive and refractory, under trade names such as Carborundum and Crystolon. (*See* SILICON CARBIDE.)

Organosilicon Compounds.—The introduction of silicon into carbon chains and rings—or, looked at from the other direction the introduction of organic radicals into the simpler silicon compounds—opens an immense variety of possible structures, some hundreds of which have been prepared and described. None of these compounds has been identified as a natural product, though some may be taking part in the growth of silicon-bearing plant and animal tissues. Among the new compounds are the industrially important silicones (*q.v.*) which derive by polymerization from the general formula R_2SiO, in which R is an organic radical such as CH_3, C_6H_5 and the like. Another group of carbon-bearing compounds comprises the silicates, such as the industrially useful tetraethyl orthosilicate, $(C_2H_5)_4SiO_4$. The term organosilicon compound is restricted to those that contain at least one silicon-carbon bond.

Sulfide.—Silicon forms a colourless crystalline sulfide, SiS_2; its structure is that of an infinitely long chain of tetrahedra, SiS_4, which share opposite edges, the parallel chains themselves being held together by weak residual forces. It therefore differs from silicates of the chain type in that SiO_4 groups share corners but not edges (*see* below).

THE SILICATES

More than 95% of the earth's rocks contain silica as their principal component. Any such rock can be seen, either by direct inspection or by examination under the microscope, to be an aggregate of minerals, each homogeneous, usually crystalline, and usually transparent. In many rocks, one of these minerals is pure silica in the form of quartz. Most of the other rock minerals contain silica in combination; they are the silicate minerals.

Many of these minerals have been artificially duplicated, usually

by high-temperature processes. New crystalline silicates, never found in nature, have also been produced by these same processes. Most of the silicates, both natural and artificial, are insoluble in water. A striking exception is the group of silicates of the alkali metals (lithium, sodium, potassium), which are obtainable as homogeneous water solutions containing silica in a wide range of proportions. These solutions can be concentrated to form glasses that still contain as much as 25% water. One such series, the sodium silicates, forms an important article of commerce, long known by the commercial name of water glass, now more commonly called silicate of soda. By fusion with two or more oxides, commonly sodium oxide (Na_2O) and lime (CaO), both introduced as carbonate, silica forms the silicate glasses.

This extremely varied lot of substances, the silicate minerals, the synthetic silica compounds and the silicate glasses, are grouped by the chemist under the name silicates. As the name itself implies, the silicates were regarded throughout the 19th century and the first quarter of the 20th century as salts of certain silicic acids, in the same sense that sodium sulfate is a salt of sulfuric acid. The analogy was temporarily useful, even though it proved to be very difficult to produce a silicic acid resembling in any way the chemist's better-known acids. Pure silica proved to be extremely insoluble in pure water, and the homogeneous phases containing only silica and water, which are obtainable by various methods, were found to contain mostly colloidal or dispersed silica, with very little truly dissolved silica. The methods of organic chemistry failed completely to throw any light on the constitution of the silicates, and the field remained one for unrestrained speculation, uncheckable hypothesis and unproductive analogy.

The advent of the X-ray methods of studying the structure of crystalline and amorphous solids made possible a resumption of progress in the chemistry of the silicates. Systematic knowledge of silicate structure and constitution at mid-20th century was based largely on X-ray work particularly on the crystalline substances, the glasses and the clays, combined with the atomic-volume principle elucidated by V. M. Goldschmidt and others. It developed that the guiding unit is a group having the formula SiO_4. In all the silicates, as well as in pure silica, there are groups of four oxygen atoms arranged in space at the corners of a tetrahedron. There is usually a silicon atom at the centre of the tetrahedron, but it can be replaced here and there by aluminum (in most cases not to exceed one out of four silicon atoms), and rarely by beryllium. At the same time, an occasional tetrahedral corner will be found occupied not by an oxygen atom but by a fluorine atom or an OH pair (hydroxyl). The systematization of all the silicates is based upon the various ways in which it is possible to interconnect the silicon-oxygen tetrahedra. The chemically basic or electropositive elements, such as Li, Na, K, Mg, Ca, Sr, Ba, Al (in part), Mn, Fe, fit into the openings between the tetrahedra in accordance with their atomic volume. The oxygen atom itself is larger than most of the silicate-forming atoms, which leads to the geometric possibility of fillable openings.

Silicate Structures.—Mineralogists and chemists are not in complete agreement as to the classification of the structures, but the main groupings are briefly characterized below, and some of the better-known examples are listed in parentheses. Some of the old names were still retained for lack of a more logical terminology.

Three-Dimensional Networks, or Silica Type.—All the tetrahedra share corners with other tetrahedra. The formulas contain the group SiO_2 or multiples thereof. The type includes silica itself; the feldspar group (orthoclase, plagioclase); the nephelite group; the cancrinite and sodalite groups in which certain electronegative units such as Cl, CO_3 and SO_4 are added to the structure; and the large zeolite family, with added molecules of H_2O as the principal structural characteristic.

Sheet Structures (Sometimes Called the Disilicate Type).—The tetrahedra are arranged in a plane, each being joined to other tetrahedra by three atoms in the plane. The type includes the aluminum disilicates (kaolin and other clay minerals); the nonaluminum member, talc; the analogous aluminum member, pyrophyllite; the chlorite family; and the micas (muscovite and biotite series).

Chain and Ring Structures, or Metasilicate Type.—These may be either chains or rings of tetrahedra, reminiscent in a way of the chain and ring structures of the chemistry of carbon. The chain structures, which are potentially infinite in length, include: single chain structures, containing SiO_3 or multiple thereof in the formula (pyroxene group, pyroxenoid family); double chains containing Si_4O_{11} in the formula (amphibole group); and other more complex chains. Ring structures include beryl and probably the tourmaline series.

Double Tetrahedral Structures, or Pyrosilicate Type.—Two tetrahedra share a single oxygen atom, giving formulas containing Si_2O_7. These silicates include the melilite group and the hemimorphite family.

Independent Tetrahedral Structures, or Orthosilicate Type.—There is no sharing of corners between tetrahedra, hence formulas contain the group SiO_4. The old formal derivation was from an imaginary silicic acid of the formula $Si(OH)_4$. These silicates include a varied assortment, such as the chrysolite group (olivine, forsterite, fayalite), garnet group and epidote group. Allied to these orthosilicates are some subsilicate minerals, in which not all the oxygen is in the independent tetrahedra (sillimanite, mullite).

Phase Equilibrium in Silicate Systems.—There is another approach to understanding the chemistry of the silicates which almost completely ignores the existence of atoms and proceeds on the thermodynamic basis of the Gibbs phase rule, determining experimentally the limits of stability of all the crystalline and liquid phases that can be produced synthetically. The work of the Geophysical laboratory of the Carnegie Institution of Washington, D.C., founded by A. L. Day in 1906, is the most extensive in this field. The method has the great advantage of proceeding from simple to complex systems and of covering the ground thoroughly as it goes, at least with respect to equilibriums readily attainable at high temperatures. It has the disadvantage of being faced with what might appear a numerically hopeless task. Hundreds and thousands of tests, short though they may be, are necessary for the complete survey of a system of three oxide components, and several such systems had been patiently completed by the 1950s, but no four-oxide system had been finished. Yet most silicate minerals contain at least five components and a great many contain ten. The total number of systems (unary, binary, ternary, etc.) that can be assembled from ten oxides is 1,023. Additional complications are: (1) the necessity, in systems containing the oxides of iron and manganese, of taking the metal and gaseous oxygen as the components, adding gaseous pressure as another variable; (2) the experimentally difficult problem of sealed high-pressure apparatus when water is one of the components. Nevertheless, the elucidation of the simpler systems had brought rich returns in such matters as the formulation of fundamental principles on the genesis of rocks, and the solving of industrial problems in ceramics.

Ten or more binary silicate systems are represented by reasonably satisfactory phase-equilibrium diagrams, together with six ternary systems (SiO_2 with the pairs Al_2O_3–CaO, Al_2O_3–MgO, CaO–MgO, CaO–K_2O, CaO–Na_2O, K_2O–Na_2O). Systems containing dissociable oxides are still fragmentary. Silicates of sodium and potassium with H_2O as a third component are fairly complete. Details will be found in articles on the individual metallic elements.

The Problem of Valence and Neutrality.—Implicit in both the atomic-structural approach and the phase-equilibrium approach to the study of the silicates is the chemist's basic assumption of the validity of Dalton's law of definite and multiple proportions. It is assumed, for instance, that the phases and the structures in a series of sodium aluminosilicates containing water can be correctly described in terms of the formulas SiO_2, Al_2O_3, Na_2O, and H_2O. But 20th-century chemistry, studying the metallic alloys and various inorganic systems containing solid solutions (*i.e.*, crystalline solutions), found that Dalton's rules are often only an approximation or a working limit, not an exact statement. In the system iron-oxygen, for example, there occurs one crystalline phase (wüstite) with a temperature range of stability from 560° to 1,424° C. and a range of composition from 23.10% to 25.60% oxygen, no point of which can be represented by a simple

integral relation between the atoms; the composition FeO (22.27% oxygen) lies entirely outside the field and is nonexistent as a phase. The wüstite phase is a crystalline solution of oxygen in iron or iron in oxygen, whose composition depends upon the pressure and temperature. This relation will certainly carry over into the iron-bearing silicates.

Furthermore, it became apparent from study of the adsorptive and base-exchanging power of many substances that not only can there be ionization in liquid solutions but also in crystalline substances as well, with the possible production of structures into which ions and polar molecules can be introduced as a part of the systematic structure, to be released again under the proper chemical environment. The silicates with sheet structure include many such substances. (*See* ADSORPTION; VALENCE.)

Silicates in Industry.—The silicates are the foundation material for the widely various ceramic industries. In the United States the word ceramic is held to embrace not only industries based upon clay, which is the traditional usage, but any industry utilizing earthy raw materials (sand, clay, feldspar, magnesite) to manufacture a product with the help of a kiln or a furnace. The ceramic industries in the United States therefore include glass, vitreous enamels, portland cement and certain electric-furnace products (abrasives); as well as structural clay products (building brick, paving brick, tile), refractories, earthenware (pottery, sewer pipe) and whiteware (china, decorative porcelain, electrical porcelain). Soluble silicates have been mentioned above.

Another class of industrial silicates, the silicate slags, is closely allied to the metallurgical industries. In the production and refining of some metals, an active reagent is the nonmetallic liquid floating on the surface of the metal, collecting and removing the gangue of the ore and also reacting with constituents in the metal. The metallurgical slag produced in greatest tonnage is blast-furnace slag from the manufacture of crude iron. It consists essentially of silica, alumina and lime, with subordinate magnesia. Some of it is utilized as raw material for portland cement. Slag from the open-hearth furnace for making steel contains principally silica, lime and oxides of iron and manganese; in some localities it is high enough in phosphorus pentoxide to be utilized as fertilizer. Slags from the acid and basic Bessemer processes and from the smelting and refining of copper, lead and tin are siliceous in varying degree.

See also references under "Silicon" in the Index.

BIBLIOGRAPHY.—J. W. Mellor, *Modern Inorganic Chemistry*, rev. and ed. by G. D. Parkes (1951); E. A. Hausner, *Silicic Science* (1955); W. L. Bragg, *The Atomic Structure of Minerals* (1937); C. F. Bell and K. A. K. Lott, *Modern Approach to Inorganic Chemistry* (1963); *Gmelins Handbuch der anorganischen Chemie*, 8th ed., System-Nummer 15, Silicium, Teil B pp. 1–226 (1959); H. Le Chatelier, *La silice et les silicates* (1914) and *Kieselsäure und Silicate* (1920); H. Berman, *Amer. Mineralogist*, 22:342–408 (1937); H. Strunz, *Mineralogische Tabellen*, pp. 36–49 (1941); Fritz Ephraim, *Inorganic Chemistry*, rev. by P. C. L. Thorne and E. R. Roberts (1948); W. M. Latimer & J. H. Hildebrand, *Reference Book of Inorganic Chemistry*, 3rd ed., ch. 14 (1951).

(R. B. SN.)

SILICON CARBIDE (CARBORUNDUM). Silicon carbide is a crystalline compound of metallic silicon and nonmetallic carbon which is of great technical and industrial importance. It was discovered accidentally by Edward G. Acheson in 1891 while he was experimenting with the use of electrical energy for the production of artificial diamonds. Acheson was heating a mixture of clay and coke in an iron bowl with an ordinary arc light carbon in an attempt to convert the carbon to diamond, and found some bright hexagonal crystals attached to the carbon electrode. He thought that he had prepared a new compound of the carbon and the alumina (aluminum oxide) of the clay. Because the natural form of fused alumina has the mineral name of corundum (*q.v.*), he called the new compound carborundum. This name was originally protected as a trade name but in many places became synonymous with silicon carbide. In 1893 Acheson was granted a U.S. patent covering his invention. About the same time Henri Moissan produced a similar compound from a mixture of quartz and carbon. However, in a publication of 1903 Moissan ascribed the original discovery to Acheson.

The significance of the discovery was immediately realized by Acheson, for he found that the crystals approximated the hardness of the diamond. His early product was offered for the polishing of gems, and sold at a price comparable with natural diamond dust. The new compound was found to be attainable with cheap raw materials and with good yields so that it soon became of industrial importance as a manufactured abrasive of wide application.

Properties.—Silicon carbide, which has the chemical formula SiC, crystallizes in the hexagonal system. Five crystalline-type habits in the hexagonal system were isolated and described in the technical literature. The commercial product has crystals up to ½ in. (13 mm.) average dimensions in the hexagonal plane and thicknesses of up to ¼ in. (6 mm.) in the other dimension. The crystals are usually found closely interlocked into a massive porous structure, but occasionally fully developed crystals are formed in voids of the furnace ingots. These are the spectacular large, shiny crystals which are often exhibited as typical silicon carbide. Pure silicon carbide is clear with a slight green tinge. The index of refraction is very high, 2.648 for the ordinary ray and 2.691 for the extraordinary ray. A slight amount of impurity will change the colour from light green to blue, gray and black. Typical impurities in the crystal are aluminum, magnesium, calcium, graphite and free silicon, all of which change the colour to black when present in amounts up to 0.5%. The crystal begins to oxidize when heated in air at around 1,000° C., forming a film of silica over the surface. This film produces iridescent and attractive colours when the proper film thickness is developed.

Some natural silicon carbide was found in the Canyon Diablo meteorite and bears the mineralogical name Moissanite after Henri Moissan.

A low-temperature form of silicon carbide (β-SiC) was discovered and isolated in 1926 by H. Ott. It has the same chemical composition but crystallizes in the cubic system and has properties different from the industrial product. The ordinary hexagonal variety of crystal has a hardness on the Mohs' scale of scratch hardness just above fused alumina or corundum and considerably below the hardness of the diamond. In 1891 silicon carbide was the hardest synthetic material which had ever been made. The discovery of boron carbide in 1929 took this distinction from silicon carbide.

Other properties of silicon carbide, in addition to its hardness, served to develop a large commercial market for it. The crystals are mechanically strong, have high thermal conductivity and desirable fracture characteristics which make them extremely useful in grinding wheels and in abrasive paper and cloth products. The high thermal conductivity, taken with the resistance to fluxing, fusion or oxidation of the material, makes it valuable in the manufacture of high-temperature bricks and other refractory products such as muffles, tubes and saggers. Silicon carbide also has unique electrical properties, including a high negative temperature coefficient of resistance. In the cold, silicon carbide might be classed as an insulator; as it is heated up to higher temperatures it becomes more conductive, so that at 2,000° C. it is as conductive as graphite. This places it in the class of semiconductors which have many special uses.

Industrial Production.—The total production of silicon carbide in the United States and Canada in the mid-1960s was in excess of 100,000 tons annually. Of this amount, roughly 60% was used for abrasives and 40% for refractories and other special uses. The total production of synthetic abrasives of all types during this period was about 400,000 tons annually, so that it is apparent that while silicon carbide was still an important abrasive material, it no longer represented the chief abrasive manufactured with electric power. This distinction had passed to fused alumina or corundum, which finds its chief use in the cutting and grinding of steel and other high tensile strength products. Silicon carbide is used mostly for the shaping of low tensile strength materials such as cast iron, bronze, glass and marble.

The commercial manufacture of silicon carbide is accomplished in much the same way throughout the world from the two basic raw materials, glass sand (pure SiO_2) and high-grade coke (low-ash petroleum or pitch coke). Most of the world production is localized in the Niagara district of the United States and Canada

where electric power from Niagara falls is available for the large requirement of electrical energy needed in the production. About 4 to 5 kw.hr. are consumed for the production of each pound of silicon carbide. The other large producing areas in the world are in Norway and in Czechoslovakia.

At mid-20th century the electric furnaces used for the production of commercial silicon carbide were similar throughout the industry and were much the same as the original furnace described by Acheson in his original patent. A typical furnace is a large brick box ranging in size from 20 to 50 ft. in length and from 10 to 20 ft. in width and 10 ft. deep. The refractory brick box serves only as a container to hold the crude mixture which consists of ten parts of glass sand and six parts of high-grade coke. Some variable amount of wood sawdust is added to the mixture to change its thermal conductivity and gas permeability characteristics. This loose granular mix is piled into the box until it is half full, and a trench is cut lengthwise in the loose mixture. This trench is filled with coarse metallurgical coke particles which are in contact with large carbon electrodes at each end of the box. This granular core is used ordinarily as a starting resistance to bring up the centre zone of the mixture to the reactive temperature which is from 2,000° to 2,600° C. Because of the long resistor and the many contacts between the carbon particles, a high voltage (up to 500 v.) is needed for starting the furnace, but as soon as some reaction takes place and a sheath of silicon carbide forms around the outside of the central resistor, the conductivity of the resistor is increased. This increase proceeds during the course of the run so that the voltage continually must be adjusted downward. As the current goes up to as high as 40,000 amp. the voltage must be adjusted down to a low of 75 v. Furnaces normally require from 1,000 to 3,000 kw. of power input to produce the abrasive efficiently. Expensive electrical equipment is necessary to control these large blocks of power through the required operating range. Production stops long before the central developing ingot of silicon carbide approaches the walls of the brick container because there is no known refractory which can stand the high temperatures involved in this furnace operation. The product is made in an envelope of itself. The chemical reactions liberate great volumes of carbon monoxide; in fact 1.4 tons of carbon monoxide is produced for every ton of silicon carbide recovered. This gas permeates the loose mix and burns on the outside of the furnace, aiding somewhat in the reduction of heat losses from the container.

See also references under "Silicon Carbide" in the Index.

BIBLIOGRAPHY.—R. R. Ridgway, "Temperature Measurements in Commercial Silicon Carbide Furnaces," *Trans. Electrochem. Soc.*, 61–19 (1932); A. H. Ballard and R. R. Ridgway, "Oxidation Rates of Silicon Carbide and Graphite Powders," *ibid.*, 61–30 (1932); R. R. Ridgway, "Manufactured Abrasives—Old and New," *Ind. Eng. Chem., News Ed.*, 21, 858 (June 10, 1943); L. S. Ramsdell, "The Crystal Structure of β-SiC, Type IV," *Amer. Min.*, 29, 431–442 (1944). (R. R. R.)

SILICONE, any one of a group of synthetic polymers (*see* POLYMERIZATION) composed of the elements silicon, carbon, hydrogen and oxygen. They are usually noted for ability to withstand comparatively high temperatures without decomposing.

Silicone polymers may take the form of oily liquids, resins or rubbery solids (elastomers), depending upon their chemical composition and the average size of the molecule. The commercial types of silicone resin and rubber are usually designed to have the greatest heat stability consistent with strength, adhesion and the other necessary physical properties, but some types are designed especially for fluidity or flexibility at very low temperatures or for some other special property such as solvent resistance. All types are quite inert chemically; they are indifferent to dilute acids and most reagents except strong alkalies and hydrofluoric acid, both of which are capable of destroying their basic structure.

The name "silicone" was coined by F. S. Kipping of University College, Nottingham, Eng., to describe compounds of silicon which were analogous to the organic ketones (*see* ALDEHYDES AND KETONES). A silicone therefore designates a substance which has a composition corresponding to two organic groups or radicals bonded to a silicon atom which in turn is linked to oxygen; it might equally well be described as an organosilicon oxide. The analogy between silicones and ketones begins and ends with the empirical composition, however, for the two groups of compounds are very different structurally. The ketones have simple molecules corresponding to their elementary composition, whereas the silicones always have polymeric structures in which the unit of composition is repeated many times in a single molecule. A silicone polymer may therefore be pictured as a chain or network of alternate silicon and oxygen atoms bearing appropriate organic groups attached to the silicon atoms:

$$\begin{array}{ccccccc} & R & & R & & R & \\ & | & & | & & | & \\ - & Si & -O- & Si & -O- & Si & -O- \\ & | & & | & & | & \\ & R & & R & & R & \end{array}$$

The organic groups R modify what otherwise would be a purely inorganic silicon oxide or silica glass, and provide the desired degree of flexibility, plasticity, or even fluidity. The kind of organic group and its size in relation to the silicon-oxygen framework thus has an important influence on the properties of the silicone polymer. In practice, the R groups are almost always methyl groups (CH_3-), but occasionally phenyl (C_6H_5-) or ethyl (C_2H_5-) groups.

Silicones usually are made by hydrolyzing the simpler organosilicon chlorides or ethers and condensing the products to polymeric form. The starting material is silica in the form of sand or quartzite, which by one of several methods is chemically combined with organic groups. In the Grignard method, silicon tetrachloride is prepared from the silica and is allowed to react with approximately two equivalents of organomagnesium chloride (Grignard reagent, *q.v.*) in an ether solvent to form a mixture of organosilicon chlorides from which the pure compounds are separated by distillation:

$$RMgCl + SiCl_4 = RSiCl_3 + MgCl_2$$
$$2RMgCl + SiCl_4 = R_2SiCl_2 + 2MgCl_2$$
$$3RMgCl + SiCl_4 = R_3SiCl + 3MgCl_2$$

In the direct method, elementary silicon is prepared by reduction of the silica and is caused to react directly with an organic chloride in the vapour phase and in the presence of a catalyst to form a mixture of organosilicon chlorides, from which the pure constituents again are distilled:

$$2RCl + Si = R_2SiCl_2, \text{ etc.}$$

In a third method, ethylene is absorbed by dichlorosilane made from silicon:

$$2HCl + Si = H_2SiCl_2$$
$$H_2SiCl_2 + 2C_2H_4 = (C_2H_5)_2SiCl_2$$

Pure organosilicon chlorides from any of these methods become the intermediates from which the silicone polymers are made.

Silicone resins are made by hydrolyzing a mixture of monosubstituted and disubstituted silicon chlorides, sometimes in a solvent. Reaction with the water replaces the chlorine with hydroxyl groups, and the hydroxides intercondense to form a polymer:

$$R_2SiCl_2 + 2H_2O = R_2Si(OH)_2 + 2HCl$$

$$xR_2Si(OH)_2 = \left[\begin{array}{c} R \\ | \\ -Si-O- \\ | \\ R \end{array}\right]_x + xH_2O$$

Silicone resins prepared in this way usually are applied in the form of a solution or varnish. After the solution has been applied to the desired surface the solvent is evaporated and the resin is cured by heat, sometimes with the assistance of curing agents or catalysts. The final product has a molecular structure in which the chains of silicon and oxygen atoms are interconnected or cross-linked by extra atoms of oxygen on the monosubstituted silicon atoms, and is therefore comparatively insoluble and infusible. Such resins are useful as insulating and protective coatings, and some are serviceable at temperatures up to 600° C. for extended periods.

Silicone oils may be made by preparing linear polymers of dimethylsiloxane of a length corresponding to the desired viscosity and vapour pressure. In order to prevent further change of molecular weight through condensation, the chains are blocked by attaching unreactive groups to the ends. Such chain-blocked oils show little change of viscosity with temperature and so are useful at low temperatures as well as high. They do not oxidize or sludge as readily as do hydrocarbon oils, and they are suitable as dielectric media because they have low electrical losses over a wide range of temperatures and frequencies.

Silicone elastomer is made by hydrolyzing pure dimethyldichlorosilane and extending the resulting linear polymers to very high molecular weight, so that the chains consist of thousands of dimethylsiloxane units. Such high polymers have elastic properties, and if they are milled with inorganic fillers and curing agents, the compounded material can be cured or vulcanized to a rubbery product.

Like silicone oil and resins, silicone rubber can be utilized at temperatures both above and below those permissible for the organic elastomers, both natural and synthetic.

Very thin films of silicone polymers also are formed *in situ* by reaction of the organosilicon chlorides with the adsorbed water on cellulose or glass, and the films so formed cause the surface to become highly water-repellent. Such a film is strongly bound and can be removed only by chemical action or by severe abrasion, but may readily be cleaned by alcohol or by solutions of wetting agents.

The process is useful wherever paper, cloth or glass is to be made water repellent.

BIBLIOGRAPHY.—E. G. Rochow, *Introduction to the Chemistry of the Silicones*, 2nd ed. (1951); R. R. McGregor, *Silicones and Their Uses* (1954); H. W. Post, *Silicones and Other Organic Silicon Compounds* (1949); E. G. Rochow, D. T. Hurd and R. N. Lewis, *Chemistry of Organometallic Compounds* (1957). (E. G. Rw.)

SILICON STEEL: *see* MAGNETISM: *Permeable Materials;* STEELS, ALLOY.

SILICOSIS is a chronic, incurable disease of the lung produced by the inhalation of fine particles of silica, the chief mineral constituent of sand and of many kinds of rock. An occupational disease, it is one of the many forms of pneumoconiosis (*q.v.*) afflicting persons who work in dust-laden air. The disease rarely occurs with exposure of less than two years' duration and rarely to dust concentrations of less than 6,000,000 particles per cubic foot (about 210,000 per litre) of air. Only particles less than 10 microns ($\frac{1}{2,500}$ in.) in diameter reach the lung and particles one to three microns in diameter are the most injurious. Combined forms of silica (silicates) do not produce damage and are dangerous only because of the percentage of free silica they may contain. Actually, inhaled inert dusts may modify the effects of silica.

The disease is characterized by whorls of fibrous tissue that form about aggregates of silica particles. These foci gradually enlarge to form fibrotic masses involving both the functional tissues of the lung and the draining lymph nodes. Symptoms are shortness of breath followed later by cough, tightness across the chest, and weakness. Diagnosis is based on the X-ray appearance of the lungs and a history of exposure to silica dust. The most frequent and important complication is superimposed tuberculosis. Although silicotuberculosis is still a serious disease, prognosis has been improved by the advent of tuberculosis chemotherapy.

Occupations in which silicosis is a hazard include mining, stonecutting, grinding, polishing, glassworking, sandblasting, and pottery, steel, and abrasive-soap manufacturing. Since there is no specific treatment for silicosis once it is established, control of the disease lies in prevention. The use of proper ventilation, wet-grinding techniques, and face masks to filter out dust has helped to lessen the incidence of silicosis. (W. R. By.)

SILISTRA (ancient DUROSTORUM), the chief town of the *okrug* of that name in northeastern Bulgaria, lies on the right bank of the Danube (where it forms the frontier with Rumania) and 65 mi. (105 km.) ENE of Ruse. Pop. (1963 est.) 28,736. To the south and southeast are the remains of the old fortifications. The town manufactures bricks and tiles, furniture, rush matting, and cotton textiles. Its river port has a large silo and exports mainly cereals. Silistra is linked by rail with the Ruse-Varna line, and the Constanța-Bucharest, Rum., road passes through it.

The Romans founded a military camp at Durostorum (later Dorostolum; medieval Dristra; Bulgarian Drstr or Drustur) in the late 1st century B.C. or early 1st century A.D. It became one of the important towns of Moesia Inferior (*see* MOESIA). Its importance was maintained under the Byzantines and Bulgars (*see* DOBRUJA). For some time in the 10th century it was apparently the seat of the Bulgarian patriarch, and the emperor John I Tzimisces defeated the Russian prince Svyatoslav I there in 971. Under the Turks (who finally gained possession of it by 1420) Silistra was an important fortress in the strategic Turkish quadrilateral Silistra-Ruse-Kolarovgrad (Shumen)-Varna. The Russians attacked Silistra several times in the 19th century. The town was assigned to Bulgaria in 1878 and to Rumania, after the second Balkan War, in 1913 as part of the southern Dobruja. It was returned to Bulgaria in 1940. (*See* also BULGARIA: *History.*)

SILISTRA OKRUG is a fertile agricultural region and produces cereals, beans, sugar beet, and grapes; stock is also raised. Area 1,108 sq.mi. (2,870 sq.km.); pop. (1962 est.) 170,300. (AN. BE.)

SILIUS ITALICUS (in full TIBERIUS CATIUS ASCONIUS SILIUS ITALICUS) (25 or 26–101), Latin epic poet, was the author of a long epic entitled *Punica,* an account in hexameter verse of the second war of the Romans against the Carthaginians (218–201 B.C.). Silius was an advocate of distinction in his early years, but he soon took to the public service and was consul in A.D. 68 (the year of Nero's death). His life thereafter is described by the younger Pliny in one of his letters: "he won repute by his proconsulship of Asia, and obliterated by the praiseworthy use he made of his leisure the stain he had incurred through his former exertions" (presumably as an assistant of Nero). "In dignity and contentment, avoiding power and therefore hostility, he outlived the Flavian dynasty, keeping to a private station after his governorship of Asia."

Silius was a rich man and able to indulge his tastes as a patron of literature and the arts. He so venerated both Virgil and Cicero that he bought for himself Virgil's tomb at Naples (which he restored) and Cicero's estate at Tusculum. In later life he imitated both, the former in writing epic, the latter in discussing philosophy with his friends, such as the poet Martial, Epictetus, who thought well of him as a philosopher, and Cornutus, the Stoic, who dedicated to him a commentary on Virgil. The Stoicism of Silius is not as marked in his epic as that of Lucan in his, but it is discernible, as one might expect, in a story of long and embittered warfare; and it was clearly to be seen in his life, for at the age of 75 he discovered that he was suffering from an incurable disease and starved himself to death.

The *Punica,* the only work he is known to have committed to writing, is the longest epic in Latin literature, comprising over 12,000 lines. It deals with the second only of the Punic wars, though the story of Regulus (*q.v.*) appears in it as a digression. Silius draws heavily on the historian Livy for his material. For the most part he sticks well to his story and conscientiously describes all the six great battles of the war. The "details of slaughter" (as his translator J. D. Duff puts it) are often repulsive and the catalogues of armies monotonous. It is very imitative of Virgil's *Aeneid* both in its form and in its mythology. The gods take part, for example, and Juno frustrates the Romans just as she does the Trojans in the *Aeneid.* It lacks the poetry and the drama of Virgil, and the metre is monotonous and the style sometimes repetitive. The epic has indeed been harshly judged by its critics and has been rarely edited since the 18th century, but Silius can tell a story well and in simple language, far purer than the overdecorated style of other Latin poets of the Silver Age; and his accounts of military matters (*e.g.,* of the operations of siege engines) often bear the stamp of experience. His characters are mostly rather wooden, but his Hannibal is drawn with some dramatic skill and steals the place of hero from Scipio. Of the minor characters the best drawn are Fabius Cunctator,

an evident copy of Lucan's Cato, and Paulus, the consul killed at Cannae.

The last three books show signs, as well they might, of fatigue; but posterity should be grateful for about half a dozen beautiful pieces of verse, mostly in dramatic scenes of war.

The poem was discovered in a manuscript in 1416 or 1417; from this manuscript, now lost, all existing manuscripts are derived. There were two *editiones principes* (1471). Other editions are by D. Heinsius (1600); J. C. T. Ernesti (1791); L. Bauer, "Teubner Series" (1890); J. D. Duff, with English translation, "Loeb Series," two volumes (1934).

For his life the authorities are Pliny, *Epistles*, iii, 7; Tacitus, *Histories*, iii, 65; with frequent references in Martial. (An. K.)

SILK, the thread and cloth made from the fibre produced by silkworms. A variety of insects, including spiders, produce silk but most of these substances—a spider's web, for example—are too adhesive and fragile to be suitable for textiles. The fibres used for commercial silk manufacturing are produced mainly by the mulberry silkworm of the orient, *Bombyx mori*, and a few other closely related members of the same family (Bombycidae), in the form of a cocoon that acts as a shell to enclose and protect the insect during its period of transformation from caterpillar to chrysalis to moth.

The breeding and cultivation of silkworms is termed sericulture. Silk manufacture includes all phases of silk processing, from the production of thread or yarn directly from the cocoon to the finishing of the woven or knitted fabric.

This article is divided into the following main sections:

I. HISTORY

The origin of silk is so ancient that the course of its early development is to be found only in legend and fable. The industry undoubtedly began in China, where, according to native record, it has existed from antiquity. It became important in the Chinese economy under the emperor Huang Ti (2640 B.C.), whose empress, Hsi Ling Shi, is said to have made sericulture fashionable by giving it her personal attention. Chinese literature testifies to the fact that royalty and nobility followed her example by cultivating mulberry trees, raising worms, and reeling silk. The empress herself is credited by the Chinese with the invention of silk reeling and the loom. Silk was used as a medium of exchange within the country and also was exported. The Chinese zealously guarded the methods of production, and under imperial decree, disclosure of any of them meant death by torture for the informer. Thus, many centuries passed before any knowledge of the source or manufacture of this fabulous fabric spread, and when it did, around A.D. 300, it was through Japanese initiative. The *Nikongi*, one of the most ancient Japanese histories, relates that some Koreans were sent by Japan to China to bring back weaving instructors. The Koreans returned to the Japanese court with four Chinese girls, who taught their art first to royalty and then to the people. A temple was built at Settsu (Osaka Province) in honour of these four pioneers of the industry that eventually helped raise Japan to a world power.

Ancient Sanskrit literature refers to an Indian silk industry. The word used could have denoted either weaving of silk fabrics or, more probably, the manufacture of silk garments. Even at that early date the rajas of that country were extremely partial to its opulence. According to an old legend, in about 140 B.C., silkworm eggs and mulberry seeds were carried to Khotan, in an old southern Chinese province near the border of India, hidden in the lining of the headdress of a Chinese princess who had gone there to be married and wanted her beautiful native silks at the new court. This may have made possible the beginning in India of the actual breeding and rearing of silkworms. Since sericulture now is known to have been established in India first between the Brahmaputra and Ganges rivers, it seems credible that it was introduced overland from China. From the Ganges Valley, sericulture moved slowly west into Persia and central Asia.

The Old Testament has two incidental references to silk, in Ezek. 16:10, 13. Aristotle, in his *Historia animalium*, described the silkworm as "a great worm which has horns and so differs from others. At its first metamorphosis it produces a caterpillar, then a bombylius and lastly a chrysalis . . . all these changes taking place within six months. From this animal, women separate and reel off the cocoons and afterwards spin them. It is said this was first spun in the island of Cos by Pamphile, daughter of Plates." Long before this raw silk probably had been imported by Cosians and woven into *coa vestis*, a gauzy, sheer fabric for which the island was famous in antiquity.

Toward the beginning of the Christian era, raw silk began to be exported from the East to Rome. When it first appeared there, silk was of enormous cost—it was literally worth its weight in gold—and its use by men was considered effeminate. Allusions to silk became common in classical literature; although these references show familiarity with the fabric itself, they are vague and inaccurate as to source. It was believed by some to be a fleece from a bush or tree, by others a fibrous substance of the inner bark of a tree. Even the naturalist Pliny the Elder knew nothing more about the silkworm than Aristotle did.

By the 2nd century A.D. the silk-weaving industry flourished in Egypt, Syria, and Palestine and by the 4th it had developed in Constantinople and Persia. When the seat of the Roman Empire was moved to Constantinople, the emperor Justinian, seeing the great commercial potentialities of silk, took steps to create a monopoly, setting up looms within the imperial palace. He also attempted to divert the thriving silk trade from the Persian route along which it was then brought into eastern Europe, but he failed in this endeavour. He was able, however, to persuade two Persian monks who had lived in China to return there and smuggle silkworms to Constantinople in the hollows of their bamboo canes (*c.* A.D. 550). These few hardy silkworms were the beginning of all the varieties that stocked and supplied European sericulture until the 19th century.

The silkworm and its product flourished in the West during the Middle Ages. Byzantine silks, widely used for ecclesiastical vestments and altar cloths, became world famous, and feudal lords adopted the cloth for themselves, their ladies, and their courts. Knight crusaders and other military adventurers carried back to Europe silk from the Middle East, with Arabic patterns and colours quite different from anything seen up to that time. More significantly, from the point of view of the European industry, the Moors, who had conquered Sicily and parts of Spain, established sericulture there. When Sicily was taken by the Normans in the 11th century, Roger I, the first Norman king, improved the quality and increased the production of Sicilian silk, the use of which spread north to the mainland. Sericulture and silk manufacture were profitable, and the industry was encouraged by the nobility of medieval Florence, Genoa, Milan, and Venice. Under royal patronage these cities became famous for silks and gained a prominence never completely relinquished. In France, silk weaving was started at Tours in 1480 under Louis XI, and in 1520 Francis I brought silkworm eggs from Milan to be reared in the fertile Rhône Valley. It remained, however, for J. B. Colbert, minister of finance under Louis XIV, to establish the industry firmly in that country by offering premiums for the planting of mulberry trees and by otherwise stimulating local French interest in sericulture.

Silk manufacture began in England in the 14th century, but

the first real impetus to the British industry came in the 16th century with the emigration of large numbers of skilled Flemish weavers, a result of the prolonged war between the Netherlands and Spain. James I was interested in the industry and made many attempts to stimulate mulberry planting and silkworm rearing in Britain and in the colonies. A second wave of highly skilled immigrants followed the revocation of the Edict of Nantes in 1685; most of these French Huguenot weavers settled in an area in London known as Spitalfields, where a guild of silk weavers had been formed in 1629. In 1713 English silk weavers petitioned parliament for a higher tariff on the importation of silk fabrics, because French silks could be bought in England at prices lower than those prevailing for the domestic product. Stressing the economic importance of their industry, the English weavers cited the manufacture of black silk for hoods and scarves, which had been unknown in England 25 years before and which had grown to an annual yield of £300,000 sterling. At this time the silk industry in England also had to depend on continental Europe for its supply of thrown silk for weaving. (Throwing is a vital intermediate process of silk manufacture.) The first English throwing mill was built in Derby in 1719 by Sir Thomas Lombe; plans for it are said to have been obtained by his half brother John Lombe, who disguised himself as a common workman, gained entrance into an Italian mill, and smuggled out drawings of the machinery used there. In 1825 the British, Irish and Colonial Silk Company, with capital of £1,000,000, was formed to introduce silk production into Ireland. It was a dismal failure. Despite these various efforts, sericulture never became important in British industry.

Sericulture was introduced into the Western Hemisphere in 1522, under Hernán Cortés, in Mexico. Mulberry trees were planted and eggs were brought from Spain, but the industry died out before 1600. In 1609 James I tried to reestablish the silkworm in the New World, but the attempt failed because of a shipwreck. Ten years later a determined effort was made to implant sericulture in Virginia. Laws were passed to stimulate the industry's development, which was bolstered further by bounties and rewards. Detailed instructions on the care of the worms and the raising of trees were sent to the colonists. One such booklet published in London in 1620, *OBSERVATIONS To Be Followed, For the making of fit roomes to keepe Silk-wormes in: As also For the Best Manner of Planting of Mulbery trees to feed them*, concluded with the following:

> It is a thing well knowne, that a few Silk-wormes, fed at large, and ease, make farre more silke than a greater number, pent in narrow and ill-favoured roomes. No ill smels must come neere them, they must be kept sweet, and oft perfumed: therefore having such store of sweet-woods in Virginia as you have there, you shall do well to make their roomes and tables of these woods: sweet sents being a thing most agreeable to them. Bee carefull to doe things curiously and thorowly well for them at the first, for your more plentifull and certain gaine after: considering the charge to you is all one: and a thing once wel done, they say, is twice done, which will thereby also bring you twice double profit, with long continuance.

Britain continued to offer rewards of various kinds to the American colonies for silk cultivation. At the onset of the American Revolution, Benjamin Franklin and others were trying to establish a silk-spinning factory at Philadelphia, Pa. After the war, bounties were offered by several of the states, the first by Connecticut in 1783. About 1838 a frenzied speculative mania for sericulture developed in the United States, caused principally by a promoter who extolled the suitability of the South Sea Island mulberry tree for feeding silkworms. So intense was the fever that all kinds of crops were uprooted to make room for plantations of the new golden mulberry. In Pennsylvania, $300,000 was paid for plants in one week; frequently, young trees were sold two and three times in a few days at steadily increasing prices. Within a year, the falsity of the claims became apparent, and the speculation collapsed.

Sporadic efforts to establish a silk industry in the United States continued. The next peak of interest followed the tremendous increase in the use of silk after World War I. Prices of raw silk rose steadily in the United States from $3.64 a pound in 1914 to $21 a pound in 1919. An intense effort was made in California to produce larger cocoons and more frequent crops. The cocoons raised were superior, but the local cost of hand reeling was prohibitive, and the volume was not sufficient to make it economically profitable to have the work done in Japan. Another attempt at sericulture was initiated in the United States by a promoter who talked of meeting world competition and spread the vision of large profits from the production of a silk filament precoloured by direct injection of dye into the silkworm. Louisiana, Florida, and Canada witnessed scenes of mounting activity in the purchase of mulberry trees and rainbow-hued silkworms. Newspaper accounts implied that fortunes were to be made. A year later, investors were unable to find buyers for their cocoons. Once more sericulture in the United States had proved impracticable.

In Iran the antinarcotic division of the League of Nations sought to replace the growing of the opium-producing poppy with sericulture; that country was one of the original hosts to the silkworm and was believed to have natural characteristics favourable for it. The industry was started, and in 1939 Iranian government representatives were sent to New York in an endeavour to sell the fabrics in the United States, which was the largest user and importer of silk, but the types of cloth, styling, and sparse production made the Iranian product unacceptable to the U.S. market. At about that time, U.S. manufacturers turned again to China, original source of silk. It was found that production methods there were antiquated, so the silk industry in the United States subscribed a large fund to initiate the study of sericulture in three of the missionary colleges in China. Land was bought for mulberry orchards, new equipment was installed in modern buildings, and courses were planned to teach the latest methods. The Chinese were disinclined to change their methods, however, since they found that even without doing so they could continue to sell their production to Europe. This, coupled with the increasing domestic upheavals in China, finally resulted in the discontinuance of attempts by the Americans to modernize the Chinese silk industry.

India was another experimental field for the commercial development of silk. Interest also was shown by officials in the Black Sea area of the U.S.S.R., in Greece, and in Spain. During World War II silk was grown and spun in Brazil, which has climatic conditions favourable for sericulture. Despite these diverse efforts, the principal source of supply continues to be Japan, which has dominated the world market since 1865.

Silk can be produced and marketed profitably where the climate is favourable and where labour is cheap, abundant, and skilled. The rising standard of living the world over makes it necessary for the industry to continue its search for additional mechanization of the various processes in order to keep the product competitively priced and a great deal of work to this end was being done in Japan in the 1960s.

Japan is by far the greatest consumer of raw silk, using approximately ten times as much as the next largest consumers, the United States, Italy, and India. (Figures are not available for China and the U.S.S.R.) In 1938, 80% of the silk consumption in Western countries went into hosiery. After World War II stocking manufacturers turned to nylon and the U.S. ceased to be the leading silk consumer. Silk consumption by Japan and certain other countries was, by the 1960s, back to prewar levels, but for cloth alone, which had thus made striking advances.

In spite of the progress in man-made fibres for clothing since the end of World War II, silk remains the first choice for luxurious and elegant clothes the world over. In addition to the traditional types of silk fabric used for women's evening, cocktail, and all "special occasion" dresses, "wild silk" fabrics such as shantung and Thailand silk enjoyed a vogue in postwar years. Silk suitings for both men and women also proved popular. In the lower price ranges, silk figures very little, since a cheap silk fabric does not wear well and does not have the easy-care characteristics of similarly priced fabrics in man-made fibres. An exception is "Jap silk," a lightweight, closely woven, strong, and relatively inexpensive silk which is widely used for linings.

II. SILK-SPINNING INSECTS

1. Cultivated Silkworms.—*Bombyx mori.*—The mulberry leaf-feeding silkworm, *Bombyx mori*, the principal source of silk,

belongs to a family of Lepidoptera that includes some of the largest and handsomest moths and butterflies. *B. mori* itself is an inconspicuous moth, ashy white in colour, with the male's body less than one-half inch long and the female's longer and heavier. Its wings are short and weak; the fore pair are slightly curved and the hind pair do not reach the end of the body. The larva, black when newly hatched, turns to cream colour, grows to 3 or 3½ in. in length, and is slender in comparison with others of its family. The second thoracic ring is humped, and there is a spinelike horn at the tail. The silkworm produces one generation annually where the seasons are defined, for example, in Europe and the Middle East. In Japan reproduction takes place two or three times during the year, and in parts of India and China it is almost continuous. The quality of silk produced by the silkworm is in inverse ratio to the number of hatchings.

The silk glands, which wind along the underside of the caterpillar's body and which produce fibroin, consist of two long, pipelike sacs called hind-silk glands. These empty into two S-shaped tanks called mid-silk glands that ripen the fluid. Toward the forward part of the body, the glands again become thin and are known as fore-silk glands. Near the head the two glands open through a common orifice, the spinneret, on the underlip of the caterpillar. As the caterpillar reaches maturity, the glands become gorged with a clear, viscous fluid, which is used to spin the cocoon. As this spinning fluid (the two strands of fibroin) is ejected through the spinneret, it is joined by a second secretion, sericin, from another set of glands located at the opening. The sericin solidifies immediately upon contact with the air, and the double fibre emerges as a single filament. The cocoons are hard, compact shells, oviform in shape and frequently constricted in the middle, with straggling, flossy filaments on the exterior and the interior layers so densely agglutinated as to resist complete unwinding. The fibres vary from white to yellow in colour. With variations according to sex and species, the cocoons measure from 1 to 1½ in. in length and ½ to 1 in. in diameter and weigh 15 to 50 gr.

Within 10 to 12 days after the completion of the cocoon, the enclosed insect is ready to emerge. It moistens one end of the cocoon with enzymes that partially dissolve the fibres, enabling the creature to push aside the filaments and emerge as a perfect moth. The sexes mate almost immediately; the female lays 500 or more eggs within four to six days and soon dies. Under improved methods, the sericulturist destroys the male immediately, as it is believed multiple matings by the male are deleterious to the quality of the ultimate silk.

Other Bombyx Species.—Numerous allies of the mulberry silkworm spin usable cocoons and produce what is known commercially as wild silk. Six species of *Bombyx*, differing from *B. mori* but also mulberry-feeding, have been domesticated, more or less, in India. Included in this group are *B. textor* of Bengal, a large species having one generation yearly and producing a soft, flossy cocoon; *B. sinensis*, producing several generations annually and making a small cocoon; *B. croesi*, *B. fortunatus*, and *B. arracanensis*, all yielding several generations annually and forming reelable cocoons. Many other mulberry-feeding Bombycidae have not come under cultivation.

2. Wild Silkworms.—The family of giant silkworms (Saturniidae) includes several species that produce a silk that is commercially marketable but not so fine in quality as that spun by *Bombyx mori*. One of the most important moths yielding wild silk is *Antheraea pernyi*, an oak-leaf feeder native to Mongolia. From it is derived the brownish tussah silk imported extensively into Europe. Closely allied to it is *A. mylitta*, which feeds on leaves of the jujube tree in India and spins a large, compact silver-gray cocoon. *A. assama* is domesticated to some extent in its native Assam in India.

Cultivation in Europe of the Japanese *A. yamamai*, another oak-leaf feeder, was disappointing: the eggs, brought to Europe in the early 1860s, were found to hatch at a time when oak leaves were not available, and the silk did not compare with that from the domesticated mulberry silkworm. The mezankoorie moth of Assam also yields a valuable cocoon, as does *Attacus atlas*, an

BY COURTESY OF THE JAPAN SILK ASSOCIATION

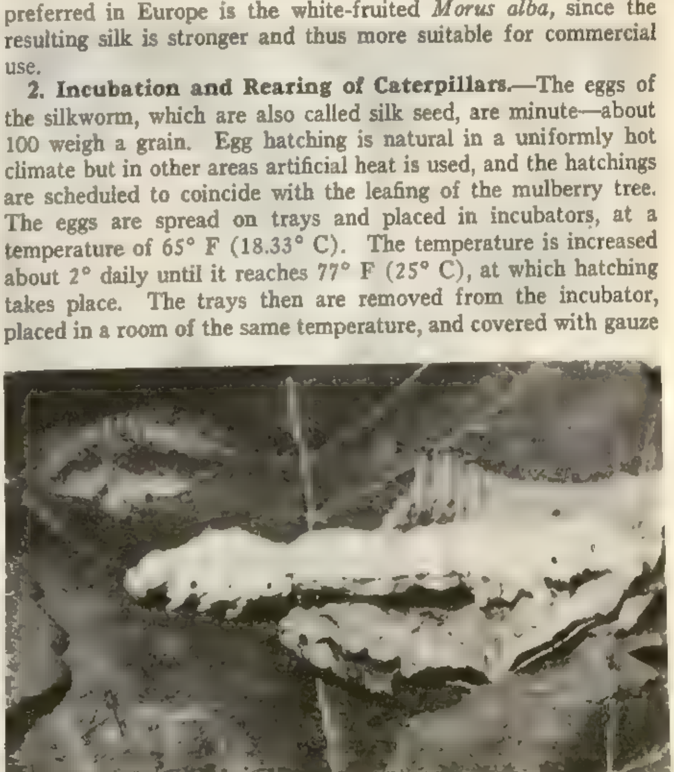
FIG. 1.—RACKS CONTAINING THE FEEDING SILKWORMS

omnivorous feeder found throughout India, Ceylon, Burma, China, and Java. *Samia cynthia*, which feeds on a variety of plants, including ailanthus and castor oil plant, spins a loose, flossy cocoon, orange-red or sometimes white.

III. SERICULTURE

1. Food Supply.—The first requirement of sericulture is a stock of mulberry trees adequate to feed the silkworms in their larval stage. The soil in which the mulberry trees grow and the age and condition of the trees are important. An elevated position with dry, friable, well-drained soil produces the best-quality leaves. Many species are cultivated, and, as they have grown for at least 3,000 years, they show the complex variations of most cultivated plants. In Japan, pruning produces more luxuriant growth and, over a period of time, has reduced the height of the trees, obviating the need of ladders in harvesting. The mulberry preferred in Europe is the white-fruited *Morus alba*, since the resulting silk is stronger and thus more suitable for commercial use.

2. Incubation and Rearing of Caterpillars.—The eggs of the silkworm, which are also called silk seed, are minute—about 100 weigh a grain. Egg hatching is natural in a uniformly hot climate but in other areas artificial heat is used, and the hatchings are scheduled to coincide with the leafing of the mulberry tree. The eggs are spread on trays and placed in incubators, at a temperature of 65° F (18.33° C). The temperature is increased about 2° daily until it reaches 77° F (25° C), at which hatching takes place. The trays then are removed from the incubator, placed in a room of the same temperature, and covered with gauze



BY COURTESY OF THE JAPAN SILK ASSOCIATION

FIG. 2.—SILKWORM FEEDING ON MULBERRY LEAVES

or perforated paper. Finely chopped mulberry leaves are spread over this covering, and the larvae wiggle through the openings to begin feeding.

The place in which the caterpillars are reared must be scrupulously clean, disinfected, lime-washed, and well ventilated. A vast number of newly hatched larvae may be kept in a small space, but as they grow, they and their food require increasing room. A common method of maintaining them during feeding is by means of trays stacked on shelves about 3 ft. wide, with adequate passage on all sides. The shelves are constructed of light scantling or bamboo, beginning about a foot from the floor and running at two-foot intervals, as high as a person can conveniently reach.

After three or four days the caterpillars are fed whole leaves. They are voracious during the feeding period, which continues for about 42 days, interrupted by four 24-hour intervals of sleep. On awakening from these sleeps the caterpillars slough their skins and then resume feeding. After the last sleep they start a final 10-day period of feeding. In the course of the total feeding period, they consume about 20 times their own weight in leaves and grow from a length of less than $\frac{1}{16}$ to about 3 in.

Studies have indicated that one ounce of seed eggs produced 30,000 to 35,000 silkworms, which consumed one ton of ripe leaves. From this seed 130 to 140 lb. of fresh cocoons were produced, with an ultimate yield of 12 lb. of reeled silk.

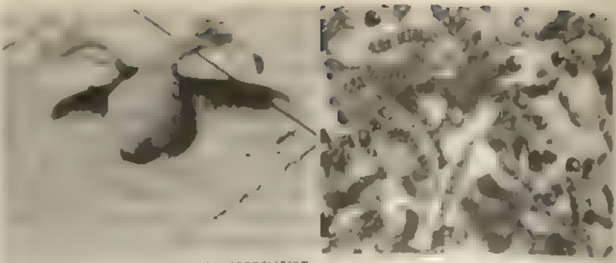
3. The Cocoons.—The caterpillars show they are ready to spin their cocoons by raising the forepart of the body and waving it slowly from side to side. Provisions now are made for proper spinning. A commonly used method involves trays with compartments about $2 \times 2\frac{1}{2}$ in. in size, each of which accommodates one caterpillar. The caterpillars are tossed lightly into the compartments, and the trays are fastened together at intervals, one above the other, to form mounts. The mounts are raised from the floor and revolved slowly for 60 hours, the time required for the caterpillar to complete its cocoon. The temperature of this room is kept at about 73° F (22.78° C) and the humidity not higher than 70%. After an additional seven days, during which period the caterpillar changes to a pupa, the cocoons are harvested.

If the silkworms were overcrowded during the time of spinning, two or more may have spun a common cocoon, which is impossible

FIG. 3.—(LEFT) SILKWORM SPINNING ITS COCOON; (RIGHT) PUPA INSIDE FINISHED COCOON

to put through the usual reeling processes because of the intermingling and interweaving of the fibres. Such cocoons, called doupions, are set aside for special handling. The silk so processed produces a slubbed yarn, and, when such textures are in demand, doupion cocoons are produced deliberately.

In the harvesting, all the cocoons, other than the doupions and those selected for reproduction of eggs, are treated to preserve them intact for reeling—that is, the chrysalis must be killed without damage to the cocoon. In primitive sericulture this was done by exposure to the sun, but in many cases the heat failed to kill the pupa before it emerged as a moth, resulting in a damaged cocoon. Another method is suffocation by steam, in which the cocoons are placed in drawers over steam for eight or ten minutes



BY COURTESY OF THE JAPAN SILK ASSOCIATION

FIG. 4.—(LEFT) FEMALE MOTH LAYING EGGS; (RIGHT) LARVAE HATCHING

and then are turned and air-dried for six weeks to two months. The process is time consuming and requires an excessive amount of hand labour. The safest, most practical, and widely used method is suffocation by hot air, using an oven built of brick or concrete and equipped with a series of eight moving belts. The temperature in this oven at the top is 210° F (98.89° C), graduated to 140° F (60° C) at the bottom. Cocoons are fed into a hopper at the top and by means of the belts pass to the bottom, where they drop into canvas bags. Passage of the cocoons through the whole process requires seven or eight hours, during which time the chrysalises are killed and the cocoons are dried, ready for reeling.

4. Selection of Eggs for Reproduction.—Production of seed eggs is carefully controlled to reduce the great risk of propagating diseases. In Japan, the largest and most important source of these eggs, producers can operate only under government licence and control; regulations are stringent and are enforced assiduously. Symptoms of any of the dreaded silkworm diseases are readily recognized by adept sericulturalists, and moths that show signs of disease at any stage of growth are destroyed. Female moths selected to lay eggs are placed in a special room, on individual, flat, eight-inch squares of cheesecloth. When the laying has been accomplished, some females, chosen at random, are destroyed and chemically or microscopically examined to ascertain that they are wholly free of disease. The eggs also are examined. If they are found to conform to government standards and regulations, they are washed, dried, and carefully packed for local distribution or shipment.

The primary aim of this phase of sericulture is the propagation of lively, disease free stock. Crossbreeding also is done to develop new varieties for specific requirements that range from higher yield of reeled silk to hardier strains for various climates.

5. Diseases.—The silkworm has been propagated for more than [illegible] years under purely artificial conditions, frequently insanitary, with the objective of securing the greatest amount of silk possible; it is, therefore, only to be expected that the [illegible] is subject to many serious diseases. Among the most fatal and disastrous of these is muscardine, caused by fungus, which is peculiarly contagious and infectious; spores of the fungus, transferred from one silkworm to the other, penetrate the skin and germinate corpuscles within the insect's body. Fortunately, muscardine has not been epidemic in many years, and in some countries is no longer considered a serious problem.

Another disease, which was first noticed near Avignon, France, in the early part of the 19th century, and was later called pébrine, manifests itself by dark spots on the skin of the larvae; eggs do not hatch perfectly; larvae are weak, stunted, fastidious in eating, and many die; cocoons are loose; and moths are feeble. In the second generation there is an increased intensity of [illegible] symptoms. The disease, thus, is hereditary, infectious, and contagious. By [illegible] French cultivators were forced to import [illegible] but, despite this measure, silk production continued to [illegible] Italy the disease was even more disastrous, and the malady [illegible] eastward to such an extent that unaffected eggs could be obtained only from Japan.

In early research on the disease, vibratory corpuscles [illegible] named for their identifier Cornelia, were found in the blood [illegible] tissue of the affected silkworms. In 1865 the French government commissioned Louis Pasteur to study the disease. He established

...these Cornelian corpuscles are parasites and are not only the ... but the cause of the sickness, and that it is transmitted ...gh heredity or contagion with diseased worms or by the ... of leaves on which the parasitic corpuscles are present. ...er concluded that the existing epidemic conditions were only ...aggeration of a normal state, and he advised selective breed... regenerate the stock. Eventually, a supply of healthy Eu... stock was reestablished, but, of all the diseases to which ...lkworm is subject, pébrine remains the most dreaded.

...ther disease of the silkworm is *grasserie*, which causes the ... to become rather thin and shiny; the body of the silkworm ... to have become distended or fattened, or become *engraissé* ...ce the name). *Grasserie* generally is attributable to cold, ... air, the use of wet leaves in feeding, or sudden temperature ...ges. It, therefore, is more accidental than contagious and ...ly reaches dangerous proportions.

...cherie is a contagious intestinal disease caused by a bacillus, ...ified by Pasteur, of the cholera group. The disease invariably ... during the most active feeding period, generally after a ...d of coarse leaves taken from mulberries pruned the same ... and grown in damp soil. Preventive measures include the ...ance of fermentation of leaves, maintenance of hygienic ...ation and proper humidity, and the use of quicklime where ...essary to facilitate the transpiration of the silkworm.

IV. PHYSICAL AND CHEMICAL PROPERTIES OF SILK

...mmon cocoons with chrysalises enclosed weigh from four-... of a gram to two grams. About one-sixth of this weight is ... Three-quarters of a cocoon can be reeled; the remainder is ...ce floss and husk. The reelable portion varies in thickness ...nges from 500 to 1,200 m. in length. The thread usually is ...er and stronger toward the middle of the reeled portion ... the end. It has been estimated that under modern seri...re one pound of silk requires from 1,500 to 2,500 cocoons.

...he raw-silk fibre consists of two cores of fibroin naturally ce...nted together and covered with sericin, besides small quantities ...ty and colouring matters. The fibroin filaments, which ...itute about 70% to 80% of the dry, raw mulberry silk, have ...haracteristic soft white appearance and lustre of pure silk ... the gum has been removed. Under the microscope the fila... appear smooth and rodlike, and, when examined by polar...ht, show the colours given by double-refracting substances. ...ross section is roughly triangular. The filaments have great ... strength, extend considerably before breaking and show ... elastic effects under limited stress. In fact, silk is said to ...lmost the equivalent tensile strength of iron wire, that of ...ire being 90,000 lb. per square inch, and that of silk 64,000. ...sticity is 20%, which means that the silk fibre can be ...ted one-fifth of its original length, and, when released, will ... its natural state. Silk owes much of its excellence as a ... fibre to its properties of tensile strength and elasticity.

...in is an amphoteric colloid protein. The formula ...N_5O_6 sometimes assigned to it should be taken to be no ...than an approximate expression of its elementary compo... X-ray examination indicates the possible presence in ... of a crystalline constituent, and there is some chemical ... that fibroin is chemically heterogeneous. Fibroin has low ...al and thermal conductivity. It is insoluble in alcohol, ... other organic liquids, and water but swells slightly therein. ... in cold concentrated solutions of mineral acids or of ...alkalies, and in an ammoniacal solution of copper oxide. ... these solutions it may be reprecipitated in a more or less ... form when the solution is neutralized. However, if the ... stands too long before being neutralized, more far-... changes will be found to have taken place. When heated, ...elts and burns, emitting a smell of burned feathers. This ... odour serves to distinguish it from the vegetable fibres. ... be differentiated from wool by its appearance under the ...pe and by certain chemical tests. The ultimate filaments ... silks are thicker and more ribbon-shaped than those of ...rry silk and exhibit longitudinal striations along which the ...nts tend to split into fibrillae under any mechanical or chemical action. The fibroin of wild silks possesses properties similar to those of mulberry fibroin but is more resistant to chemical action. Sericin, also a protein, is more active chemically than fibroin, from which it may be separated by the solvent action of hot water, acid, or alkaline solution. Its composition is similar to that of fibroin, but it contains more oxygen. Like fibroin, sericin may be chemically heterogeneous. The colouring matter of yellow silk is xanthophyll, a yellow pigment related to carotene.

Silk has great powers of absorption and takes up dyestuffs at a temperature much lower than does wool. This greater capacity of absorption, perhaps higher than that of any fibre, permits silk, when dyed, to take on the brilliant colours for which it is famed. Its absorptive powers are so great that the weight of silk can be increased as much as 30% in moisture content without any change in appearance. This characteristic made it necessary to establish an international standard of 11% as the permissible limit of regain in moisture over dry weight. (E. T. A.)

V. SILK MANUFACTURE

Cocoons are manipulated to produce continuous filaments of raw-silk yarn through a process, distinctive to the silk industry, called reeling. Silk waste, the short silk fibres that are a by-product of reeling or other manufacturing operations, is spun like cotton or wool into a yarn called spun silk. Filament silk is prepared for weaving by throwing (twisting) to the extent required. After weaving, the fabric is degummed (boiled-off) and dyed or printed, unless it has been woven from yarns dyed before weaving.

1. Reeling.—The purpose of reeling is to bring together the threads from two or more cocoons (depending on the final thickness or size of thread desired) to form them into one continuous, uniform and regular strand that constitutes commercial raw silk. The process, which is extremely delicate, has been mechanized and automated; the establishment in which it is done is called a filature.

In the reeling process, the cocoons are first sorted, and doupions, pierced cocoons and any other unreelable cocoons are put aside. Those to be reeled are boiled in water for 10 to 12 minutes to soften the natural gum (sericin), which holds the filaments together. The ends of the filaments are caught, unwound and kept together by coagulation of the natural gums, so as to form a single uniform, rounded thread of raw silk that is wound on reels.

When the silk is removed from the reels, it is rewound into skeins of the desired circumference, length and weight, which then are twisted and doubled over into one-half their length and packed into small bundles called books. Some silks like doupioni are packed in full-length skeins, flat in bundles. These books or bundles are packed to form a bale weighing about 133 lb.

The size or "count" of thread of raw silk in the industry is designated by the term denier, which is the relationship of length to weight based on a standard of 4,464,531 yd. of thread that would weigh exactly one pound and is the equivalent of one denier. The denier count rises as the size or weight of the yarn increases and the yards per pound decrease in relation to this standard; therefore, a thread reeled into a size with a range of 9 to 11 denier with a mean of 10 denier would contain 446,453 yd. to a pound. The sizes of silk thread used commercially are 9/11 denier, 13/15, 14/16, 18/20, 20/22, 24/26, 28/30, 30/32 and 40/44, covering a range from approximately 446,000 yd. to the pound as noted for 9/11 denier through approximately [illegible] yd. to the pound for 40/44 denier. A thread reeled from six cocoons would result in a size of 13/15 denier having 318,000 yd. of yarn to the pound which could easily stand strains imposed by the manufacturing process.

2. Spun Silk.—The silk used to manufacture spun silk includes all kinds of unwindable raw silk. Before the introduction of machinery for spinning spun silk, nearly all the waste from reeling and silk winding was destroyed as being useless. The exception was that which could be hand combed and spun by means of the spinning wheel, a method still practised in the farm homes in India and other far eastern countries.

The raw material for spun silk is obtained principally from the lustreless and uneven thread emitted by the silkworm when it commences to spin; from the extreme outside and inside layers of each cocoon; from pierced cocoons (those from which the moth has emerged); and from waste resulting from breaking threads during the reeling and other processes.

There were, up to the beginning of the 20th century, two distinct types of spun silk: schappe silk, produced on the European continent (but never in Great Britain) and containing as much as 20% of the original gum; and spun silk, which was entirely free of the gum. Now that the use of fermentation processes has entirely died out, the term schappe silk, as used in France and Germany, is the exact equivalent of the term spun silk, neither containing more nor less gum than the other.

Spun silk, like continuous filament raw silk, is pure silk containing no extraneous matter. Spun silk yarns are used for shirtings, blouse and dress material, silk "tweed," and sewing and embroidery threads. The combings left after the processing of the spun silk fibres are called noils and are blended with other fibres or spun into coarse yarns. One cloth woven from these yarns is used for artillery powder bags; because silk is a protein fibre, it burns to a hard, beadlike ash that is easily removed from the guns, leaving no powdery ash to foul the mechanisms. Noil silk fabrics are also used for furnishing fabrics.

3. Throwing.—The silk thread, one continuous filament that has been reeled from the cocoon, does not need the combing and preparation to make the fibres parallel that is necessary in order to spin the shorter fibres of cotton and wool. The only processing needed before weaving is that of producing different effects in the finished yarn by twisting the silk thread varying degrees either to the right or left. This process is called throwing (from the Anglo-Saxon *thrawan*, "to twist"); the plant in which it is done is a throwing mill. Each twist is called a turn, and the greater the number of turns per inch, the greater the foreshortening and contraction of the thread. Many fabrics are woven with yarns containing little or no twist.

The first step in readying silk yarn for throwing is soaking the skeins in an oil-based emulsion for about eight hours at a beginning temperature of 100° F (37.78° C), to soften the natural gum, thus making the thread more pliable and workable. This done, the skeins are removed, dried and, unwound from the skein in the winding process; that is, the silk is unreeled from the skeins onto spools or bobbins. The bobbins are wound with a rapid reciprocating motion, so as to lay the thread in a crossing diamond-like pattern on the bobbin.

There are four fundamental steps in throwing: winding; doubling or plying two or more threads together with low twist, usually less than five turns per inch; twisting, increasing the turns per inch beyond the natural or that put in doubling; and, finally, skeining or coning or tubing. The twisting machinery used to perform this operation has been perfected and refined so that it is possible to twist a thread composed of six cocoon threads and which has approximately 318,000 yd. to the pound, designated as 13/15 denier, at the rate of 18,000 rpm.

The most used types of thrown yarns are tram, organzine, and crepe. Tram is two or more threads twisted together with a few turns per inch and is used for the filling or crosswise thread. Organzine is the result of doubling two or more threads, previously low twisted as single threads, and twisting them in the opposite direction, two turns per inch less or more than the twist in the single. This type of yarn is used chiefly for the warp or lengthwise thread, after being degummed and dyed in the skein. Crepe is a high-twist yarn with a range of 40 to 80 turns per inch.

Crepe twisted yarns are used in both warp and filling for such fabrics as chiffons, or as filling with raw silk, no-twist warps for opaque fabrics such as crepes. Crepe twisted yarns are twisted either to the right or left. Threads of each twist are used alternately in the fabric in both filling or warp. The crepe effect is achieved by the contrary pull of the two tightly twisted threads released from their set condition by the degumming process that causes their contraction and, therefore, shrinkage of the fabric in which they are woven. (All crepe twisted yarns are degummed after weaving.) For easier identification, the threads are tinted different colours for the left and right twists. The colours used are fugitive tints that completely disappear in the degumming of the woven fabric before dyeing.

Silk thread for sewing is manufactured as part of the throwing operation but is considered a distinct and separate branch, because some additional equipment must be used, and great care must be taken in every operation, especially spooling, to make a smooth, knotless thread.

4. Making the Fabric.—*Woven Silk.*—The first operation after throwing, or if no throwing is required, the first step after soaking the skeins, is the winding of the silk thread onto bobbins, cones, or tubes so that it may be handled in the next operation, warping.

Warping is the operation that prepares the yarns that run lengthwise in a fabric for the weaving process. It must be done most carefully, as the condition of the warp determines the efficiency with which the manufacturer can run his looms. The number of threads per inch multiplied by the width of the cloth desired determines the total number of threads that have to be laid one next to the other and controlled in this parallel position throughout the weaving operation. During warping each thread must be kept separate and not overlay an adjacent thread, as this would cause streaks in the woven fabric. The warps can be made in any desired length to more than 2,000 yd. for piece-dyed fabrics. The number of threads per inch will range from 90 to more than 400, depending on the construction or weave of the fabric, in widths ranging from very narrow for ribbons to 50 in. for broad silks.

The entering operation is the process of threading each warp end through the eye of a heddle. In modern mills, heddles are made of flat stainless steel, 11 to 13 in. long and $\frac{1}{16}$ to $\frac{1}{8}$ in. wide, with the eye in the centre of their length. The heddles are distributed on bands of steel held top and bottom in frames called shafts, at least wide enough to accommodate the width of the warp. The type of weave desired determines the minimum number of shafts to be used, and the number of heddles must correspond to the number of weaving ends in the warp. Since each weaving end must have a heddle to control it, a warp of 15,000 ends, if running each end single, must have 15,000 heddles, and, if this is to be a satin made on five shafts, the 15,000 heddles would be distributed equally on the five shafts, with 3,000 heddles per shaft. After the warp ends are entered into the heddles, they must be controlled further to achieve the desired density of the fabric. This is accomplished by the use of a reed. The reed is made of flat steel wire, $\frac{1}{16}$ to $\frac{1}{8}$ in. deep and 4 to $4\frac{1}{2}$ in. high, soldered into strips of steel top and bottom to hold them in place equidistant. Each space between two adjacent wires is referred to as a dent. If the construction calls for 300 ends per inch in the warp and they are distributed on five shafts, five ends would be threaded in a dent of a 60 count reed (that is a reed with 60 dents per inch), thus fixing the thread count per inch in the fabric. The shafts, heddles and reed when grouped together are termed harness because they control the threads in the warp during the weaving operation.

The entering operation just described is employed when a new type or design of fabric is to be produced. However, when a warp has been woven out and the same or related fabric is to be continued, the threads in a new warp are twisted or tied onto the previous warp in an operation called twisting. Originally, this operation was accomplished by hand, each end of the previous warp being twisted onto an end of the new warp. It required great skill and was very time consuming; an excellent operator could average only about 700 to 1,000 twists per hour. The operation now is accomplished by a machine called a twister or knotter, which will twist or tie individual knots at a rate of 17,000 per hour.

The quilling operation places the filling yarn on a suitable bobbin or quill. Modern quilling machines are high speed, sensitive, and almost fully automatic. The quilling spindle is started and an empty quill is automatically positioned, rotating at more than 12,000 rpm. When the quill is full, the spindle automatically stops,

ejects the full quill, returns to the starting position and positions a new empty quill. The full quills are automatically packaged, ready to be transported to the weaving operation for insertion in the shuttle in the loom. All of these operations are accomplished with very little tension, so as not to stretch or damage the fine silk threads.

Silk looms operate on the same principles as other textile looms. There are plain looms running one shuttle and box looms with two to seven shuttles. The up and down motion of the harness shafts are controlled in the plain and box looms by means of rotating cams that raise and lower each shaft in a restricted, fixed pattern, or by a device called a dobby head, which has greater flexibility in the control of the up and down movement of the harness shafts. In general, the plain looms weave the flat fabrics such as taffetas and linings; the box looms make the crepes and novelties. Jacquard weaving, using dyed yarns, produces elaborate, multi-coloured patterns in such fabrics as tapestry, brocade, brocatelle, damask, and figured dress goods (*see* also WEAVING).

All parts of the silk loom are balanced carefully to minimize vibration and give smoothness to the weaving and control of the fine silk threads. Precision that would be relatively unimportant in weaving other fibres has to be closely maintained in silk weaving.

Silk woven from the thread in its natural state is referred to as raw goods or greige goods. Because of the sericin in silk, the fabric has a lustreless, stiff, and harsh appearance that would be unacceptable for most practical uses. One exception is bolting cloth, a unique type of silk fabric used primarily by millers in sifting fine flour and also as a filter for women's powder compacts. Another exception is stencil cloth, used for screen printing and in the preparation of printed circuits in the electronics industry. Bolting cloth originally was made on hand looms in the cellars of mountain homes in the Sankt Gallen and Appenzell sections of Switzerland, where the humid air helped to overcome the static electricity produced during the weaving operation. Although some of the finer grades are still hand loomed, after 1940 most bolting cloth was manufactured on power looms.

Knitted Silk.—There are two types of warp knitted silks, tricot and milanese. Tricot is characterized by fine vertical wales on the face and crosswise ribs on the back. Its construction permits the making of designs such as stripes. Milanese is a fabric in which the threads are interknitted at every course in which the threads run diagonally. A fine rib on the face and a faint diagonal or diamond effect on the back are characteristic. These fabrics are run resistant and are used for women's gloves, lingerie, etc., and, after printing, for women's lightweight, crease-resisting dresses. In women's hosiery, silk has been displaced almost entirely by nylon.

5. Degumming, Dyeing, Printing, and Finishing.—Silk fabrics woven as raw goods must be degummed or scoured before they are dyed and finished. These fabrics are termed piece-dyed to differentiate them from the yarn-dyed, which are woven with yarns dyed in the thread before weaving. The degumming process is done as part of the dyeing and finishing operation.

Silk is usually boiled-off in a solution of 30% of its own weight of neutral soap dissolved in sufficient water to give a 1% solution. If degummed in skeins, the raw silk is immersed in this solution which is maintained at a temperature just below the boiling point, the skeins being turned to expose all parts equally to the solvent action of the hot solution. If woven in the gum, the fabrics are degummed in the same type of solution. After one hour the silk usually is given a second bath with half the quantity of soap. The moisture is removed in a centrifuge.

Care must be exercised not to overdo the scouring. Perfect scouring reduces the weight approximately 20% to 27% and renders the silk a brilliant pearly white with a delicate, soft, flossy texture, since the filaments that were agglutinated are separated from each other and reflect light in a beautiful, deep lustre. Silks to be finished white are bleached at this point, usually with hydrogen peroxide.

Silk has an excellent affinity for dyestuffs and finishing materials, and the handling of the fibre either as yarn or fabric has been the pride of dyers throughout the ages. The finishing of silk fabrics to give them smoothness, lustre, stiffness, or rustle also has been the subject of constant experimentation by master dyers. Although formulas have been published for dyeing and finishing preparations, every dyer treasures his own method of handling silk, and pride of individual artisanship continues.

Silk, for centuries, has been treated with salts of tin and lead to increase its weight and give pure silk more body and close texture. This weighting of pure silk also reduced its cost. Practically all U.S. silk fabrics are unweighted pure silk, pure dyed, but the practice of weighting is carried on to a small extent in Europe and Japan within strict limits laid down by the International Silk Association. Governmental bodies in the United States and some other countries require that, where any weighting is present in silk fabrics, the percentage of weighting to the weight of the silk fibre content in the finished fabric must be stated on the label.

Until shortly after World War I, almost all silk was printed by the hand-block process. This is a highly skilled operation and the craft has practically died out, though it is still carried on by one or two firms in England for the luxury men's wear trade.

Most printing of silk fabrics is done by the same general methods followed in other textiles. Roller printing is done by engraved copper rollers, with a separate roller for each colour in the design. Designs run from one to ten colours and can be printed by application or discharge. In the first, the design is applied to the cloth; in the second, the design and colours are chemically developed after printing.

Screen printing, a more expensive process, gives a hand-printed effect and is preferred especially for silk; more brilliant colour effects are achieved, and there is wider latitude in the type of design. A separate screen is used for each colour, and the number of colours that can be employed is almost unlimited. Screen printing is slower than roller printing and frequently is used where a smaller quantity of a given design is wanted (*see* also TEXTILE PRINTING).

Finishing gives the dyed and printed silk smoothness, deep lustre, and suppleness, and provides protection against certain characteristics of wear such as creasing, perspiration stains, and water-spotting. In the finishing operation, minute quantities of starches, sulfonated oils, glycerin, resin, and various organic chemical compounds are used individually or in combination as softeners and penetrants to achieve the desired effects of drape and feel. Fabrics must be finished wrinkle free and to a uniform width. The natural strength and elasticity of silk provides ample opportunity for manipulation in processing.

6. Silk Cloth.—Some of the fabrics long identified in the public mind with silk are taffeta, satin, chiffon, crepe, crepe de Chine, shantung, surah, organza, and faille.

Taffeta is made either piece-dyed or yarn-dyed and is considered a medium-weight fabric. It is opaque and has a semibright lustre, with a smooth, flat, firm, crisp texture. Garments made of this fabric usually have a semibouffant silhouette.

Satin is made either piece-dyed or yarn-dyed and may be either a light- or heavyweight fabric. It is opaque, with a bright lustre on the face side and duller on the back, with a slick, smooth touch and soft or firm texture, as desired.

Chiffon is a piece-dyed, lightweight, sheer fabric. It is translucent, with a semidull lustre and a sheer, gauzelike, grainy texture, with a soft drape. It is printed also.

Crepe de Chine is a lightweight, piece-dyed fabric. It is opaque, with a semidull lustre and a smooth, slightly crinkly texture, with a soft drape. It is printed also.

Shantung most frequently is piece-dyed and can be either a very light- or heavyweight fabric. It is opaque, with a semidull lustre, being woven with doupioni silk filling, or, more usually, with tussah silk. It has an uneven, soft texture, with random thick and thin portions in the yarn called slubs. It is an extremely popular fabric, much imitated in man-made fibres.

Surah is made either piece-dyed or yarn-dyed. In the U.S. this term is synonomous with tie silk. In Britain tie silk indicates Jacquard woven tie cloths; printed surah is known as foulard in the men's trade. It can be either a light- or heavyweight fabric.

It is opaque, with a semibright lustre, firm drape, and a smooth texture, with a clearly defined diagonal weave. This is the base fabric for much printed silk.

Organza is a lightweight, sheer fabric, which is yarn-dyed in the gum. It is almost transparent, with a dull lustre and a meshlike, flat, firm, starchy, stiff texture, and in printed form is used a great deal by the haute couture.

Faille is usually yarn-dyed and a medium- to heavyweight fabric. It is opaque, with a semibright lustre and has a thick, smooth, firm texture, with a clearly defined rib running crosswise from selvage to selvage.

These and other silk fabrics long have been valued for their aesthetic appeal, but silk is not merely a fashion fabric; it also has many utilitarian uses. Until 1942 it was the major fabric used for parachutes, a usage preserved in the expression "hit the silk," meaning to parachute from an airplane. Silk remains the standard of excellence by which nonsilk fabrics and, indeed, many other items are judged. "Smooth as silk" and "the silken touch" have not without reason become everyday descriptive phrases.

See also references under "Silk" in the Index. (M. H. Ru.)

Bibliography.—L. Duran, *Raw Silk* (1921); P. Carboni, *Silk*, Eng. trans. by K. Walter (1952); C. H. C. Cansdale, *Cocoon Silk* (1937); R. C. Rawlley, *Economics of the Silk Industry* (1919); W. P. Seem, *Raw Silk and Throwing* (1929); S. Matsui, *The History of the Silk Industry in the United States* (1930); C. Singer *et al.* (eds.), *A History of Technology*, vol. iv (1958); E. R. Flemming, *Encyclopedia of Textiles*, trans. and rev. by R. Jaques (1958); Z. Bendure and G. Pfeiffer, *America's Fabrics* (1947); G. G. Denny, *Fabrics*, 4th rev. ed. (1936); G. E. Linton, *Modern Textile Dictionary* (1961); I. Wingate, *Textile Fabrics and Their Selection* (1952); Silk and Rayon Users' Association, *The Silk Book* (1951); P. W. Gaddum, *Silk* (1964).

SILK SCREEN PRINTING. The silk screen printing process is the modern commercial form of stencil printing. Since the screen may be made from wire gauze, cotton fabrics and other materials as well as from silk bolting cloth, the process is generally called screen printing by commercial printers. Artists use specialized techniques and call their method serigraphy to distinguish it from the commercial variety.

Principles and Uses.—In screen printing the design is painted on, or affixed to, a sheet of silk bolting cloth or other screen material that is stretched tightly over a wooden or metal frame. The frame, which is hinged to a base or table, is dropped to make contact with the paper or other material to be printed. Ink that resembles paint is poured on one end of the screen and is forced through the open mesh by the scraping action of a squeegee, a thin but firm strip of rubber set in a wooden form that serves as a handle. The open areas in the screen allow the ink to pass through, but the solid areas are impermeable. The design thus is transmitted in terms of open and blocked areas of the screen. The screen is then swung back, the paper is removed to dry and another sheet is positioned for another impression. Each colour requires a separate screen but, since one colour can be printed in register over another, any number of colours can be used in a single design. A reducing varnish is mixed with the colours to permit easy passage through the mesh, and a transparent base (a Vaselinelike substance) is mixed in to achieve transparent effects.

Compared with other printing processes, screen stencil printing is slow, since it usually is employed as a hand method, but it has the advantage of requiring relatively cheap equipment and comparatively little preliminary work. It is economical especially for runs of up to 3,500 impressions; therefore, it is used widely for show cards, posters, signs, charts and decalcomanias. It permits printing on almost any surface, including paper, cloth, glass, wood, plastic, rubber, metal, linoleum and oilcloth. Areas of relatively great size can be stenciled; *e.g.*, frames 50 × 60 in. are used frequently in textile printing. (*See* Textile Printing: *Modern Industrial Processes: Stencil and Screen Printing*.)

As the objects to be decorated need not be limited in thickness, nor even be flat, the screen method is used widely for printing on such objects as toys, novelties, bottles and machinery. The process is not suitable, however, for work where fine details and delicate gradations of tone are needed, as in the reproduction of photographs and in colour printing. While such work can be done, the results are comparatively coarse.

Mechanical methods have been applied to screen printing; automatic presses with speeds up to 1,800 impressions per hour are in general use.

Developments.—The stencil method is very old. By A.D. 500 it was used commonly in China and Japan. A Japanese screen with stenciled decorations, dated about A.D. 700, is in the collection of the Fogg museum at Harvard university. A highly developed method was used in Japan from the late 17th century to recent times. The designs, often very intricate, were cut in separate pieces of parchment. These were tied or glued in place by numerous crossing strands of silk or of human hair.

The first consistent use of stencils in Europe dates from the late 15th century. Playing cards and religious subjects printed from wood blocks were coloured by stencil. In the late 17th century, wallpaper was produced in France by the same method, notably by Jean Papillon. Stenciling was widely used in the United States after 1800 for decorating walls, floors, furniture, clocks and trays. The painting of patterns on velvet with stencils became a genteel accomplishment for young ladies.

The use of silk screens is a 20th-century development. Samuel Simon of Manchester, Eng., received a patent for this process in 1907, after which the method was adopted slowly until, in the late 1920s, it came into limited use as a commercial printing method. After that time it grew rapidly into a large industry.

Serigraphy is used widely by artists for creating original designs in colour. Although a few prints were produced in the early 1930s, the process first attracted general attention in 1938 when Anthony Velonis organized a group of artists on the New York city Work Projects administration (WPA) art project. The new medium was named serigraphy by the art historian Carl Zigrosser.

Preparation of Screens.—Silk screen stencils can be prepared by a variety of methods. Photography is employed increasingly, but a number of hand techniques are still in use.

Aside from such primitive methods as painting with shellac and stopping out with pasted paper, the oldest and simplest is the cutting of a film stencil with a knife. In this process a thin prepared film with a translucent backing (*e.g.*, paper acetate, vinyl or Mylar) is placed over the design to be reproduced. The cutter, using a special knife, follows the outlines of the design, cutting through the film but not through the backing. The cut film is removed and the remaining film is laid face-up on a flat support. The underside of the screen is laid on the film, which is made to adhere by applying a solvent. The screen is then turned and the backing is stripped away.

In serigraphy the design is drawn on the silk with lithographic tusche, a viscid, greasy ink, and with lithographic crayon for delicate shadings. The screen is coated with a water-soluble glue solution and, after the screen dries, the tusche image is washed out with an organic solvent such as kerosene. This lays the screen bare in the design areas. Since the original colour sketch is visible through the screen, the artist can prepare a separate screen for each colour and superimpose one over another.

Many techniques are used in photographic silk screen work. The most versatile involves the use of a special carbon tissue composed of a paper base coated with a pigmented (usually red) gelatin solution. This tissue is immersed in a solution of light-sensitive ammonium or potassium dichromate (bichromate), after which a squeegee is used to press the tissue onto a thin sheet of clear plastic. A photographic positive is placed over the plastic and exposed to an arc light. The gelatin thus is hardened under the transparent areas of the positive, while the areas under the design remain soft and are washed away with warm water. The paper backing of the carbon tissue then is stripped off, the image is affixed to a screen and the temporary plastic support is removed. The design thus is open to the passage of ink, while the hardened gelatin areas act as stop outs.

Bibliography.—B. Zahn, *Silk Screen Methods of Reproduction*, a handbook showing early commercial methods (1929), *Screen Process Methods of Reproduction*, a textbook of modern commercial methods (1956); J. I. Biegeleisen and E. J. Busenbark, *The Silk Screen Printing Process*, a manual stressing commercial methods (1938); H. Sternberg *Silk Screen Color Printing*, a simplified handbook for artists (1942); J. B. Stephenson, *From Old Stencils to Silk Screening*, a history of

stenciling and a handbook for craftwork on tin, furniture, etc. (1953); J. I. Biegeleisen and M. A. Cohn, *Silk Screen Stenciling as a Fine Art*, mainly devoted to art methods (1942); H. Shokler, *Artists Manual for Silk Screen Print Making*, describing art and commercial methods (1960). (J. Ka.)

SILL, ANNA PECK (1816–1889), U.S. educator, a pioneer in women's education in the United States who sought to bring the education of girls to the level of that for men, was born Aug. 9, 1816, at Burlington, N.Y. After attending the district school, she began to teach at a school near Albion, N.Y., and in 1837 entered Miss Phipps' Union seminary, where she shortly became a member of the faculty. There she remained until 1843, when she established a seminary for girls at Warsaw, N.Y.

After three years Miss Sill abandoned this project and went to teach at Cary Collegiate institute, Oakfield, N.Y. She was invited to Rockford, Ill., in 1849 to open a girls' seminary and was made principal of the school (Rockford Female seminary) in 1852. Thereafter she spent much of her time raising funds for the young institution. It expanded gradually until in 1882 it began to confer collegiate degrees. The name was changed to Rockford college in 1892.

Miss Sill had retired in 1884, but she remained with the institution as principal emerita and taught occasional classes. She died in Rockford on June 18, 1889.

SILL. In geology, a tabular body of igneous rock formed by intrusion of magma (molten rock) along a structural break (bedding or foliation) in the earth. The tabular form and parallelism with surrounding structure are the diagnostic features. To many geologists there is another essential; injection must have occurred along relatively flat-lying breaks. The magma, therefore, displaced the adjacent rock vertically by lifting the roof. This mechanism is quite different from that involved in the formation of steep intrusive sheets (most dikes). In the latter, magma entered passively as the walls of the steep fractures were spread apart by horizontal movements.

It is believed that, as the injected magma comes into contact with the cooler adjacent rock, it becomes chilled and rapidly solidified to form a thin, dense selvage or layer along the sill roof and floor. Somewhat slower cooling of the more internal portions of the sill results in coarser textures. Margins of many sills show a streaked and banded appearance, believed to have been produced by liquid flow along the smooth sill roof and floor. Large crystals commonly show parallel arrangement of their long dimensions near sill boundaries. These are believed to have formed early and to have been turned into the lines and layers of flow parallel to sill contacts.

Some sills, after congealation, have been split open and injected by fresh magma. Such bodies are called multiple sills if the successive injections were of one kind of magma, and composite sills if different kinds of magma were involved.

Sills may range up to several thousand feet in thickness and may extend across country for hundreds of miles. They are commonly associated with dikes and larger intrusive bodies (*see* Batholith; Laccolith; Geology). Most described sills are composed of diabase or dolerite and closely related rocks. Granitic, rhyolitic, dioritic, and andesitic sills are locally abundant; and the rocks comprising them commonly belong to the class known as porphyries (*see* Porphyry).

Some thick sills, such as the Palisades sill in New Jersey, furnish striking evidence that an essentially homogenous magma became differentiated during consolidation and yielded a diverse, layered series of rocks. In general, such sills show a dense chill zone at the top and bottom; and heavy rocks, near the base, grade upward to somewhat lighter rocks near the top. It seems clear that early developed heavy crystals (olivine and pyroxene) settled in the slowly congealing magma, whereas lighter minerals (feldspar) may have remained more or less stationary or possibly were buoyed upward.

The rocks immediately in contact with sills commonly appear to have been profoundly changed. This phenomenon, known as contact metamorphism, is believed to have been brought about largely by the heat of the intrusive magma. In the case of large sills, the metamorphic zone may extend a few hundred feet from the intrusive margin. Sedimentary rocks may show a baked or indurated and bleached appearance in contact with a sill roof or floor. Shale, for example, has commonly been reconstituted to a tough, brittle rock called hornfels, in which many new minerals (cordierite, feldspar, garnet, pyroxene, etc.) have formed. Blocks and fragments of the adjacent rock, accidentally caught up in the intrusive mass, may also show these metamorphic effects (*see* also Metamorphism).

See M. P. Billings, *Structural Geology* (1954); R. A. Daly, *Igneous Rocks and the Depths of the Earth* (1933). (C. A. Cn.)

SILLANPÄÄ, FRANS EEMIL (1888–1964), the first Finnish writer to win the Nobel Prize for Literature (1939). The son of a peasant farmer, he was born at Hämeenkyrö, in southwest Finland, on Sept. 16, 1888. At Helsinki University, where he studied natural science, he was influenced by the monistic philosophy of Wilhelm Ostwald (*q.v.*). In 1913 he broke off his studies, returned to the country, married, and began to write. His first short stories were published in journals in 1915. From 1924 to 1927 he worked for a publishing company in Porvoo. A personal crisis was followed by a new creative period in the early 1930s, when several of his best works were written. He died at Helsinki on June 3, 1964.

Sillanpää's first novel, *Elämä ja aurinko* (1916; "Life and Sun"), the story of a young man who returns home in midsummer and falls in love, is characteristic both in title and subject. In all his work, life itself is the dominant force. People are seen as essentially part of nature, which is not merely a lyrical background. Instinct, through which life's hidden purpose is revealed, rules human actions.

Shocked by the Finnish civil war of 1918, Sillanpää wrote his most substantial novel, *Hurskas kurjuus* (1919; *Meek Heritage*, 1938), describing how a humble cottager becomes involved with the Red guards without clearly realizing the ideological implications. The novelette *Hiltu ja Ragnar* (1923) is the tragic love story of a city boy (Ragnar) and a country servant-girl (Hiltu). After several collections of short stories in the late 1920s, Sillanpää published his best-known, though not his most perfect work, *Nuorena nukkunut* (1931; *Fallen Asleep While Young* [U.S. title *The Maid Silja*] 1933). This tells poetically the story of an old peasant family. In *Miehen tie* (1932), describing a young farmer's growth to maturity, realistic and lyric elements are blended. *Ihmiset suviyössä* (1934; "People in a Summer's Night"), is stylistically his most finished and poetic novel. His reminiscences, *Poika eli elämäänsä* (1953) and *Päivä korkeimmillaan* (1956), throw new light on his personality as a writer.

See studies of Sillanpää by T. Vaaskivi (1937), R. Koskimies (1948), and A. Laurila (1958). (K. L. K. L.)

SILLIMAN, BENJAMIN (1779–1864), U.S. chemist and geologist, best known as the founder (1818) and, until 1838, sole editor of the *American Journal of Science and Arts*, popularly called *Silliman's Journal*. He was born at North Stratford (now Trumbull), Conn., on Aug. 8, 1779. The son of a Revolutionary War general, he graduated from Yale College in 1796, becoming a tutor in 1799 and professor of chemistry and natural history in 1802. He taught chemistry, mineralogy, geology, and pharmacy, and aided greatly in expanding Yale's educational resources, especially the mineralogical collections, the Sheffield Scientific School, the Trumbull Gallery, and the Medical School. His clear, interesting expositions and skilful illustrative experiments made him a popular scientific lecturer.

After visiting Europe in 1805, Silliman wrote *A Journal of Travels in England, Holland and Scotland* (1810), and a similar publication resulted from another European tour in 1851. His other books included *Elements of Chemistry* (1830), and editions of William Henry's *Chemistry* (1808) and Robert Bakewell's *Geology* (1827). He retired in 1853 as professor emeritus, and died at New Haven on Nov. 24, 1864. The mineral sillimanite was named for him.

Benjamin Silliman (1816–1885), his son, achieved distinction in applied chemistry and, like his father, as a popular lecturer. Born at New Haven on Dec. 4, 1816, he graduated from Yale in 1837, and assisted his father in setting up a chemical laboratory

(1842) out of which later grew the Sheffield Scientific School. He taught medical chemistry and toxicology at the University of Louisville, Ky., from 1849 to 1854. His lectures on agricultural chemistry at New Orleans in 1845–46 are believed to have been the first such course in the United States. In 1854 he returned to Yale to succeed his father in the chair of chemistry. He relinquished most duties of this professorship in 1870, but continued lecturing until his death, at New Haven, on Jan. 14, 1885.

The younger Silliman prepared numerous reports on mines and on technical manufacturing processes, and conducted fundamental research on meteorites, photography, gases, etc. His now famous "Silliman report," (April 1855) gave impetus to plans for drilling the first producing oil well, near Titusville, Pa., in 1859. The document was a 20-page report of Silliman's analysis of a sticky substance which had been found in Pennsylvania and sent to him for examination. The substance was petroleum, and he indicated its exceptional suitability for lubricating, lighting, and other purposes and emphasized its potentially great economic possibilities. Few researchers have so profoundly influenced industrial development and human lives. (*See* CHEMICAL INDUSTRY: *Petrochemical Industry.*)

In 1846 Silliman published *First Principles of Chemistry*, and in 1858 *First Principles of Physics or Natural Philosophy*, both widely used. In 1853 he edited *The World of Science, Art and Industry*, and in 1854 *The Progress of Science and Mechanism*. For the Centennial of Chemistry in 1874 he presented a notable address, "American Contributions to Chemistry." He assisted his father with the *American Journal of Science* from 1838 to 1845, and was its co-editor in chief until his death. (F. M. FL.)

SILLIMANITE, an aluminum silicate mineral that, like the closely related aluminum silicates andalusite and kyanite (*q.v.*), may be used to form mullite, an important refractory material (*see* also MULLITE). It crystallizes in the orthorhombic system and has the composition $Al_2O_3.SiO_2$. It occurs in metamorphic rocks as long slender crystals and as rosettes. The hardness range is from 6 to 7 and specific gravity from 3.23 to 3.24. Its colour ranges from brown to gray to olive-green. In the laboratory it may be synthesized over a limited range of temperatures and pressures and is presumably so formed in nature. At pressures under approximately 3,000 atm., in the temperature region of 600°–900° C., the stable phase of essentially the same composition is the mineral mullite. Sillimanite occurs at pressures from 3,000 to 12,000 atm. Above these pressures the mineral kyanite appears. At temperatures below approximately 600° C. the polymorph of sillimanite, andalusite, appears. It is an important geological mineral and serves as an indicator both of temperature and pressure of environment in which rocks containing it were formed. It is present in metamorphosed kaolin-rich sediments. It was named after Benjamin Silliman (*q.v.*). (G. C. KY.)

SILO: *see* ENSILAGE.

SILONE, IGNAZIO (pseudonym of SECONDO TRANQUILLI) (1900–), Italian author whose novels reflect his belief in the need for social reform, was born at Pescina dei Marsi, Abruzzi, on May 1, 1900. One of the founders of the Italian Communist party, he left the party in 1930 while in exile in Switzerland. Returning to Italy in 1944, he continued to influence Italian socialism, although without holding any important political office.

Silone's first novel, *Fontamara* (1930; Eng. trans., 1934), is a poignant description of poverty and Fascist oppression in an Abruzzi village. Its peasant hero, who struggles for freedom but dies in prison, believes that political and social regeneration will be accomplished by a Communist revolution. The hero of *Pane e vino* (1937; Eng. trans., *Bread and Wine*, 1937) and *Il Seme sotto la Neve* (1940; Eng. trans., *The Seed Beneath the Snow*, 1943) has turned from Communism to a Christian belief in universal brotherhood.

In *Una Manciata di More* (1952; Eng. trans., *A Handful of Blackberries*, 1954) Silone denounces Communism though still urging social reform and the development of the south of Italy, which is also the subject of *Il segreto di Luca* (1956).

See bibliography in *Fiera letteraria* (Aug. 14, 1955); an essay by Edmund Wilson in the *New Yorker* (Sept. 8, 1945). (F. DI.)

SILURIAN SYSTEM. This article is divided into several broad sections and subsections dealing with the Silurian system and period of geologic time. In addition to the cross references to related material given under the various headings of this article, for general background information *see* GEOLOGY: *Historical Geology*; PALEOBOTANY; PALEONTOLOGY.

For additional information on the forms of life referred to *see* separate articles on these forms, as ECHINODERMATA; FISH; etc. *See* also references under "Silurian System" in the Index.

Following are the main divisions of this article:

I. The Silurian System and Period
  1. Physical Events and Features
  2. Climate and General Ecology
  3. Facies, Biogeography and Migration Routes
II. Silurian Life
  1. Characteristics and Evolution
  2. Invertebrates and Associated Fossils
  3. Fishes, Land Plants and Associated Fossils
III. Regional Summary
  1. Great Britain
  2. European Continent
  3. U.S.S.R.
  4. Asia
  5. Australia
  6. South America
  7. North Africa
  8. North America

I. THE SILURIAN SYSTEM AND PERIOD

Rocks of the Silurian system are those that were formed during the Silurian period, which is roughly the interval from 405,000,000 to 435,000,000 years ago, a middle division of the Paleozoic era (*q.v.*), as indicated on the accompanying geologic time chart. Many European geologists outside of England and the U.S.S.R. also include under the term Silurian the older rocks and the earlier time more generally called Ordovician. To them the Silurian as

Geologic Time Chart

System and Period	Series and Epoch	Distinctive Records of Life	Began (Millions of Years Ago)
	CENOZOIC ERA		
Quaternary	Recent (last 11,000 years)		
	Pleistocene	Early man	2+
Tertiary	Pliocene	Large carnivores	10
	Miocene	Whales, apes, grazing forms	27
	Oligocene	Large browsing mammals	38
	Eocene	Rise of flowering plants	55
	Paleocene	First placental mammals	65–70
	MESOZOIC ERA		
Cretaceous		Extinction of dinosaurs	130
Jurassic		Dinosaurs' zenith, primitive birds, first small mammals	180
Triassic		Appearance of dinosaurs	225
	PALEOZOIC ERA		
Permian		Reptiles developed, conifers abundant	260
Carboniferous			
Upper (Pennsylvanian)		First reptiles, coal forests	300
Lower (Mississippian)		Sharks abundant	340
Devonian		Amphibians appeared, fishes abundant	405
Silurian		Earliest land plants and animals	435
Ordovician		First primitive fishes	480
Cambrian		Marine invertebrates	550–570
	PRECAMBRIAN TIME		
		Few fossils	more than 3,490

here defined is Upper (or Late) Silurian or Gotlandian.

Rocks of the Silurian system have yielded about 10% of U.S. iron, 20% of its salt and a little oil and gas. They also serve, especially where mainly limestone and dolomite, as host for many metallic mineral deposits, including important gold and tin ores in eastern Australia. And they are exploited industrially for cement, agricultural lime and building stone.

Silurian volcanoes dumped thousands of cubic miles of debris into linear subsiding zones and basins in eastern Australia and other regions, announcing related mountain building of this age. Contemporaneous deposits provide the earliest reasonably varied occurrences of fishes and land plants. Although both originated tens of millions of years before, neither is widely recorded until

Devonian time. An aberrant vertebrate relative, *Monograptus*, the last surviving bladelike graptolite (*q.v.*), burst into a flowering of short-lived species whose flexible external skeletons are used to divide the Silurian marine deposits of the world into time-equivalent zones. Archaic reef-building corals and the brachiopods advanced less speedily but over broader fronts. The evidence of all these creatures, and the sediments, is interpreted to signify mild climates; not markedly zoned latitudinally, but with local, long-lasting, Late Silurian arid regions.

The responses of Silurian rocks to local conditions of erosion and weathering account for many of the world's distinctive landscapes. Undercutting of weak sedimentary layers beneath massive Middle Silurian dolomites at the outlet of Lake Erie created and maintains Niagara falls. The picturesque glade country of Tennessee consists of isolated grassy areas on infertile shales surrounded by wooded areas on more calcareous shales or limestones of Middle Silurian Age. Beautiful Wenlock Edge in Shropshire is an east-dipping Middle Silurian limestone ridge, and much of Wales and the Lake district of England is underlain by Silurian rocks. The Swedish island of Gotland is the type site of the Gotlandian; picturesque Visby, its capital and ancient Hanseatic harbour, is built around a Silurian reef limestone.

Rocks are identified as Silurian (or a specific part of it) from comparison of their contained fossils with those of the type marine sequence in south Wales and Shropshire, or on relationship to or correlation with other rocks so dated. By definition, therefore, the Silurian includes all deposits formed during the same time as those in the type sequence, regardless of origin, location, or local relationships. The principal Silurian subdivisions for selected regions are outlined in the accompanying correlation table, emphasizing marine sequences as the basic local standards.

The Silurian is an enduring monument to British persistence. The Roman emperor Tiberius Claudius vowed (*c.* A.D. 43) to exterminate even the name of the warlike Silures of areas now in the Welsh borderland for their stubborn resistance under Caractacus (Caradoc). Beaten and dispersed by Julius Frontinus for Vespasian (A.D. 78), they were, in fact, almost forgotten until 1835, when Scotsman Sir Roderick Impey Murchison (*q.v.*), an old soldier himself, delighted in naming his new rock system for them. In his epochal *The Silurian System* (1839), Murchison reflects how:

> In ancient days
> The Roman legions and great Cesar found
> Our fathers no mean foes, and Cressy plains
> And Agincourt, deep tinged with blood, confess
> What the Silures' vigour withstood
> Could do in rigid fight.

He and his British successors have done a job worthy of their ancestors in unraveling the complicated rock succession of the type Silurian, subdividing it and establishing its graptolitic and shelly faunas as independent world-wide standards for correlation. In most of the rest of the world, however, the Silurian is one of the most neglected and thus most poorly known sequences—ironic contrast to the wide attention given to the lower part of Murchison's original Silurian, the Ordovician, which was not split off as an independent system until 1879.

1. Physical Events and Features.—Mountain-building deformation older than the oldest local Silurian deposits is indicated by Silurian overlap across folded Ordovician rocks in the Acadian region of east central North America (Taconic folding or orogeny) and eastern Australia (Benambean orogeny). Similar relationships indicate later deformation between Silurian and overlying Devonian in Australia (Bowning orogeny), in northwestern Europe (Caledonian orogeny) and apparently in the western U.S. Thus it is frequently said that the Silurian began and ended with important, if local, compressive movements of the earth's crust. Strong evidences of volcanism throughout the Silurian of eastern Australia, the Pacific border of North America and the Acadian region, however, and the thick sequences of coarse detrital sediments in Australia, south Scotland and the eastern Appalachian trough seem to imply continued crustal instability in the borderlands. As a consequence of such events, some seaways that were well defined during Ordovician time were strongly modified or abandoned in the Silurian, and new directions of marine invasion were found.

2. Climate and General Ecology.—Widespread reef formation by Silurian corals and stromatoporoids (hydrozoan relatives) couples with the evidence for high calcium metabolism furnished by laterally equivalent limestones and dolomites to imply extensive warm shallow seas in present continental areas and little latitudinal or seasonal temperature variation. Interference of mountainous tracts with prevailing winds, or other special meteorologic circumstances, created local arid conditions in the middle latitudes such as those under which the Silurian salt and gypsum deposits of New York state and the Michigan basin originated. If deposits which some interpret as glacial in southeastern Alaska, British Columbia and northern Norway prove to be both glacial and Silurian they would suggest mountains high enough to support valley glaciers.

Principal ecologic variations were presumably connected with physical and chemical changes due to increasing depth at sea or altitude on land, or special geographic configuration.

3. Facies, Biogeography and Migration Routes.—As those of all other geologic ages, Silurian rocks formed under differing local physical or chemical conditions, or with different sources of supply and biologic migration routes, take on a characteristic stamp, or facies, which may tell a great deal about their origin and ecology. Facies variations of rocks and accompanying fossil assemblages may be lateral (in space), vertical (in time) or both. Similar sedimentary facies that do not differ drastically in age, and some that do, are likely to contain similar fossils; and different facies of the same age ordinarily contain different assemblages of fossils. (*See* also FOSSIL.)

The stratigraphic correlation of unusual facies is worked out from field relations to deposits of known age, or is based on ubiquitous fossils of restricted time span. Species of graptolites, so characteristic of Silurian as well as Ordovician beds, have proved especially useful in indicating the probable contemporaneity of contrasting sedimentary facies. They are, in fact, such generally excellent index fossils, that they are logically interpreted as animals which actually floated at the surface of the water all or much of their lives instead of living somewhere below the surface like most planktonic species or floating only temporarily like the larval stages of so many aquatic organisms. This would explain their being buried, not only with the open-sea sediments of their normal living environment, but also in many different shallow water deposits, together with bottom-living shelly fossils. In eastern Australia, for example, species of *Monograptus* establish the Silurian age and particular stratigraphic level of the important early land plant *Baragwanathia* and associated paleobotanical remains.

Classification of Silurian biotal facies (faunas and floras) shows that they follow sedimentary facies that appear to be related to distance from land, depth of water and nature and amount of detritus brought to the sedimentary site. The usual Silurian deposits and faunas (plants are not numerically important) are divisible into three principal facies of marine sediments: (1) graptolitic shale and flaggy sandstone; (2) detrital limestone and calcareous shale with shelly faunas; (3) carbonate rocks (limestone and dolomite) with mainly coral and stromatoporoid faunas. To these three primary facies types must be added evaporite sequences (salt and gypsum) with highly restricted planktonic faunas in shaly interbeds, slabby argillaceous limestones that in part seem to represent special conditions of tide-flat sedimentation, Late Silurian continental (nonmarine) deposits with plants and fishes, and the volcanic facies, which commonly grades to or alternates with graptolitic shale. The three principal facies are found in most regions with thick and extensive Silurian sections; the others represent special local conditions.

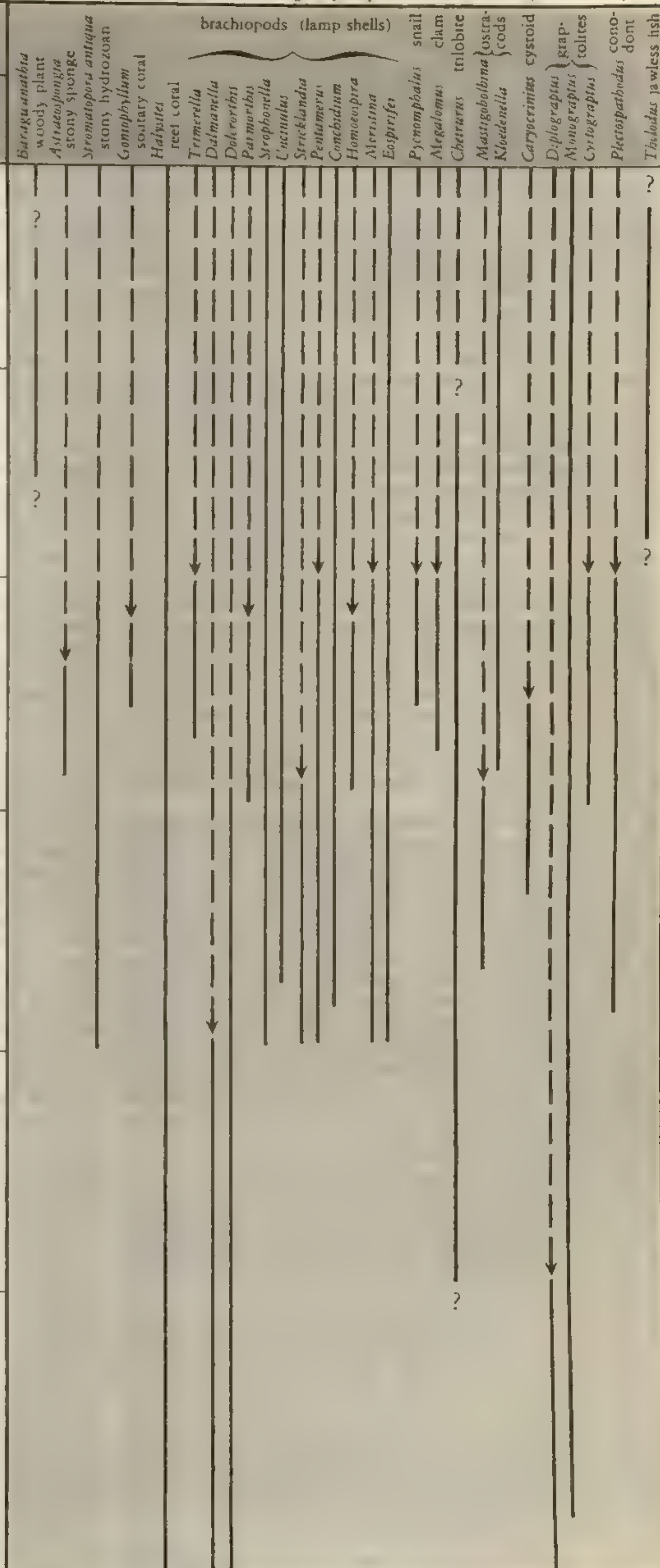
The wide dispersal of similar Silurian faunal facies implies good communication channels. Highly generalized possible connections at times of maximum marine expansion are suggested by the distribution map. The distinctive limestone and dolomite sequences of the western U.S. are very exceptional. They represent an extreme development of the carbonate facies, characterized by predominance of dolomite, with a sparse but widely distributed fauna

Correlation Table Showing General Silurian Subdivisions, Local Rock Sequences and Ranges of Representative Genera

Series	Upper Silurian Middle Silurian Lower Silurian
Stage	Ludlovian Wenlockian Upper Llandoverian Lower Llandoverian (Salopian: Ludlovian and Wenlockian; Valentian: Upper and Lower Llandoverian)
North America: South Central Canada	— — — Attawapiskat limestone Ekwan River limestone — — — ? — — — Severn River limestone Port Nelson dolomite — — —
North America: Nevada	? Lone Mountain dolomite — — — Roberts Mountain limestone and dolomite — — —
North America: New York state	Rondout limestone Cobleskill dolomite Salina shales with salt beds Guelph dolomite Lockport dolomite Rochester shale Irondequoit limestone Williamson shale Wolcott limestone Sodus shale Reynales limestone Maplewood shale Thorold sandstone Grimsby sandstone Cabot Head shale Whirlpool sandstone ? Upper Queenston ?
North America: Northern Nova Scotia	Stonehouse limestone Moydart limestone McAdam limestone Ross Brook shale Beechhill sandstone and limestone ?
Britain: Lake district	Kirkby Moor flags Bannisdale slates Coniston grits Upper Coniston flags Lower Coniston flags Stockdale graptolitic shales
Britain: S.E. Wales and Shropshire	Upper Ludlow shale and sandstone Aymestry limestone Lower Ludlow shales and flags Wenlock limestones and shales Wenlock shales and flags Woolhope limestones and shales Upper Llandovery shales and lesser sandstone Middle Llandovery shales Lower Llandovery shales and sandstones
Gotland	*Ascoceras* limestone — — — Ostracod limestone — — — *Sphaerocodium* limestone *Dayia* limestone ? — — — Marlsand limestone — — — ? — — — *Stricklandinia* marl — — — *Arachnophyllum* shale — — —
Bohemian massif	*Monograptus hercynicus* limestones and shales (f_1) Crinoidal limestone and shales (e_2) Brachiopod limestones and shales (e_2) Cephalopod sandstones (e_2) *Monograptus dubius* shales and limestones $e_{1\gamma}$ *Monograptus priodon* shales and limestones $e_{1\beta}$ *Diplograptus* shales $e_{1\alpha}$ (e_1)
U.S.S.R.: Eastern Urals	*Atrypa marginalis* limestones — — — *Conchidium* limestones — — — *Brooksina* strata limestones and shales — — — Bobrovka limestones, shales and sandstones — — — Tura limestones — — — Shales and sandstones (partly graptolitic) ?
U.S.S.R.: Altai mts	Reef limestones and shales — — — ? Shales and limestones with *Halysites* and graptolites ?

Correlation Table Showing General Silurian Subdivisions, Local Rock Sequences and Ranges of Representative Genera (Continued)

Southwest China: East and central Yunnan	East Australia: Melbourne district	East Australia: Yass district	South America	North Africa: Central Morocco
Miao-kao limestones and shales with ostracods and fish	Lower Yering sandstones, shales and conglomerates		Widely present but undivided	Shales and limestones
		Upper Hume shales, sandstones and conglomerates		Shales with *Monograptus nilssoni*
Apparently missing	Melbourne graptolitic sandstones and shales	Lower Hume shales and limestones		Shales with *Monograptus riccartonensis*
		Laidlaw tuffs and grits		
		Yass sandstones, shales and limestones		
Mien-tien shales, sandstones and limestones with *Monograptus* and *Climacograptus*	Keilor graptolitic shales, sandstones and conglomerates	Douro tufts		Shales with *Monograptus spiralis* and *turriculatus*
		Bango limestones, shales and tuffs with *Halysites*		Shales with *Monograptus regularis*
		Hawkins tuffs		Shales with *Monograptus* and *Climacograptus*



of corals and large, thick-shelled pentameroid brachiopods. Thin graptolitic shales have been found only locally. The sea in which these deposits formed may have connected with the eastern North American seaways at its south end, although proof of this is lacking.

II. SILURIAN LIFE

1. Characteristics and Evolution.—The record of Silurian invertebrate life shows essentially an elaboration of Ordovician lineages (*see* ORDOVICIAN SYSTEM). It has, however, novel and distinctive features of its own, such as expansion of the corals and echinoderms, and a general substitution of the dominant groups of brachiopods. And it differs from the Devonian record in the types of corals and brachiopods present, the continued importance of graptolites and many less striking features. These differences, of course, are due partly to accidents of preservation and vagaries of collecting and recording.

Silurian plants and fishes are the oldest preserved, except for isolated occurrences of fish fragments in the Middle and Upper Ordovician of the North American Rocky mountains and Cambrian plants reported from eastern Siberia, Estonia and north India. This distribution pattern is apparently attributable to the rarity of preserved pre-Silurian terrestrial deposits.

The characteristics and evolution of Silurian or any other fossil life are of interest not only from an aesthetic and scholarly point of view, but also because they provide the only practicable means of consistently telling fossil time. Radiogenic methods of absolute age determination do not yet suffice for general dating or correlation of the Paleozoic rocks, even though a very few apparently reliable dates are used to set, approximately, the hands of the paleontological clock in terms of years.

From the fossils also can be deduced much of the history, ecology and geography of ancient times, always in connection with the characteristics and positions of the rocks themselves. Ecologic deductions, for instance, are greatly aided by the tracks, trails and burrows made by different organisms. Indeed these are almost the only traces of life in some deposits, such as the Silurian of the eastern Appalachian trough.

2. Invertebrates and Associated Fossils.—In the Silurian, as with other fossiliferous rocks, the invertebrate and noncellular (or unicellular) organisms are the true indices of time, on the basis of which nearly all meaningful stratigraphic correlation with the standard marine sequence is founded. The form sequence of the monograptid graptolites illustrates well how such correlation may be accomplished either by relating stage of evolution to time or by the empirical knowledge of the ranges of specific morphologic types. Correlation, or paleoecological analysis, in practice, however, is usually made by analysis of the total biotal characteristics of the problem sample, considered in relation to lithology and sequence of the enclosing sediments. Ranges of selected index genera at the right side of the correlation table illustrate how Silurian fossiliferous sediments can be dated, using apparent terminal ranges of a variety of fossils.

The graptolites, despite their suggested close alliance to the vertebrates, are at best invertebrate chordates, and, in this sense, the most conspicuous and distinctive Silurian invertebrates. The monograptids are the evolutionary culmination of the simple sawblade-like forms of graptolites and appear to be exclusively Silurian—no Ordovician or Devonian report has ever been unequivocally substantiated.

Moreover, only a few species of two genera (*Diplograptus* and *Climacograptus*) of their Ordovician ancestors carried over into the Silurian, and these only into the lower part. The Silurian graptolite assemblage thus consists almost entirely of *Monograptus*, of short-lived derivative genera from it or older forms, and of occasional reticulate colonies such as persisted into the Devonian. The apparent time subdivision of Silurian rocks is based largely on short-ranged species of *Monograptus*—18 of the 26 graptolite zones in the type Silurian sequence are so characterized, and many of the same species floated around the world.

Corals and stromatoporoids are subordinate to the graptolites as Silurian time indicators but far ahead of them in range of evolutionary diversification. Compound corals particularly flourished, including the tabulates, which now reached their greatest expansion, and fasciculate tetracorals which did not attain their acme until the Devonian. Horn corals of simple types were more abundant and more complicated than Ordovician, less so than Devonian relatives. Stromatoporoids had also to await the Devonian for their climax, and no Silurian or other common Paleozoic coelenterate is anything like its modern relatives. Presumably however, they did have the same alternation of sessile and free-floating generations as typifies the living representatives of the phylum, which would account for wide distribution of identical species in favourable environments. Unlike the graptolites, which apparently could drift even after death, they are not found far from their place of life except as a result of unusual events.

Bryozoa are still about as common as they were in the Ordovician, with a slight dwindling of stony types and an increase of the more delicately ornamented twiglike cryptostomes. Locally they made small mounds at the bottom of the Silurian seas, as represented by the ball reefs of the Rochester shale in western New York state.

Among Silurian brachiopods, the globular, short-hinged and commonly large pentameroids were the most distinctive forms. Short-hinged spire-bearing brachiopods were essentially new, although there are common Ordovician forms. The wide-hinged spiriferoids began the expansion that was to flower in Devonian and Lower Carboniferous time.

The pelecypods continued to expand slowly from Ordovician beginnings, and the prosobranch gastropods became important fossils, in continuance of a long, progressive expansion to modern times. Among the cephalopods, ammonoid progenitors appeared and nautiloids dwindled. Some distinctive Ordovician groups of nautiloids were missing completely (*e.g.*, Endoceratida) and others died out during the Silurian (*e.g.*, Ascoceratida). As if to emphasize their decline, no important new nautiloid groups appeared.

Among the arthropods, trilobites continued their slow decline from a Cambrian maximum, and ostracods and merostomes both expanded slightly. Among the latter, eurypterids became much more common, while the aglaspids were missing; but xiphosurans ancestral to living *Limulus* are recorded. Ostracods are locally abundant and useful Silurian index arthropods. Trilobites, despite their general decline, showed an expansion of bizarre spiny types, perhaps as an adaptation to floating or camouflage, forced by increasing competition for the nutrient-rich bottom niches or predation. Allied developments may have caused the eurypterids to enter fresh water in later Silurian time, from which, according to one theory, they eventually chased the fishes. It is an old law of life that each time new or better offensive or competitive equipment or tactics are evolved compensation is made, or the hunted or competitive inferior perishes or takes exile.

Crinoids were the most successful Silurian echinoderms, enjoying their greatest recorded expansion of numbers and kinds. Unlike their commoner living relatives, these lilies of the sea were stalked forms; if they were equally colourful their colonies were the showy marine gardens of Silurian time. Cystoids were also common, and, although much more primitive, were similar in living habits and appearance to the crinoids. Other echinoderms were not important.

Noncellular invertebrates, or the Protista (Protophyta and Protozoa), are abundant in some Silurian deposits. Chitinous and agglutinated Foraminifera have been described from North America, the Baltic and England, and Radiolaria are occasionally reported, but little is known in detail of these Silurian records.

An important group of Silurian problematica is the Chitinozoa, consisting of minute, black, chitinous, vase-shaped to club- or ball-shaped organisms. Revelation of such tiny creatures or parts of creatures by chemical treatment of apparently unfossiliferous rocks sometimes permits the correlation of deposits whose position cannot be closely established by other methods.

3. Fishes, Land Plants and Associated Fossils.—Silurian occurrences of fish and plant fossils are more numerous and complete than in older rocks. They are, however, mostly, though not entirely, restricted to the upper part of the system and the fossils are not demonstrably more advanced than their predecessors.

DISTRIBUTION OF SILURIAN SEAS AND LANDS

Conversely, they are greatly surpassed in number and variety by Devonian records. In the Late Silurian, in fact, the arthropods appeared to be making more headway than the vertebrates in colonizing the lands. The arachnids were represented by the scorpionid *Paleophonus*, from the Upper Silurian of both sides of the Atlantic. Millipedes are reported to occur in association with eurypterids (water-dwelling scorpionlike creatures) in Wales. Neither scorpion nor millipede was surely air breathing, but both were at least evolving in that direction. In the estuarine to questionably marine or fresh water deposits, where the early fishes are most commonly found, is also an abundance of eurypterid remains. Indeed the Silurian and Devonian saw the climax of this group, which comprised the greatest arthropod predators of all time, perhaps including principal foes of the early fishes, and possible ancestors of the arachnids.

The simple, externally armoured, mainly backboneless earliest fishes and the primitive early plant assemblages are discussed in a separate article on the Devonian system (*q.v.*).

III. REGIONAL SUMMARY

The distribution map shows where Silurian marine deposits of different ages might have been laid down on the present land areas. Naturally a map for any instant of the 30,000,000 years or so of Silurian time would look very different from this highly generalized scheme, particularly if the Paleozoic geology of the ocean basins were known. Silurian nonmarine deposits and plutonic rocks are not well enough known to consider on a composite map of this scale. This map and the correlation table together give a general view of Silurian stratigraphy and paleogeography that supplements the condensed discussion of outcrop regions below.

1. Great Britain.—Silurian sediments of the type region in Wales and western England, and extending into southern Scotland, were deposited in a northeast-southwest trending subsiding area (the Caledonian geosyncline) and bordering shelf seas to the east. The basin or geosynclinal facies is characterized by graptolite-bearing slates and shales, flaggy and shaly silt stones, sandstones and mixed coarse detritus about 5,000 to 10,000 m. thick. This facies is typically developed in the Lake district of northwestern England. The shelf facies consists of shales (including mudstones), limestones and silt stones with a shelly fauna and a maximum thickness of perhaps 1,000 m. in typical development. It thins eastward to disappear toward the inferred Silurian seashore. Complete gradation in sedimentary facies, thickness and fauna occurs from shelf to basin along the margins of the subsiding trough, with some persistent zones of graptolitic shale that correlate between facies and record the history of down warping. The type area of south Wales and the Welsh borderland is happily intermediate between extreme developments of either shelf or basin facies. Northward, to the south of Scotland, the sediments suggest approach toward the northwest side of the geosyncline.

At many places in the Welsh borderland the oldest Silurian deposits in Britain are separated from underlying rocks by unconformity or abrupt upward reduction of volcanic components. Preceding crustal movements, of which the profuse Ordovician volcanicity is one indication, may have brought the southeastern part of the British subsiding area generally above water in the waning stages of the Ordovician. Subsidence, with sharply waning volcanism, was almost immediately renewed, however, and continued intermittently to near the end of the Silurian or beyond. Onset of the Caledonian orogeny forever reversed the process and made mountains where the sea had been, from Wales to Scandinavia.

Subdivisions of the British sequence are shown on the correlation table. The oldest interval, the Llandoverian (or Valentian, or May Hill), saw the sharpest distinctions between basin graptolitic shale facies and calcareous and in part sandy shelf facies. This distinction was less marked in the Wenlockian or middle division, when the importance of the calcareous facies was increased by growth of coral-stromatoporoid reefs and colonies on the shelf and spread of shelly fossils into the graptolitic shales. In

Ludlovian or Upper Silurian time the distinction became progressively less marked. It finally disappeared in the last half of the Ludlovian (Aymestry) along with general faunal impoverishment and disappearance of the graptolites from Britain.

2. European Continent.—The same Caledonian depression in and along which were deposited the Silurian sediments of Britain, continued northeastward to Scandinavia, with a bordering shelf sea that locally extended as far eastward as Estonia. Best-known Scandinavian sequences are in southern Norway, southern Sweden and Gotland. That of south Sweden (Skåne) is a somewhat deformed but essentially complete basin succession of graptolitic shales with limestone interbeds increasing toward the top. This grades northwestward to the fine-textured and richly fossiliferous, but also structurally disturbed, intermediate to platform facies of limestones and shales north of Oslo. Southeastward it goes into the thin, calcareous, relatively flat lying shelf-facies of Gotland and Estonia. The Caledonian trough of pre-Silurian time thus seems now to have shifted its axis to the east and to have become a broad but not very deep embayment bulging northeastward into but not across Scandinavia. Later Caledonian deformation from a northwest direction folded the Norwegian sequence and that of south Sweden but did not noticeably affect Gotland or the eastern Baltic.

The Silurian section of Gotland, important because it comprises the type Gotlandian (synonym of Silurian as here used) is a thin but richly fossiliferous succession of limestones and calcareous shales. Many of the hills of Gotland are coral-stromatoporoid reefs, marvelously exposed for study. Despite its importance and fine shelly faunas, agreement has not been reached as to names, limits, or number of subdivisions for the Gotlandian section.

Between the Baltic and Brittany, the Silurian is represented by poorly developed and little understood shales, sandstones and limestones, commonly strongly metamorphosed and discontinuous. Only eastward in the Harz and Thüringer uplifts does the section improve, with Bohemian aspects. The Armorican massif itself (Brittany) displays a basinal Silurian facies—500 m. or so of graptolitic black shales, with sandstone at the base.

In the Bohemian (or Czechoslovakian) massif around Prague is the famous Silurian sequence of Joachim Barrande and J. Perner; somewhat deformed by Caledonian as well as by later movements. This is prevailingly of graptolitic shales with at first inconspicuous limestone bands. The calcareous fraction increases toward the top until the whole sequence goes over to gray and black limestones, with both bedded and reef facies. Total thickness is somewhat under 200 m. A thin sandstone with land plants near the top is followed by shale and limestone with the last known uniaxial graptolites. This sequence represents a large span of Silurian time during which sediments presumably accumulated in a relatively shallow or slowly subsiding basin which connected through Brittany to the Caledonian seas. It also spilled over through the Montagne Noire of south central France and had connections to Elba, Sardinia, Spain, the eastern Bavarian Alps and the Balkans, where Silurian rocks generally resemble those of central Czechoslovakia.

3. U.S.S.R.—The so-called Tethyan seaway connected through the present site of the Alps, Mediterranean and the northern middle east to the Uralian trough and Siberia during parts of Silurian time. It left about 400 m. of richly fossiliferous marine Wenlockian and Ludlovian calcareous shales and limestones in Podolia, southwest Ukraine (formerly southeast Poland). The beds dip slightly westward along the walls of deeply entrenched streams beneath a cover of younger deposits, to disappear under Downtonian (basál Devonian) marine deposits and the Old Red Sandstone.

The north-south trending Uralian subsiding belt lay on the eastern side of the north European heartland and extended beyond Spitzbergen presumably to the north of Greenland and the northern islands of the Canadian Arctic archipelago. It developed independently, and its Wenlockian and Ludlovian shelly fauna is more like that of southeastern Alaska and northern California than to anything else well known. Presumably there was a direct connection across central Siberia and the Aleutian area at this time. Graptolites occur in Llandoverian shales and sandstones of the Urals. The Silurian of the Altai mountains of south central Siberia consists of shales and limestones with graptolites and corals, at places reefy.

4. Asia.—Limestones and sandstones with corals, brachiopods and graptolites are known from Spiti and Kashmir at the north of the Indian continent, and in the Shan states of Burma, where graptolites establish Early and Middle Silurian age. Total thickness may not exceed 300 m.

Connections are from the Tethyan or Mediterranean area to southwest China, where (in east and central Yünnan) there are about 1,000 m. of fossiliferous shales, sandstones and limestones of Llandoverian and Ludlovian age. Llandoverian deposits, in fact, occur widely in western and central China, but the Ludlovian is known only in the southwest, and the Wenlockian nowhere at all. The record in north China is obscure.

The Kitakami mountains of north Japan (north Honshu) reveal 200 to 300 m. of Middle and Upper Silurian limestones and slates, with corals and stromatoporoids in the limestones and radiolarians in slates near the series boundary. Silurian is also reported from central Honshu and the islands of Shikoku and Kyushu.

In Celebes is a Silurian outpost between Yünnan and eastern Australia, identified by the coral *Halysites*.

5. Australia.—The known Silurian of Australia is limited to the eastern coast and Tasmania. A thick (3,000 m.) conglomerate in the centre (Macdonnell ranges) has been tentatively referred to the Silurian for obscure reasons, but no fossils are known from it. Known Silurian sections consist dominantly of coarse clastic and volcanic deposits. Interbedded with these locally are graptolitic shales and limestones or calcareous shales with a shelly fauna. Northeast of Melbourne are 8,000 m. of clastic beds that may represent the whole of the Silurian. In the Yass district of New South Wales the section is only about 1,000 m. thick, but is famed for the rich and beautifully preserved fossils of its Yass beds and lower Hume limestones and shales.

Silurian crustal unrest, indicated by prolonged and widespread volcanic activity, culminated with compressive movements and deep-seated igneous intrusions locally termed the Bowning orogeny, and considered a phase of the more inclusive Caledonian orogeny. When the Devonian seas reinvaded eastern Australia their deposits frequently came to rest on the folded and erosionally truncated edges of Silurian rocks.

6. South America.—A compilation of South American paleogeographic data by Kenneth Caster shows broad invasion of that continent by the Silurian seas. The details, however, are little known and difficult to interpret, because Silurian was lumped with Ordovician in the reconnaissance studies that so long characterized South American geological exploration. Between South America and Antarctica, detrital and volcanic deposits at the base of the Cumberland bay series of South Georgia Island are questionably referred to the Silurian on doubtful fossil evidence.

7. North Africa.—Silurian deposits throughout north Africa are prevailingly in the basin facies. They imply a broad, open, but not very rapidly subsiding sea that connected freely with the Tethyan (Mediterranean) and north European seaways. In central Morocco the entire succession, save the top fraction, consists of graptolitic shales with characteristic species of most of the British stages. Altogether it is about 1,000 to 1,500 m. thick. This facies continues not only through the Great Atlas, where an axis of subsidence might be expected, but also southward, to the middle of the Sahara shield.

8. North America.—The Silurian rocks of North America are mainly the deposits of warm platform seas, which deepened in the eastern U.S. to form the Appalachian trough. Detrital sediments originating in eastern borderlands were largely restricted to this trough, and some (with fishes and eurypterids) along its eastern side may be nonmarine or very near shore. Leaving aside the Acadian deposits, with their European affinities, the gross form of the Appalachian succession is that of a wedge of sandstones, shales, carbonate rocks and some conglomerates from 300 to 2,000 m. thick—like a giant composite delta. By contrast, equivalent strata on the west side of the trough and the east flank

of the Cincinnati arch are greatly reduced and are here already within the interior platform seas. Prevailing aridity over and adjacent to the Late Silurian seas from western New York state to southern Michigan led to the accumulation of commercial deposits of gypsum and salt in arms of the sea or intermittently connected salt basins. These evaporites are interbedded with green and red shales and grade eastward to coarser and partly red detritus. During parts of the Middle Silurian unusual (and uncertain) conditions led to deposition of the famous hematitic Clinton iron ores from New York to Alabama. The iron replaces formerly calcareous shell fragments and is deposited between them in stratigraphically limited zones of varying thickness. Graptolitic shales are also found in some of the Appalachian trough deposits, although not as common as in the good basin developments of Europe and north Africa.

Little mud or sand entered the Silurian seas west of the Cincinnati arch, and the sediments there were thin, almost entirely calcareous, and locally include reefs. Thus, sequences of limestone generally less than 300 m. thick are prevalent in the Mississippi valley. Limestone is also a common rock of the Silurian in central Canada, southwest of Hudson bay and in the Acadian region at the eastern apex of the continent; for instance, in eastern Quebec (Chaleur bay), northern Nova Scotia (Arisaig) and Anticosti Island. The interior limestones grade to dolomites in the shallower parts of the ancient seas, especially in the Great Lakes region. The Silurian has restricted development in the eastern Great Plains and the subsurface of north central Texas, but richly fossiliferous limestones and shales are exposed in the Arbuckle mountains of southern Oklahoma.

Development in the Cordilleran region is almost restricted to the U.S., with an arm along the Mexican border to west Texas. There it is preponderantly dolomite, with limestone and shale subordinate and with identifiable fossils only locally. In the Eureka district of central Nevada the section is as much as 1,400 m. thick. The Pacific border deposits of California and Alaska are coarse, impure, detrital and volcanic sediments with occasional fossiliferous limestone lenses and with Siberian and Uralian faunal affinities. Those of the North American arctic seem to be mainly calcareous, dolomitic or shaly. *See* also references under "Silurian System" in the Index.

Bibliography.—*Great Britain:* Roderick Impey Murchison, "On the Silurian System of Rocks," *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 7:46–52 (1835), *The Silurian System*, parts i and ii (1839); British Geological Survey and Museum, "The Welsh Borderland," "Northern England," *Brit. Reg. Geol.* (1948); O. T. Jones, "The Geology of the Llandovery District: Part I., the Southern Area." *Quart J Geol. Soc. Lond.* 81:344–388 (1925).

Europe and U.S.S.R: Roland Brinkman, *Abriss der Geologie*, pp. 50–72 (1954); Maurice Gignoux, *Stratigraphic Geology*, pp. 71–103 (1955); Emanuel Kayser, *Lehrbuch der Geologischen Formationskunde*, pp. 103–175 (1923); A. N. Khodalevich "Upper Silurian Brachiopoda of the Eastern Urals," *Transactions Ural Geological Service* (1939); J. Kiaer, "Das Obersilur im Kristianiagebiete," i, *Math-Naturw.* (1908); J. C. Moberg, "Historical-Stratigraphical Review of the Silurian of Sweden," *Sverig. Geol. Unders. Afh.*, series C, no. 299 (1911); D. V. Nalivkin, *Major Divisions of the Paleozoic of the Union of Soviet Socialist Republics*, Report of 16th International Geological Congress, Washington, D.C., 1933, pp. 573 583 (1936); D. V. Nalivkin *et al.*, *The Geological Map of the USSR* (scale 1.5,000,000), Ministry of Geology of the U.S.S.R., 8 sheets (1955); J. Perner and O. Kodyn, "Correlation of the Silurian of Bohemia," *Amer. J. Sci.*, series 5, vol. 4, pp. 53–72 (1922); W H. Twenhofel, "Silurian and High Ordovician Strata of Estonia, Russia, and Their Faunas," *Bull. Mus. Comp. Zool. Harv.*, vol. 5, no. 4, pp. 289–340 (1916).

Asia, Africa, Australia: A. W. Grabau, *Stratigraphy of China*, part i, "Paleozoic and Older," pp. 103–139, Geological Survey of China (1923–24); J. S. Lee, *The Geology of China*, pp. 108–115 (1939); D. N Wadia, *Geology of India*, pp. 154–169 (1953); Henri Termier, *Etudes Géologiques sur le Maroc Central et le Moyen Atlas Septentrional*, 4 vol., Service des Mines et de la Carte Géologique de Maroc, Notes et Mémoires, no. 33, pp. 219–299 (1936); T. W. Edgeworth David and W. R. Browne, *The Geology of the Commonwealth of Australia*, pp. 183–228 (1950).

North America: T. W. Amsden, "Lithofacies Map of Lower Silurian Deposits in Central and Eastern United States and Canada," *Bull. Amer. Ass. Petrol. Geol.*, 39.1:60–74 (1955); K. E. Caster, "The Age and Relations of Columbian Devonian Strata," Proceedings of the 8th American Science Congress, vol. 4, *Geological Sciences*, pp. 27–67 (1942); C. O. Dunbar, *Historical Geology*, pp. 181–201 (1949); H. M. Duncan, "Ordovician and Silurian Coral Faunas of Western United States," *Bull. U.S. Geol. Surv.*, 1021-F, pp. 207–236 (1956); Edwin Kirk and T. W. Amsden, "Upper Silurian Brachiopods from Southeastern Alaska," *Prof. Pap. U.S. Geol. Surv.*, 233C (1952); T. B. Nolan, C. W. Merriam and J. S. Williams, "The Stratigraphic Section in the Vicinity of Eureka, Nevada," *Prof. Pap. U.S. Geol. Surv.*, 276, pp. 36–40 (1956); C. K. Swartz and Collaborators, "Correlation of the Silurian Formations in North America," *Bull. Geol. Soc. Amer.*, 53, pp. 533–538 (1942).

(P. E. Cd.)

SILVA, ANTÓNIO JOSÉ DA (1705–39), Portuguese playwright, also known as "O Judeu" (The Jew), who was an important figure in the revival of drama in 18th-century Portugal. He was born at Rio de Janeiro, Brazil, the son of Portuguese Jews, with whom he went to Lisbon in 1712. In spite of initial hardship, resulting from persecution under the Inquisition, he began to study law at the University of Coimbra. In 1726 he was imprisoned with his mother and brothers, and, under torture, abjured his Jewish faith. He was set free and in 1728 completed his studies. He returned to Lisbon, joined his father's legal practice, and married a cousin, who had also suffered religious persecution. They were both charged with heresy by the Inquisition and imprisoned on Oct. 5, 1739. António was first condemned to an auto-da-fé but on Oct. 18 he was strangled and his body burned.

Silva wrote eight plays, all produced at the Bairro Alto theatre in Lisbon. These were *A Vida do grande D. Quixote de la Mancha* (1733), *Esopaida* (1734), *Los Encantos de Medeîa* (1735), *Anfitrião* (1736), *Labirinto de Creta* (1736), *Guerras do Alecrim e da Mangerona* (1737), *As Variedades de Proteu* (1737), and *Precipício de Faetonte* (1738). They were written for the *ópera dos bonecos* (puppet theatre) in a new style, developed under the influence of Italian opera and Spanish and French comedy, which consisted of a mingling of prose dialogue and popular lyrics. They form a skilled and witty satire against the pretensions of a society based on caste and privilege.

Bibliography.—Three of da Silva's plays were published anonymously, *Labirinto de Creta* (1736); *Variedades de Proteu* (1737); *Guerras do Alecrim e da Mangerona* (1737), as were also the first two volumes of the collection *Teatro Cómico Português* (1744). *Obras Completas* were ed. by José Pereira Tavares, 4 vol. (1957–58). *See* also Teófilo Braga, *História do Teatro Português; a baixo comédia e a ópera* (1871); articles by Claude Henri Frèches in *Bulletin d'Histoire du Théâtre Portugais*, vol. I, no. I (1950); vol. II, no. I (1951); vol. IV, no. I (1953), and in *Bulletin des Études Portugaises* (1952); João Lúcio de Azevedo, "O Poeta A. J. da Silva e a Inquisição," in *Novas Epanáforas* (1932); Luís de Freitas Branco in *A Evolução e o Espírito do Teatro em Portugal*, vol. II, pp. 114–115 (1949). (L. de S. R.)

SILVANUS, an ancient Roman god of the countryside, similar in character to Faunus (*q.v.*), normally depicted in the guise of a countryman, as befits one whose worshipers were mostly peasants. Initially the spirit of the unreclaimed woodland fringing the settlement, he had some of the menace of the unknown. But he had a footing in the cultivated land as well as in the wilderness, and when clearings pushed back the forest he evolved into a god of woodland pastures, of boundaries, and of villas, parks, and gardens. He never enjoyed a state cult or temple, priesthood, or official festival, but the simple ritual of his private worship beneath a sacred grove or tree, with its frugal offerings or sacrifice at an often improvised altar, had a universal appeal, particularly to humble folk in all parts of the Roman Empire. In Italy, where little woodland remained, Silvanus was the patron of smallholders or small-scale cattle farmers. Outside Italy he was especially venerated by those connected with forests—huntsmen, lumbermen, timber merchants. In the provinces his cult often fused with that of local divinities; thus the Illyrian rural god equated with Pan by the Greeks became Silvanus under the Romans. In Latin literature his character tended to merge with that of Silenus or Pan (*see* Pan; Satyrs and Sileni) and to be assimilated to the Greco-Roman mythological tradition. (D. E. W. W.)

SILVER, a metallic element known from very early times. Its chemical symbol is Ag (from the Latin *argentum*), atomic number 47, and atomic weight 107.870; its melting point is 960.5° C, and its density at 20° C is 10.49. Together with gold, iridium, palladium, and platinum it is one of the so-called precious metals and, because of its comparative scarcity, brilliant white colour, malleability, and resistance to atmospheric oxidation, has long been used in the manufacture of articles of value such as coins, ornaments,

and jewelry. Apart from these uses, silver has a number of applications, most of which depend on its high thermal or electrical conductivity and corrosion resistance. Silver which has been hardened by elements such as copper or gold is used in electrical contacts. In electronic engineering conductors are sometimes coated with silver for radio-frequency currents, where the skin effect is important.

Silver apparatus is used to a limited extent in the chemical engineering industry, for example, in the manufacture of certain medicinal chemicals, in processing foodstuffs and beverages, and in handling organic acids such as citric and lactic acids. Silver functions as a mild oxidation catalyst in the vapour phase oxidation of certain organic compounds such as the lower aliphatic alcohols.

The silvering of glass mirrors is an old-established industry and depends on the action of a chemical reducing agent such as invert sugar, Rochelle salt, or formaldehyde on an ammoniacal silver solution. It is also possible to produce mirrors by evaporating silver on to a surface from an electrically heated filament in high vacuum.

HISTORY

Early Roman records show that before the use of the term *argentum*, the word "Luna" and a crescent moon symbol were used for silver. The actual use of silver dates far back into antiquity; however, it is believed that gold and copper were the first metals to be worked by man. Silver ornaments and decorations have been found in royal tombs dating back as far as 4000 B.C. In the Code of Menes, who is supposed to have reigned in Egypt about 3100 B.C., it was decreed that "one part of gold is equal to two and one-half parts of silver in value." This may have been the first gold standard. By 800 B.C. it is probable that both gold and silver were used as money in all countries between the Indus and the Nile. The Romans probably advanced the art and science of metallurgy of silver further than any other people up to their time. The Romans had several large and important silver workings, using perhaps the first fire metallurgy (pyrometallurgy) to produce silver metal and then converting the metal into priceless silver ornaments.

According to Pliny in his *Historia naturalis*, "the ore was washed and sieved five times, fused with lead and then cupelled for pure silver." Cupellation is the process of oxidizing the lead to lead oxide, which is quite volatile at high temperatures (*see* ASSAYING). Several centuries later it was found that certain ores of gold and silver could be amalgamated by grinding in the presence of iron or copper salts, water and mercury. This method finally evolved into the patio, or Mexican process (*see* AMALGAMATION). The cyanide process (*q.v.*) for the treatment of gold and silver ores, developed in the latter part of the 19th century, was largely superseded in the 20th century by flotation combined with smelting. From about 1850, 65% or more of the world production of silver was derived as a by-product in the smelting and refining of copper, lead, and zinc ores.

OCCURRENCE

Silver is widely distributed in nature, but the total amount present in the lithosphere is quite small when compared with other metals. For every 10,000,000 parts of iron found in the lithosphere, there are about 2 parts of silver. Practically all sulfides of lead, copper, and zinc contain some silver, and in general if the three different metal sulfides occur in the same ore, the largest per cent of the silver will be found in the lead sulfide (galena). These are called argentiferous ores.

Argentiferous ores may contain amounts of silver from a trace to several thousand troy ounces per avoirdupois ton or about 10%.

Unlike gold, silver tends to form many naturally occurring minerals. Among the most important are argentite (silver glance), Ag_2S, and cerargyrite, or chlorargyrite (silver chloride, or horn silver), $AgCl$; also polybasite (silver antimony sulfide) $Ag_{16}Sb_2S_{11}$, proustite (silver arsenic sulfide) Ag_3AsS_3, pyrargyrite (silver antimony sulfide) Ag_3SbS_3, and stephanite (also a silver antimony sulfide) Ag_5SbS_4. Deposits of native (chemically free, or uncombined) silver are also commercially important. Sylvanite

BY COURTESY OF AMERICAN SMELTING AND REFINING CO

DRILLING UNDERGROUND IN A SILVER MINE. WIRE NETTING PROTECTS MINERS DURING BLASTING

(gold, silver telluride) (Au, Ag) Te_2 has been mined as an ore of both metals. (*See* also ORE DEPOSITS.)

The world's largest silver mine, the Real de Monte y Pachuca, is in Hidalgo, Mex. In the United States, the Comstock Lode discovered in 1859 near Virginia City, Nev., proved to be one of the richest gold and silver deposits in the world; other important occurrences include those in the Coeur d'Alene district of Idaho, and in Utah, Arizona, Colorado, and a number of other states. In Canada, silver is found in the Canadian Shield, as at Sudbury and Cobalt in Ontario, and in the west in British Columbia and the Yukon Territory. Silver deposits in the highlands of Peru, as at Cerro de Pasco, have been worked since the 17th century. Important silver-lead-zinc ores occur in Australia at Mt. Isa and North Broken Hill. Smaller but still important occurrences are widespread.

METALLURGY

Several methods are employed for the extraction of silver, or gold and silver from the ores. It must be realized that the majority of ores that contain silver also contain the base metals, lead or copper or zinc, or any combination of these three metals, and are treated metallurgically mainly for the three. The silver they contain is removed as a by-product. Under certain economic conditions it may be more profitable to send such ores to a lead or copper smelter.

Amalgamation and the cyanide process have been mentioned above. The cyanide process is used mainly on ores in which gold is the main valuable constituent and silver a by-product. However, a large amount of the ore mined in Mexico contains only silver as the main valuable constituent and is treated by the cyanide process (*q.v.*). Other processes use flotation and smelting (*see* METALLURGY). Except for the treatment of the Mexican ores, smelting is the most widely used process. Sometimes silver ores or concentrates containing silica, lime, or iron are smelted with lead ores to help produce the desired slag and to take advantage of the properties of lead as a collector (*see* LEAD).

Refining of crude silver bullion or gold-silver bullion may be carried out by a number of methods, the method used depending to a large extent on the purity and the amount of bullion to be refined. Bullions containing a large amount of impurities may require smelting in a small furnace with lead oxide and other materials to form a slag (which can be sent to a lead smelter) and the lead bullion, which contains most of the gold and silver. The lead bullion is then cupelled. The resulting alloy, mainly gold and silver, is called doré. The most important methods of treating high-purity gold and silver bullions are electrolysis and parting.

In electrolysis the gold and silver bullions and dorés are used for the anode (positive pole). Silver and the more basic impurities are dissolved in a dilute nitrate water solution at the anode. Pure silver is deposited at the cathode (negative pole). The resulting cathode silver is melted and cast into bars. (*See* also ELECTROMETALLURGY: *Electrorefining: Silver.*)

Parting, or wet method, is generally applicable to doré or very high-purity gold and silver bullion in which the silver to gold ratio must be higher than two parts of silver to one part of gold. If the ratio is not two to one or better, then the silver must be added and the resulting alloy remelted. The doré or gold and silver bullion is placed into a bath of hot concentrated sulfuric or strong nitric acid, and the silver is allowed to dissolve. The silver forms water-soluble sulfates or nitrates. The residue is filtered off and treated for its gold content. The clear solution containing the silver is treated with ferrous sulfate or copper or iron to precipitate the silver. The resulting precipitate is filtered off, washed, dried, and melted down with flux to give a product that averages 99.5% silver or 995 fine of silver.

Mechanical and Working Characteristics.—In the molten state silver dissolves oxygen (at the melting point about 20 vol. are dissolved per volume of molten silver at one atmosphere pressure of oxygen). Most of this is evolved on solidification, a phenomenon known as the splitting of silver; this is avoided by bringing the oxygen to a low level before pouring.

The metal can be formed by hot or cold methods; the density is lowered by cold working.

It is the most malleable and ductile of all metals except gold, and it can be beaten into leaves of less than 0.00025 mm. thickness. The pure annealed metal has a Vickers pyramid hardness of 26: it is harder than gold but softer than copper. Pure silver is too soft for use as a coinage metal or in the manufacture of jewelry, and for these purposes it is generally alloyed with a small percentage of copper, which while increasing its hardness has no appreciable effect on its colour, but does lower the melting point and thermal and electrical conductivities.

Silver has an excellent thermal conductivity. Taking it as 100 the relative values for other metals are:

Ag	Cu	Au	Zn	Sn	Fe	Pt	Pb	Bi
100	73.6	53.2	19.0	14.5	11.6	8.4	8.1	1.8

The electrical conductivity of silver is also slightly superior to that of copper. Resistivity is appreciably increased by cold working. The optical reflectivity of the freshly polished metal is about 98% for infrared light and 95% in the visible region of the spectrum. It decreases very rapidly in the ultraviolet below 3,500 Å and is only 10% at 3,200 Å.

USES

Historically, a major use of silver has been monetary, in the form of reserves of silver bullion and in coins (*see* BIMETALLISM). By the 1960s, however, the use of silver for industrial purposes exceeded total annual world production. Because of the resulting world-wide shortage, the United States in 1965 eliminated silver from its 10-cent and 25-cent coins and sharply reduced the silver content of its 50-cent pieces from 90% to 40%, marking the first major change in U.S. coinage since 1792. The U.S. Treasury announced it would use its reserve of 1,000,000,000 oz. to hold the world price at or below \$1.29+ an ounce, at least until new cupronickel and copper coins were available in quantity. No change was made in the U.S. silver dollar; the U.S. in 1963 had repealed its silver purchase laws and transactions tax, and authorized the replacement of one-dollar silver certificates by Federal Reserve notes.

Use of the metal for sterling and plated silverware, ornaments, jewelry, and similar products has continued to be important. Alloys of silver with copper are harder, tougher, and more fusible than pure silver and are used for jewelry and coinage. The proportion of silver in these alloys is stated in terms of fineness, which means parts of silver per 1,000 of the alloy. Sterling silver contains 92.5% of silver and 7.5% of another metal, usually copper; *i.e.*, has a fineness of 925. An average sterling teaspoon contains about 1 oz. of silver. Jewelry silver is an alloy containing 80% silver and 20% copper (800 fine). Gold dental alloys contain about 75% gold, 10% silver, 10% copper, and the remaining palladium, platinum, and zinc. Dental amalgams are alloys of silver, tin, and mercury. Yellow gold used in jewelry is composed of 53% gold, 25% silver, and 22% copper.

In the photography industry, silver is used in the form of halides such as silver bromide for making film and photographic plates. (*See* PHOTOGRAPHY: *Theory of Photography.*) Photochromic glass that darkens as light intensity increases and lightens when it decreases contains submicroscopic crystals of silver halide; these turn to metallic silver when exposed to light, but unlike the permanent darkening of photographic film, the silver atoms reunite when the light source is removed and the glass becomes transparent again.

Prior to World War II, manufacturers of airplanes and diesel locomotives began to use pure silver as a bearing material; such bearings have a seizure resistance of 1 compared with about 3 for babbit. After restriction on silver use was removed at the end of World War II, nearly all aircraft and diesel locomotives were equipped with silver-plated bearing sleeves.

Silver has a relative electrical conductivity of 100 compared to about 95 for copper. Silver and its alloys are used extensively in the electrical industry, mainly because of its resistance to oxidation, for switches and contact points, with expanding use of silver-copper alloy contacts in relays to control circuits in computers and tabulators. Silver-magnesium-nickel alloys are used for high-thermal-conductivity applications and in electrical instruments and relays operated at high temperatures. Silver coatings applied to ceramics, glass, and mica find wide use in electronic devices. Intricate electrical circuits, particularly for portable radios, may be made by stamping or printing a given circuit on panels with silver ink. Silver solders or silver brazing alloys make strong joints that have high resistance to corrosion. Among the soldering and brazing alloys are those based on the silver-copper-zinc and the silver-copper-phosphorus systems (*see* also SOLDERING). Special brazing alloys, including silver-aluminum, have been developed for applications requiring high strength at elevated temperatures. Silver also is used in epoxy cold solders to bond microminiature electronic devices without heat, and in conductive compounds used as an adhesive in electronic applications.

The catalytic properties of silver and its salts are used extensively to influence chemical reactions of ammonia for making fertilizers and other oxidation chemical reactions. Silver foil and plates are used for replacement of missing bone fragments. Certain silver compounds have fungicidal properties; a weak silver nitrate solution can be used to disinfect certain plants.

(S. L. S.; X.)

PROPERTIES AND COMPOUNDS

Physical and Chemical Properties.—The principal physical and chemical properties of silver are given in the accompanying table. The natural isotopes are Ag^{107} and Ag^{109} which are roughly equally abundant. Its chief radioactive isotopes are Ag^{108} (half-life, 225 days), Ag^{110} (half-life, 24 seconds), and Ag^{111} (half-life, 7.5 days). Silver is more electropositive than copper but less so than gold. The standard electrode potential is −0.7978 v. at 25° C. The metal does not react with moist or dry oxygen but is oxidized superficially by moist ozone. It is quickly tarnished by sulfur at room temperature and a surface film of silver sulfide is

Physical and Chemical Properties

Chemical symbol	Ag
Atomic number	47
Atomic weight	107.870
Crystal system	face-centred cubic
Melting point (° C)	960.5
Boiling point (° C)	2,212
Electrical resistivity at 20° C, microhm-cm	1.59
Thermal conductivity cal/cm/cm²/sec/° C	0.934(100°)
Density at 20° C, g per cc	10.49
Coefficient of linear thermal expansion /° C × 10^{-6}	19.86(0–100°)
Specific heat at 0° C, cal/g	0.0559
at 100° C, cal/g	0.0568
Periodic classification	IB

formed. A similar effect is obtained with hydrogen sulfide in presence of air or oxygen or with materials such as vulcanized rubber, which can give free sulfur. Solutions of sulfides also blacken silver.

Fluorine, chlorine, bromine, and iodine all react with silver at elevated temperatures but attack on the solid tends to be limited by protective halide films. Aqueous hydrofluoric acid and fluoride solutions do not attack silver appreciably but hydrochloric, hydrobromic, and hydriodic acid react superficially. Silver dissolves readily in either strong or dilute nitric acid, forming silver nitrate ($AgNO_3$) and nitric oxide. Hot strong sulfuric acid dissolves the metal and forms silver sulfate (Ag_2SO_4) and sulfur dioxide. When the acid strength is less than 60% there is no action up to the boiling temperature except in presence of oxidizing substances.

In the absence of air fused alkali hydroxides have no action. It is also unaffected by fused alkali carbonates or cyanides though aqueous cyanide solutions in presence of oxygen form the argentocyanide ion.

Colloidal Silver.—In addition to the highly reactive forms of silver sometimes called molecular silver, which are obtained as finely divided powders by precipitating the metal from its solutions with a reducing agent, or by reducing silver chloride with a metal such as iron or zinc in an acid solution, true colloids are readily obtained. Reducing agents commonly used in preparing colloidal silver are: formaldehyde, hydrazine salts, and salts such as tartrates or citrates. Protective colloids are sometimes added. In addition good silver colloids are obtained by passing an arc between pure silver electrodes under water. The colour of the colloids may vary from lilac to red and is largely conditioned by the particle size. Red and yellow solutions contain the smallest and blue solutions the largest particles. Colloidal silver exhibits reducing properties and, in common with some other colloids, is able to catalyze the decomposition of hydrogen peroxide.

Chemically Pure Silver.—Silver of the highest purity is of special interest to the chemist because it was made and used in all the classical work on the determination of atomic weights. A procedure used by T. W. Richards and H. Wells in 1905 may still be quoted, though various other methods are now available. Silver nitrate is recrystallized, precipitated by hydrochloric acid and the resulting chloride thoroughly washed, and reduced to metal with pure invert sugar and caustic soda. The reduced silver is fused in a block of lime in the reducing flame of a blowpipe. The product is further purified electrolytically and again melted in a lime boat in an atmosphere of hydrogen. Silver of at least 99.999% has been obtained.

Compounds of Univalent Silver.—In the majority of silver compounds the element has a valence of one. These so-called argentous compounds include such familiar substances as silver chloride, bromide, iodide and nitrate. A few argentic compounds are known in which the element has a valence of two (see below).

Silver oxide (Ag_2O) is prepared as a dark brown precipitate by adding an excess of sodium hydroxide solution to a solution of silver nitrate. It is appreciably soluble in water (0.02 g. per litre at 25° C.) and this solution has an alkaline reaction and behaves as if it contained silver hydroxide. It will precipitate many insoluble metal hydroxides from solutions of their salts, absorbs carbon dioxide from the atmosphere to form silver carbonate, and is used in organic chemistry for replacing a halogen atom by a hydroxyl group. The oxide exerts one atmosphere oxygen pressure at equilibrium at 187° C, loses oxygen slowly at 250° C, and is rapidly decomposed to silver and oxygen at [illegible] C. In its reactions it behaves as a moderately strong oxidizing agent. The oxide dissolves in aqueous ammonia probably forming the complex $[Ag(NH_3)_2]OH$. The solution on exposure to air forms a black precipitate of fulminating silver, which is dangerously explosive. Its composition is uncertain but it may contain either the nitride Ag_3N or the imide $AgNH_2$.

Halogen Compounds.—The halides of univalent silver, the most useful compounds of silver, are all well known and stable compounds. Silver fluoride (AgF) differs from the chloride, bromide and iodide in that it is readily soluble in water (182 g. of AgF in 100 g. of water at 25° C). It is obtained by dissolving silver carbonate in hydrofluoric acid and evaporating the solution in va… It forms several hydrates. The fluoride has a rock salt str… in which it resembles silver chloride and bromide but differs… the iodide. The melting point is 435° C; the fused salt de… poses slightly and contains a little free silver. A silver sub… with the formula Ag_2F is formed by the interaction of silver … ride and metallic silver at 50° C to 90° C, or in the cathodic … tion of silver fluoride. It is a solid with a bronze colour … curious crystal lattice in which pairs of layers of silver ions … nate with single layers of fluoride ions. It is thus, in a sense, … mediate between a metal (silver) and an ionic crystal (… fluoride), which is in keeping with the high electrical conduc… of the solid. Decomposition into silver and silver fluoride o… above 100° C or when the solid reacts with water.

Silver chloride (AgCl) is prepared by the addition of hy… chloric acid or a soluble chloride to a solution of a silver salt … silver nitrate). It is formed as a curdy white precipitate wh… coagulates on heating or shaking. The salt may also be prep… by direct union of the elements as, for example, when chlori… passed through molten silver. It melts at 455° C and b… 1,564° C. It is found in nature as the mineral cerargyrite. Its solubility in water at 25° C is 1.91 mg. per litre. It is, howe… much more soluble in concentrated hydrochloric acid and the … tion is believed to contain the acid $HAgCl_2$. Silver chlorid… also dissolved by solutions of soluble chlorides, the solubility … greater the greater the concentration of the dissolved salt. It … changed to silver sulfate by boiling with concentrated sulf… acid and dissolves readily in dilute aqueous ammonia. The … tion, which contains the complex cation $[Ag(NH_3)_2]^+$ yields … ver chloride again on acidification with nitric acid. Silver chlor… is soluble in alkali cyanide solutions and the solutions contain … complex anion $[Ag(CN)_2]^-$. On the addition of sodium t… sulfate solution to silver chloride, a solution is obtained wh… contains the complex anion $[Ag(S_2O_3)_2]^{3-}$. Silver bromide … iodide dissolve similarly and the formation of this complex … reaction which occurs in the fixing of the emulsion of a ph… graphic plate or film after first developing and washing it. … fixing removes all unchanged silver halide and so makes … emulsion insensitive to light. Silver chloride forms the comp… $AgCl3NH_3$ when it is treated with gaseous ammonia below … All the ammonia is driven off at 65° C but there are several … ammoniates which are stable in the intermediate tempera… range.

Silver bromide is found in nature as the mineral bromyrite (… bromargyrite). In the laboratory, it is formed by the action … bromine on silver or, as a curdy pale yellow precipitate, … hydrobromic acid or a soluble bromide is added to a solution … silver salt. It melts at 434° C. The solubility in water is … than that of silver chloride (0.11 mg. per litre at 21° C). … only sparingly soluble in dilute ammonia solution, the solu… being less than that of silver chloride but greater than that … iodide. This difference may be used as a means of distinc… between the three halides. The reactions of silver bromide … aqueous solutions of alkali cyanide and thiosulfate are simil… those of silver chloride. Ammoniates are also formed with … ammonia.

Silver iodide is found native in the mineral iodyrite (iod… rite). It is prepared by methods exactly analogous to those … silver chloride and bromide, and is a yellow solid, which me… 552° C. The solubility in water at 25° C is only 0.002… mg … litre. The salt is soluble in a strong solution of potassium … Other reactions are very similar to those of silver chloride … bromide.

When silver chloride, bromide, or iodide is exposed to lig… colour changes from white or yellow to pink, violet, and fin… black. In this process halogen is lost and, with prolonged … tense irradiation, weight losses of up to about 10% may be … corded. It is not yet certain what is formed but it is thou… be either silver or a silver subhalide or both. When the … halides are used in photography very much smaller exposure… those referred to above are involved; indeed they are such … no visible change occurs in the halide grains. Only when

[illegible] is subsequently developed, *i.e.*, submitted to the action [illegible] chemical reducing agent, are those grains which have been [illegible] to light blackened because of the production of metallic [illegible] (*see* PHOTOGRAPHY).

[illegible] *Compounds.* Silver nitrate ($AgNO_3$) is the chief silver compound which is of technical importance. It is made in large [illegible] by dissolving silver in nitric acid of density 1.25–1.30. [illegible] in transparent plates which melt at 212° C and can [illegible] cast. The solubility at 20° C is 222 g. per 100 g. of water. [illegible] moderately soluble in methyl and ethyl alcohols and, to a [illegible] extent in various other organic solvents. When heated at about 320° C silver nitrate loses oxygen and forms silver nitrite. At a red heat silver is formed.

The chief use of silver nitrate is as an intermediate in the preparation of silver halides for incorporation in photographic [illegible]. When used for this purpose the technical salt must first be purified. In analytical chemistry aqueous solutions of [illegible] nitrate are used in the volumetric determination of halides, [illegible] and thiocyanates, as well as for the detection of reducing [illegible] and of the cations of various acids which form insoluble silver salts. Silver nitrate has a causticizing action on the skin and produces a blackening because of the formation of silver. Cast sticks of silver nitrate are used in pharmacy under the name lunar caustic. A similar blackening is produced when the salt is in contact with other types of organic matter and, in aqueous solution, it has been used for marking linen. The free nitric acid tends, however, to destroy the fabric and it is preferable to use the silver salts of weaker organic acids.

Silver sulfide (Ag_2S) is found native in the mineral silver glance or argentite. It is also found associated with the sulfides of other metals (*e.g.*, with lead sulfide). Although it is not one of the thermodynamically stable sulfides, Ag_2S may readily be formed by the direct union of the elements or by the action of hydrogen sulfide on a solution of a silver salt. It is a black-brown [illegible] which is insoluble in cold dilute mineral acids, though it dis[illegible] in hot dilute nitric acid and also in alkali cyanide solutions. It melts without decomposition at 842° C.

Silver sulfate (Ag_2SO_4) is produced when the metal is dis[illegible] in hot sulfuric acid, or by adding dilute sulfuric acid to a fairly strong solution of silver nitrate. It forms white rhombic prisms which are isomorphous with anhydrous sodium sulfate and are only sparingly soluble in water. The salt dissolves in [illegible] ammonia, the solution yielding the complex salt [illegible]SO_4.

[illegible] selenide (Ag_2Se) occurs as the mineral naumannite. It is [illegible] heating silver powder with selenium. It is a black crystal[illegible] material and a semiconductor. Silver telluride (Ag_2Te) resembles the selenide and is prepared by a similar method. Both [illegible] and tellurium may occur with silver minerals and if they [illegible] their way into the metal as trace impurities, they have a [illegible] effect on its mechanical properties as well as on its use[illegible] as a catalyst.

[illegible] cyanide is obtained as a curdy white precipitate when [illegible] cyanide is added to a silver salt solution. It is soluble [illegible] of cyanide solution and forms the complex anion [illegible]. This is important in electroplating, a process in [illegible] the object to be plated is made the cathode, the anode being [illegible] of pure silver. The electrolyte is, as a result, a solution [illegible] the complex $K[Ag(CN)_2]$, the anion of which is dis[illegible]ted to a slight extent into silver cations and cyanide anions. On electrolysis the former are deposited on the cathode to be [illegible] the latter react with the anode, causing fresh silver [illegible] into solution as the complex and so replenish the bath. [illegible] thiocyanate ($AgCNS$) is a very insoluble salt which is also [illegible] in nitric acid. It is precipitated by the addition of am[illegible] thiocyanate solution to a silver nitrate solution. Silver [illegible] (AgN_3) is formed as a white precipitate when sodium azide [illegible] added to silver nitrate solution. It resembles silver chloride [illegible] many of its reactions, but is very explosive in the dry state and [illegible] been used as a detonator.

[illegible] addition to the silver salts already mentioned there are nu[illegible] others, among which are the phosphates, arsenates, and chromates, which are insoluble and are used in testing qualitatively for the presence of these acid radicals in solution. The silver salts of many organic acids have been prepared and are of interest in that they yield silver when they are ignited in air and so may be used in determining the equivalent weight of the acid.

Silver also yields a large number of co-ordination compounds. A few of the more important, such as the complex cyanide and thiosulfate, have already been mentioned. In certain of these compounds the stability of the cation is increased by co-ordination. Thus for example co-ordination of the silver ion with ethylene thiocarbamide (etu) yields the complex chloride and bromide $[Ag(etu)_3]Cl$ and $[Ag(etu)_3]Br$, neither of which is darkened by light.

Compounds of Bivalent Silver.—One of the few simple compounds of bivalent silver is argentic fluoride (AgF_2), which is a dark brown solid formed when fluorine reacts with argentous flouride. It melts at 690° C and is a powerful oxidizing agent and also a good fluorinating agent. The other halides of bivalent silver are unknown. Anodic oxidation of silver or, alternatively, oxidation of an aqueous solution of silver nitrate with a persulfate gives a black compound of the empirical formula AgO which is very probably an oxide of bivalent silver.

There is some indication that anodic oxidation of silver may give a higher oxide, possibly Ag_2O_3. The stability of the argentic ion is much increased by coordination with various ligands. Thus, for example, although argentic persulfate is unknown, coordination of the metal with pyridine enables the relatively stable compound $[Ag\ (pyr)_4]S_2O_8$ to be prepared. It is isomorphous with the corresponding cupric compound. Other coordinating molecules which give products of a similar nature are *o*-phenanthroline and dipyridyl.

Medical Aspects.—Much of the silver which is used in medicine is in the colloidal form in association with proteins. The chief uses of silver compounds are as antiseptics, astringents, and caustics. Silver ions in water have a germicidal action, even at very low concentration, and this has been utilized to a limited extent in water sterilization.

Applied externally, silver nitrate has a limited caustic action, destroying the superficial tissues and separating the part acted on as a slough. It may be employed to destroy warts. In granular lids and various forms of ophthalmia, solutions of silver nitrate are employed.

The nitrate is both astringent and stimulating as well as bactericidal, and solutions of it have been used to paint indolent ulcers and in chronic pharyngitis or laryngitis.

External or internal medicinal administration of silver, except in large doses, causes no harmful systemic effects but continued exposure may produce a chronic form of silver poisoning which is known as argyria, and in cases of which silver is deposited in the tissues. The most marked symptom is the dark slate-blue colour of the lips, cheeks, gums, and later, of the skin.

In large doses silver nitrate is a poison, causing violent abdominal pains, vomiting and diarrhea, with the development of gastroenteritis. The treatment consists in the use of solutions of common salt, followed by copious drafts of milk or white of egg and water, and soap in water, in order to dilute the poison and thus protect the mucous membranes of the esophagus and stomach.

(H. J. Es.)

PRODUCTION

Mexico has been the leading producer of silver in the world since about 1900, followed by the United States, Canada, and Peru; these countries generally account for more than [illegible] of the world total. The U.S.S.R. and Australia also are major producers. Other producers (averaging more than 1,000,000 oz. a year) include, in Central America, Honduras; in South America, Bolivia, Chile, and Argentina; in Europe, Germany (East and West), Spain, Yugoslavia, Sweden and Czechoslovakia; in Asia, Japan and Burma; and in Africa, Republic of South Africa and Republic of the Congo.

See also references under "Silver" in the Index.

(S. L. S.; X.)

BIBLIOGRAPHY.—J. Newton, *An Introduction to Metallurgy*, 2nd ed. (1948); J. A. Babor and A. Lehrman, *General College Chemistry*, 3rd ed. (1951); L. Addicks (ed.), *Silver in Industry* (1940); American Society for Metals, *Metals Handbook*, 8th ed. (1961); *Thorpe's Dictionary of Applied Chemistry*, vol. x (1950); J. R. Partington, *General and Inorganic Chemistry*, 3rd ed. (1958).

SILVER AND GOLD WORK includes personal ornaments, utensils, vases, decorative objects, and the like, made of silver or gold, with their various alloys. This article is a historical survey and is divided into the following main sections:

For a more detailed discussion of personal ornaments of silver and gold (and other materials) *see* JEWELRY. *See* also METALWORK, DECORATIVE, especially for metalworking processes.

I. THE ANCIENT WORLD

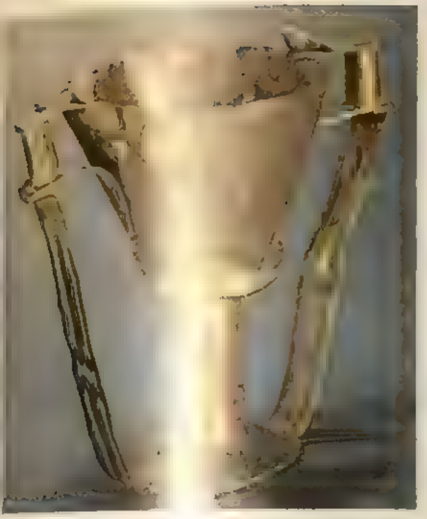
1. Pre-Mycenaean.—Gold and silver and their natural or artificial mixture, called electrum or white gold, were worked in ancient Greece and Italy for personal ornaments, for vessels, arrows and weapons, for coinage, and for inlaid and plated decoration of baser metals. Pliny notes that gold is generally found mixed with silver, and says that when the proportion of silver amounted to one-fifth the alloy was called electrum. The source of native electrum was the River Pactolus in Lydia.

Aegean lands were rich in precious metals. The considerable deposits of treasure found in the earliest prehistoric strata on the site of Troy are generally assigned to the second city; they are earlier than the sixth ("Homeric") city, and are not likely to be later than 2000 B.C. The largest of them, the so-called Treasure of Priam, is a representative collection of jewels and plate. The gold ornaments were packed in a large silver cup. They consist of elaborate diadems or pectorals, six bracelets, 60 earrings or hair-rings, and nearly 9,000 beads. The Trojan vases have bold and simple forms, mostly without ornament, but some are lightly fluted. Many are wrought from single sheets of metal. The characteristic handle is a heavy rolled loop soldered or riveted to the body. Some silver flasks with inverted cup-covers have small shoulder-studs pierced vertically for hanging. Bases are sometimes round or pointed, sometimes fitted with separate collars, but more often slightly cupped to make a low ring-foot. An odd shape in gold is an oval bowl or cup with a broad lip at each end and two large roll-handles in the middle. The oval body has Sumerian affinities, and it seems likely that Trojan arts at this time were Asiatic rather than European. Asiatic influence had indeed invaded Europe, for the oval shape occurs in the contemporary pottery of the Greek mainland and islands (Helladic and Cycladic). A plain spouted bowl of usual early Helladic shape in the Louvre is the typical specimen of goldsmith's work from pre-Mycenaean Greece, and the scarcity of precious metals points to lack of wealth as prime cause of the artistic backwardness of these regions. Silver seems to have been more plentiful in the Cyclades, but only a few simple vessels, headbands, pins and rings survive. Conditions were different in Crete.

2. Minoan and Mycenaean.—A profusion of gold jewelry was found in early Minoan burials at Mochlos, three silver dagger-blades in a communal tomb at Kumasa; silver seals and ornaments of the same age are not uncommon. An elegant silver cup from Gournia belongs to the next epoch (Middle Minoan I, c. 2000 B.C.); it is unique, but numerous imitations of its cusped and carinated form in clay, and of its metallic sheen in glazed and painted decoration, prove that such vessels were common. Minoan plate and jewelry are amply represented in the wealth of the mainland tombs at Mycenae and Vaphio. The vases from Mycenae are made indifferently of silver, gold and bronze; but gold is generally reserved for drinking-cups, small phials and boxes; silver is used for jugs as well. Much of the funeral furniture is gold, notably the masks that hid the faces or adorned the coffins of the dead. It has been thought that the small gold discs, which Heinrich Schliemann found in prodigious quantities (700 in one grave), were nailed on wooden coffins, but they may have been sewn on clothes. They are impressed with geometrical designs based on circular and spiral figures, stars and rosettes and natural forms such as leaves, butterflies and octopods. Smaller bossed discs bearing similar patterns may be button-covers. Models of shrines and other amulets are also made of gold. A splendid piece of plate is a silver counterpart of the black steatite libation-vase from Knossos in the form of a bull's head, with gold horns, a gold rosette on the forehead, gold-plated muzzle, ears and eyes. The gold here and in other Mycenaean plating is not laid on the silver, but on inserted copper strips. The gold cups from Mycenae are of two main types: plain curved or carinated forms related to the silverware and pottery of Troy, and embossed conical vessels of the Minoan tradition.[a] Some of the plain pieces have handles ending in animals' heads, which bite the rim or peer into the cup. The embossed ornament consists of vertical and horizontal bands of rosettes and spiral coils, floral, foliate, marine and animal figures. The designs are beaten through the walls and are consequently visible on the insides of the vessels; but the finest examples of their class, the two gold cups from the Vaphio tomb near Sparta, have a plain gold lining which overlaps the embossed sides at the lip. The reliefs on the Vaphio cups represent men handling wild and domesticated cattle among trees in a rocky landscape. The handles show the typical Minoan form: two horizontal plates riveted to the body at one end and joined at the other by a vertical cylinder. Steatite vases carved with similar pictorial reliefs were evidently made to imitate embossed gold. A fragment found at Palaikastro had part of its original plating attached.

Cretan and mainland tombs have produced many examples of weapons adorned with gold. Modest ornaments are gold caps on the rivets joining hilt and blade, but the whole hilt is often cased in gold. An example from Mycenae has a cylindrical grip of openwork gold flowers with lapis-lazuli in their petals and crystal filling between them; the guard is formed by similarly inlaid dragons. The most splendid Mycenaean blades are bronze inlaid with gold, electrum, silver and niello. Here again the work is done on inserted copper plates. This kind of flat inlay seems to have been originally Egyptian; it occurs on daggers from the tomb of Queen Aah-Hotep, which are contemporary with the Mycenaean (*c.* 1600 B.C.), and it is significant that two of the Mycenae designs have Egyptian subjects, though their style is purely Minoan. These are the scenes of cats hunting ducks among papyrus-clumps beside a river in which fish are swimming. Another blade bears Minoan warriors fighting lions, and lions chasing deer. A dagger from Thera has inlaid axe-heads; one from Argos, dolphins; and fragments from the Vaphio tomb show men swimming among flying fish. These are masterpieces of Minoan craftsmanship; in the long decadence of the Mycenaean age there seems to have been no invention, and later pieces of goldsmith's work repeat conventional forms and ornaments.

3. Greek and Etruscan.—The period of transition from the Bronze to the Iron Age, when Aegean external relations were violently interrupted, was not favourable either to wealth or art, and



Mycenaean gold cup, the so-called Cup of Nestor, from the royal graves at Mycenae, Greece; 1600–1500 B.C. National museum, Athens

Etruscan gold pendant in the form of a mask representing the river-god Achelous; about 500 B.C. The Louvre, Paris

Greek silver disk, orig[illegible] in-side lid of a jewel box [illegible] tomb at Canosa di Puglia, It[illegible]ating from the 2nd century B.C. Mu[illegible]ionale, Taranto, Italy

Gold mask of Pharoah Tutankhamen inlaid with lapis lazuli and coloured glass; 14th century B.C. Cairo museum

Roman silver pitcher with relief decoration from the villa at Boscoreale near Pompeii, Italy; about 1st century B.C. The Louvre, Paris

ANCIENT SILVER AND GOLD WORK

SILVER AND GOLD WORK

Altar cross of Hugo d'Oignies made of silver-gilt stamped openwork, set with amethysts and carnelians and decorated with miniature paintings. French, 13th century. Victoria and Albert museum

Silver-gilt reliquary with three towers. German Gothic. From the Aachen cathedral, c. 1375

The royal gold cup of the kings of France and England. Made about 1380 in Paris for the duc de Berry (Jean de France). The British museum

Gold reliquary of the Holy Thorn, enameled and set with jewels, representing the theme of the Last Judgment. Made for the duc de Berry, c. 1400. The British museum

EUROPEAN MEDIEVAL SILVER AND GOLD WORK

BY COURTESY OF (TOP LEFT) THE VICTORIA AND ALBERT MUSEUM, (TOP RIGHT) FOTO ANN BREDOL—LEPPER, AACHEN, (BOTTOM LEFT, BOTTOM RIGHT) THE TRUSTEES OF THE BRITISH MUSEUM

Gold and enamel cup, known as the Rospigliosi cup, in the mannered style of the late Italian Renaissance. Height: 7¾ in. The Metropolitan Museum of Art, New York

Gold and enamel saltcellar, wrought by Benvenuto Cellini (1500–71) for Francis I of France, who was his patron during his exile in that country. Overall height: 10¼ in. The Kunsthistorisches museum, Vienna

ITALIAN RENAISSANCE GOLD WORK

BY COURTESY OF (TOP) THE METROPOLITAN MUSEUM OF ART, BEQUEST OF BENJAMIN ALTMAN, 1913, (BOTTOM) PHOTOGRAPHISCHES ATELIER DER KUNSTHISTORISCHEN SAMMLUNGEN, VIENNA

SILVER AND GOLD WORK

Silver-gilt cup with rock crystal bowl and stem. Designed by Sir Martin Bowes; London, 1554

Silver-gilt ice pail in the classical style. Designed by John Flaxman and made by Paul Storr; London, 1812

Steeple cup of silver gilt. London, dated 1627 Victoria and Albert museum

Standing silver-gilt saltcellar set with panels of painted glass. London, 1592

Flemish silver-gilt ewer made to commemorate the expedition of Emperor Charles V against Tunis in 1535. Antwerp, 1558. The Louvre

EUROPEAN SILVER: 16TH, 17TH AND 19TH CENTURIES

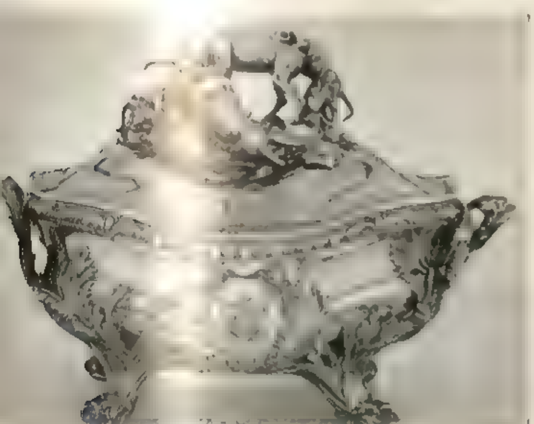
Silver centrepiece illustrating the duties of the Goldsmiths' company. Made by Hunt and Roskell; London, 1854

One of a pair of silver fruit dishes, heavily embossed in the *art nouveau* style; English, 1897. Designer and maker, Gilbert Marks

American colonial silver cup by Jacob Hurd (1702–58), bearing cipher of the original owner. The Metropolitan Museum of Art

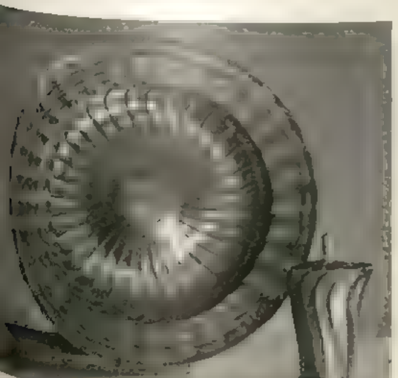
Gold and jasper ewer made by the Dutch goldsmith, Paul van Vianen, for the emperor Rudolph II in 1608. The Kunsthistorisches museum, Vienna

American colonial silver tea set made by Paul Revere (1735–1818). The Metropolitan Museum of Art, New York city



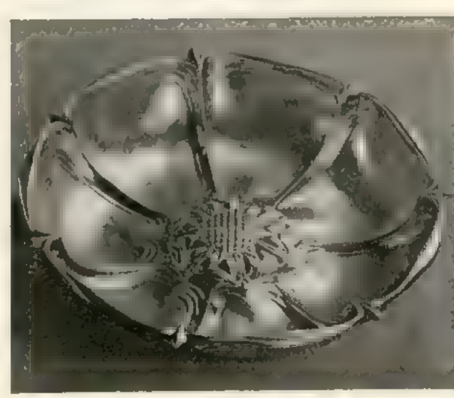
Silver tureen, bearing the arms of the duke of Orleans, by Edmé Pierre Balzac, Paris; French, 1757. The Metropolitan Museum of Art, New York city

Silver tea set from Trinity college, Cambridge; English, 1931. Designer and maker, Harold Stabler



Fluted silver rosewater dish and beaker presented to Corpus Christi college, Cambridge; English, 1938. Designer, R. M. Y. Gleadowe; makers, Wakely and Wheeler, Ltd., London

American colonial silver grace cup by John Coney (1655–1722) of Boston, gift to Harvard college in 1701. The Museum of Fine Arts, Boston



Silver bowl in the shape of a Tudor rose; English, 1926. Designer and maker, Omar Ramsden

EUROPEAN, ENGLISH AND U.S. SILVER AND GOLD: 17TH TO EARLY 20TH CENTURY

BY COURTESY OF (TOP ROW CENTRE, SECOND ROW LEFT, THIRD ROW LEFT) THE METROPOLITAN MUSEUM OF ART, (TOP ROW CENTRE) MORRIS K. JESSUP FUND, 1952, (SECOND ROW LEFT) BEQUEST OF JUDGE ALPHONSO T. CLEARWATER, 1933, (THIRD ROW LEFT) BEQUEST OF CATHERINE D. WENTWORTH, 1948, (TOP ROW RIGHT) KUNSTHISTORISCHES MUSEUM, VIENNA, (BOTTOM ROW CENTRE) HARVARD UNIVERSITY; PHOTOGRAPHS, (OTHERS) THE WORSHIPFUL COMPANY OF GOLDSMITHS, LONDON

SILVER AND GOLD WORK

Mexican half gold, half silver pectoral, found at Teotitlán del Camino, Oaxaca; pre-Columbian. Height: just under 1¾ in. National museum, Mexico City

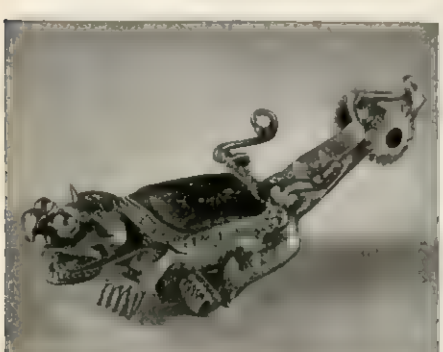
Mexican silver ring found in Tomb 7 at Monte Albán, Oaxaca; pre-Columbian. Maximum height: 1½ in., diameter: ⅝ in. Regional museum, Oaxaca

Calima sheet gold headdress from the San Juan valley, Colombia; pre-Columbian. Maximum height: 11¼ in. Gold Museum of the Bank of the Republic, Bogotá

Peruvian silver effigy beaker raised from a flat sheet of metal; pre-Columbian. Height: 4¾ in., diameter at base: 2⅞ in.

Mexican hollow gold casting from Achiutla, Oaxaca; pre-Columbian. Dimensions: ⅞ in. by ¾ in. American Museum of Natural History

Quimbaya ceremonial gold flask found in Cauca valley, Colombia; pre-Columbian. Gold Museum of the Bank of the Republic, Bogotá



Panamanian cast gold ornament in the form of a mythological animal, inset with a large emerald, found at Coclé; pre-Columbian. Maximum over-all length: 4⅝ in. University of Pennsylvania museum, Philadelphia

Navaho silver necklace with squash blossoms; after 1880. Museum of the American Indian, New York city

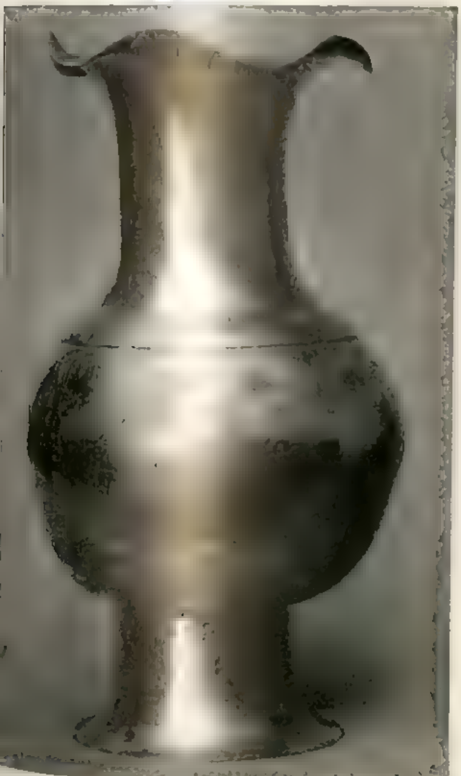
Mexican cast gold war shield with turquoise mosaic inlay, found at Yanhuitlan, Oaxaca; pre-Columbian. Diameter of shield: just over 2⅛ in., length of arrows: just under 3¼ in. National museum, Mexico City

NORTH AND SOUTH AMERICAN SILVER AND GOLD WORK

BY COURTESY OF (TOP LEFT, BOTTOM RIGHT) NATIONAL MUSEUM, MEXICO CITY, (TOP CENTRE) REGIONAL MUSEUM, OAXACA, (TOP RIGHT, SECOND ROW RIGHT) GOLD MUSEUM OF THE BANK OF THE REPUBLIC, BOGOTÁ, (SECOND ROW CENTRE) AMERICAN MUSEUM OF NATURAL HISTORY, (BOTTOM LEFT) UNIVERSITY MUSEUM, PHILADELPHIA, (BOTTOM CENTRE) MUSEUM OF THE AMERICAN INDIAN; PHOTOGRAPHS, (TOP LEFT, TOP CENTRE, SECOND ROW LEFT, SECOND ROW CENTRE) D. T. EASBY, JR., (TOP RIGHT, CENTRE RIGHT) NATIONAL GALLERY OF ART, WASHINGTON, D.C., (BOTTOM RIGHT) MEXICAN NATIONAL INSTITUTE OF ANTHROPOLOGY AND HISTORY

Ancient Persian gold boss. Achaemenid period, 5th–4th centuries B.C. From the Oxus treasure in the British museum

Chinese silver parcel-gilt dish with *repoussé* ornament and engraved with the story of Chiang Tzu-Ya. From Pei Huang Shan; T'ang dynasty, A.D. 618–906. The British museum



Chinese silver-gilt vase engraved with 12 creatures of the zodiac. From Pei Huang Shan; T'ang dynasty, A.D. 618–906. The British museum

Indian silver dish, embossed and chased, showing a Yaksah, perhaps King Kuvera, drinking; Kushan, 3rd or 4th century A.D. Found at Buddaghara, near Tank, Dera Ismail Khan district, West Pakistan. The British museum

Ancient Persian gold vase in the form of a fish; Achaemenid period, 5th–4th centuries B.C. From the Oxus treasure in the British museum

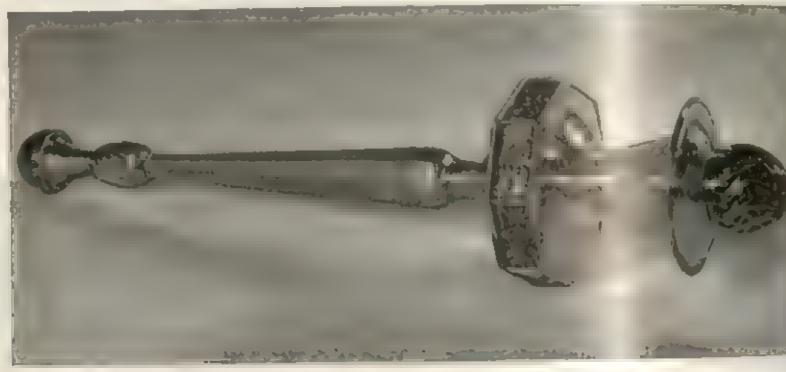
Persian silver bowl showing a king, perhaps Firuz I, slaying lions; Sassanian or neo-Persian empire, c. A.D. 224–651. The British museum

MIDDLE- AND FAR-EASTERN SILVER AND GOLD WORK

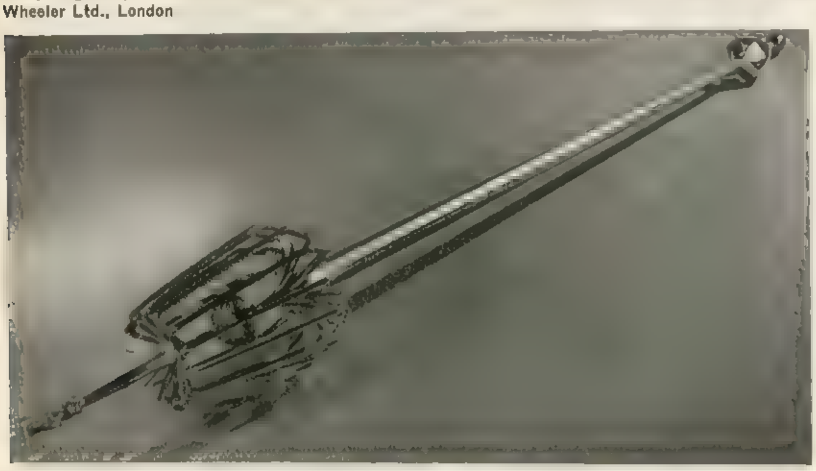
SILVER AND GOLD WORK

Loving cup presented to the City of London by the Worshipful Company of Goldsmiths; 1963. Designer, Atholl Hill; engraver, T. C. F. Wise; makers, Wakely and Wheeler Ltd., London

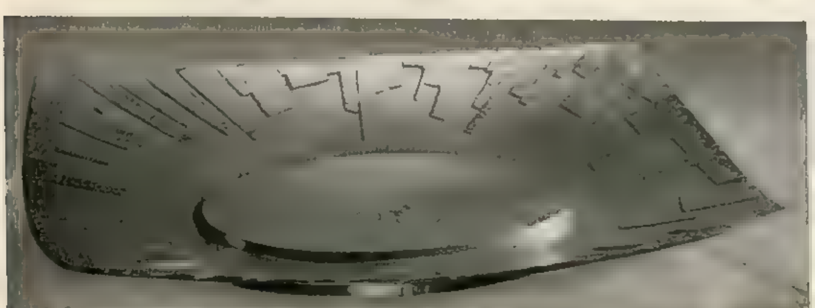
Salt and pepper pots from Corpus Christi college, Oxford, England; 1946. Designer and maker, Leslie Durbin. Given by Sir Alan Barlow



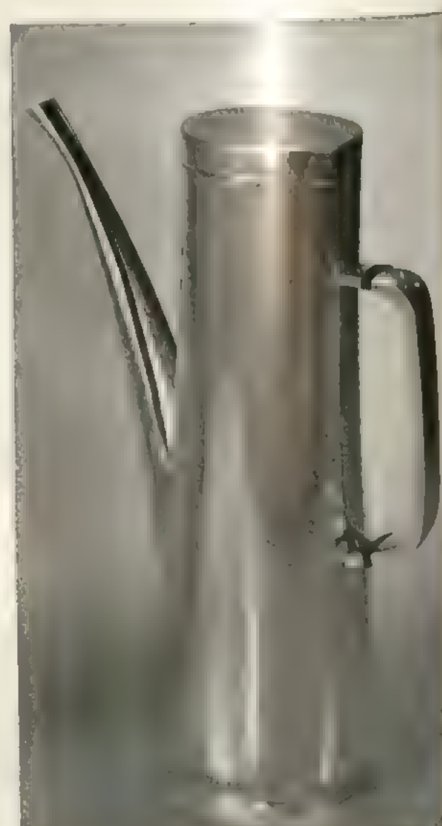
Torch commemorating the XIV Olympiad held in London in 1948. Designer, Bernard Cuzner; makers, Bernard Cuzner and Stanley G. Morris; engraver, William Biddle. Collection: Goldsmiths' hall, London



Silver and gilt mace made for Cornell university, Ithaca, N.Y.; 1962. Designer, Eric Clements; makers, Wakely and Wheeler Ltd., London



Dish presented to Lincoln hall, University of Nottingham, Eng., by Sir Francis Hill; 1962. Designer and maker, Desmond Clen-Murphy



Large coffee pot; 1962. Designer, Keith Redfern; maker, Tom Boucher. Collection: Goldsmiths' hall, London

20TH-CENTURY SILVER WORK

PHOTOGRAPHS, THE WORSHIPFUL COMPANY OF GOLDSMITHS, LONDON

the only considerable pieces of plate that have come from Greece are the embossed and engraved silver bowls made by Phoenicians. Most of them bear elaborate pictorial designs of Egyptian or Assyrian character, and are evidently foreign to Greece; but some simpler types, decorated with rows of animals in relief or wrought in the shape of conventional flower bowls, can hardly be distinguished from the first Hellenic products. Early Greek work is rare. A severe and elegant silver bowl in the Metropolitan Museum, New York, represents the flower type in its finest style. It is cast and chased, and probably belongs to the 5th century B.C. A gold treasure from Panagyurishte in Bulgaria gives a more complete idea of Greek goldsmiths work.

Silver vases and toilet instruments have been found beside the commoner bronze in Etruscan tombs. A chased powder-box of the 4th century is in the Metropolitan Museum, New York. The bronze reliefs of the archaic chariot in the same collection have their opulent counterparts in some hammered silver and electrum fragments in London, Munich and Perugia. The electrum details are attached with rivets.

4. **Roman.**—About the 4th century B.C. there was revived the fashion of ornamenting silver vessels with relief, and this type of work, elaborated in the Hellenistic age and particularly at Antioch and Alexandria, remained the usual mode of decoration until the end of the Roman Empire. Various fabrics of molded pottery correspond to the successive styles of metalware. A silver vase in the British Museum, bearing a frieze of chariots between floral bands, is nearly a reflex of an earthenware Calene bowl (3rd century B.C.) in the same collection. Pliny names Greek silversmiths whose work was valued highly at Rome, and laments the disappearance of the art in his own day. He must refer only to its quality, for Roman silverware has been abundantly preserved. Many rich hoards in modern collections were buried by design during the calamitous last centuries of the ancient world, and the most sumptuous, the Boscoreale treasure, was accidentally saved by the same volcanic catastrophe that destroyed Herculaneum and killed Pliny. This treasure is mostly in the Louvre. A slightly smaller hoard found at Hildesheim, and now in Berlin, also belongs to the early Empire. The acquisition and appreciation of silver plate was a sort of cult at Rome. Technical names for various kinds of reliefs were in common use (emblemata, sigilla, crustae), weights were recorded and compared and ostentatiously exaggerated. Large quantities of bullion came to Rome with the spoils of Greece and Asia in the 2nd century B.C., and Pliny says that even in republican times there were more than 150 silver dishes in the city of a hundredweight apiece. Weights of vessels are often marked on their bases.

Cups and jugs of Augustan style are usually covered with ornament in high relief. The subjects are very diverse: historical, mythological and mystic scenes, formal and naturalistic designs of flowers and foliage, graceful studies of animals and birds. Others have conventional fluting, petals or gadroons, Bacchic instruments and masks, embossed or engraved wreaths, gilt or inlaid with niello. Silver and niello inlay was commonly applied to bronze plates. A singular type of silver bowl (patera clipeata) has a central ornament in high relief or even in the round: portrait busts are not uncommon in this place. In course of time the ornament was restricted, and later Roman plate is largely plain with narrow border-friezes, small central medallions, and handles embossed in low relief. One of the very few gold pieces that survive, a shallow bowl found at Rennes and now in the Bibliothèque Nationale of Paris, is exceedingly elaborate. It measures 25 centimetres across and weighs 1,315 grams. The central medallion and its surrounding frieze contain scenes of a bibulous contest between Bacchus and Hercules; between these and the edge is a row of 16 gold coins each framed in a foliate wreath. The coins range from Hadrian to Caracalla. In the same collection are several examples of very large silver plates (clipei or *missoria*), in which the whole field is embossed with mythological or historical subjects. The largest (called the Shield of Scipio) is 72 centimetres in diameter and weighs 10,300 grams. Another bears the name of Gelimer, king of the Vandals and Alans (6th century). The "Shield of Theodosius" at Madrid shows the emperor seated between Valentinian and Arcadius among his guards, with an allegorical group in the exergue. (E. J. F.; M. C. Rs.)

II. EUROPE

A. Early Christian and Medieval

1. **Early Christian Period.**—*Silver.*—The earliest Christian silver work closely resembles the pagan work of the period in its naturalistic grace, ornament, and use of the traditional techniques of embossing and chasing. Even the subject matter is sometimes classical: the late-4th-century casket of Projecta, part of the Esquiline treasure found at Rome and now in the British Museum, London, is decorated with pagan scenes, and only the inscription shows that it was made for a Christian marriage. Among the few pieces with Christian subjects are the small Roman cruets from Taprain, Scotland (Royal Scottish Museum, Edinburgh, and in the British Museum) and the small pyx (casket for the reserved sacrament) from Pula, Yugos. (Kunsthistorisches Museum, Vienna).

Most of the silver of the latter part of the period has been found in the Christian east, in Syria, Egypt, Cyprus, Asia Minor, and Russia and is mostly "church" plate (chalices, censers, candlesticks, and bowls and dishes probably used as patens). Secular plate was also decorated with religious subjects, for example the dishes depicting the life of David (Cyprus Treasure, Nicosia Museum, and Metropolitan Museum, New York); and both dishes and vessels were produced with pagan subjects, like the Concesti Amphora, or the Silenus Dish (both in the Hermitage, Leningrad). The figure style is often harder and flatter than previously, characterized by strictly frontal positions and symmetry. The techniques of chasing and embossing still predominated, but abstract patterns and Christian symbols inlaid in niello (a black silver alloy) were used increasingly. The appearance of imperial "control stamps," early forerunners of hallmarks, show most of this material to be of the 6th and 7th centuries. It is not known which cities were important centres of production, but the Eastern capital, Constantinople, must have been foremost among them.

Gold.—Of work in gold of the earliest Christian period only personal jewelry has survived, but from the 6th and 7th centuries onward other pieces are also extant. Among the most important of the latter are the votive crowns and crosses offered to the churches in Spain and Italy by royal patrons, the finest being those found at Guarrazar (Toledo Province), which are inlaid with garnets and set with jewels (National Archaeological Museum, Madrid, and Musée Cluny, Paris); the cross of King Agilulfo, and a pair of gold book-covers inscribed by Queen Theodolinda (both at Monza Cathedral, Italy) set with pearls, gems and cameos and decorated with gold cloisonné work (compartments divided by narrow strips of metal filled with inlay) inlaid with garnets, a popular style for jewelry among the Germanic peoples. This cloisonné jewelry, inlaid with garnets, paste, or millifiori glass, reached an especially high standard of workmanship in Britain, as is shown by a purse lid, a sword and jewelry from the cenotaph to an East Anglian king (A.D. 650–660) discovered at Sutton Hoo, Suffolk, and now in the British Museum. Major works in silver and gold were also produced in the northern Hiberno-Saxon school and in the service of the Celtic church; work in precious metal, such as the buckle on the Moylough belt-shrine and the Ardagh Chalice (both in the National Museum, Dublin), displays the same masterly synthesis of the northern arts and humanist Mediterranean tradition as is exemplified in the Lindisfarne Gospels (*c.* 700; British Museum).

2. **Carolingian and Ottonian Period.**—The earliest works of the Carolingian renaissance, made in the last quarter of the 8th century, resemble Hiberno-Saxon art of the 8th century in their abstract treatment of the human figure, their animal ornament and their use of niello and "chip-carving" technique; examples are the Tassilo Chalice (Kremsmünster, Austria) and the Lindau book-cover (Pierpont Morgan Library, New York). From about A.D. 800 onward, however, the influence of the Mediterranean tradition gained strength at Charlemagne's court at Aachen and later spread through the whole empire. The triumphal arches (now lost) given by the emperor's biographer Einhard

(*q.v.*) to Maastricht cathedral were typical of this movement; they were miniature versions (22 cm. high) of the great marble triumphal arches of antiquity embossed in silver with Christian subjects. The bulk of work in precious metals which survives from the Middle Ages is ecclesiastical. Golden altars, like that of S. Ambrogio in Milan (*c.* 850), where scenes from the life of Christ and St. Ambrose are framed by panels of cloisonné enamel and filigree; reliquaries and book-covers in gold and silver, set with gems and decorated by embossed figures and scenes—for example, the cover of the Codex of St. Emmeram of Regensburg (*c.* 870; Staatsbibliothek, Munich)—all these testify to the magnificence of Carolingian work whose techniques were to dominate the goldsmith's craft until the 11th century. Patronage throughout this period was mainly in the hands of the emperors and the great princes of the church, and the form of liturgical plate and reliquaries, altar crosses, etc., underwent no fundamental change; Ottonian work of the later 10th and 11th centuries can be distinguished from that of the 9th only in the development of style. The larger, more massive figures on the golden altar (*c.* 1023) given by Henry II to Basel Minster (Musée Cluny), for example, with their strict pattern of folds, are markedly different from the nervous, elongated figures of the Carolingian period.

3. Romanesque Period.—In the 12th century the church supplanted secular rulers as the chief patron of the arts and the work was carried out in the larger monasteries. Under the direction of great churchmen, for example, Henry of Blois, bishop of Winchester, or Abbot Suger of Saint-Denis, near Paris, a new emphasis was given to subject matter and symbolism. The techniques of the time are expounded in a book of instructions, *De diversis artibus,* written by Theophilus, a priest who worked in Germany in the early 12th century. Craftsmen were no longer anonymous; work by Roger of Helmarshausen, Reiner of Huy, Godefroid de Claire, Nicholas of Verdun and others can be identified, and the parts they played as leaders of the great centres of metalwork on the Rhine and the Meuse are clearly recognizable. Their greatest achievement was undoubtedly the development of the brilliant champlevé enameling (in which compartments are hollowed out of the surface of the metal and filled with enamel), a method which replaced the earlier cloisonné technique. Gold and silver continued to be used as rich settings for enamels, and as the framework of portable altars or small devotional diptychs or triptychs, and in reliquary shrines with their embossed figure work, as well as in liturgical plate.

The masterpieces of the period are the great house-shaped shrines made to contain the relics of saints; for example, the shrine of St. Heribert at Deutz (*c.* 1160), or Nicholas of Verdun's shrine of the Three Kings at Cologne (*c.* 1200). In the latter, the figures are almost free-standing, and in their fine rhythmic draperies and naturalistic movement they approach the new Gothic style.

4. Gothic Period.—The growing naturalism of the 13th century is notable in the work of Nicholas' follower, Hugo of Oignies, whose reliquary for the rib of St. Peter at Namur (1228) foreshadows the partly crystal reliquaries in which the free-standing relic is exposed to the view of the faithful; it is decorated with Hugo's particularly fine filigree and enriched by the naturalistic cut-out leaves and little cast animals and birds. The increasing wealth of the royal courts, the aristocracy and, later, the merchants led to the establishment of secular workshops in the great cities and the foundation of the confraternities or guilds of gold- and silversmiths, the first being that of Paris in 1202. As in architecture, monumental sculpture and ivory carving, the lead held by Germany and the Low Countries during the Romanesque period, now passed to France. The use of architectural forms continued to be the basis of design in precious metal; for example, the silver shrine of St. Taurinus at Évreux (*c.* 1250), which is a Gothic chapel in miniature, with saints under pointed arches, clustered columns and small turrets. In England, the few pieces that survived the dissolution of the monasteries in the 16th century follow the same architectural pattern; *e.g.,* the 14th-century Ramsay Censer (Victoria and Albert Museum, London) or the magnificent crozier made for William of Wykeham (1324–1404) at New College, Oxford. Germany first produced work in the Gothic style in the second half of the 14th century with the large Gothic head-reliquary of Charlemagne and the splendid "Three-Tower" reliquary, both still at Aachen. In Italy, in spite of the undercurrent of classical taste, the Gothic style predominated in the 14th century, especially at Siena; also it was probably in Italy *c.* 1280 that *basse-taille* enamel—a technique in which intaglio relief carving in the metal below its surface is filled with translucent enamel—originated, whence it spread rapidly through the upper Rhine region to France and England. The Parisian school of enamelers predominated in the latter half of the 14th century. For the first time enough secular plate survives to show it equaled the ecclesiastical in opulence: two fine pieces are the Royal Gold Cup made in Paris *c.* 1380 (British Museum) and the so-called King John's Cup, probably English work of *c.* 1340 (King's Lynn, Norfolk).

In the late Gothic period, as well as court treasures such as the *Goldene Rössel* (Altötting, Bavaria) or the Thorn reliquary (British Museum), both early 15th century, there was an increased output of secular silver due to the rise of the middle classes; the English mazers (wooden drinking bowls with silver mounts) and the silver spoons with a large variety of finials are examples of this more modest plate. Numerous large reliquaries and altar plate of all kinds were still produced. At the end of the Middle Ages their style and that of secular plate developed more distinctive national characteristics strongly influenced by architectural style: in England by the geometric patterns of the "perpendicular," in Germany by heavy and bizarre themes of almost Baroque exuberance, and in France by the fragile elegance of the "Flamboyant."

The purity standards of silver became rigorously controlled and "hall-marking" was enforced, the marking of silver in England being especially carefully observed. *See* HALLMARK.

(P. E. LA.)

B. THE RENAISSANCE TO MODERN TIMES

1. 16th Century.—The Italian goldsmiths preceded the rest of Europe in reverting to the style of Roman antiquity but, in the absence of antique goldsmiths' work, vases of marble or bronze had to serve as models. Goldsmiths often worked from very free interpretations of the antique made by artists in other media, such as Giulio Romano (d. 1546) and Francesco Salviati (1510–63). Many of these designs but very few of the actual pieces have survived; the most famous is the enameled gold salt cellar (Kunsthistorisches Museum, Vienna) which was made for Francis I by the celebrated Florentine Benvenuto Cellini (*q.v.;* 1500–71). In the second half of the 16th century many gifted Italian and immigrant goldsmiths worked at the court of Cosimo I, grand duke of Tuscany, specializing in vessels of hardstone mounted in enameled and jeweled gold; their work is well represented in the Museo degli Argenti in the Pitti Palace, Florence, and in the Kunsthistorisches Museum, Vienna; similar work was done by the Sarachi family in Milan.

Little French goldwork is extant and most of the surviving material is in the Galerie d'Apollon in the Louvre, Paris. Among the most sumptuous pieces are a sardonyx and gold ewer, the gold St. Michael's Cup, both at Vienna, and a sardonyx-covered cup in the Louvre, which all display northern features. The massive plate of the Order of the Saint-Esprit (Louvre), dating from 1581–82, is of quite individual character and the enameled gold helmet and shield of Charles IX (1550–74) in the Louvre have no parallel for either quality or opulence.

In other parts of Europe the goldsmiths clung to Gothic forms till well into the first half of the century, especially in the provincial towns. Still immensely rich in ecclesiastical silver, Spain has little early domestic silver; Spanish silversmiths (*plateria*) gave their name to the heavily ornamented style of the period, Plateresque. Using precious metal from the New World, goldsmiths such as Juan d'Arfe (1535–1603) produced vast containers for the Host, known as *custodia.* The most important Portuguese work, Gil Vicente's monstrance (1506) for Belém Monastery near Lisbon, is still Gothic in style; later Portugal developed its

own style related to Spanish work but not copied from it.

Some of the finest 16th-century goldsmiths' work was executed in Antwerp and elsewhere by Flemish goldsmiths, such as Hans of Antwerp, goldsmith to Henry VIII, or Jacopo Delfe, called Biliverti, goldsmith to Cosimo I. The Flemish masters showed particular sympathy for the Mannerist style, derived from Italy but transformed by such native engravers as Cornelis Floris and Frans Floris the Elder. By about 1580 Dutch goldsmiths began to rival the Flemish; the van Vianen family of Utrecht won international renown, especially Adam (*c.* 1570–1627), who excelled at embossing, and his brother Paul (*c.* 1555–*c.* 1614), who worked in Italy, Munich and the workshop of Rudolph II at Prague.

The principal centres in the north were Nürnberg and Augsburg, the former particularly notable for the exuberant Mannerism of the Jamnitzer family, the latter for its ebony caskets with silver-gilt mounts. Many German princes, especially the dukes of Bavaria, maintained their own court workshops. Production was on a vast scale and great quantities survive. Characteristic German forms are columbine cups (the trial piece of the Nürnberg Goldsmiths' Guild), pineapple cups and double cups.

England is rich in 16th-century secular silver but church plate was mostly destroyed at the Reformation. The Renaissance style was introduced by Hans Holbein the Younger (*q.v.;* 1498–1543), who designed vessels for the court, and follows that of the Low Countries and Germany. Certain individual forms were also produced, such as standing salts with tiered covers and "steeple" cups which had a tall finial on the cover.

2. Baroque Period.—In the first half of the 17th century Dutch goldsmiths, such as the van Vianens and, later, Johannes Lutma of Amsterdam, developed a fleshy form of ornament, known as auricular, which became common in northern Europe, including England, where Christiaen van Vianen worked as court goldsmith to Charles I, and Germany, where the Thirty Years' War (1618–48) reduced both the quantity and quality of production. After the mid-century bold Dutch floral ornament—usually embossed in thin metal as though the pieces were for display rather than use—was characteristic and influential; France, however, undoubtedly led fashion with its state workshops at the Gobelins, the refined French acanthus ornament contrasting sharply with the coarser Dutch designs. Since Louis XIV melted the royal plate to pay his troops, no French work of this period remains but its quality is demonstrated in the work of the Huguenot silversmiths who left France after the revocation of the Edict of Nantes (1685). Mostly provincials, they brought new standards of taste and craftsmanship wherever they settled, particularly in England where the foremost names of the late 17th and earlier 18th centuries were of French origin: P. Harache, P. Platel, D. Willaume, S. Pantin, P. de Lamerie, P. Crespin, to mention but a few.

Silver furniture, a feature of the state rooms at Versailles, became fashionable among kings and noblemen: it was constructed of silver plates attached to a wooden frame and each suite contained a dressing table, looking glass and a pair of candlestands. In France such furniture did not survive the French Revolution, but much remains in England, Denmark, Germany and Russia.

After the Thirty Years' War Germany did not regain its eminence: even the enameled goldwork from the court workshops at Prague and Munich became larger and more ostentatious in colour but inferior in design and finish. In Scandinavia, particularly Sweden, goldsmiths evolved forms of beakers and tankards showing strong German influence. Spanish silver was of massive architectural design, oval champlevé enameled bosses being set at intervals over the surface of the larger pieces. The few extant Italian pieces suggest that the goldsmiths worked their material with the skill of sculptors.

3. 18th Century.—Early 18th-century English work combined functional simplicity with grace of form while the work of Dutch and German goldsmiths is in a similar style but of less pleasing proportions. The preeminence of the English work is, however, due to the destruction of all but a fraction of French silver of the same period. What survives is outstanding in originality of design and fineness of finish. Little remains of the work of Louis XV's goldsmith, Thomas Germain (1673–1748); however, there are at Lisbon three services and part of a fourth commissioned in 1756 from his son, François Thomas Germain (1726–91); at Leningrad, another service by François Thomas Germain, the famous Orloff service by Jacques Roëttiers (1707–84), and four services ordered from Robert Joseph Auguste (1725–1805) in the 1770s for Catherine II; at the Metropolitan Museum, New York, a fine collection of French provincial silver. The superiority of French work lay in its excellence of design and the high quality of the cast and chased work; where other goldsmiths worked in embossed metal, the French modeled and cast their ornament and then applied it, a technique that consumed much more of the precious material. In France provincial goldsmiths competed successfully with those of the capital, but in England all the best artists went to London. In the early 1730s the French Rococo style was imported to England and adopted by goldsmiths of both Huguenot and English descent, one of the latter being Thomas Heming, goldsmith to George III.

English Adam silver is of unequal merit owing to the introduction of industrial methods by some large-scale producers. In France Robert Auguste created pieces of great refinement in the Neoclassical style which was copied in Turin and in Rome, for example by L. Valadini. A notable workshop was founded in Madrid in 1778 by D. Antonio Martinez, who favoured severely Classical designs. In both the northern and southern Netherlands local production followed French precept, but more individuality survived in Germany. In Augsburg excellent table silver was produced and, more important, pictorial panels embossed in the highest relief by members of the Thelot family and silver furniture made by the Billers and the Drentwetts. At Dresden Augustus the Strong established under Johann Melchior Dinglinger (1664–1731) a court workshop that produced jewels and enameled goldwork unequaled since the Renaissance, and the gold snuff boxes made by Johann Christian Neuber (1735–1808) rivaled those of the Parisian goldsmiths.

4. 19th Century.—The Napoleonic adventure brought French fashions back into prominence and the Empire style developed by J. B. Odiot and M. G. Biennais was widely followed on the continent. In England the Regency goldsmiths, of whom Paul Storr was the foremost, created their own more robust version of the Empire style. Perhaps the most impressive monument of this period is the service made in Lisbon between 1813 and 1816 and presented to the duke of Wellington for his liberation of Portugal (now in Apsley House, London).

By mid-century most of the earlier styles had been revived fleetingly and a recognizably "Victorian" style evolved, based on details drawn from diverse sources. Craftsmanship was at its best, but the design of domestic silver was trivial and that of presentation pieces strove too consciously for naturalistic effect. In the latter half-century the craft became an industry and the goldsmith a factory worker. In this respect Matthew Boulton (*q.v.*) was the great pioneer: his Soho manufactory near Birmingham, which dominated the British "toy" industry from the 1770s, produced high quality steel buckles, buttons, coins, sterling silver, Sheffield plate and steam engines, establishing standards of design, of factory management and welfare services which even in the mid-20th century are seldom surpassed. However, during the 19th century these standards deteriorated and at the close of the century a second pioneering movement started—the craft revival associated with William Morris (*q.v.*) and the *art nouveau* (*q.v.*) style which led to the production of original pieces, some of doubtful merit. In England the most interesting work has been done by the sculptor Sir Alfred Gilbert (*q.v.*) who, following the lead of William Burges (1827–81), the architect and designer, combined silver with ivory and semiprecious stone in romantic confections. *See* also SHEFFIELD PLATE; ENAMEL. (J. F. H.)

C. THE 20TH CENTURY

The structure of trade, following the drastic social changes which have taken place since 1914, is similar in all industrial countries. First, a few artist-craftsmen maintain independent studio workshops, producing commercially unprofitable but artistically significant work. Many of them also teach in art schools or work

part-time in factories as industrial designers. Second, factories using modern equipment—*e.g.*, stamping, pressing, spinning, casting, vacuum evaporation and mechanical polishing—account for nearly all the financial turnover, but seldom break new ground artistically. The retail shops buy stock almost entirely from factories and wholesalers and usually sell it anonymously. The evolution of style is thus impeded by the cost of new machinery; by the natural caution of wholesalers and retailers; by the buying public which prefers precious ornaments to be timeless; and by the consideration that buying is an investment for value, rather than for beauty. In consequence the most lively designs are often those for costume jewelry, and the best modern work usually has been on a tiny scale, making little impact on the trade.

In Paris, designs by René Lalique (1860–1945) inspired the *art nouveau,* which spread to Belgium and then through Europe and the United States. In Moscow, Peter Carl Fabergé (1846–1920) set a superb standard of craftsmanship for small ornaments. In Denmark, Georg Jensen (1866–1935) with Johan Rohde (1856–1935) and others, under the enlightened patronage of the Pedersen family, not only achieved an individual Danish style but built up several large factories with retail outlets all over the world, thus proving that good modern design in silver and jewelry need not be confined to artists' studios; their influence has spread throughout Scandinavia. Wiwen Nilsson in Sweden, Raymond Templier and Jean Fouquet in Paris evolved the angular style of the 1920s. After 1935 the big factory, Gense A/B, at Eskilstuna, Swed., pioneered the use for tableware of stainless steel, first invented in England and there exploited on a smaller scale by J. and J. Wiggin at Walsall, Staffordshire. In the 1960s Germany was second only to Scandinavia in the number and quality of its artist-craftsmen, while W. M. F. (Württembergische Metallwaren Fabrik) at Geislingen is probably the biggest silverware factory in Europe. In England, notable for the most varied work, the Worshipful Company of Goldsmiths has helped a vigorous group of designers to emerge since 1945, including Gerald Benney, Eric Clements, David Mellor and Roger King. *See* also JEWELRY.

(G. McK. H.)

III. NORTH AND SOUTH AMERICA

1. Pre-Columbian.—When the Spaniards arrived in the New World in the 16th century they found a wide range of well-developed technical skills in fine metalwork in Mexico, Costa Rica, Panama, and the Andean region. They could offer little to the Indian smiths, who had already mastered cold hammering and annealing; *repoussé* decoration and chasing; pressing sheet gold over or into carved molds to make a series of identical forms; sheathing wood, bone, resin, and shell ornaments with gold foil; decorating with metal inlays and incrustation with jade, rock crystal, turquoise, and other stones; joining by clinching, stapling, and soldering; in Ecuador and Western Mexico, drawing gold wire; casting by the cire perdue method of both solid and hollow ornaments, often with false filigree or false granulation decoration; wash gilding; and colouring alloys containing gold by "pickling" in mild plant acids. There was some regional specialization, hammer work in "raising" a vessel from a flat disc of sheet gold or silver having reached its apogee in Peru, while cire perdue casting was highly developed in Colombia, Panama, Costa Rica and Mexico. The miniature, hollow cire perdue castings of the Mixtec goldsmiths in Mexico have never been surpassed in delicacy, realism and precision, and the tiny solid-cast frogs from Panama must be viewed through a magnifying glass to be appreciated. In Mexico bimetallic objects of gold and silver were made by two-stage casting; the gold part was cast first and the silver, which has a lower melting point, was then "cast on" to the gold in a separate operation. (A famous example is the pectoral of Teotitlán del Camino in the National Museum in Mexico City.) A silver llama in the American Museum of Natural History in New York indicates that the Peruvian smiths had taken the first step toward cloisonné, the cloisons being filled with cinnabar instead of enamel.

A truly great technological and artistic triumph of the pre-Hispanic workers in Ecuador was the making of complex beads of microscopic fineness from an alloy of gold and platinum, achieved by "sintering" gold dust and small grains of alluvial platinum. Platinum was not used in Europe until the latter half of the 19th century, about 500 or 600 years later.

As in other early cultures, the pre-Hispanic goldsmiths were a privileged and highly respected group, sometimes having their own patron deity such as *Xipe Totec* in Mexico or *Chibchachun* in Colombia. In Peru just before and at the time of the Conquest the goldsmith (*kori-camayoc*) is said to have been a full-time government worker, supported by the state and producing exclusively for the Inca. According to the early Mexican picture writings (codices) and the accounts of the Spanish chroniclers the craft was hereditary, the secrets being passed on from father to son.

Most of Montezuma's treasure and Atahualpa's futile ransom went directly into the melting pot, but even the conquistadors were constrained by admiration to send a few outstanding objects back to Spain intact. Some examples of the loot seen by Albrecht Dürer in Brussels in 1520 evoked his oft-quoted "Never in all my born days have I seen anything that warmed my heart as much as these things . . . ," the first professional criticism of pre-Columbian art on record. Unfortunately, what he saw then has since disappeared.

Despite Spanish greed for gold and subsequent centuries of tomb robbing and melting down the finds, many masterpieces of the goldsmith's art have come to light and survive in public and private collections. The major ones are those in the Museum of the American Indian and the American Museum of Natural History in New York, the University of Pennsylvania Museum, the Peabody Museum of Harvard University, the Cleveland Museum of Art, and the Robert Woods Bliss Collection on loan to the National Gallery of Art in Washington. Many fine examples may be seen also in the British Museum, the Musée de l'Homme in Paris, the ethnographic museums of Berlin and Hamburg, and the archaeological museum in Madrid. Mexico, Costa Rica and South America have superb regional collections, such as those in the National Museum in Mexico City, the Regional Museum of Oaxaca with the famous Tomb 7 jewels from Monte Albán, the National Museum in Costa Rica, the unique Gold Museum of the Colombian Bank of the Republic in Bogotá, and the archaeological museum in Lima, to say nothing of the private collections of Emilio Estrada and Carlos Zevallos in Ecuador, and those of Miguel Mujica Gallo and Rafael Larco Hoyle in Peru.

The earliest examples of goldwork in the New World are the sheet-gold adornments with *repoussé* decoration from Chongoyape, Peru, in the Museum of the American Indian and the Bliss Collection. They were made some time between 1000 and 500 B.C. Casting seems to have begun in Mochica times early in the Christian era in northern Peru, whence it is thought to have spread northward into Ecuador, Colombia, Panama, Costa Rica and finally Mexico. Dating in the intervening areas is problematical, but it is generally accepted that fine metalwork in gold, silver and copper did not reach the valley of Oaxaca in Mexico until about A.D. 900. Some finds in western Mexico suggest an earlier beginning date there, and also that knowledge of the craft came by sea rather than overland from South America.

It is said that the Spaniards saw some pre-Columbian goldwork when they first arrived in Florida, but none seems to have survived. Ornaments made from native copper have been found in the Ohio Valley, Arkansas and the Great Lakes region, but no gold or silver work.

2. Southwest Indian.—The famed Indian silverwork in the Southwest does not begin until 1853, when the craft was introduced to the Navaho by Mexican smiths. Although this Mexican origin is recognized by Arthur Woodward in his historical study of Navaho silversmithing (see *Bibliography*), he points out that certain ornament types and modes of decoration among the Navaho trace back to earlier Indian silverworking in the eastern woodland, the plains, and the Rocky Mountains. It was not until 1872 that the first Zuñi smith learned the craft from the Navaho. The Zuñi had been carving turquoise long before the introduction of silversmithing, so it is not surprising that the most prominent characteristic of Zuñi work is the extravagant use of turquoise

insets. Navaho work is distinguished by die-stamped designs, whereas die-work is very rare in Zuñi silver. Authentic Navaho and Zuñi pieces are still being made, but the tourist market has been flooded with cheap, commercial imitations.

3. **Colonial.**—Silversmithing in the New World in the colonial period is more or less derivative from Europe and England. In North America it was first brought to New England by English craftsmen in the 17th century. The most important centres were Boston, Newport, New York, Philadelphia, Baltimore and Annapolis. Outstanding collections include the Garvan Collection at Yale University, and those in the Boston Museum of Fine Arts, the American Wing of The Metropolitan Museum of Art in New York, and the Philadelphia Museum of Art. North American colonial silver is distinguished for its simplicity and graceful forms, copied or adapted from English silver of the period. On the other hand, the colonial silver of Mexico, Brazil, Colombia, Peru, Chile and Bolivia, while European in concept, shows a blending of Iberian designs and forms with indigenous influences that trace back to pre-Hispanic times. Most of these relics survive in churches as sacramental vessels, but there are some notable private collections such as that of Pedro de Osma and the collection of silver stirrups, spurs and horse trappings belonging to Constante Larco Hoyle, both in Lima, Peru.

4. **Modern.**—The outstanding centre for fine handwork in silver in the Western Hemisphere is the little village of Taxco in the state of Guerrero, Mexico. An American resident, William Spratling, revived the ancient craft there in 1931, and trained a whole generation of talented silversmiths. As of a census made in 1955 there were about 300 workshops, large and small, each dedicated to artistic handwork in silver. (D. T. E.)

IV. THE MIDDLE AND FAR EAST

1. **Iran.**—The Persians have been skilful metalworkers since the Achaemenid period (550–330 B.C.) when they were already acquainted with various techniques such as chasing, embossing, casting, and setting with precious stones. Statuettes of gold and silver are known from the 5th century B.C. and vessels of silver and gold from this time take the form of phials, conical cups, vases and rhyta (drinking cups in the shape of an animal's head). The Oxus treasure in the British Museum and the Susa find in the Louvre, Paris, are good examples of such work. During the Parthian period (247 B.C.–A.D. 224) silver and gold work was strongly influenced by Hellenistic predilection for richly decorated bowls and dishes. The zenith of old Iranian metalwork, however, was reached during the Sassanid period (A.D. 224–651) when there was great variety of shape, decoration and technique. Drinking vessels—stem cups and cups with handles—ewers, oval dishes, platters and bowls are the dominant forms; hunting scenes, drinking scenes and animals are represented in high relief. The patterns were cut out of the solid silver, or made separately in sheets and then soldered to the vessel. In some cases parts of the decoration are traced on a ringmatted ground and parcel gilt. From this time onward cloisonné enamel is used for jewelry.

During the Islamic period gold and silver were used either as an inlay on vessels of bronze, brass and iron or for jewelry. Earrings, bracelets and necklaces were often made in filigree work.

2. **India.**—In India gold jewelry has been found from the Indus culture. Excavations at Takshasila (Taxila) have revealed gold and silver drinking vessels and jewelry of Hellenistic types dating back to about the 1st century A.D. From the same time is the important Buddhist gold reliquary from Bimaran, Afg., set with rubies and decorated with *repoussé* figures in Gandhara style (*see* INDIAN ART); it is now in the British Museum, London.

During the Gupta period (A.D. 320–647), vessels of Hellenistic and Persian shapes were evidently made, for they are represented in the sculpture and frescoes of the period. More Indian in style are a silver dish of the 3rd or 4th century A.D., decorated with a Bacchanalian scene, and a silver bowl of the 7th century from northern India embellished with medallions in low relief (both in the British Museum). Jewelry played a very important role and can also be studied in the frescoes at Ajanta and on contemporary sculptures, though no original pieces have survived.

In spite of the fact that gold and silver vessels have been common in India since classical times there is very little material extant before the 17th century when all kinds of vessels were produced in bronze, brass, copper, and, for the royal houses, in silver. The shapes and decorations vary in the different regions. Delhi was famous for its craftsmen, especially in the time of Akbar the Great (1556–1605), Jahangir (1605–27) and Shah Jahan (1627–58). Much work was done in precious metal, and vessels and ornaments of jade were inlaid with gold and gems. Northern India is famous for its enamels. Enamelers from Lahore were brought to Jaipur in the 16th century by Man Singh, and enamel was employed extensively in combination with gold and silver work in the 17th and 18th centuries there and elsewhere. The Punjab, Lucknow, and the districts of Chanda, and Cutch in Gujarat State were long celebrated for their metalworkers. In the south the silver work in *svamin*-style is characterized by religious figure scenes in relief which were executed in three different techniques. Tirupathi put silver sheet on copper; Madras, Bangalore, and Tiruchirappalli are known for hammered vessels with traced decoration; and Thanjavur (Tanjore) produced a more Baroque effect with inlays of silver in copper. From the former Travancore state, Mysore and Bijapur in the southwest come chased vessels with floral patterns, the lotus being the dominant motif. In the north the Hindu style is well represented by works from Varanasi (Benares).

Persian-Islamic influence is found in several shapes of the vessels; *e.g.*, ewers and basins for water, and smoking furniture such as hookas also with Islamic patterns. Jewelry from the later periods employs precious stones, pearls, gold and silver in great variety. The old types are repeated with symmetrical arrangements of rosettes and leaves for bracelets, necklaces, pendants, rings and foot ornaments. Very fine work in silver filigree was executed at Cuttack in Orissa and was employed on jewelry and various larger items.

3. **Nepal, Tibet, Burma, Thailand, Annam.**—Indian styles and techniques spread to the neighbouring countries. In Nepal precious metals were used in architecture; pagodas, temples and palaces sometimes had facades richly decorated with ornaments embossed in gilt copper with settings of precious stones.

In Tibet copper and brass were usually used for vessels but these metals were often decorated with applied silver or gold ornaments, and in eastern Tibet especially teapots were made of silver with gilt appliqué. While many of the ornaments are Chinese, Buddhist shapes and patterns of Indian origin were used for ritual vessels. Other ritual objects were sometimes made of silver or, more rarely, of gold, though bronze is again the common material. Silver is used for the amulets and jewelry with rich setting of turquoises, carnelian and lapis lazuli.

In Thailand the Buddhist vessels were made out of chased silver very often in the shape of a lotus flower whose petals are decorated with other floral and figure motifs in *repoussé*.

Burma is known for its chased silver vessels heavily decorated with figures and floral patterns in relief, related to the south Indian *svamin* work. The use of gold and silver vessels for domestic purposes was denied to all but those of royal blood. Good examples of earlier golden regalia are in the Victoria and Albert Museum.

In Annam gold and silver work of the Cham culture is preserved from the 10th century, exemplified by the crown and heavy jewelry made for a life-size statue found in the ruin of a temple at Mi-son. From later times are the royal treasure with four crowns, various amulets, armrings, and table services of gold richly decorated with *repoussé* and open work.

4. **China.**—In ancient China gold and silver were rare. Gold was used as an inlay for bronzes in the Chou dynasty (1027–256 B.C.), and between the 6th and the 2nd centuries vessels and small bronzes for daily use were very often decorated with geometrical and zoomorphic patterns hammered into the bronze surface. Gilding and silvering are also common at the same period. Dresshooks and small items of jewelry were sometimes cast in gold and silver and imitated the more usual bronze forms. Granular work is used for jewelry, a technique which probably has an Indian origin.

Silverwork first became important during the T'ang period (A.D. 618–906) when the Chinese had learned from the Sassanid Persians how to chase the silver. To begin with they followed their teachers very closely in the forms of the bowls and larger vessels as well as in the patterns. The T'ang drinking vessels such as stem cups, cups with handles, ewers, trays, and lobed oval dishes on a stem are Persian shapes transformed by Chinese taste. Among the patterns are vine and palmette scrolls of great variety, hunting scenes and landscapes of symmetrical flowers and trees with birds and animals; all these have parallels in Persian silver and textiles but are more delicate in their Chinese version. The techniques used by the Sassanid silversmiths are adopted by the Chinese; for example, double sheets for a bowl and tracing of the patterns on ringmatted ground. T'ang jewelry is made of gold or gilt silver mainly as hair ornaments in granular work and filigree ornamented with set turquoises. Collections of T'ang silver and jewelry can be seen in the Shosoin Imperial Treasure House at Nara, Japan; in the Carl Kempe Collection, Stockholm; and in the British Museum and the Victoria and Albert Museum, London.

During the Sung dynasty (960–1279) silver work declined in technical quality but jewelry played a more dominant role. Hair ornaments became increasingly intricate with elaborate naturalistic flowers and various auspicious symbols.

During the Yüan (1280–1368) and Ming (1368–1644) periods skill in silver work revived and once again the smiths followed many Near Eastern styles. Drinking vessels (ewers and cups), boxes, and even large ceremonial gold vessels have been found in the Ming tombs. During the excavation of the tomb of Emperor Wan Li (1573–1620) in 1956–58 a series of gold vessels set with precious stones was found together with porcelain and textiles. All the gold items are decorated with incised patterns of dragons, phoenixes and similar subjects.

During the Ch'ing period (1644–1912) both silver and gold were used lavishly and gold filigree work especially is common in the 18th century. However, most of the forms and ornaments employed are borrowed from lacquer and porcelain ware and only jewelry has its own style, rich combinations of kingfisher feathers glued to the metal.

5. Korea.—The Chinese colonists who settled in Korea during the Han Empire (202 B.C.–A.D. 221) first brought gold- and silversmiths to Korea. By the 5th–6th centuries A.D. Korean work, as exemplified by large gold crowns and various pieces of jewelry excavated from tombs at Kyongju, was beginning to develop distinctive characteristics. At the time of the United Silla (668–935) and Koryo (918–1391) kingdoms Chinese influence was strong but the Korean style persisted in silver and gold work. Several vessels with floral patterns in relief are preserved from these periods.

6. Japan.—Knowledge of metalwork seems to have spread to Japan by way of Korea during the Yayoi Period (*c.* 1st century B.C.) but gold and silver never played any important role in this country. In the Nara period (A.D. 710–784) the Chinese T'ang style was dominant and most of the gold and silver work preserved in the Shosoin at Nara was made under Chinese influence or by Chinese workmen. Silver vessels were used extensively among the aristocracy in the Heian period (784–1185), although not many of these vessels have survived, and both gold and silver were often used for applied reliefs or as inlay on bronze. In the later periods the use of precious metals was practically confined to inlays in bronze or iron and the highest technical skill is shown by the artists who made the sword fittings. (B. V. Gy.)

See also references under "Silver and Gold Work" in the Index.

BIBLIOGRAPHY.—*Ancient Work:* British Museum Catalogues: F. H. Marshall, *Greek, Roman and Etruscan Jewellery* (1911, bibl.); H. B. Walters, *Silver Plate* (1921); H. Schmidt, *Schliemanns Sammlung Trojanischer Altertümer* (1902); V. Stais, *Collection Mycénienne* (1909). C. R. Williams, *Catalogue of Egyptian Antiquities*, The New York Historical Society (1924); R. Siviero, *Gli ori e le Ambre del Museo Nazionale de Napoli* (1954); B. Svoboda and D. Concěv, *Neue Denkmaler Antiker Toreutik* (1956); P. Jacobsthal, *Greek Pins and Their Connections with Europe and Asia* (1956); L. H. Brom, *De Kantharos van Stevensweert* (1951). (E. J. F.; M. C. Rs.)

Medieval: H. Swarzenski, *Monuments of Romanesque Art* (1954); C. R. Dodwell (trans.), *The Various Arts* (1961); J. Braun, *Meisterwerke der deutschen Goldschmiedekunst der vorgotischen Zeit*, 2 vol. (1922); S. Gevaert, *L'Orfèvrerie Mosane* (1943).

Renaissance to Modern Period: H. Leitermann, *Deutsche Goldschmiedekunst* (1953); F. Rossi, *Italian Jeweled Arts* (1957); J. W. Frederiks, *Dutch Silver* (Renaissance-18th century), 4 vol. (1952–61); H. Thoma, *Kronen und Kleinodien* (1955); H. Nocq, *L'Orfèvrerie civile française* (16th–19th centuries), 2 vol. (1926); S. Brault and Y. Bottineau, *L'Orfèvrerie française du XVIII*e *siècle* (1959); M. H. Gans and D. de Wit-Klinkhamer, *Dutch Silver* (1961); C. C. Oman, *English Domestic Silver*, 4th ed. (1959), *English Church Plate* (1957); J. F. Hayward, *Huguenot Silver in England* (1959); A. J. Collins, *Jewels and Plate of Queen Elizabeth I* (1956); M. Rosenberg, *Der Goldschmiede Merkzeichen*, 4 vol., 3rd ed. (1922–28); C. J. Jackson, *English Goldsmiths and Their Marks*, 2nd ed. (1921); E. Kris, *Goldschmiedearbeiten* (1932); R. Dos Santos, *Ourivesaria Portuguesa*, 2 vol. (1959); F. Dennis, *Three Centuries of French Domestic Silver*, 2 vol. (1960). (J. F. H.)

Modern: Worshipful Company of Goldsmiths, *Modern British Silver* (1951, 1954, 1959, 1964), *Modern British Jewellery* (1964); E. Hiort (ed.), *Modern Danish Silver* (1955); G. A. Jensen, *Fifty Years of Danish Silver in the Georg Jensen Tradition* (1956); O. Untracht, *Enameling on Metal* (1957); B. Cuzner, *A Silversmith's Manual*, 2nd ed. (1958); H. Maryon, *Metalwork and Enamelling*, 4th ed. (1959); A. K. Snowman, *The Art of Carl Fabergé*, rev. ed. (1962); Graham Hughes, *Modern Jewellery, 1890–1964*, 2nd ed. (1964). (G. McK. H.)

North and South America: Pre-Columbian: General: Dudley T. Easby, Jr., "Ancient American Goldsmiths" in *Natural History Magazine*, vol. lxv, no. 8 (1956); Pál Kelemen, *Medieval American Art*, 2 vol. (1946); S. K. Lothrop, W. F. Foshag, and Joy Mahler, *Pre-Columbian Art, Robert Woods Bliss Collection* (1957); P. Rivet and H. Arsandaux, *La Métallurgie en Amérique précolombienne* (1946).

Colombia: José Pérez de Barradas, *Orfebrería Prehispánica en Colombia*, 4 vol. (1954 and 1958).

Ecuador: Paul Bergsøe, "The Metallurgy and Technology of Gold and Platinum among the Pre-Columbian Indians," *Ingeniørvidenskabelige Skrifter*, no. A-44 (1937); Carlos Zevallos Menéndez, "Orfebrería Prehistórica del Ecuador" in *Cuadernos de Historia y Arqueología*, vol. vi, no. 16–18 (1956).

Mexico: Alfonso Caso, "La Orfebrería Prehispánica" in *Artes de México*, no. 10 (1955); Dudley T. Easby, Jr., "Sahagún y los orfebres precolombinos de México" in *Anales del Instituto Nacional de Antropologia e Historia*, vol. ix (1957); Marshall H. Saville, *The Goldsmith's Art in Ancient Mexico* (1920).

Peru: Wendell C. Bennett and Junius B. Bird, *Andean Culture History*, 2nd ed. (1960); D. F. Rubín de la Borbolla, *Los Tesoros Artísticos del Perú* (1961) (its coverage extending to the modern period); Dudley T. Easby, Jr., "Los Vasos Retratos del Perú" in *Revista del Museo Nacional*, vol. xxiv (1955); S. K. Lothrop, "Gold Ornaments of Chavín Style from Chongoyape, Peru" in *American Antiquity*, vol. 6, no. 3 (1941); William C. Root, "Metallurgy" in vol. 5 of *Handbook of South American Indians*, Smithsonian Inst. (1949).

Southwest Indian: John Adair, *The Navajo and Pueblo Silversmiths* (1944); Frederick J. Dockstader, *Indian Art in America* (1961); Arthur Woodward, *A Brief History of Navajo Silversmithing* (1938).

Colonial: Mexico: Lawrence Anderson, *The Art of the Silversmiths in Mexico, 1519–1936*, 2 vol. (1941); Jorge Olvera, "La Plata Colonial de Chiapas" in *Artes de México*, no. 10 (1955); M. Romero de Terreros y Vinent, *Las Artes Industriales en la Nueva España* (1923).

United States: Vincent D. Andrus, *Early American Silver*, A Picture Book of The Metropolitan Museum of Art (1955); C. Louise Avery, *Early American Silver* (1930); Kathryn C. Buhler, *American Silver* (1950); John Marshall Phillips, *American Silver* (1949), containing an excellent bibliography.

Modern: Mexico: William Spratling, "25 Years of Mexican Silverware" in *Artes de México*, no. 10 (1955), English text with good illustrations.

United States: See the magazine *Craft Horizons*, published by The American Craftsmen's Council; and the exhibition catalogue *Designed for Silver*, Museum of Contemporary Crafts (1960). (D. T. E.)

Middle and Far East: G. C. M. Birdwood, *Industrial Art of India*, 2 vol. (1880); H. L. Tilly, *The Silver Work of Burma* (1902); P. Eudel, *L'Orfèvrerie algérienne et tunisienne* (1902); A. P. Charles, *Gold and Silver Ware of the United Provinces* (1905); O. M. Dalton, *The Treasure of the Oxus*, 2nd ed. (1926); A. K. Coomaraswamy, *Mediaeval Sinhalese Art* (1909), *The Art and Crafts of India* (1911); H. L. Roth, *Oriental Silver, Malay and Chinese* (1910); V. A. Smith, *A History of Fine Art in India and Ceylon* (1911); F. Sarre, *Die Kunst des alten Persien* (1922); *The Journal of Indian Art and Industries* (1884–1916); P. Ackerman, "The Art of the Parthian Silver- and Goldsmiths," E. Margulies, "Cloisonné Enamel," J. Orbeli, "Sasanian and Early Islamic Metal Work" in *A Survey of Persian Art*, ed. A. U. Pope, vol. i (1938); H. Bossert, *Geschichte des Kunstgewerbes aller Zeiten und Völker*, vol. iii and iv (1930–31); J. Orbeli and C. Trever, *Orfèvrerie Sasanide* (1935); J. G. Andersson, "The Goldsmiths in Ancient China," *Museum of Far Eastern Antiquities Bulletin*, no. 7 (1935); V. Griessmaier, "Die granulierte Goldschnalle," *Wiener Beiträge zu Kunst- und Kulturgeschichte Ost Asiens*, vii (1933); B. V. Gyllensvärd, *Chinese Gold and Silver in the Carl Kempe Collection* (1953), *T'ang Gold and Silver*

(1957); R. L. Hobson, "A T'ang Silver Hoard," *British Museum Quarterly*, vol. i, no. 1 (1926); Y. Okada, "History of Japanese Ceramics and Metalwork," *Pageant of Japanese Art*, vol. iv (1952); M. Feddersen, *Japanisches Kunstgewerbe* (1960). (B. V. Gy.)

SILVER CITY, a ghost town of Owyhee County in southwestern Idaho, U.S., was founded March 10, 1863, and quickly displaced adjacent Ruby City as the centre of the Owyhee mines. Exceptionally rich silver lodes in nearby War Eagle and Florida mountains took years to develop but were the subject of great excitement and bitter rivalry for control. The Poorman mine, although entangled in litigation, publicized the district particularly well; silver crystals from it won a special gold medal at the Paris Exposition of 1866. In 1868 armed conflict among claim owners forced the governor to dispatch troops from Ft. Boise to restore order. An important mining industry with a strong labour union movement had developed prior to financial collapse of the silver market accompanying the failure of the Bank of California on Aug. 26, 1875; by then Silver City had Idaho's earliest daily newspaper—the *Owyhee Daily Avalanche*—and sufficient stability to survive the economic panic. More efficient mining technology, extensive British investment, and improved railway transportation near that remote mountain district (elevation over 6,000 ft. [1,800 m.]) brought a long revival after 1886. The gradual decline of mining activity led to loss of population and, in 1935, of the county seat. Silver City's peak population had been several thousand; in the 1960s it was less than 50. It can be reached by road from Nampa (44 mi., part paved) or by gravel road (about 20 mi.) from the highway between Boise and Reno, Nev. There are a number of other ghost towns in the county. *See* also IDAHO: *History*. (M. D. B.)

SILVERFISH, a slender, flat, wingless insect, *Lepisma saccharina*, with three tail bristles, so called because of its covering of silvery scales, flexible body, and quick movements. It normally lives indoors and is worldwide in distribution. Like many domestic insects, its original habitat is unknown; its nearest wild relatives are in Africa and western Asia. Occasionally, it damages paper and fabrics by eating the farinaceous sizing and glue; usually, it does little or no damage, unlike its similar-looking relative, the firebrat (*Thermobia domestica*).

The silverfish belongs to the family Lepismatidae of the order of bristletails (Thysanura) and is in the primitive subclass Apterygota (*q.v.*), none of whose members was ever winged. It is plain gray with the antennae and tail bristles shorter than the body. Four common species of domestic Lepismatidae are mistaken for the silverfish: the firebrat has conspicuous dark brown and cream markings; *Ctenolepisma longicaudata*, like the silverfish, is plain gray but has antennae much longer than the body; *Ctenolepisma lineata*, also with long antennae, has dark longitudinal stripes on the gray background; and *Acrotelsa* (*Stylifera*) *gigantea* is evenly dark in colour and lives only in warm regions.

The males superficially resemble the females. Unlike all other groups of true insects, the silverfish and its closest relatives have, instead of actual copulation, courtship movements with the male depositing a packet of sperm that the female then places in her vagina. The oval, whitish eggs probably are inserted into cracks and litter. The young, which hatch in several days, are scaleless and have short appendages; they molt every few days and gradually acquire the adult characters.

The silverfish, unlike most insects, continues to molt throughout the two or more years of its life, long after reaching sexual maturity. Control can be accomplished with poisonous baits or insecticides. *See* also INSECT; ENTOMOLOGY: *Principles of Insect Control*. (C. L. RE.)

SILVERIUS, SAINT (d. 537), pope from 536 to 537, was a legitimate son of Pope Hormisdas, born before his father entered the priesthood. He was consecrated successor to Agapetus I on June 8, 536, having been pressed on the Roman clergy by the Ostrogothic king Theodahad. Six months afterward (December 9) he was one of those who admitted Belisarius' Byzantine forces into the city.

Silverius opposed the restoration of the patriarch Anthimus, whom Agapetus had deposed, and thus brought upon himself the hatred of Theodora, who desired to see Vigilius made pope. He was deposed by Belisarius in March 537 on a charge of treasonable correspondence with the Goths, degraded to the rank of monk and sent to Lycia. Justinian, who entertained his complaint, sent him back to Rome, but Vigilius was ultimately able to banish his rival to the island of Palmarola, off Naples, where he died, probably about December 537. He is venerated as a martyr; feast day June 20. *See* also PAPACY.

SILVER SPRINGS, one of the largest limestone springs in the world, is in Marion County, Fla., U.S., 6 mi. E of Ocala. The average flow of the springs is about 530,000,000 gal. per day and more than 480 tons of minerals are carried off in solution each 24 hours. The water maintains a constant temperature of 72° F (22° C). The source of the main spring is a cavern 65 ft. long and 12 ft. high. Most of the water discharged from the more than 100 springs in the group is from the abundant rainfall of north-central Florida, which passes through the porous surface soil and sand of the region and enters the limestone lying immediately beneath. This filtration causes the extreme clearness of the spring water.

Silver Springs is the source of Silver River, a 7-mi.-long navigable stream which eventually reaches the Atlantic Ocean via the Oklawaha and St. Johns rivers. There are more than 30 varieties of fish, as well as turtles and shellfish in Silver Springs. Remains of mastodons, manatees, and extinct elephants have also been found in the springs or in the Silver River.

Swimming facilities, glass-bottomed boat rides, cruises on the Silver River, a zoo, and a snake farm are provided for visitors. Thousands of tourists visit the springs each year. Because of the clarity of the water, Silver Springs is a favourite location for motion-picture companies that wish to film underwater scenes.

The explorer Hernando de Soto was probably the first European to see Silver Springs when he camped at Ocali, an Indian village near the springs, in 1539. The Indian name for Silver Springs is Sua-ille-oka, which means "sun-glinting water." (J. E. Jo.)

SILVESTER (SYLVESTER), the name of three popes and an antipope.

SAINT SILVESTER I (d. 335), pope from 314 to 335, succeeded Melchiades. The most important event of his pontificate was the Council of Nicaea (325), which condemned Arius. Silvester did not attend personally but was represented by two legates. Associated with his name is the famous legend (which developed after his own time) according to which he is supposed to have baptized the emperor Constantine I in the Lateran and cured him of leprosy. Supposedly, it was gratitude for this that caused Constantine to give him dominion over the four Eastern patriarchates and over almost all the Western world, the so-called Donation of Constantine. This legendary gift, which since the early 16th century has been acknowledged by many scholars to be a forgery, was of importance in the development of the medieval theory of church and state (*see* DONATION OF CONSTANTINE). Silvester's feast day is Dec. 31. (R. E. McN.)

SILVESTER II (Gerbert) (d. 1003), pope from 999 to 1003, was born in the county of Auvergne. Educated in the abbey of St. Géraud at Aurillac in Auvergne under the abbots Gerald and Raymond, he was taken beyond the Pyrenees in 967 by Borel, count of Barcelona, who entrusted his further education to Atto, bishop of Vich (Ausona). Accompanying these two patrons to Rome in 970–971, Gerbert was presented to the emperor Otto I, to whom he admitted his skill in the quadrivium (*i.e.*, the medieval liberal arts of arithmetic, music, geometry and astronomy) while deploring his comparative ignorance of logic. About 972 he went to Reims to study under Archbishop Adalbero, under whom he seems to have lectured for many years, having among his pupils Robert, the future king of France, and the chronicler Richer. His growing fame roused the envy of Otric of Saxony, who, suspecting him of error in his classification of the sciences, accused him of this before the emperor Otto II. By Otto's command the two rivals then had a disputation at Ravenna, which is said to have lasted a whole day (about Christmas, 980). Otto subsequently gave Gerbert the abbey of Bobbio (982 or early 983), but Gerbert found such difficulty in collecting his dues that he returned to Reims in 984 to resume

service under Adalbero. The latter involved him in his intrigues against the last Carolingian kings of France.

Adalbero died in Jan. 989, having, according to Gerbert, designated him his successor as archbishop. The appointment, however, was given instead to Arnulf, an illegitimate son of the late king Lothair. Arnulf took an oath of fealty to Hugh Capet, who had taken the place of Lothair's son Louis V as king of France, and Gerbert was persuaded to remain with him. Then, in the late summer or autumn of 989, Reims was taken by Lothair's brother Charles of Lorraine, now the head of the Carolingian house. Gerbert, falling into Charles's hands, continued for a time to serve under Arnulf, who had gone over to his uncle's side. He was able, however, to return to the Capetian allegiance before the treachery of Adalberon (Ascelin) of Laon put both Charles and Arnulf at the mercy of Hugh Capet (March 991). At the synod of St. Basle, near Reims, Arnulf was degraded from the archbishopric and Gerbert, in recognition of his services, appointed his successor (June 991).

The vigour and the activity of Gerbert as a metropolitan made themselves felt as far away as Tours, Orléans and Paris. But meanwhile Arnulf's friends were active, and Pope John XV objected to the summary deposition of Arnulf. Gerbert had to appear before a papal legate at Mouzon in 995 and went himself to Rome to put his case before the new pope, Gregory V, in 996 (he was present at Otto III's coronation). Finally, in 997, after Gregory's synod of Pavia had suspended all the bishops concerned in the proceedings of St. Basle, Gerbert despaired of his archbishopric and left France for the court of his former pupil the young emperor Otto III.

Otto welcomed him as an old supporter of the imperial family as well as a great scholar and quickly procured his appointment to the archbishopric of Ravenna (about April 998). The death of Gregory V during the following year enabled the emperor to do even better for his protégé, by sponsoring his elevation to the papacy. Gerbert was consecrated as Silvester II on April 2, 999. It is he who is generally credited with having encouraged the splendid vision of a restored empire that then began to fill Otto's mind.

In ecclesiastical administration Silvester was by no means inactive. He confirmed his old rival Arnulf in the see of Reims (999), rebuked Adalberon (Ascelin) of Laon for his crimes (1000), and settled the dispute between Willigis, archbishop of Mainz, and Bernward, bishop of Hildesheim (1001). Of more lasting importance was the erection of Gniezno to metropolitan status in a Polish church independent of the German hierarchy (1000). The genuineness of the letter to St. Stephen of Hungary, accepting his kingdom as a fief of the Holy See, is contested. Silvester's plans for the advancement of the church and empire together were halted by Otto III's death on Jan. 23, 1002. Little more than a year later, on May 12, 1003, the pope himself died. He was buried in the church of St. John Lateran.

Besides being the most distinguished statesman, Gerbert was also the most accomplished scholar of his age. Richer has left a detailed account of his system of teaching at Reims. So far as the trivium (grammar, rhetoric and logic) is concerned, his textbooks were Victorinus' translation of Porphyry's *Isagoge,* Aristotle's *Categories* and Cicero's *Topics* with the commentaries of Boëthius. From dialectics he urged his pupils to the study of rhetoric; but, recognizing the necessity of a large vocabulary, he accustomed them to read Virgil, Statius, Terence, Juvenal, Horace, Persius and Lucan. More remarkable still were his methods of teaching the quadrivium. To assist his lectures on astronomy he constructed elaborate globes of the terrestrial and celestial spheres, on which the course of the planets was marked. For facilitating arithmetical and perhaps geometrical processes he constructed an abacus with 27 divisions and a thousand counters of horn. A younger contemporary speaks of his having made a wonderful clock or sundial at Magdeburg. It is known from his letters that Gerbert was accustomed to exchange his globes for manuscripts of those classical authors that his own library did not contain.

More extraordinary still was his knowledge of music—an accomplishment which seems to have been his earliest recommendation to Otto I. Gerbert's letters contain more than one allusion to organs which he seems to have constructed, and William of Malmesbury preserved an account of a wonderful musical instrument still to be seen in his days at Reims. The same historian says that Gerbert borrowed from the Saracens the abacus with ciphers (namely, for numerals from one to nine, instead of mere units, but without the zero).

Perhaps Gerbert's chief claim to the remembrance of posterity is to be found in the care and expense with which he gathered together manuscripts of the classical writers. His love for literature was a passion. In the turmoil of his later life he looked back with regret to his student days, and "for all his troubles philosophy was his only cure." Everywhere (Rome, Trier, Montier-en-Der, Gerona, and Barcelona) he had friends or agents to procure copies of the great Latin writers for Bobbio or for Reims. To the abbot of Tours he writes that he is "labouring assiduously to form a library," and "throughout Italy, Germany and Lorraine [Belgica] is spending vast sums of money in the acquisition of mss." It is noteworthy, however, that Gerbert never wrote for a copy of one of the Christian fathers, his aim being, seemingly, to preserve the fragments of a fast-perishing secular Latin literature.

So remarkable a character as that of Gerbert left its mark on the age, and fables not unlike those later accumulated around Faust soon began to cluster round his name. Toward the end of the 11th century Cardinal Benno, the opponent of Hildebrand, is said to have made Gerbert the first of a long line of magician popes. William of Malmesbury adds a love adventure at Córdoba, a compact with the devil, the story of a speaking statue that foretold Gerbert's death at Jerusalem—a prophecy fulfilled, somewhat as in the case of Henry IV of England, by his dying in the Jerusalem church of Rome—and that imaginative story of the statue with the legend "Strike here" which found its way into the *Gesta Romanorum.*

Collections of Gerbert's works were edited by A. Olleris, *Oeuvres de Gerbert, pape sous le nom de Sylvestre II* (1867); by J. Havet, *Lettres de Gerbert* (1889); and by N. Boubnov, *Gerberti opera mathematica* (1899).

Bibliography.—E. Amann and A. Dumas, *L'Église au pouvoir des laïques,* vol. vii of A. Fliche and V. Martin (eds.), *Histoire de l'Église* (1942); F. Picavet, *Gerbert, un pape philosophe* (1897); I. Eichenbrun, "Gerbert (Silvester II) als Persönlichkeit," in *Beiträge zur Kulturgeschichte des Mittelalters und der Renaissance,* vol. xxxv (1928); J. Leflon, *Gerbert, humanisme et chrétienté au Xe siècle* (1946); P. E. Schramm, *Studien zur Geschichte des Kaisers Otto III* (1923).

Silvester III (John), bishop of Sabina, became pope on Jan. 20, 1045, after the Romans had driven out Benedict IX; but he resigned the papacy on March 10, 1045. Retiring to his old bishopric, he recognized Gregory VI as pope. The synod of Sutri (Dec. 1046), under the presidency of the emperor Henry III, formally passed sentence of deposition on him, as is shown by the documents issued in his capacity as bishop of Sabina during 1046.

Silvester IV (Maginulf), antipope from 1105 to 1111, a man of dubious antecedents, was elected in opposition to Paschal II by the dissident nobles under Werner, margrave of Ancona, on Nov. 18, 1105. Though Paschal's troops expelled him from Rome the next day, his supporters continued to give trouble during the following year; and the emperor Henry V accorded him some recognition. When Paschal and Henry came to terms, however, the antipope's pretensions were dismissed (April 1111).

See also Papacy.

SILVICULTURE is the technical branch of forestry (*see* Forests and Forestry) which is concerned with the establishment and maintenance of the forest. *See* Arboriculture.

SIMANCAS, a town of northern Spain in Valladolid Province, on the right bank of the Pisuerga River and 8 mi. (13 km.) SW of Valladolid on the road to Zamora. Pop. (1960) 1,582 (mun. It was the Roman Septimanca and has a citadel dating from the Moorish occupation in the 9th century. In 939 it was the scene of a battle between the Moors and Christians. The citadel is now the Archivo General del Reino, to which the national archives of Spain were removed by order of Philip II in 1563. Their transfer to Simancas was said to have been first suggested by his secretary Francisco de los Cobos (d. 1547), who was the governor of this fortress. The extensive alterations were made by several celebrated

16th-century architects including Juan de Herrera. The arrangement of the papers was entrusted by the king to the archivist Diego de Ayala in 1561. They contain important private as well as state documents. The archives of the Indies were transferred in 1785 to the Lonja of Seville. (C. D. L.)

SIMBIRSK: *see* UL'YANOVSK.

SIMCOE, JOHN GRAVES (1752–1806), British soldier and first lieutenant governor of Upper Canada (1791–96), was born at Cotterstock, Northamptonshire, Eng., on Feb. 25, 1752. As commander of a Loyalist corps, the Queen's Rangers, he served with distinction during the American Revolution.

Appointed lieutenant governor of the newly established Upper Canada (later Ontario) in 1791, he proved an energetic administrator, although seldom moderate in his language about the United States. His enthusiasm for the new Canadian colony and for aristocratic government in it quickly waned, the more so as he greatly disliked the governor in chief, Lord Dorchester. In 1796 he agreed to go as governor and commander in chief to Saint-Domingue (Haiti), half-conquered by the British from the French. In seven months he reformed its administration and waged a successful campaign, but returned to England and resigned his colonial appointments. He was unable to get more active service, but was given command of the Western District with headquarters in Exeter (1799–1806). Then a new British government appointed him commander in chief in India; diverted on a diplomatic mission to Lisbon, Portugal, he fell ill and, returning to England, died at Exeter on Oct. 26, 1806.

BIBLIOGRAPHY.—A. L. Burt, *United States, Great Britain and British North America* (1940); Canadian Historical Association *Annual Report* for 1953 and 1958; E. A. Cruikshank (ed.), *Correspondence of John Graves Simcoe*, 5 vol. (1923–31); W. R. Riddell, *Life of John Graves Simcoe* (1926); J. R. Robinson (ed.), *Diary of Mrs. John Graves Simcoe* (rev. ed., 1934); J. G. Simcoe, *Military Journal of the Operations of the Queen's Rangers* (1787; reprinted 1844). (S. R. M.; X.)

SIMEON, SAINT, STYLITES (*c.* 390–459), Syrian monk, the originator of a fearful form of asceticism that consisted of standing ceaselessly, day and night, on top of a column (*see* STYLITE). Born near the border of Syria and Cilicia, and at first a shepherd, he spent several years in a monastery near Antioch, then remained three years walled up in a cell before he began to live in the middle of a circular enclosure. There his incredible austerity and the fame of his miracles attracted crowds of pilgrims. To escape from them he built taller and taller columns on which he could lead in peace a life of contemplation. His tallest column, on which he spent more than 20 years, was about 50 ft. high, railed round on top. He spent most of the day and night in prayer, with frequent deep obeisances. He ate practically nothing, and stood all the time, without any shelter from the sun, the cold, the wind, and the rain. Always exposed to public gaze, he seemed a superhuman example of spiritual strength, of penitence, and of closeness to God. Twice a day he preached to the people and concerned himself with their affairs, consoling the sick and miserable, and settling quarrels. He also converted pagans, of whom large numbers, especially Arabs, joined the Christian pilgrims who came to him from all parts of Syria and neighbouring countries and even from the West. The Greeks keep his feast on July 26, the date of his death, and principally on Sept. 1, when he was buried in Antioch.

BIBLIOGRAPHY.—H. Lietzmann, *Das Leben des hl. Simeon Stylites* (1908), including the three chief sources, with critical studies: Theodoret's *Historia religiosa* 26, the Greek life by the monk Anthony (with Lat. trans.) and a German trans. of the Syriac life; French trans. of Theodoret and Anthony, with rehabilitation of the latter, in A. J. Festugière, *Antioche païenne et chrétienne*, pp. 347–401 and 493–506 (1959). *See* also H. Delehaye, *Les Saints stylites* (1923); P. Peeters, *Orient et Byzance: Le tréfonds oriental de l'hagiographie byzantine*, pp. 93–136 (1950). (F. HA.)

SIMEON, SAINT, THE NEW THEOLOGIAN (949–1022), Byzantine abbot and mystic whose works had a formative influence on Orthodox spirituality, was born in Paphlagonia, of an aristocratic family, and came to Constantinople as a young boy. He insisted on entering the monastery of Studios (977), but differences with the abbot resulted in his migration to the neighbouring house of St. Mamas in the same year. There he became abbot (980) and imposed strict discipline. His vigorous cult of his spiritual father, Simeon the Studite, and other reasons brought him into conflict with the patriarch and he was exiled to the Asian shore of the Bosporus (1009). There he built up the monastery of St. Marina near Chrysopolis (Scutari) and, though eventually reconciled to the patriarch, refused to return to the capital. He was a well-known and popular figure and was called the Younger or the New Theologian, perhaps to distinguish him from the two other "theologians" par excellence in the Orthodox Church—John the Evangelist and Gregory of Nazianzus. Some of his works have been translated into modern Greek, Russian, and Latin; much has not been published in the original Greek.

His numerous writings fall into three main groups: sermons, the *capita*, and the "hymns of the divine loves." One group of the sermons, the catecheses, originally preached to the monks of St. Mamas, were later drastically revised and combined with other writings of Simeon to provide a compilation suitable for a wider public. Both his sermons and his hymns are illuminated by his own personal experience, and his underlying theme is the means, particularly obedience and tears, whereby through grace union with God may be attained in this life. His emphasis was often Christo-centric, though the climax of his mystical experiences was the indwelling of the Trinity revealed as divine light. His contribution is highly individual; he provides a link between earlier mystics, as John Climacus, and later Byzantine spirituality.

BIBLIOGRAPHY.—Hymns published by Dionysios Zagoraios with modern Gr. trans. of selected works, 2nd ed. (1886); critical ed. of *Capita* with Fr. trans. by J. Darrouzès (1957); critical ed. of catecheses with Fr. trans. by B. Krivocheine (1961); some works, mostly in Lat. trans. only, in J. P. Migne, *Patrologia Graeca*, vol. 120 (1864); Eng. trans. of hymns by J. M. Hussey (1961). For list of all works, with discussion of manuscript tradition, *see* B. Krivocheine, "The Writings of St. Symeon the New Theologian," *Orientalia Christiana periodica*, vol. 2, pp. 298–328 (1954). *See* also life by Simeon's disciple Nicetas Stethatus, ed. by I. Hausherr and G. Horn, *Orientalia Christiana analecta*, vol. 12 (1928); K. Krumbacher, *Geschichte der byzantinischen Litteratur*, 2nd ed. (1897); J. M. Hussey, *Church and Learning in the Byzantine Empire 867–1185* (1937); H. G. Beck, *Kirche und theologische Literatur im byzantinischen Reich* (1959). (J. M. HY.)

SIMEON (SYMEON) OF DURHAM (d. *c.* 1130), English chronicler, author of the *Historia Dunelmensis ecclesiae* and probably author of part of the *Historia regum*, was a monk at Durham.

The *Historia Dunelmensis ecclesiae* is a useful history of the see of Durham from its establishment (635) at Lindisfarne (*see* HOLY ISLAND) to 1096. The *Historia regum* consists of two overlapping English histories, covering the years 731–957 and 848–1129 respectively. It was long attributed to Simeon, but it now seems certain that he had no part in compiling the first history, which contains unique and valuable Northumbrian annals for the 8th and 10th centuries. He was, however, probably responsible for the second history, a chronicle for the years 848–1118 based on Asser and Florence of Worcester, and a narrative of the years 1119–29, partly following Edmer but partly original.

BIBLIOGRAPHY.—Both works are in *Symeonis monachi opera omnia*, ed. by T. Arnold, 2 vol. (1882–85); for a translation, *see* J. Stevenson, *Church Historians of England*, vol. ii, part ii (1855). *See* also H. S. Offler, *Medieval Historians of Durham* (1958); P. H. Blair, "Some Observations on the 'Historia Regum' Attributed to Symeon of Durham," *Celt and Saxon*, ed. by N. K. Chadwick (1963).

SIMEON, CHARLES (1759–1836), Anglican churchman, a leader of the evangelical party and a representative of the evangelical tradition at its best, whose influence left an indelible mark on religious life at Cambridge, was born at Reading on Sept. 24, 1759, and educated at Eton and King's College, Cambridge, where he was a fellow from 1782 (vice-provost, 1790–92). He came under the influence of Henry and John Venn and in 1782 was presented to the living of Trinity Church, Cambridge, where he remained until his death. He was at first unpopular with his parishioners who had wanted the curate, John Hammond, in his place and his services were noisily interrupted. But his pastoral care, especially in sickness and in the famine of 1788, and his quiet, peaceable nature slowly won the regard of the town and the university, and he became widely known as an evangelical leader who was yet a convinced member of the Church of England and lover of its liturgy and discipline.

He was concerned for missionary work, especially in India, and persuaded his best pupils to the venture, above all his curate. Henry Martyn; he helped to found the Church Missionary Society (1799) and assisted the newly founded (1804) British and Foreign Bible Society and the Society for Promoting Christianity Among the Jews; he gave to the support of missions from his own pocket, including part of the payment for his *Horae Homileticae*, discourses on the Bible published in 21 vol. in 1832–33. Following the example of Henry Thornton, in order to ensure the continuity of evangelical teaching, he administered the Simeon Trust (founded 1816) to purchase the right to appoint clergymen to livings. He died on Nov. 13, 1836, and is buried in the chapel of King's College.

BIBLIOGRAPHY.—*Memoirs of the Life of Charles Simeon*, ed. by W. Carus (1847); A. W. Brown, *Recollections of Simeon's Conversation Parties* (1863); H. C. G. Moule, *Charles Simeon* (1892); C. Smyth, *Simeon and Church Order* (1940); *Charles Simeon (1759–1836)*, ed. by A. Pollard and M. Hennell (1959). (W. O. C.)

SIMEON, in the Old Testament, a tribe of Israel, named after the second son of Jacob by Leah (Gen. 29:33). According to Gen. 34, the brothers Simeon and Levi massacred the males of Shechem to avenge the violation of their sister Dinah ("judgment") by Shechem the son of Hamor. Jacob disavowed the act, and on his deathbed solemnly cursed their ferocity, condemning the two to be divided in Jacob and scattered in Israel (49:5–7). Subsequently the priestly Levites (*q.v.*) are found distributed throughout Israel without portion or inheritance (Deut. 18:1, Josh. 13:14). Simeon, however, is assigned a territory—in the north, according to II Chron. 15:9, 34:6, but in southern Palestine according to Josh. 19:1–9, its cities being otherwise ascribed to Judah. Simeon is not mentioned in the "blessing of Moses" (Deut. 33), in the stories of the "judges," or in the earlier books of Samuel and Kings. It is conjectured that Dinah represents a clan or group that settled in Shechem and was exposed to danger (*e.g.*, oppression or absorption); the tribes Simeon and Levi intervened on its behalf, the ensuing massacre was avenged by the Canaanites and the two were broken up. These events are supposed to belong to an early stage in the invasion of Palestine by the Israelites (15th–13th century B.C.). (*See* TWELVE TRIBES OF ISRAEL.)

"Simeon" occurs in the New Testament, as the name of the seer who recognized the infant Jesus as the Redeemer (Luke 2:25–35) and as an alternate form of Simon.

SIMEON BEN YOHAI (2nd century A.D.), Galilean rabbinic teacher (*tanna*), one of the most eminent disciples of Akiba ben Joseph (*q.v.*) and the author, or editor, of the *Sifre*, a halakic midrash on Numbers and Deuteronomy, and of a midrash on Exodus, *Mekilta di Rab Shimeon ben Yohai* (*see* MIDRASH). The latter work, used by Maimonides, was lost for centuries and was not rediscovered and printed till the 20th century. Simeon's most eminent pupil was the patriarch Judah ha-Nasi, editor of the *Mishna*. Little is known of Simeon's life, but he became the focus of a cycle of legends centring about his concealment in a cave for 13 years, when he was pursued by Roman authorities. He advocated the ascetic ideal of total devotion to Torah learning. Later talmudic authorities concluded that "many tried to follow the ways of Simeon, but they did not succeed" (*Berachoth* 35b).

In the development of Jewish Law, Simeon stressed the importance of seeking "the reasons of the verses"—*i.e.*, the social purpose of the Torah (*Baba Metsia* 115a). He followed this rationalist procedure in ritual as well as in civil matters—for example, in exempting Ammonite and Moabite women from the prohibition "to enter into the assembly of the Lord" (Deut. 23:4–5 [3–4]; *Yebamoth* 77a, *Sifre* 249), since clearly the women could not be blamed for the inhospitality that had occasioned the prohibition.

Possibly because of Simeon's reputation as a miracle worker and an ascetic, the classic work of Cabala, the *Zohar* (13th century), was attributed to him. All modern scholars except the Orthodox are agreed that the *Zohar* was the work of Moses de León (*see* CABALA). Simeon's grave at Meron, near Safad in Galilee, became a shrine for the *Hasidim* and Oriental Jews, and even today *Lag Beomer* (33rd day after the second day of Passover), the traditional anniversary of his death, is observed as a festive day of remembrance, with thousands participating in the burning of garments and other ceremonies at the grave.

For Simeon's *Mekilta*, *see* M. M. Kasher (ed.), *Horambam Vehamechilta* (1943); for the school of Akiba, *see* A. J. Heschel, *Torah min Hashomayim* (1964). (J. B. A.)

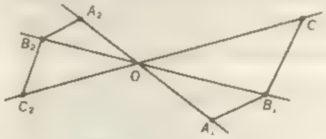
SIMEON METAPHRASTES (also surnamed LOGOTHETE and MAGISTROS) (second half of 10th century A.D.) was the most famous Byzantine hagiographer. His *Menologion* in ten volumes contains 148 lives of saints, arranged according to their feast days from Sept. 1 to Aug. 29. Most of these texts had existed previously but were not well written; Simeon gave them a stylistic revision (Gr. *metaphrasis*). This new *Menologion* was so popular that older collections ceased almost entirely to be copied; hence the disappearance of so many pre-metaphrastean lives. Simeon also wrote letters, religious poetry, and a universal chronicle which raises particularly complicated problems of textual criticism. His feast day is Nov. 28 in the Greek Church.

BIBLIOGRAPHY.—H. Delehaye, *Bibliotheca hagiographica Graeca*, 2nd ed., pp. 237, 267–292 (1909); A. Ehrhard, *Überlieferung und Bestand der hagiographischen und homiletischen Literatur der griechischen Kirche*, vol. 2, pp. 306–709 (1938); H. G. Beck, *Kirche und theologische Literatur im byzantinischen Reich*, pp. 570–575 (1959). For the chronicle *see* G. Moravcsik, *Byzantinoturcica*, 2nd ed., vol. 1, pp. 515–518 (1958). (F. HA.)

SIMEULUE, the northernmost island of the chain off the west coast of Sumatra, Indonesia, and administratively part of Atjeh (*q.v.*) province. Pop. (1960 est.) 25,000. The island, about 61 mi. (98 km.) long, is hilly, the coasts being rocky and reefbound. Sinabang, on a bay in the southeast, is the capital and port, from where there is irregular connection with Tapaktuan on the Sumatran mainland. There is also a cable between Simeulue and Singkil, in Atjeh. Other small places on the coast are Sibigo, Laehobang, Lassihing, and Urung. The Banjak islands, 67 in all, lie southeast of Simeulue; they are prolific in coconut palm and are a source of copra. The inhabitants of the island were formerly a prey of the Achinese (*q.v.*, Atjehnese) slave traders. (J. O. M. B.)

SIMFEROPOL, a town and the administrative centre of the Krymskaya (Crimea) Oblast' of the Ukrainian Soviet Socialist Republic, U.S.S.R., stands on the left bank of the Salgir River, where it emerges from the Crimean Mountains. Pop. (1959) 186,167. Although there has been settlement on the site since ancient times, the modern town of Simferopol was founded by the Russians in 1784, adjacent to the Tatar town of Ak-Mechet, after the annexation of the Crimea. There are large engineering industries producing machine tools and equipment for chemical and food-processing factories, power stations, and railways. Tinned foods, jam, wine, tobacco and cigarettes, perfume, and knitwear are also produced. Simferopol's many educational and cultural institutions include agricultural, medical, and pedagogic institutes, a research institute of mineral resources, a picture gallery, and three theatres. (R. A. F)

SIMILAR FIGURES in geometry have the same shape but not necessarily the same size; *e.g.*, any two squares are similar: if they are of the same size, they are also congruent, or identically equal. Any two circles are similar, but not any two ellipses, for ellipses may have different shapes (*see* ELLIPSE). Solids may also be similar, as in the case of cubes or of spheres. Similarity of figures is the basis of trigonometry (*q.v.*) and of indirect measures in general. Speaking more precisely, two systems of points $A_1, B_1, C_1, \ldots$ and $A_2, B_2, C_2, \ldots$, are said to be similar when they can be so placed that all lines $A_1A_2, B_1B_2, C_1C_2, \ldots$ joining corresponding points form a pencil whose vertex O divides each line into segments having a constant ratio r. In the figure that is shown here, the constant ratio is 2/3. Two figures are said to be similar when their systems of points are similar. The point O is called the centre of similitude. If $r = 1$, the figures are said to be symmetric with respect to the centre O.



SIMILAR FIGURES

See also ANALYSIS: *Analysis and Space;* ANALYTIC GEOMETRY: *Transformation;* CONGRUENCE; GEOMETRY: *Geometrical Transformations.*

SIMLA, a town and district in the Ambala division, and the headquarters of the union territory of Himachal Pradesh, India. Town pop. (1961) 42,597. Simla, a cosmopolitan town and one of the most popular hill stations in India, lies 175 mi. (282 km.) N of Delhi, with which it is linked by both road and rail. It occupies a ridge of the lower Himalayas, running east and west for about 6 mi. (10 km.). The ridge culminates in the east in the summit of Jakko (8,053 ft. [2,455 m.]) and in the west in the Observatory Hill (7,050 ft. [2,149 m.]). To the north a beautiful wooded spur, known as Elysium, branches from the main ridge. The pattern of the settled area, which extends from 6,500 to 8,000 ft. (2,000 to 2,500 m.) above sea level, is like an irregular crescent. The higher parts are occupied by wealthy families, who live in villas and bungalows around Kaithu, Elysium, Jakko, and the central belt. On Observatory Hill stands the President's Lodge, formerly the summer residence of the British viceroy. The focal point of the town is beneath Jakko. There are the main libraries, cinemas, hotels, the town hall, and Christ Church. The main road and the only real shopping centre is the Mall, which runs from end to end of the town just beneath the ridge. Below the eastern part of the Mall is the tin-roofed Bazaar, a densely built area, with houses clinging to a steep slope, which is approached by a road at each end, and is otherwise accessible only by means of steps and tortuous alleys.

Near the railway station, in the southwest, are military and government offices, and three miles to the west is the cantonment of Jutogh. Besides several sanatoria the town has two undergraduate colleges and a women's training college attached to Panjab University. Simla has splendid scenery with steep valleys on either side, thickly forested with deodar; rhododendrons clothe the slopes up to the limit of perpetual snow. To the north is the network of the Himalayan ranges rising one above the other, with snow-covered crests in the distant north. To the south are the Kasauli and Sabathu hills. The town has a mean temperature of 4° C (39° F) in January and 19.5° C (67° F) in July. Average annual rainfall (most of it in July and August) is 63 in. (1,600 mm.). Snow often falls in winter.

Simla was used as a rest area for troops after the Gurkha War, 1814–16. The first English house was built there in 1819. The governor general, Lord Amherst, spent a part of the summer in Simla in 1827 and after that it began to grow in popularity as a health resort, government departments, commercial firms, the army, and tourists making increasing use of it. From 1865 to 1939 Simla was the summer capital of the government of India. The British set the fashion of a definite seasonal migration to this hill station to escape the summer heat of the plain, and wealthy Indians followed their lead. After the outbreak of World War II, the essential government departments remained in Delhi and those of lesser importance in Simla. From 1947 until 1953 Simla was the headquarters of the East Punjab Government, subsequently becoming the headquarters of the government of Himachal Pradesh.

SIMLA DISTRICT, the smallest district of the former Punjab, has an area of 254 sq.mi. (658 sq.km.) and a population (1961) of 112,653. (O. P. B.)

SIMMEL, GEORG (1858–1918), German philosopher and sociologist whose fame as an original scholar rests on several lengthy and brilliantly written essays on sociological methodology and analysis, although during the last decade of his life he devoted himself mainly to metaphysics and aesthetics. He became well known to U.S. sociologists, in part because of the translations and commentaries on his works by A. W. Small, one of the important early sociologists.

Simmel was born in Berlin, March 1, 1858, and was appointed a lecturer in philosophy at Berlin University in 1885, where he remained until called to the chair of philosophy at Strasbourg University in 1914. He died there on Sept. 28, 1918. Simmel is noted for his analytical study of the forms of social interaction. He sought to isolate the general forms or recurrent regularities of social interaction from the specific content of association in concrete types of activity, such as political, economic, and aesthetic. Hence, his sociological writings were abstract rather than realistically descriptive of life.

In his studies of social interaction Simmel gave special attention to the problem of authority and obedience. His methodological contributions did much to clarify the scope of sociology, to give it greater precision, and to establish it as a basic social science in Germany. Simmel made one important effort to apply his abstract principles to the interpretation of social behaviour in a specific field, economics, in his essay *The Philosophy of Money.* He stressed the role of a money economy in specializing social activity and depersonalizing individual and social relationships, a concept later developed more fully by Werner Sombart in his famous work on capitalism. Simmel's writings on the philosophy of history, ethics, religion, and art are also important.

See Rudolf Heberle, "The Sociology of Georg Simmel," ch. xi, in H. E. Barnes *et al., An Introduction to the History of Sociology* (1948). (H. E. BAR.)

SIMMS, WILLIAM GILMORE (1806–1870), U.S. regional novelist, was born in Charleston, S.C., April 17, 1806.

Edgar Allen Poe stated in 1844 that Simms had "more vigor, more imagination, more movement and more general capacity than all our novelists (save Cooper) combined." A 20th-century critic, J. B. Hubbell, stated that Simms "rather than Poe is the central figure of the literature of the Old South. He knew personally most of the southern writers of the time, and he more than any other man stimulated them to write and to publish. But he was a national as well as a sectional figure, and he was the most important literary link between North and South." (J. B. Hubbell, *The South in American Literature*, p. 572, Duke University Press, 1954.) Of Simms' 82 volumes, his historical novels are most important. His field of fiction is the frontier of the lower South—Spanish explorations of the 16th century, the settlement of South Carolina in the 17th, the revolution in South Carolina in the 18th, and southwestward migration in the 19th, preserving for the social historian an authentic and salty cast of woodsmen, Indians, Negro slaves, partisan fighters, outlaws, adventurers, loyalists, shyster lawyers, land speculators, gamblers, half-breeds, and typical ladies and gentlemen of what to Simms was a heroic age.

Motherless at two, Simms was reared by his grandmother, while his Scotch-Irish father fought in the Creek wars and under Jackson at New Orleans. Simms lived a vicariously adventurous childhood through his father, while absorbing history through his storytelling grandmother who had lived through the Revolution. He attended public schools four years. Upon entering the College of Charleston at the age of ten he knew enough French, Latin, German, and Spanish to dabble in translations. At the age of 12 he completed the study of materia medica, and leaving college became a druggist's apprentice. He began publishing poetry in Charleston papers at 16, edited a magazine and published a volume of poetry at 19, married at 20, and was admitted to the bar at 21.

Simms was a prodigious worker, whether at Woodlands Plantation in winter, Charleston in summer, or on yearly publishing trips north. As state legislator and magazine and newspaper editor (the *Charleston City Gazette*), he became embroiled in political and literary quarrels. From Charleston and the South he nevertheless received lifelong praise approaching adulation; from the North, wide audience and eminent literary friendships. Though he was shadowed by the defeat of the Confederacy, the death of his second wife, poverty, and the destruction of his home and library by Gen. William T. Sherman's stragglers, the five volumes of his published letters attest a gallant figure, the richness of whose performance was too long underestimated by literary historians. He died on June 11, 1870, in Charleston.

Simms' chief fault was writing too much, too carelessly, and with too frequent use of stock devices; he was at his best the master of a racy and masculine English prose style. His gift as a charming teller of tales in the oral tradition, and the care of an antiquarian in preparing historical materials, triumph over hasty composition, whether in *Vasconselos* (1853) in a 16th-century setting; *The Yemassee* (1835) in a colonial setting; his revolutionary series—*The Partisan* (1835), *Mellichampe* (1836), *The Kinsmen* (1841), *Katherine Walton* (1851), *Woodcraft* (1854), *The Forayers* (1855), *Eutaw* (1856); the best of his border romances—*Richard Hurdis* (1838) and *Border Beagles* (1840); his

short story collection *The Wigwam and the Cabin* (1845); or in his *History of South Carolina* (1840). Of his 19 volumes of poetry the collected *Poems* (1853) deserve mention. Most popular of his biographies were *The Life of Francis Marion* (1844) and *The Life of Chevalier Bayard* (1847). His literary criticism is represented in *Views and Reviews of American Literature* (1845).

See A. S. Salley and Donald Davidson (eds.), *Letters of William Gilmore Simms*, introd., vol. i (1952); Vernon L. Parrington, *Main Currents in American Thought*, vol. ii (1927). (M. C. S. O.)

SIMNEL, LAMBERT (*c.* 1475–1535), impostor and claimant to the English crown, the son of an Oxford joiner, was a cat's-paw in the conspiracies hatched to restore the Yorkist line after the victory of Henry VII (1485). A young Oxford priest, Richard Symonds, seeing in the handsome boy some alleged resemblance to Edward IV, determined to exploit him. In 1486, the rumour that the "princes in the Tower," Edward's children, were still alive, suggested that Simnel might be passed off as one of them; the year after, the false report of the death in the Tower of another young Yorkist, Edward, earl of Warwick, changed the impersonation. Symonds took his charge to Ireland where the Yorkist interest was strong, especially with the powerful house of Fitzgerald, earls of Kildare. In May 1487 Simnel was accepted as earl of Warwick and crowned at Dublin as King Edward VI, and despite Henry VII's efforts (which included parading the real earl through the streets of London) the conspiracy spread. In June 1487, Simnel landed in Lancashire, supported by 2,000 German mercenaries provided by Edward IV's sister Margaret, duchess of Burgundy, and in company with a genuine Yorkist claimant, John de la Pole, earl of Lincoln, who had abandoned his temporary loyalty to the Tudors. Although Lancashire gave them no support, the threat was, nevertheless, very real: Henry VII had to face it with what strength he could muster in the presence of a nation used to such disorders and ready to wait upon the outcome. This was settled in the Battle of Stoke (June 1487); Lincoln was killed, Simnel and Symonds were taken, the other Yorkist leaders disappeared. Symonds was kept in prison, but Henry VII displayed only sardonic contempt for Simnel whom he recognized to have been a harmless dupe. Employed in the royal kitchens, the pretender made a modest career for himself, dying soon after 1534.

See J. Gairdner, *Henry VII* (1889); J. D. Mackie, *The Earlier Tudors, 1485–1558* (1952). (G. R. E.)

SIMOCATTES, THEOPHYLACT (or THEOPHYLACTUS; Gr. THEOPHYLAKTOS SIMOKATTES) (fl. 6th–7th century A.D.), Byzantine writer whose history is a valuable source for the reign of Maurice (582–602). He was an Egyptian, who after legal training held office at Constantinople under Heraclius (610–641). His history, for which he had access to official sources of information, shows keen awareness of the Greek tradition of historiography, for he was acquainted with the work of Herodotus and Diodorus as well as later historians including near contemporaries. He also wrote a work on natural history (*Quaestiones naturales*) and a number of literary exercises in epistolary form.

His *Historia* was edited by C. de Boor in the "Teubner Series" (1887) and A. I. Bekker in the Bonn corpus (1834). Other works edited include *Quaestiones naturales*, ed. by J. L. Ideler, *Physici et medici graeci minores*, vol. i, pp. 168–183 (1841), and, together with the *Epistulae*, by J. F. Boissonade (1835); *Epistulae* ed. by R. Hercher, *Epistolographi Graeci*, pp. 763–787 (1873).

See K. Krumbacher, *Geschichte der byzantinischen Litteratur*, pp. 247–251, 2nd ed. (1897); G. Moravcsik, *Byzantinoturcica*, vol. 1, pp. 544–548, 2nd ed. (1958). (J. M. HY.)

SIMON, SAINT, one of the 12 apostles, always mentioned in the last of the three groups of four names in the apostle lists (Mark 3:18; Matt. 10:4; Luke 6:15; Acts 1:13). In Mark and Matthew he bears the surname *Kananaios*, or the Cananaean, often wrongly interpreted as meaning "from Cana" or "from Canaan." It is in fact the Greek transliteration of an Aramaic word *qan'anaya*, meaning "the Zealot," which is precisely the title given him by Luke both in his Gospel and in Acts. Whether he was one of the group of Zealots (the Jewish nationalistic party prior to A.D. 70) cannot be answered for lack of evidence. Nothing further is known about him from the New Testament. He is said to have preached the Gospel in Egypt and then joined St. Judas (Thaddaeus) in Persia, where both were martyred (according to the apocryphal Passion of Simon and Jude). Oct. 28 is the feast of SS. Simon and Jude. *See* also APOSTLE. (J. A. FI.)

SIMON OF ST. QUENTIN (fl. 1247), Dominican mission-traveler and diplomatist who accompanied the Dominican embassy, under Friar Ascelin or Anselm, which Pope Innocent IV sent in 1247 to the Mongols of Armenia and Persia. Large sections of Simon's history of this embassy have been preserved in Vincent of Beauvais's *Speculum historiale*. The embassy proceeded to the camp of Baiju, or Bachu *Noyan* (*i.e.*, "General" Baiju, *Noyan* signifying a commander of 10,000), at Sitiens in Armenia, 59 days' journey from Acre. The papal letters were translated into Persian, and thence into Mongol, and so presented to Baiju. But the Tatars were irritated by the haughtiness of the Dominicans and on July 25, 1247, the latter were dismissed with the *Noyan's* reply, dated July 20, which complained of the high words of the Latin envoys and commanded the pope to come in person and submit to the Master of all the Earth (the Mongol emperor). The mission thus ended in complete failure.

BIBLIOGRAPHY.—W. W. Rockhill, *Rubruck*, pp. xxiv–xxv (Hakluyt Society, 1900); C. R. Beazley, *Dawn of Modern Geography*, vol. ii, p. 277 (1901), and *Carpini and Rubruquis*, pp. 269–270 (1902). (WM. C. B.)

SIMON, SIR JOHN (1816–1904), English sanitary reformer, was born in London on Oct. 10, 1816. He studied medicine at St. Thomas's Hospital, and from 1840 held minor surgical posts at King's College Hospital. In 1847 he became surgeon and lecturer in pathology at St. Thomas's, retaining these posts until 1876. His pathological research led to his election to the Royal Society at the early age of 29, and his inventive and resourceful surgery soon secured him an outstanding position in that art. These scientific pursuits provided the groundwork for his chief occupation as a sanitary administrator. From 1848 to 1855 he held the position of first medical officer of health to the City of London. The much-publicized success of his pioneering efforts there made him a national figure and established the character of the medical officership in English local government thereafter.

Between 1855 and 1876 Simon was the first medical officer to the central government and his most fruitful years, 1858–71, were spent in this post. He gradually created a medical department of state to superintend the national health, established state scientific research, investigated all health concerns and published magisterial reports thereon, perfected the vaccination system, supervised the medical profession, and procured much legislation, including the Sanitary Act of 1866 which for the first time rendered a public health law universal, scientific, and compulsory. On being transferred in 1871 to the local government board, Simon found that the dominant lay secretariat frustrated his aims, and he resigned in 1876. His later life was spent in literary and professional work. He was created knight commander of the Bath in 1887. He died in London on July 23, 1904.

See R. Lambert, *Sir John Simon 1816–1904* (1963). (R. J. LA.)

SIMON, JOHN ALLSEBROOK SIMON, 1ST VISCOUNT (1873–1954), British lawyer and Liberal politician, who held four of the great offices of state: home secretary (1915–16 and 1935–37), foreign secretary (1931–35), chancellor of the exchequer (1937–40), and lord chancellor (1940–45). He was born in Manchester on Feb. 28, 1873, the son of a Congregational minister. Educated at Fettes College, Edinburgh, and at Wadham College, Oxford, he was elected a fellow of All Souls College, Oxford, in 1897, and was given the honorary degree of doctor of civil law by Oxford University in 1926. So devoted was he to Oxford that, at his express wish, his body was cremated in his D.C.L. gown.

Simon possessed exceptional intellectual gifts and a capacity for public affairs that raised him, through a successful career at the bar, to the front rank in politics. Yet, as a public figure, he earned a name for coldness and lack of humanity that both puzzled and wounded him. The explanation was the grief he suffered through the death in 1902 of his first wife Ethel Venables, less than four years after their marriage. In his autobiography, *Retrospect*

(1952), he wrote "The preachers are wrong who speak as though grief and disappointment are calculated to refine character. Far more men and women are soured by misfortune than are injured by prosperity." He married again in 1917 but those 15 years of loneliness made a permanent mark on his character.

Simon was called to the bar in 1899 and was elected Liberal M.P. for Walthamstow, in Essex, in 1906. He was made solicitor general in H. H. Asquith's government in October 1910, when he was knighted. Promoted attorney general in October 1913, he became home secretary in May 1915. There seemed nothing to stop his steady political advance, but in January 1916 he took a decision which excluded him from office until 1931: he resigned the post of home secretary as a protest against the government's decision to introduce conscription during World War I. This was an honorable act but, as Simon admitted years afterward, it was an error of judgment. He lost his seat in 1918 but returned to the House of Commons in 1922 as Liberal member for Spen Valley, in Yorkshire, which he represented until 1940 when he was created a viscount.

Simon was far from idle while out of office. His legal practice earned him more than £50,000 a year. He was chairman of the Indian Statutory Commission which, between 1927 and 1930, examined the constitutional future of India. During the world economic crisis of 1931, Simon abandoned his belief in free trade and advocated protection. Some other Liberals followed him and from 1931 to 1940 he was leader of a separate Liberal National Party that worked closely with the Conservatives. After the election of the National government (October 1931) Simon returned to office in November as foreign secretary. He was unable, acting through the League of Nations and in consultation with the United States, to stop Japan's occupation of Manchuria (1931). Simon became home secretary once more in 1935, and chancellor of the exchequer in 1937. He was created Viscount Simon in May 1940 and was lord chancellor from 1940 to 1945. He died in London on Jan. 11, 1954. (J. F. B.)

SIMON, JULES FRANÇOIS (1814–1896), French philosopher and political leader, the theorist of Radicalism who as prime minister in 1876–77 became a central figure in the formative crisis of the Third Republic, was born on Dec. 27, 1814, at Lorient (Morbihan). His father sent him first to school at Vannes and then (1833) to the École Normale in Paris. He taught philosophy at Caen and Versailles, and became a pupil at the Sorbonne, where he later lectured. He was elected to the National Assembly of 1848 as a liberal, devoted philosophically to the cause of freedom of speech, worship, and thought. He opposed the *coup d'état* of Louis Napoleon in December 1851 and was suspended from his academic post. Until 1863 he devoted himself to historical and philosophical work, producing several books of a semipopular kind. Then he took the necessary oath of allegiance to the empire and was able to secure election to the *Corps Législatif* in 1863. The coalition to which he belonged, the Union Libérale, polled some 2,000,000 votes, most of which came from the large towns. Simon was now firmly set on the path to a political career.

At first it was, for so staunch a liberal, a career of opposition and criticism. In religion a deist and (after 1870) a freemason, Simon brought to the support of the liberal opposition to Napoleon III a store of energy, subtlety, and ability which he concealed behind a serene and detached manner. In 1868 he published *La Politique radicale*, which later became, along with Léon Gambetta's Belleville manifesto of 1869, the basis of the Radical Party program. Reelected in 1869, he became a member of the Government of National Defense set up in Paris after the defeat of Napoleon III's army by the Germans at Sedan had destroyed the Second Empire. Sent to Bordeaux to get Gambetta to agree to peace, Simon was defied by that enthusiast for "a people's war" against the Germans; but on Feb. 18, 1871, Adolphe Thiers made him minister of education, religion, and fine arts in his emergency government.

Conditions were not favourable for Simon's favourite schemes of educational reform. He was bitterly opposed by the clerical parties, and contrived only to make certain reforms in secondary schools, though he deserves credit for achieving the basic task of getting schools in working order again after the upheavals of the war, the invasion, and the Commune of Paris. The author or editor of numerous works on philosophical subjects, and a member of the Académie des Sciences Morales et Politiques from 1863, he was elected to the Académie Française in 1875.

Simon fell from office with Thiers on May 18, 1873. His second tenure of power occurred in very different circumstances. The elections of 1876 returned a strong Republican majority to the Chamber of Deputies. President MacMahon, who succeeded Thiers, was committed, by temperament and affiliations, to a conservative and clerical policy. But the Republican gains obliged him, on Dec. 12, 1876, to invite Simon to form a ministry. Although relatively moderate, the ministry soon became involved in violent conflict with the clerical movements, and on May 16, 1877, MacMahon wrote Simon a letter which was tantamount to dismissal. Simon, although opposed to the more rabidly anticlerical measures of Gambetta or Jules Grévy, was undefeated in the chamber and could have defied the president. He weakly resigned, thereby precipitating the constitutional crisis of *le Seize Mai* (May 16), centring on the question whether ministerial responsibility was owed to the president or to the Chamber. Because events determined that it should be owed to the Chamber, MacMahon himself resigned on Jan. 30, 1879, and the Third Republic became essentially a parliamentary system.

Simon had played a mainly negative part in the crisis. He never regained office and rapidly lost political influence. He continued to engage in political journalism, but became a senator for life. He died in Paris on June 8, 1896, a half-forgotten figure of the formative years of the republic. He had published *Le Gouvernement de M. Thiers* in 1878. *Le Soir de ma journée* appeared posthumously (1902).

See L. Séché, *Jules Simon* (1898). (D. Tn.)

SIMON, RICHARD (1638–1712), French biblical critic, whose clear vision and accurate formulation of biblical problems and especially his insistence on the historicocritical method, pioneered modern biblical study, was born at Dieppe, May 13, 1638, and received his formal education under the Oratorians and the Jesuits, and at the Sorbonne. He joined the Oratorians in 1662, was ordained priest in 1670, was expelled from the Oratorians in 1678, and served as curé in Bolleville until 1682. Thereafter he continued his research and writing privately in Rouen, Paris, and Dieppe, where he died April 11, 1712.

Simon was too advanced for his age, and his method and its results were assailed by Roman Catholics (ten of his writings were placed on the *Index*) and Protestants alike. On his part, Simon opposed Spinoza's attempt to imbue the study of Scripture with rationalism, attacked the Protestant abandonment of tradition, and struck at Roman Catholic neglect of historical (scriptural and traditional) theology.

Paradoxically, the Lutheran biblical critic J. S. Semler coloured Simon's method and views with rationalism, and in this guise they became the staple of rationalistic higher criticism, thus rendering Simon's work further suspect for Roman Catholics. Since 1900, however, he has generally been accorded due recognition as a Catholic scholar.

Simon's best-known work is the *Histoire critique du Vieux Testament.* The first edition (1678) was confiscated; a faulty edition made from the few surviving copies appeared in Amsterdam in 1680, but the best edition, supervised by the author himself, was published at Rotterdam in 1685. Other works include: *Du Texte du Nouveau Testament* (1689); *Des Versions du Nouveau Testament* (1690); *Des Principaux commentateurs du Nouveau Testament* (1693); and *Nouvelles observations sur le texte et les versions du Nouveau Testament* (1695).

Bibliography.—A complete critical study of Simon and his work is still wanting, but the following may be profitably consulted: A. Bernus, *Richard Simon et son Histoire Critique du Vieux Testament* (1869); H. Margival, *Richard Simon et la critique biblique au XVII^e siècle* (1900); F. Stummer, *Die Bedeutung Richard Simons für die Pentateuchkritik* (1912); A. Molien, in *Dictionnaire de théologie catholique*, vol. xiv, col. 2094–2118 (1941); R. Deville, "Richard Simon, critique catholique du pentateuque," in *Nouvelle revue théologique*, 73:723–739 (1951); J. Steinmann, *Richard Simon et les origines de l'exégèse biblique* (1960). (J. P. W.)

SIMONIDES of Ceos (*c.* 556–*c.* 470 B.C.), ancient Greek lyric poet and epigrammatist who appears to have originated the literary form of the epinician ode (*i.e.*, ode in honour of victors in the great games at Olympia) and to have been the first Greek author to write on commission, for fees. A native of Iulis on the Aegean island of Ceos (now Kea [Keos]), he studied music and poetical composition at home, but left the island as a young man and seems to have lived mostly in Athens thereafter. Quotations and allusions in later writers connect him with the "tyrants" of Athens, of Crannon and Larissa in Thessaly, and of Syracuse, where he is said to have died; but he did not depend entirely on patronage—his epinician for an Olympic victor of 520 B.C. is the earliest such poem recorded, and later writers refer to his professionalism.

Though he was often successful in the Athenian competitions for dithyrambic poetry (*q.v.*), he seems to have preferred men to gods as subjects for poetry: the surviving remains of his choral lyrics include fragments of songs in honour of those who fought at Thermopylae and Salamis, and his poem of advice to Scopas of Crannon is extensively quoted in Plato's *Protagoras;* the extant epigrams —for which he was so famous that many which he cannot have written are ascribed to him—were mostly intended as epitaphs or for dedicatory inscriptions. The fragments illustrate his reported comparison of poetry to verbal painting and of painting to mute poetry. Ionian by birth and adaptable and progressive by temperament, he was better fitted than Pindar, the conservative Theban, to express the panhellenic ideals of the new age which developed after the Greek victories over the Persians. Bacchylides (*q.v.*) was his nephew.

Bibliography.—Lyric poems in D. L. Page, *Poetae melici Graeci*, pp. 237–325 (1962); epigrams in E. Diehl, *Anthologia lyrica Graeca*, vol. ii, pp. 84–118 (1925); allusions and anecdotes in J. M. Edmonds, *Lyra Graeca*, vol. ii (1924). *See* also C. M. Bowra, *Greek Lyric Poetry*, 2nd ed. (1961), *Pindar* (1964). (Jn. A. D.)

SIMON MAGUS (1st century A.D.), according to Acts 8:5–24, a magician regarded by many inhabitants of Samaria as "that power of God which is called Great"; after being converted and baptized by the evangelist Philip, he offered money to two of the apostles for the power to impart the Spirit (hence the word "simony"). Rebuked by Peter, Simon requested the apostles' prayers. The conclusion of this story in Acts is probably intended to imply that Simon's repentance was genuine; but later Christian writers, such as Justin, Irenaeus, Hippolytus, and Epiphanius, treat him as the originator of Gnosticism (*q.v.*). Practically nothing is known of the process which produced Simonianism, though presumably before and after his Christian period Simon was regarded as an incarnation of some god's power because of his ability to perform miracles. Apparently he was a Christian long enough to learn something about Jesus and to develop doctrines about himself in relation to christological thought. According to the *Clementine Homilies* (*see* Clementine Literature) he was originally a disciple of John the Baptist; he may conceivably have had some relation to the covenanters of Qumran (for which *see* Dead Sea Scrolls).

At a later time Simon seems to have journeyed to Rome in the company of a Tyrian ex-prostitute named Helen, who he claimed was his First Thought, through whom he himself created the universe. She had been imprisoned by angels and powers who had emanated from her, and had passed through a succession of incarnations, once as Helen of Troy. Simon, the Father, had descended to rescue this "lost sheep" (cf. Luke 15:6). Parts of this myth recall the Oriental myths about the imprisonment of Inanna, Sumerian queen of heaven, and other goddesses. In Rome Simon probably made some converts, since in the second century at Rome it was claimed that a statue bearing an inscription to the god Semo Sancus had been erected *Simoni deo sancto* (Justin, *Apol.*, i, *26, 2*).

Probably in the 2nd century, Simonians stated that Simon had appeared among the Jews as Son and had seemed to suffer, had descended in Samaria as Father, and had come among the rest of the nations as Holy Spirit. This quasi-Trinitarian picture is clearly Christian in origin, as is the Simonian statement that those who have placed their hope on Simon and Helen are "saved by his grace, not by just works" (cf. Eph. 2:8–9).

The Simonians regarded Zeus and Athena as equivalent to Simon and Helen, and the myth of the birth of Athena is not unlike that of Helen's origin. The Simonians made use of all kinds of materials and developed their theology (or rather theologies; the "Great Pronouncement" which Hippolytus ascribes to Simon has connections with Stoicism and with other Gnostic systems) in highly imaginative fashion. How much of their thought goes back to their founder is impossible to determine, but it cannot be demonstrated that the essentials of the system are pre-Christian.

See R. P. Casey, "Simon Magus," in F. J. F. Jackson and K. Lake, *The Beginnings of Christianity*, vol. 5 (1933); R. M. Grant, *Gnosticism and Early Christianity* (1959). (R. McQ. G.)

SIMONSTOWN (Simonstad), a town and naval base of Cape of Good Hope Province, S.Af., lies 22½ mi. (36 km.) S. of Cape Town by road. Pop. (1960 census) 8,635, comprising 3,518 Europeans, 3,584 Coloureds, 1,414 Bantu, and 119 Asians. Among places of interest are the Admiralty House (1814), the church of St. Francis of Assisi (1814, rebuilt 1834—the first English church founded in South Africa), the still older naval cemetery, and the Residency (1776), with its former slave quarters. There are facilities for bathing, golf, and tennis. An industrial site has been established and there is a fish-oil refinery, fishing being one of the main occupations. The town is linked by bus and electric train with Cape Town.

Named after Simon van der Stel, governor of the Cape during 1679–99, Simonstown was a Dutch military and naval depot from 1741. In 1814 it became the headquarters of the British South Atlantic Naval Squadron. In 1957 it was transferred to South Africa, Britain being guaranteed continued use of its facilities.

SIMONY is defined as the buying or selling for a temporal (*i.e.*, material) price something spiritual or closely connected with the spiritual. More widely, it is any contract of this kind forbidden by divine or ecclesiastical law. The name is taken from Simon (Acts 8:18), who endeavoured to buy from the apostles the power of conferring the gifts of the Holy Ghost (*see* Simon Magus). Simony in the form of buying holy orders or church office was virtually unknown in the first three centuries of Christian life but became familiar when the church had positions of wealth and influence to bestow. The first legislation on the point was the second canon of the Council of Chalcedon (451); thenceforward prohibitions and penalties were reiterated against buying or selling promotion to the episcopate, priesthood, and diaconate. Later the area of the offense was held to cover all traffic in benefices, and later still it was widened by ecclesiastical law to cover pecuniary transactions on Masses (apart from the authorized offering), blessed oils, and other consecrated objects, and finally included any lucrative action, such as resignation or withdrawal in favour of another, connected with an ecclesiastical benefice. In the case of holy orders the penalty was always suspension of both the bestower and the recipient and deprivation of any benefice concerned.

From an occasional scandal, simony became endemic in Western Europe in the 9th and 10th centuries, when regalian and other proprietary rights over prelacies and churches prevailed throughout Europe in feudal society, and the "heresy" of simony became a major object of attack on the part of the papacy in the age of Gregory VII. The traditional canonical sanctions were rigorously applied, and simony at the high levels of church life became once more occasional rather than normal, though throughout the Middle Ages numerous cases occurred in the episcopate and even in the papacy. More common still were the purchase of benefices and the organized traffic in offices both at the Curia and in the various regional churches.

Simony in its most flagrant forms gradually disappeared from the 16th century onward with the disendowment and secularization of church property, but in the Church of England, where the medieval system has continued in a modified form to the present day, simony has remained a real, if rare, source of offense. In the modern Roman Catholic Church relevant legislation has chiefly dealt with abuses connected with Mass stipends.

Bibliography.—A. Bride in A. Vacant *et al.* (eds.), *Dictionnaire de théologie catholique*, vol. 14, col. 2141–60 (1941); E. Lesne, *Histoire de la propriété ecclésiastique en France*, vol. 1–4 (1910–38); J. Leclercq,

"Simoniaca heresis," *Studi Gregoriani* 1:523–530 (1947); R. A. Ryder, *Simony: an Historical Synopsis and Commentary, Catholic University of America Canon Law Studies*, vol. lxv (1931). (M. D. K.)

SIMOOM: *see* WIND.

SIMPLICIUS, SAINT (d. 483), pope from 468 to 483, was born at Tivoli. During his pontificate the Western Empire was overthrown, and Italy passed into the hands of the barbarian king Odoacer (476). In the East, the usurpation of the empire by Basiliscus (475–476), who supported the Monophysites, gave rise to many ecclesiastical troubles. The emperor Zeno, having procured the removal of Basiliscus, endeavoured to compound with the Monophysite party; and Acacius, bishop of Constantinople, who had previously been on the pope's side in defense of the Council of Chalcedon, abandoned Simplicius and subscribed to the *Henoticon*, the conciliatory document promulgated in 482 by the emperor. Simplicius died on March 2, or perhaps March 10, 483. He is commemorated on March 2. *See* also PAPACY.

SIMPLICIUS (6th century A.D.), Greek philosopher, a native of Cilicia, was a pupil of Damascius at Athens and of Ammonius at Alexandria; his work displays the influence of both these schools of Neoplatonism (*q.v.*). The school of philosophy at Athens having been disendowed and its teaching forbidden (529), Damascius, Simplicius, Priscianus, and four others resolved in 531 or 532 to seek refuge with Chosroes, king of Persia, but within two years they returned to Greece.

After his return Simplicius wrote commentaries on Aristotle's *De coelo*, *Physica*, *De anima*, and *Categoriae*, which, with a commentary on the *Enchiridion* of Epictetus, have survived. These contain many valuable fragments of the older philosophers as well as of his immediate predecessors. One of his general principles is that Aristotle is in harmony with Plato, being opposed only to a superficial understanding of his words. (D. J. A.)

SIMPLON PASS, Switzerland, lies in the Lepontine Alps at 6,578 ft. (2,005 m.) on the watershed between a north-flowing tributary of the Rhône and a south-flowing tributary of the Toce. It was not till the mid-13th century that it attained any importance as a routeway and it was only with the building (1800–07) of a carriage road (which negotiates the Gondo Gorge) by Napoleon that it began to compete with the other Alpine passes as a major link between central and southern Europe. Near the summit of the pass is the hospice, first mentioned in 1235 as in the charge of the Order of St. John; it is now occupied by the Augustinians, having been given over to that order by Napoleon in gratitude for their hospitality on his passage (1800) over the Great Saint-Bernard Pass. Bellevue Simplon-Kulm Hotel stands near the hospice. A railway tunnel opened in 1906 passes 4,430 ft. (1,350 m.) beneath the pass. It is about 12 mi. (20 km.) long and connects Brig, Switz., with Iselle, Italy. When the pass is closed to road traffic, usually from mid-October to early April, cars are transported through the tunnel by train. (A. F. A. M.)

SIMPSON, SIR JAMES YOUNG (1811–1870), Scottish obstetrician, whose fame rests principally upon his pioneer work in the introduction of anesthesia into midwifery practice, was born at Bathgate, Linlithgowshire, Scot., on June 7, 1811. He was sent at the age of 14 to Edinburgh University and obtained his M.D. degree in 1832. He was elected senior president of the Royal Medical Society of Edinburgh in 1835 and in 1837–38 acted as deputy for John Thomson, professor of pathology. In 1840 he was appointed to the chair of midwifery in the university. When news of the first trials of ether reached Scotland in 1846, Simpson immediately realized its potential value and tried it in obstetric practice. In March 1847 he advocated its use in a paper read before the Royal Medical Society. He continued, however, to make systematic search for more efficient anesthetics, carrying out many self-experiments with his assistants, and on Nov. 4, 1847, he discovered the anesthetic properties of chloroform. He published a full account of chloroform (1847) in which he stressed its advantages over ether and strongly advocated its use in surgical operations and in obstetric practice.

The use of chloroform for the relief of pain in childbirth was violently opposed both on medical and on theological grounds, but Simpson's uncompromising advocacy and, finally, the administration of chloroform to Queen Victoria at the birth of Prince Leopold in 1853 silenced all opposition.

Apart from his great work on anesthesia, Simpson made contributions of permanent value to many departments of obstetrics and gynecology. His principal writings are *Obstetric Memoirs and Contributions* (2 vol., 1855–56); *Selected Obstetrical and Gynaecological Works* (1871); *Anaesthesia, Hospitalism, Homoeopathy, etc.* (1871); *Clinical Lectures on the Diseases of Women* (1872). He was also greatly interested in archaeology and in the history of medicine; two volumes of his *Archaeological Essays* were published in 1873. He was appointed one of the queen's physicians for Scotland in 1847 and in 1866 was created a baronet. He died on May 6, 1870, in London.

BIBLIOGRAPHY.—John Duns, *Memoir of Sir J. Y. Simpson, Bart.* (1873); E. B. Simpson, *Sir James Young Simpson* (1896); H. L. Gordon, *Sir James Young Simpson and Chloroform* (1897). (W. J. BP.)

SIMROCK, KARL JOSEPH (1802–1876), German literary scholar and poet, was born at Bonn on Aug. 28, 1802. Simrock's great achievement was to make available for his generation a vast body of earlier Germanic literature, either by translation into modern German (*Das Nibelungenlied*, 1827), by rewriting (*Das Amelungenlied*, 1843–49), or by editing (*Die deutsche Volksbücher*, 18 volumes, 1839–67). He also translated many of Shakespeare's plays and in 1855 published a *Handbuch der deutschen Mythologie*.

As a young man Simrock studied law at Bonn and Berlin but also attended literary lectures by A. W. Schlegel and Karl Lachmann (*qq.v.*). In 1830 he was dismissed from his judicial appointment for writing a poem in praise of the July Revolution, but the death of his father shortly after left him in comfortable circumstances and he retired to Bonn to devote himself to literary studies. He died there on July 18, 1876, having been made an honorary professor in 1850.

See *Ausgewählte Werke*, ed. by G. Klee, with biography, 12 vol. (1907); N. Hocker, *K. Simrock* (1877). (G. T. HU.)

SIMS, WILLIAM SOWDEN (1858–1936), U.S. naval officer who by his persistent efforts to improve ship design, fleet tactics, and naval gunnery influenced the Navy, so it was said, "more than any other man who ever wore the uniform." He was born on Oct. 15, 1858, at Port Hope, Ont., where his father, a U.S. engineer, was then employed. The family moved to Pennsylvania in 1872 and William, after being rejected once, entered the U.S. Naval Academy. After graduating in 1880 he served almost continuously on sea duty for the next 17 years. While attached to the nautical school ship "Saratoga" (1889–93) he prepared a text on navigation that was used for many years in both the naval service and the merchant marine.

At the end of the century (1897–1900) Sims served as naval attaché to the U.S. embassies in Paris and St. Petersburg. His observation of foreign navies convinced him of the comparative inferiority of the U.S. Navy. From the China Station, where he served from 1900 to 1902, he sent to the Navy Department a series of remarkable reports setting forth the defects in U.S. ships and naval marksmanship. In several papers he described a new system of gunnery that had been developed by Capt. Percy Scott, the great British officer who was also serving on the China Station. Receiving no satisfactory response from the Navy Department, Sims wrote directly to Pres. Theodore Roosevelt, who brought him to Washington as inspector of target practice. After seven years in this position Sims returned to sea duty with the reputation of "the man who taught the Navy how to shoot."

In the next eight years (1909–17) he was captain of two different battleships and commander (1913–15) of the Atlantic Torpedo Flotilla where he developed an imaginative tactical doctrine for destroyers, then a new type of naval vessel. After the United States entered World War I he became commander of the U.S. Naval Forces Operating in European Waters, with headquarters in London. His most notable contributions to the victory at sea were his part in promoting adoption of the convoy system and his great success in working cooperatively with the naval commands of the Allied powers.

Returning to the United States in 1919 Sims was reassigned,

at his own request, to duty as president of the Naval War College in Newport, R.I. Resenting awards of decorations to many officers whom he considered undeserving, he refused to accept his own Distinguished Service Medal. From Newport he wrote a letter to the secretary of the navy severely criticizing the manner in which the Navy Department had administered the naval effort in the war. There followed an extended congressional investigation which produced inconclusive and partisan judgments. Shortly thereafter Sims, who had earlier reached the temporary rank of admiral, retired from the Navy on Oct. 15, 1922. He died in Boston, Mass., on Sept. 28, 1936.

See Elting E. Morison, *Admiral Sims and the Modern American Navy* (1942). (E. E. Mo.)

SIN, name of the moon god of the Babylonian-Assyrian pantheon, derived from old-Akkadian *zu-en* (usually written *en-zu*), "lord of wisdom." His Sumerian name is Nanna. Son of Enlil, he was head of the secondary cosmic triad of the Akkadian pantheon, its two lesser members being his son and daughter, Shamash and Ishtar (*qq.v.*). (In Sumer this triad was absorbed into the corresponding Sumerian triad of Nanna, Utu, and Inanna.)

The chief seats of Sin's worship were Ur in southern Babylonia and Harran in northern Assyria, where he was styled "lord of Harran," but temples are found in all the large cities of Sumer, Babylonia, and Assyria in his honour. His chief function was to be the "lamp of heaven and earth." He and his son Shamash open the gate of Anu, the chief god of heaven. He is "lord of the months."

Sin is represented with flowing beard and with the crescent above his horned tiara, on seal cylinders and on the great stela of King Ur-Nammu (*c.* 2130 B.C.). Later, his symbol is often the crescent alone. In the astral-theological system he is represented by the number 30 (obviously connected with the 30 days' average duration of the moon's course); while the planet Venus (*i.e.*, his daughter Ishtar) is represented by the number 15. As "lord of signs" and of "decisions" (*e.g.*, at the waning of the moon) he is named in omen literature.

In Assyria the cult of the moon god seems to have been less prominent than in Babylonia. He occurs frequently in inscriptions alongside Shamash and Ashur, the head of the Assyrian pantheon; but his shrines do not seem to have great local or national importance, except at Harran. Nabonidus, the last king of Babylon (556–539), whose mother was a priestess of Sin at Harran, inaugurated a movement to elevate the cult of Sin to the supreme place in the pantheon.

At an earlier time the emphasis placed upon moon worship by Sargon of Akkad (*c.* 2350–2300 B.C.) may have been due to his western Semitic origins. In Arabia and among all the Semitic peoples of western Asia the moon god was a most important deity.

The consort of Sin was Ningal to whom a special temple was built at Ur, and her cult was widely known in Syria. The cult of the Babylonian Sin seems to have been particularly favoured by the Assyrian colony in Cappadocia early in the 2nd millennium B.C., and there is evidence of his cult among the Hittites of Anatolia and Syria.

See also BABYLONIA AND ASSYRIA: *Religion.*

BIBLIOGRAPHY.—A. Deimel, *Pantheon Babylonicum* (1914); E. (Paul) Dhorme, *Les Religions de Babylonie et d'Assyrie* (1949); A. Falkenstein and W. von Soden, *Ṣumerische und Akkadische Hymnen und Gebete* (1953). (T. FH.)

SIN is the name given to moral evil when considered from the religious point of view. In ancient Greek thought sin was looked upon as in essence a failure on the part of man to achieve his true self-expression and to preserve his due relation to the rest of the universe; it was attributed mainly to ignorance. Christianity, while not denying this aspect, sees sin as a deliberate violation of the will of God attributable to man's pride, self-centredness, and disobedience. In this respect the Christian view is in direct continuity with the Old Testament and with Judaism. While insisting more strongly than most religions upon the gravity of sin, both in its essence and in its consequences, Christianity has emphatically rejected the Manichaean doctrine that either the created world as a whole or the material part of it is inherently evil (*see* MANICHAEISM). It holds that evil is the result of the misuse of their free will by created beings, both angelic and human, and that the body, with its passions and impulses, is to be neither ignored nor despised but sanctified; in the Bible, the "flesh" which is spoken of disparagingly is not the human body but human nature in rebellion against God.

Theologians divide sin into "actual" and "original." Actual sin is sin in the ordinary sense of the word and consists of evil acts whether of thought, word, or deed. Original sin (the term can be misleading) is the morally vitiated condition in which we find ourselves at birth as members of a sinful race. In Gen. 3 this is depicted as an inherited consequence of the first human sin. Theologians differ as to the interpretation of this narrative, but it is agreed that original sin, however mysterious its origin and nature may be, arises from the fact that we come into the world not as isolated individuals but as members of a corporate human race inheriting both good and evil features from its past history.

Actual sin is subdivided, on the basis of its gravity, into mortal and venial. This distinction is often difficult to apply but can hardly be avoided. Mortal sin is a deliberate turning away from God, sin in a grave matter committed in full knowledge with full consent of the will; until it is repented it cuts the sinner off from God's sanctifying grace (*see* GRACE). Venial sin, while it weakens the sinner's union with God, is not a deliberate turning from him and so does not wholly block the inflow of sanctifying grace.

Actual sin is also subdivided into material and formal. Formal sin is both wrong in itself and is known by the sinner to be wrong; it therefore involves him in personal guilt. Material sin consists of an act which is wrong in itself (because contrary to God's law and man's moral nature) but is not known to be wrong and so is not culpable.

See also SINS, SEVEN DEADLY; and references under "Sin" in the Index.

See R. C. Mortimer, *Elements of Moral Theology*, ch. iii–v (1947); N. P. Williams, *Ideas of the Fall and of Original Sin* (1927). (E. L. M.)

SINAI, the name given to the triangular peninsula lying between Egypt, Israel, and Saudi Arabia, with the Mediterranean forming its northern boundary, the Suez Canal and Gulf of Suez on the west, and the Gulf of Aqaba to the southeast. Sinai is usually regarded geographically as part of Asia, but politically is a desert governorate of the United Arab Republic (Egypt). Area 23,200 sq.mi. (60,088 sq.km.). Pop. (1960) 126,000.

The oldest rocks and the highest mountains in Sinai are to be found in the southern third of the peninsula, an extension of the Red Sea Hills of the African mainland. These form a complex of Precambrian crystalline rocks: gneisses and schists into which were intruded masses of igneous plutonic rocks with varied igneous dikes veining the whole great complex. They form high, rugged mountains predominantly red in colour, from which the Red Sea takes its name. In the west, basaltic lava flows, probably of Miocene age, rest upon the old rocks. Along the Gulf of Suez coast a narrow plain of recent rocks separates the mountains from the sea except just south of the Wadi Fayran (Feiran) where for 20 mi (32 km.) abrupt cliffs of red granite rise from the shore. All along the shore of the Gulf of Aqaba the Sinai Mountains rise in a precipitous wall. North of the Wadi Fayran outcrops of red (Nubian sandstone terminate the ancient core of the peninsula. Against them to the north in a series of dissected scarps successively younger strata outcrop, mainly of limestone of Cretaceous and Eocene age. The gravel-covered central part of the limestone plateau, known as the wilderness of At Tih ("the wandering") averages 3,000 ft. (914 m.); its highest part, Jabal al 'Ajmah (Egma) in the south, attains 5,335 ft. (1,626 m.). To the north the land slopes down to the Mediterranean, its surface broken by inlier of older rocks creating marked limestone and sandstone hill masses, principally Yu'alliq (3,570 ft. [1,088 m.]), Hilal, both of Middle Cretaceous series, and Magharah of Jurassic age. Near the Mediterranean shoreline is a broad tract of sand dunes, some over 300 ft. (91 m.) high. This northern shoreline, fringed with salt marsh is sinking. In the west this is compensated for locally by the eastward drift of Nile sediment.

The granite peaks in the south, the highest mountains in the governorate of Sinai, include Mt. Catherine or Jabal Katrinah (8,651 ft. [2,637 m.]), Umm Shawmar (8,482 ft. [2,585 m.]), Ath Thabt (7,997 ft. [2,437 m.]), and Musa or Mt. Sinai (7,482 ft. [2,228 m.]). The area is difficult of access: deep, rocky gorges hem in the maze of sharp ridges and gaunt peaks. This mountainous area receives a few inches of rain each year, mostly in irregular, heavy downpours, insufficient for agriculture but sustaining springs and wells, and in a few bordering wadis small streams flow and nourish oases, such as at Feiran. The greater part of the scanty drainage of the northern two-thirds of the peninsula is ultimately to the Wadi al 'Arish which, rising in Al 'Ajmah Plateau, eventually reaches the Mediterranean. For most of the year the plateau is waterless except for a few permanent springs, as at An Nakhl. The Wadi al 'Arish rarely carries water to the sea, but water is obtained from wells dug in its bed and a small dam has been built. Underground water is also tapped by wells in the hollows between dunes along the northern coast where, four or five feet down, a foot or so of fresh water rests on salt.

The climate is healthful. In the mountain regions the nights are usually cold, but over most of the peninsula the heat of summer is intense. The rainy period of winter often results in serious floods. In a normal year there is sufficient rainfall to support scrub and tamarisk bushes and even a few poor crops. There are date palms at Al 'Arish, Feiran, and At Tur.

Animals are rare. They include ibex, gazelle, sand fox, leopard, wildcat, jackal, hare, hedgehog, and mole. Falcons and eagles are indigenous, and there are also the seasonal migrants such as quail, partridge, and grouse, which are shot for food as are the ibex and gazelle.

Sinai is sparsely populated by *bedu* (Bedouin *q.v.*) tribes, which are divided into two main groups. In the south the most numerous are the Sawalihah; in the north, the Terabin. The people of the coastal area are an admixture, and after 1948 there was a great influx of refugees encamped in the Gaza strip. A small group, the Jebeliyeh, living round St. Catherine's Monastery, are said to be the direct descendants of Bosnian and Wallachian serfs settled there by Justinian as guards for the monastery and forcibly converted to Islam at the time of the Arab invasion. The majority of the sedentary population lives near the north coast, especially at Al 'Arish, the administrative centre (pop. [1957 est.] 15,400) and at Al Qantarah ash Sharqiyah on the eastern side of the Suez Canal (pop. 13,384).

The railway between Palestine and Egypt was inoperative after 1948. The main paved road runs from Ismailia (Al Isma'iliyah) to Abu 'Uwayjilah, where it forks, one branch going north to Gaza and the other to Al 'Awja'. The old pilgrim route, still in use, runs from Suez via An Nakhl, Ath Thamad, and Aqaba to Medina and Mecca. Most of the country is not suitable for wheeled vehicles but it is possible to pass along the wadi beds when these are not in flood. Inaccessibility has delayed exploration and development of mineral resources. Manganese is worked at Umm Bugma, 65 mi. (105 km.) S of Suez; and after World War II oil fields were developed beside the Gulf of Suez, at Sudr, Wadi Fayran, and Al Bala'im. These fields account for four-fifths of Egypt's total production. (A. B. M.)

History.—The name Sinai is thought to be derived from the Akkadian moon-god Sin. Mt. Sinai is famous as the scene of the giving of the law to Moses and the entering of the Israelites into a covenant with the Lord, but there is considerable doubt as to which of the mountains of Sinai was the relevant site. Jabal Sirbal was originally so accepted, with the result that the city of Paran was built at its foot and in time became a cathedral city and a bishopric; but the claims of Mt. Sinai were later preferred, and on this mountain was built St. Catherine's Monastery.

In the Wadi al 'Arish, at 'Ayn al Qusaymah, Bi'r Hasanah, and An Nakhl, implements of Acheulo-Levalloisian type were discovered, and it is therefore thought that Paleolithic man was present in north, central, and southwestern Sinai. Neolithic and Chalcolithic remains were found on the Ismailia-Bi'r Hasanah track. Stone circles, rectangular stone ruins, and circular tombs, all of uncertain date, occur in the Wadi Sulaf and in many other regions of the peninsula. The ancient Egyptians mined turquoise and copper in the mountains, in the Wadi Magharah, and in the Feiran Oasis; their inscriptions date from the 1st dynasty. At Sarabit al Khadim, Sir W. M. Flinders Petrie found a temple dedicated to Hathor. It was used by the mine workers and contained records and inscriptions of many of the Middle and New Kingdom rulers. The last Egyptian king to make recorded offerings at this place was Ramses V of the 20th dynasty.

The Sarabit inscriptions collected by Petrie and published by A. H. Gardiner and E. T. Peet may be the earliest alphabetic script, but the matter is still undecided. The passage of the Israelites through Sinai and the route and date of the Exodus are still matters for argument. At Shaykh Zuwaid in the north of Sinai there is an Egyptian frontier fortress dating at least from the time of the 18th dynasty, and it is likely that there was a chain of these fortresses extending along the northern coastal route. There is evidence from the texts and from surface finds that the Egyptians held all this region at the time when they were effectively controlling Palestine and Syria. After the decline of the Egyptian Empire, Nabataeans from Petra controlled the trade routes of the Wadi 'Arabah and the peninsula during the first two centuries B.C., until they were defeated by the Romans in A.D. 106. The region then became the Arabian province of the Roman Empire, and a capital was established at Bozrah (modern Bosra).

During the early Christian period Sinai became the home of a large number of hermits, particularly in the southern mountain region.

The building of the monastery of St. Catherine dates from A.D. 530 when Justinian I, after complaints of robber incursions from the monks who had settled there, fortified the traditional site of the burning bush observed by Moses on the lower slopes of Mt. Sinai. This provided a centre for the scattered communities of Christians, and was spared by the invading Muslims; to conciliate the latter, the monks, according to tradition, erected the small mosque that stands within the walls. The monastery was a pilgrimage centre during the Middle Ages. It still retains much of its original appearance and has had an unbroken history since the 6th century. The original gray granite walls (280 x 250 ft. [85 x 76 m.]) still stand, and so does the church dedicated to the Virgin Mary, which was built at the same time. In the apse is a mosaic of the Transfiguration, probably also dating from this period, though it has been restored. The monastery's greatest treasures are its icons, some of which are prior to the 8th century, and its manuscripts. These, housed in the library built in 1945, are mainly in Greek and Arabic. In 1949–50 most of these manuscripts were microfilmed by the American Foundation for the Study of Man, acting on behalf of the Library of Congress and with the assistance of the University of Alexandria.

The most important manuscript now in the monastery is the Codex Syriacus, a Syriac text of the Gospels written about 400. The Codex Sinaiticus, now in the British Museum, is a Greek manuscript of the Bible dating to the 4th century.

After 1517 Sinai formed part of the Ottoman Empire and the country was administered by an official from Constantinople. When Egypt became independent of Turkey, conditions in Sinai deteriorated and traveling became difficult. Order was not restored until 1831, when Ibrahim Pasha, son of Mehmed Ali, advanced through Sinai and defeated the Turks at Acre. At the Treaty of London in 1840 Egypt was nominally restored to Turkish rule, and remained in this position until the outbreak of war in 1914. In October of that year Turkish forces seized Al 'Arish, and in 1915 and 1916 they advanced to the Suez Canal, but were driven off. In January 1917 the Battle of Rafah in northern Sinai marked the final advance of the British Army through the country. At the end of World War I Sinai was returned to Egypt and has since been administered by that country. After 1948 the Sinai-Israeli border was the scene of frequent frontier incidents. In 1956 the Israeli Army advanced through Sinai to the Suez Canal, defeating units of the Egyptian Army. They withdrew in 1957 at the behest of the UN General Assembly.

As the land bridge between Asia and Africa, Sinai retains considerable importance in Middle Eastern affairs. (M. V. S.-W.)

BIBLIOGRAPHY.—W. F. Hume, *The Topography and Ecology of the Peninsula of Sinai* (1906); W. M. F. Petrie, *Researches in Sinai* (1906); L. Eckenstein, *A History of Sinai* (1921); H. J. L. Beadnell, *The Wilderness of Sinai* (1927); C. S. Jarvis, *Yesterday and Today in Sinai* (1931); R. Devreesse, *Revue Biblique*, vol. xli (1940); S. R. K. Glanville, *The Legacy of Egypt* (1942); F. W. Albright, *Bul. Amer. Schls. Or. Res.*, no. 109 and 110; T. Askenazzi, *Edoth*, vol. i and iii (1946 and 1948); J. Daumas, *La Péninsule du Sinaï* (1951); H. Field, *Contributions to the Anthropology of the Faiyum, Sinai, Sudan and Kenya* (1952); K. W. Clark, "Exploring the Manuscripts of Sinai and Jerusalem," *Biblical Archaeologist*, vol. xvi (1953); G. W. Murray, "The Land of Sinai," *Geog. J.*, vol. cxix (1953); E. O'Ballance, *The Sinai Campaign of 1956* (1959). (A. B. M.; M. V. S.-W.)

SINAI, ORTHODOX CHURCH OF MOUNT. The Greek Orthodox Church of Mt. Sinai (often incorrectly called the Sinai Independent Greek Orthodox Church) consists of the fraternity of the monastery of St. Catherine on Mt. Sinai, which is situated more than 5,000 ft. (1,500 m.) above sea level in a narrow valley north of Jabal Musa in the Sinai Peninsula of Egypt. It is the smallest of the autonomous churches that together compose the Orthodox Eastern Church (*q.v.*). The abbot of the monastery, who is also the archbishop of Sinai, Paran, and Raithu, is elected by the brotherhood and consecrated by the Greek Orthodox patriarch of Jerusalem. Founded by the Byzantine emperor Justinian I in 527, the monastery became an important centre of Orthodox spirituality. One of its early abbots was St. John (*q.v.*) Climacus. The monastery was at first under the jurisdiction of the Jerusalem patriarch; its independence was recognized by Constantinople in 1575. There are about 20 monks belonging to the monastery (whose numbers are limited to 36); this figure includes those living in its annexes (*metochia*) elsewhere, chiefly at Cairo and Suez. The laity of the Church of Sinai are some Christian Arabs employed by the monastery and fishermen on the Red Sea coast at At Tur (Tor; formerly Raithu). The Bedouin Arabs who live near the monastery have always acted as its guards and have in their turn been supported by it; although Muslims, they are devoted to the monastery, within whose walls is a small mosque where they worship. The monastery's greatest treasures are its manuscripts (which once included the Codex Sinaiticus now in the British Museum) and its icons, some from the pre-iconoclastic period of Byzantine art (*see* further SINAI).

BIBLIOGRAPHY.—E. Eckstein, *A History of Sinai* (1945); K. Amantos, *Syntomos historia tes mones Sina* (1953); A. Guillou, "Le Monastère de la Théotocos au Sinai," *Mélanges d'Archéologie et d'histoire*, 67:217–258 (1955). (H. S. AL.)

SINAIA, a health resort town and scenic spot in the *regiune* (administrative and economic region) of Ploeşti, Rumania, lies on the Prahova River, 2,770 ft. (844 m.) above sea level, in the southeastern foothills of the Bucegi Mountains in the Transylvanian Alps, and 78 mi. (126 km.) NNW of Bucharest. Pop. (1960 est.) 9,717. It has modern hotels and rest houses, mountain refuges, and winter-sports facilities. The Peleş Museum contains collections of paintings, furniture, and carpets. There are food-processing and metallurgical industries, and the town is connected by rail and road with Ploeşti and Braşov.

SINALOA, a Mexican state on the Gulf of California and Pacific Ocean, established in 1830. Pop. (1960) 838,404; area 22,429 sq.mi. (58,092 sq.km.). Sinaloa is bounded north by Sonora, east by Chihuahua and Durango, and south by Nayarit, with about 400 mi. of coast on the west. The state is coastal, barren and tropical, with mountains on its eastern border. Five main streams and river valleys run from these mountains to the coast. Each river valley system is relatively isolated and is the scene of extensive irrigation, hydroelectric, and communications projects. The Miguel Hidalgo Dam on the Fuerte River, completed in 1956, waters about 123,550 ac.

Sinaloa is primarily an agricultural area which raises wheat, chick-peas, cotton, tobacco, sugarcane, tomatoes, fruits, and winter vegetables on irrigated lands. Local manufactures include beer, ice, cigarettes, cooking oils, soap and cotton textiles; there is an iron foundry and a plating and galvanizing establishment. The most important coastal industry is fishing, chiefly for sharks, the livers of which yield vitamin products, and for shrimps, processed locally. Tourism and game fishing are also significant.

The backbone of Sinaloa is the Mexican Southern Pacific Railway, acquired from its U.S. owners by Mexico in 1951. It runs from Nogales to Guadalajara and, in its course through Sinaloa, parallels the coast and ties the river valleys together. A spur of the Southern Pacific line extends to the water's edge at Mazatlán, where there are improved port works and a large airport. The Pacific highway, which reaches Nogales, ties Culiacán (*q.v.*), the state capital and commercial centre, to Mazatlán (*q.v.*), and both to Guadalajara. A railway between Mazatlán and Durango opened agricultural and mining outlets. The latter are of considerable importance, as Sinaloa produces salt, graphite, manganese, gold, silver, copper, iron and lead. (J. A. CW.)

SINANTHROPUS (PEKING MAN). Remains of this significant fossil were found in a cave or fissure filling at Chou-k'ou-tien, near Peking (*see* CHINA: *History*). The site was discovered in 1921 by J. G. Andersson, W. Granger, and O. Zdansky; the latter conducted the first excavations in 1923. In a Middle Pleistocene faunal assemblage, including rhinoceros, deer, horse, pig, bear, sabre-toothed tiger, hyena, and giant beaver, several human teeth were found (*Homo* sp.): a large lower molar of unusual pattern was discovered by B. Bohlin in October 1927. This single tooth led Davidson Black (*q.v.*) to infer a previously unknown hominid genus and species, which he named *Sinanthropus pekinensis*.

By the 1960s remains of about 45 individuals of *Sinanthropus* had been found, including many isolated teeth, parts of lower jaws, and the remains of 14 or more skulls; but of the rest of the skeleton little was known. The skulls had been damaged apparently to extract the brains—in all cases the surrounding of the foramen magnum was damaged and was the only part of the skull that could not be reconstructed; the long bones had been broken presumably to extract and eat the marrow (*see* WEIDENREICH, FRANZ). The industry associated with the bones consisted of crude tools, largely of quartz. Layers of ashes showed that *Sinanthropus* also used fire (*see* FIRE: *Early Control of Fire*).

The thick skull of Peking Man is typically long and low with a prominent continuous supraorbital ridge and a cranial capacity of about 915 to 1,225 cc. Dentition is robust, the pattern of the molars being complicated by secondary wrinkles; chinless lower jaws show strong sexual dimorphism. Body height is estimated as about 5 ft. for the males and 4¾ ft. for the females.

Black (1932) pointed out the close relationship between *Sinanthropus* and *Pithecanthropus erectus* from Java, and contemporary scholars commonly use the term *Pithecanthropus pekinensis* or even *Pithecanthropus erectus pekinensis*, putting *Sinanthropus* into the same genus as Java Man, now often called *Homo erectus pekinensis* (*see* PITHECANTHROPUS; MAN, EVOLUTION OF: *Pithecanthropus*).

BIBLIOGRAPHY.—D. Black, "On a Lower Molar Hominid Tooth from the Chou Kou Tien Deposit," *Palaeontologica Sinica*, ser. D., vol. vii (1927); F. Weidenreich, *The Skull of Sinanthropus Pekinensis* . . . (1943); popular treatments include M. Boule and H. V. Vallois, *Fossil Men* (1957), and R. Moore, *Man, Time, and Fossils*, rev. ed. (1961) (G. H. R. v. K.)

SINCLAIR, a celebrated Scottish family, members of which have held the titles of jarl of Orkney and earl of Caithness. The name is a variant of Saint Clair. SIR WILLIAM SINCLAIR (*c.* 1260–*c.* 1303), younger son of Robert, comte de St. Clair of Normandy, married Amicia, apparently heiress of the barony of Roslin, and the descendant of a line of barons, one of whom (traditionally also of St. Clair descent) obtained Roslin from Malcolm III in the 11th century. Sir William was a leader of the Scots in their war against the English. His grandson SIR WILLIAM SINCLAIR of Roslin (d. 1330) was killed by the Moors in Spain with Sir James Douglas ("the Good"), who had set out to take the heart of Robert I the Bruce to the Holy Land. His son SIR WILLIAM (d. *c.* 1358) married Isabella, daughter of Malise, earl of Strathearn, Caithness, and Orkney, and their son SIR HENRY (d. *c.* 1400) as heir was confirmed in the jarldom of Orkney by Haakon VI, king of Norway, in 1379. He conquered the Faeroe Islands and rediscovered Greenland with the Venetian travelers Niccolò and Antonio Zeno.

Henry's grandson WILLIAM (d. *c.* 1480), 3rd jarl of Orkney of his line, a lord of Parliament as Lord Sinclair (*c.* 1449) and

chancellor of Scotland (1454–58), was created earl of Caithness in 1455. He was celebrated as a patron of the arts and in 1446 founded the collegiate Church of Roslin—"Roslin Chapel." In 1468 Norway ceded the Orkney Islands to James III of Scotland who in 1470 forced William to exchange all his rights there for Ravenscraig Castle, in Fife, and he retained thenceforth only the title of earl of Caithness. In 1476 he disinherited his eldest son, WILLIAM (d. 1487) "the Waster," and settled the earldom on WILLIAM (d. 1513), 2nd earl, the second son of his second marriage, who was killed at Flodden, and the Roslin estate on SIR OLIVER, the first son of his second marriage. By an agreement (1481) with Oliver, William "the Waster" received Ravenscraig castle and fief, renouncing all interest in Roslin. From Sir Oliver descended the house of Sinclair of Roslin, whose direct line ended in 1778.

GEORGE (*c.* 1526–1582), 4th earl of Caithness, grandson of the 2nd earl, was a supporter of Mary Stuart, queen of Scots. Chancellor of the jury which acquitted the earl of Bothwell of the murder of Henry, Lord Darnley, in 1567, he maintained great state in Girnigoe Castle, where he starved to death his eldest son JOHN (d. *c.* 1578). John's son GEORGE (1566–1643), 5th earl, was outlawed for deeds of violence, fled to Shetland, and was succeeded by his great-grandson GEORGE (d. 1676), 6th earl. Heavily in debt, he settled his estates in 1672 on a creditor, Sir John Campbell of Glenorchy (later earl of Breadalbane), who got himself created earl of Caithness in 1677. But in 1681 the privy council decided for a grandson of the 5th earl, GEORGE SINCLAIR (d. 1698) of Keiss, 7th earl. The title then passed to JOHN (d. 1705), 8th earl, grandson of James Sinclair of Murchil, brother of the 5th earl. His son ALEXANDER (*c.* 1684–1765), 9th earl, was last peer to take his seat in the Scots Parliament.

WILLIAM SINCLAIR (d. 1779) of Ratter, a descendant of Sir John Sinclair of Greenland, the third brother of the 5th earl, succeeded as 10th earl, but on the expiry of the Ratter branch in 1789, SIR JAMES SINCLAIR (1766–1823) of Mey, 7th bart., a descendant of George Sinclair of Mey, third son of the 4th earl, became 12th earl. On the death of his great-grandson GEORGE PHILIPS ALEXANDER (1858–1889), 15th earl, Barrogill (or Mey) Castle passed from the family and JAMES AUGUSTUS (1827–1891), 16th earl, descending from Robert Sinclair of Durran, third son of the 1st bart. of Mey, inherited the earldom. From him is descended MALCOLM IAN (1948–), 20th earl, and chief of clan Sinclair.

The line of Lords Sinclair derives from the 1st earl's eldest son WILLIAM "the Waster." His son HERNY (d. 1513), 3rd lord, received parliamentary ratification of his title in 1489. JOHN (1610–1676), 9th lord, got a resettlement on the sons of his only daughter Catherine, mistress of Sinclair, and failing them on the heirs male of her husband, John St. Clair of Herdmanston. Under this, on the death of her son HENRY (1660–1723), 10th lord, and his two sons (of whom JOHN, master of Sinclair, was attainted as a Jacobite in 1715), the right passed to a "stranger in blood" CHARLES ST. CLAIR (1692–1775) of Herdmanston, *de jure* 11th lord, to whose grandson CHARLES (1768–1863), 13th lord, the peerage was restored in 1782. The present holder CHARLES MURRAY KENNEDY ST. CLAIR (1914–), 17th lord, is descended from him.

Sir James Erskine (1762–1837), of Alva, 6th bart., grandson of Catherine, second daughter of the 10th lord and Sir John Erskine of Alva, 3rd bart., inherited the Roslin estate in 1789 as heir of entail, assumed the name St. Clair-Erskine, and in 1805 became 2nd earl of Rosslyn. The earldom had been created in 1801 for the lord chancellor Sir Alexander Wedderburn (maternal uncle of Sir James) with a special remainder in an heir of entail in the Roslin estate. From Sir James are descended the earls of Rosslyn.

See R. A. Hay, *Genealogie of the Sainte Claires of Rosslyn* (1835). (T. I.)

SINCLAIR, UPTON BEALL (1878–), U.S. novelist and polemicist, author of *The Jungle*, a landmark among naturalistic, proletarian novels. Sinclair's fiction was largely topical in nature and in the U.S. most of his books did not survive their occasion; but they were widely translated and continued to be read in other countries. Born in Baltimore, Md., on Sept. 20, 1878, he graduated from the College of the City of New York in 1897 and did graduate work at Columbia University, supporting himself by various journalistic endeavours. *The Jungle* (1906), his sixth novel and first popular success, was written when he was sent by the Socialist weekly newspaper *Appeal to Reason* to Chicago to investigate conditions in the stockyards. His report, cast in fictional form, tells the story of a family of Lithuanian immigrants employed in the stockyards; the chief character, Jurgis Rudkus, the sole member of the family not completely destroyed by the environment, eventually finds some hope for a better life through membership in the Socialist Party. Intended to create sympathy for the workers, *The Jungle* rather ironically aroused indignation at the quality of processed meat, and thus helped bring about passage of food inspection laws; as Sinclair said at the time, "I aimed at the public's heart and by accident I hit it in the stomach." (*See* also MUCKRAKERS.)

A long series of other topical novels followed, none as popular as *The Jungle;* among those of some lasting interest may be mentioned *Oil!* (1927), on events of the 1920s, and *Boston* (1928), on the Sacco-Vanzetti case. Sinclair again reached a wide audience with the Lanny Budd series, 11 contemporary historical novels beginning with *World's End* (1940).

Sinclair ended his activity in the Socialist Party in 1933 by organizing the EPIC ("End Poverty in California") movement; in 1934 he was defeated as Democratic candidate for governor. Of his autobiographical writings, *American Outpost: a Book of Reminiscences* (English title, *Candid Reminiscences: My First Thirty Years;* 1932) was reworked and extended in *The Autobiography of Upton Sinclair* (1962); *My Lifetime in Letters* (1960) is a collection of letters written to Sinclair.

SIND, a region of West Pakistan lying east and southeast of Baluchistan and south of the Punjab. It consists mainly of the alluvial plain formed and watered by the Indus River, but also includes, westward, Kohistan—the uplands north of Karachi—and the fringe of the Kirthar Range, and, eastward, the Registan or Thar Desert, extending from the Nara Channel into Indian Rajasthan. Politically it comprises the Khairpur and Hyderabad divisions, the former consisting of the districts of Jacobabad (formerly Upper Sind Frontier), Sukkur, Larkana, Nawabshah, and Khairpur, and the latter the districts of Hyderabad, Tatta, Dadu, Thar Parkar, and Sanghar. Total area 57,114 sq.mi. (147,925 sq.km.). Pop. (1961) 6,424,668.

The movements of population following the 1947 partition of India led to the influx into Sind of large numbers of refugees who were gradually absorbed into the population. By 1961 Hindus, who at one time formed about one-quarter of the population, numbered less than one-tenth. The prevailing language is Sindhi, but Rajasthani, Baluchi, Punjabi, and Gujarati are also spoken. In Karachi, Hyderabad, and Jacobabad, Siraiki, a dialect of Sindhi, is found.

Soil, Vegetation, and Climate.—A large area of the region is covered by fine fertile alluvium, which becomes coarser toward the north, where it retains moisture better and is easier to plow; the south rarely yields such rich harvests. All parts of the region are liable to the deposition of salt, which so reduces fertility that land seriously infested is useless, the worst effects being observed in the fine-textured soils of southern Sind. Many parts have also become waterlogged. Almost the whole of the central plain from the Panjnad to the Arabian Sea is marked by the changes in the courses of the Indus or its distributaries. The silt deposited by the Indus in its bed has raised its level in many parts above the surrounding country and there are artificial embankments on either side. The soil is fertile. Canals from the Lloyd (Sukkur) and Ghulam Muhammad barrages serve 7,500,000 ac. (3,035,250 ha.), producing good crops of wheat, rice, and cotton. Gudu barrage, which was under construction in upper Sind in the mid-1960s, commands a cultivable area of 2,500,000 ac. (1,011,750 ha.). There are areas in the northwest which are too high to be irrigated by flow from flood canals, and are clothed with scrub. Parts of the flood plain are covered with forests.

Aridity limits forest to about 1,200 sq.mi. (3,108 sq.km.) near the Indus from Ghotki to the mid-delta. Babul (*Acacia arabica*) is most characteristic in lower Sind and yields fuel and timber for boats. Its bark is used for tanning and the leaves and pods as fodder for camels and goats. Kandi (*Prosopis spicigera*) provides fuel and fodder.

The climate is arid and extreme. The winters are mild and the summers very hot, the temperature frequently rising to 45.5° C (114° F) and occasionally to 49° C (120° F). The mean maximum temperature for June for Jacobabad is 45.5° C, with an absolute maximum of 53° C (127° F), the highest recorded in the subcontinent. Rainfall is both scanty and irregular, the mean annual for Hyderabad being 9 in. The climate of the coastal strip is equable and humid.

History.—Excavations at Mohenjo-Daro, 180 mi. northeast of Karachi, have revealed the remains of a large city which existed between 2500 and 1500 B.C., and whose inhabitants enjoyed a high degree of civilization. Sind was invaded by Alexander the Great in 325 B.C. On his death two years later it passed to Seleucus I Nicator, who yielded it in 305 to Chandragupta, the founder of the Maurya dynasty. After a phase of Buddhist influence under Asoka, Chandragupta's grandson, there were incursions from the west and north. A Sudra dynasty with its capital at Aror (Alor) was followed by a period of Brahman rule (7th century A.D.). In 712 Sind was conquered by the Arabs and for nearly three centuries was nominally subject to the Arab caliphs. Though conquered by Mahmud of Ghazni (11th century), Sind long remained semi-independent under local Muslim dynasties, and it was not till the time of Akbar (16th century), who was born at Umarkot in Sind, that the province was incorporated into the Mogul Empire. When that empire broke up on the death of Aurangzeb (1707), local dynasties again arose. The Talpurs, of Baluch descent, were ruling Sind under the title of mirs, with their capital at Hyderabad, when during the First Afghan War (1839–42) the British army marched through the territory, using it as a base for its operations against Afghanistan. The mirs were compelled to accept a treaty by which they paid a tribute to Shah Shuja of Afghanistan, surrendered the fort of Bukkur to the British, and allowed a steam flotilla to navigate the Indus. In 1842 Sir Charles James Napier (*q.v.*) arrived in Sind and fresh terms were imposed on the mirs. The Baluch army resented this loss of independence and attacked the British residency near Hyderabad. After two decisive victories by Napier, Sind was annexed to British India as part of the Bombay Presidency. In 1936 it became a governor's province and in the following year it was constituted an autonomous province. After the creation of Pakistan in 1947 it again became a governor's province. In 1948 Karachi and its suburbs were separated from Sind and the seat of the province was shifted to Hyderabad. When West Pakistan was consolidated into a single province in 1955, the separate Sind government ceased to exist, and the region was divided between the two divisions of Hyderabad and Khairpur.

BIBLIOGRAPHY.—N. M. Billimoria, *Bibliography of Publications on Sind and Baluchistan* (1930); H. T. Lambrick, *Sir Charles Napier and Sind* (1952); M. B. Pithawalla, *Historical Geography of Sind*, 2 vol. (1936–37), and *Introduction to Sind* (1951); J. H. Marshall (ed.), *Mohenjo-daro and the Indus Civilization* (1931); S. P. Chablani, *Economic Conditions in Sind, 1592–1843* (1951). (K. S. AD.)

SINDBAD THE SAILOR, a hero in the *Thousand and One Nights* (*q.v.*), who recounts his adventures on seven voyages. These are based on the experiences of navigators, probably in the early Abbasid period (*c.* A.D. 750–850). Fortunes were made by merchants of Basra trading with the East Indies and China; but the risks were great. In these stories, the dangers are exaggerated by a strong infusion of the miraculous which offers parallels with the earlier literature of several nations.

The stories follow a pattern. Sindbad sails from Basra with merchandise. He is marooned or shipwrecked, and encounters terrible dangers which he only survives by a combination of resourcefulness and luck. He finally returns home with a fortune. On the sixth voyage, he reaches Sarandib (Ceylon). The king gives him presents to take back to the caliph, Harun al-Rashid, in Baghdad. In the Calcutta text of the *Nights*, the normal seventh voyage is replaced by the story of how the caliph induces him to go back to Ceylon with return gifts for the king.

Considerable light is shed on seafaring and trade in the East. For example, though Sindbad does not specify the goods which he takes from Basra, it is stated that he obtains diamonds and other precious stones, sandalwood, camphor, coconuts, cloves, cinnamon, pepper, aloes, ambergris, and ivory during his voyages. Among the natural hazards of seafaring, the captain loses his course on the third and sixth voyages. The ship is wrecked on the fourth to seventh voyages, though in one case (the fifth voyage), this is due not to storms, but to the fabulous roc (*q.v.*), a bird which drops huge rocks on the ship. In one instance (the third voyage) hairy apes swarm over the ship and leave the crew on an island. Perhaps the activities of pirates are suggested here. In the seventh voyage (Calcutta version), Sindbad and his shipmates are captured by savages in canoes, possibly in the Andamans.

Miraculous hazards include the giant roc bird (second and fifth voyages) whose egg resembles a huge white dome. The whale which is mistaken for an island (first voyage) has parallels with the great whales of Pliny and Solinus. The valley of diamonds (second voyage) has parallels in al-Qazwini (the 13th-century Persian geographer), Marco Polo, and St. Epiphanius, bishop of Salamis in Cyprus (d. A.D. 403). The cannibal giant of the third voyage suggests Cyclops. In the fourth voyage, Sindbad's companions are fattened by cannibals with food which causes them to lose their reason. This calls to mind the lotus-eating of the *Odyssey*. Sindbad's burial in the cavern of the dead is foreshadowed in a reference by St. Jerome to a supposed Scythian custom of burying alive with the dead those who have been dear to them. On the fifth voyage, the "old man of the sea" who compels Sindbad to carry him has been identified with the huge apes of Borneo and Sumatra.

P. Casanova suggests that Sindbad's adventures may have influenced Defoe's *Robinson Crusoe* and Swift's *Gulliver's Travels*. Sindbad the Sailor should not be confused with Sindbad the Wise, hero of the story of the Seven Wise Masters (*q.v.*).

BIBLIOGRAPHY.—Sir R. Burton, *Arabian Nights* (Nights 538–566), in Eng. trans., vol. vi (1886); P. Casanova, "Notes sur les voyages de Sindbad," *Bulletin de l'Institut français d'archéologie orientale du Caire*, vol. xx, pp. 113 ff. (1922); G. F. Hourani, *Arab Seafaring in the Indian Ocean in Ancient and Early Medieval Times* (1951). (J. A. HO.)

SINDHIA, the name of the Maratha ruling family of Gwalior which for a time in the 18th century dominated the politics of northern India. One branch held the headship of a village near the Maratha capital, Satara, in the 17th century. The founder of the Gwalior house was Ranoji, who was said to have been a personal attendant of the peshwa (title of the hereditary Maratha rulers), Balaji Baji Rao. His military abilities brought him promotion; in 1726 he was sent to Malwa to collect revenue for the peshwa. He died in 1745, having fixed his capital at Ujjain. The Sindhia capital was later moved to the rock fortress of Gwalior.

After Ranoji's grandson's death at the Battle of Panipat in 1761 his illegitimate son Mahadaji succeeded (1761–94). One of the ablest men of his time, he created a north Indian empire, though he always gave nominal allegiance to the peshwa. He emerged from war with the East India Company (1775–82) as the recognized ruler of northwest India. With the aid of French officers he defeated the Rajputs, took the Mughal emperor Shah Alam under protection, and finally won control of the peshwa by defeating Holkar, the peshwa's chief general, in 1793. He died at the height of his fame in 1794.

His grandnephew Daulat Rao, his chosen heir (ruled 1794–1827), was a man of lesser stature. He was worsted by Tukoji Holkar in the struggle for control of the peshwa. He came into collision with the British in 1803, was defeated in four battles, saw his French-trained army broken up, and had to make peace. He lost Delhi and the emperor but retained control of Rajputana until 1817, when Lord Hastings gave him the choice of cooperating against the Pindaris (*q.v.*) or war. He submitted and died in 1827.

The Gwalior State continued intact until India gained independence in 1947. Its territory was incorporated in Madhya Pradesh in 1956. The most distinguished Sindhia of modern times was Sir Madhava Rao Sindhia (1876–1925), who, succeeding in

1886, ruled with vigour and enlightenment for 31 years. *See also* GWALIOR; MADHYA PRADESH.

See H. G. Keene, *Mádhava Ráo Sindhia* (1892); J. Grant Duff, *History of the Mahrattas*, 2 vol., rev. ed. (1921). (T. G. P. S.)

SINDHĪ LANGUAGE is spoken by about 7,000,000 people (mid-1960s) in the province of Sind, West Pakistan. With Lahndā and Kāshmīrī it belongs to the northwestern group of Indo-Aryan languages (*q.v.*). Sindhī-speaking people are also found in neighbouring districts, in the Rann of Cutch, and in Kathiawar. Many Sindhī-speaking Hindus from Pakistan have migrated to Bombay and other Indian cities, *e.g.* Delhi, Ajmer, Jaipur, and Gwalior.

Linguistically, Sindhī is bounded on the west by the Balōchī (*q.v.*) language, an Iranian language, and by the Lahndā language (*q.v.*), to which it is closely akin, on the north. Of the four dialects of Sindhī, Vichōlī (Hyderabad) is standard, being employed for literary purposes. Another important dialect, Kacchī, is spoken in Cutch, and is strongly influenced by Gujarātī.

Muslim influence, over 1,000 years, gave a large number of Persian and Arabic borrowings, and the written character used for Sindhī is a variety of the Persian, with necessary modifications for sounds peculiar to the language; Hindus use a form of Devanāgarī.

Phonology.—The Sindhī system corresponds, in general, to that of Indo-Aryan. It shares one peculiarity with Kāshmīrī, viz., instead of a single final consonant, it has the consonant plus a voiceless vowel—for example, *chōkar*ᵘ "boy" and *chōkar*ᵉ "girl." Another feature is the series of "recursive" consonants, transcribed $\bar{g}$, $\bar{j}$, $\bar{ḍ}$, $\bar{b}$ (or *gg, jj, ḍḍ, bb*), which are pronounced by drawing in (instead of expelling) the breath, with the larynx lowered and the glottis closed. These may occur initially. Still another feature is the occurrence of cerebrals where dentals are found in other Indo-Aryan languages. (*Cf.* Sindhī *ḍ̄iaṇ* and Hindī *dēnā* "give.")

Morphology.—Sindhī has two genders, masculine and feminine, and two numbers, singular and plural. Generally, case relationship is indicated by means of postpositions added to the oblique forms. An adjective agrees in gender, number, and case with its noun. Sindhī, like Lahndā and Kāshmīrī, uses pronominal suffixes attached, as in Lahndā, to nouns as well as participial forms, while in Kāshmīrī they are used only with verbs.

The Sindhī verb has two conjugations, intransitive and transitive. Compound tenses consist of combinations of participles plus auxiliary. (For Sindhī *see* the *Linguistic Survey of India*, vol. viii, pt. i, 1919.)

Literature.—Until recently little Sindhī literature had been printed; activity quickened during the 1920s with the formation of literary societies and availability of publishing outlets. The Sindi Sahitya Mandal (Sindhī Literary Circle) was founded in 1949 in Bombay; others are active in other Sindhī cities.

BIBLIOGRAPHY.—G. Stack, *A Dictionary of English and Sindhī* (1855), *A Grammar of the Sindhī Language* (1849); E. Trumpp, *Sindhī Literature: The Divan of Abd-ul-Latif known by the name of Shāh jō Risālō* (1866), *Grammar of the Sindhī Language* (1872); J. Beames, *A Comparative Grammar of the Modern Indo-Aryan Languages* (1872–79); G. Shirt, Udharam Thavurdas and S. F. Mirza, *Sindhī-English Dictionary* (1879); Sir G. A. Grierson, "Vrācaḍa and Sindhī," in the *Journal of the Royal Asiatic Society*, 1902, pp. 47 ff.; R. L. Turner, "The Sindhi Recursives," in the *Bulletin of the School of Oriental Studies*, vol. iii (1921); "Cerebralization in Sindhī," *Journal of the Royal Asiatic Society* (1924); L. A. Jagtiani, "Sindhī," in the *Indian Literatures of Today* (1947); M. U. Malkani, "The Sindhi Short Story and its Development," in *Indian Writers at Chidambaram* (1957); Lachman M. Khubchandani, *The Phonology and Morphophonemics of Sindhi* (1961), *The Acculturation of Indian Sindhi to Hindi: the Study of Language in Contact* (1963). (ET. B.)

SINGAPORE, an independent sovereign republic, a member of the United Nations and of the Commonwealth of Nations, includes one main island and many small islands lying at the southern end of the Malay Peninsula, with which it is connected by a rail and road causeway (¾ mi. [1 km.]) across the Johore Strait. Singapore Island, bounded south by the Singapore Strait, is roughly diamond shaped, 26 mi. by 14 mi. (42 km. by 23 km.), with an area of 209.6 sq.mi. (543 sq.km.) and a midpoint of latitude 1°22′ N, longitude 103°50′ E. The approximately 40 small islands total 15 sq.mi. (39 sq.km.).

SINGAPORE

Singapore owes much of its wealth and continued prosperity to its focal position in Southeast Asia on international sea and air routes. Its strategic position and deepwater harbour have enabled the city of Singapore (*q.v.*) to become the largest port in Southeast Asia, the natural outlet for the products of the Malay Peninsula, and one of the world's greatest commercial centres. The island also serves as an important naval and air base.

PHYSICAL CHARACTERISTICS

An outlier of rocks and structures running through the Malay Peninsula, Singapore has a central granite zone culminating in Bukit ("hill") Timah (581 ft. [177 m.]), west of which Triassic quartzites and shales dip to the southwest and form low degraded scarps which are heavily eroded into hills aligned in a northwest-southeast direction with the steepest ridge behind Pasir ("sand") Panjang. East of the granite zone lies a mass of weathered detritus averaging 100 ft. (30 m.) in height. Weathered red and yellow lateritic soils and thick iron pan cap all rocks. A quarter of the island is lower than 25 ft. (8 m.). Only Pasir Panjang has a steep coast.

The deepwater channel of Keppel Harbour is sheltered from the south by the hilly islands of Blakang Mati and Brani. Elsewhere the coast is silting, and dangerous coral reefs and sandbanks lie offshore. At the eastern end of the Johore Strait is a deep scoured channel, which permits large vessels to approach the naval dockyard at the northern corner of Singapore Island. The solid causeway across the strait prevents the through movement of vessels and hastens silting at the western end.

On the central granite hills of Singapore Island is the only remnant of its forest, a stand of dense, mixed evergreen trees which protects three reservoirs formed by damming small streams. They are now inadequate for the island, and water is piped from south Johore. The climate is equatorial, with temperatures throughout the year varying a few degrees either side of the 28° C (82° F) mean. There is no well-defined wet or dry season, and the annual rainfall of 95 in. (2,413 mm.) is evenly distributed throughout the year. Prevailing winds are from the northeast from November to February and from the south during May to August, but strong daily land and sea breezes maintain agreeable physiological conditions. Heavy rains and high seas come during the northeasterlies, from which the harbour is well protected. (E. H. G. D.)

HISTORY

Malay tradition confirms that the first colonizer of Singapore was a prince from Palembang, the Sumatran capital of the empire of Sri Vijaya. It may have been the Chola (*q.v.*) ruler Rajendracola Deva I (1016–44) who named the settlement Singapore ("city of the Singhs" or "lions"). He almost certainly attacked the island in 1025 during his descent on Sri Vijaya and its colonies.

There was another Chola raid in 1068. In Javanese inscriptions and Chinese records down to the end of the 14th century the commoner name for the island is the Javanese Temasek or Tumasik from *tasek*, "sea." In 1275 the Javanese king Kritanagara, when he raided Pahang, probably attacked Temasek. According to a Chinese traveler, Wang Ta-yuan, just before 1349 about 70 Siamese war boats besieged Temasek for a month but had to withdraw. A Javanese work, the *Nagarakritagama*, written in 1365, includes Temasek among the conquests of the Javanese empire of Majapahit.

At the end of the 14th century, Temasek fell into decay and was supplanted by the port of Malacca. Yet in 1552 it was still a port of call from which St. Francis Xavier dispatched letters to Goa, and João de Barros, in the *Décadas da Asia* (published 1553), describes it as "a resort not only of Indian shipping but of traders from China, Siam, Champa, Cambodia, and the Malay Archipelago."

East India Company.—On Jan. 28, 1819, Sir Thomas Stamford Raffles (*q.v.*) of the East India Company, searching for a factory site, forestalled by the Dutch at Riouw (Riau, Rhio), and, having found the Carimon Islands unsuitable, landed at Singapore. He found only a few Chinese planters, some aborigines, and a few Malays, and was told by the hereditary chief, the *temenggong* (direct ancestor of the sultans of modern Johore), that there were no Dutch there and that the company could purchase land. The *temenggong*, however, was a subordinate of his cousin Abdul Rahman, sultan of Riouw, Lingga, Pahang, Johore, and Singapore, who was under Dutch surveillance. Abdul Rahman was a younger son and not a sultan *de jure*. Raffles, disobeying instructions not to offend the Dutch, withdrew his own recognition of Abdul Rahman as sultan of Singapore and installed Abdul Rahman's elder brother, Husain, to validate the purchase of land there on behalf of the company. The Dutch protested. In London the court of directors, though it decided Raffles had contravened instructions, took no action.

In 1824 an Anglo-Dutch treaty left Malaya and Singapore in the British sphere, and on Aug. 2 the whole of Singapore Island was ceded to the British for a money payment. In 1826 Singapore, Penang, and Malacca were combined as the Straits Settlements to form an outlying presidency of India. In 1830 they were reduced to a residency under Bengal, and two years later Singapore became their capital. When the East India Company lost its monopoly of the China trade (1833) it also lost its interest in Malaya. The settlements were transferred to the direct control of the governor-general of India in 1851. In 1867 they were made a crown colony under the Colonial Office.

Development of the Port.—Meanwhile, the trade of Singapore had suffered from British development after 1842 of a rival port, Hong Kong, as later it was to suffer from the French occupation of Indochina, the development of Saigon and Haiphong, and from the establishment of Dutch ports and shipping lines in the Netherlands East Indies. With the opening of the Suez Canal in 1869 and the advent of steamships, however, an era of prosperity began that led eventually to the construction of three miles of wharves at Tanjong Pagar and finally in 1921 to a naval base. The economic growth of the Malay states after they became British protectorates enlarged transit trade. (R. O. WT.)

It was the demand of the industrial West for tin and rubber that made Singapore one of the greatest ports in the world. After World War I steps were taken to modernize Malayan defenses and, with the lapsing of the Anglo-Japanese alliance, to build a large naval base in Singapore.

World War II.—In World War II the Japanese landed in Kelantan and Thailand on Dec. 7, 1941 (Dec. 8 west of Hawaii). By the end of January 1942 they had overrun the peninsula and were opposite the island of Singapore. The sinking of the battleship "Prince of Wales" and the battle cruiser "Repulse" had given them command of the sea, and a superior air force covered the advance down the peninsula. The Japanese crossed the Johore Strait on Feb. 8, 1942, and the British command surrendered the island and city on Feb. 15. Singapore remained in Japanese occupation until September 1945.

Political Developments after the War.—British political plans for Malaya excluded Singapore from the Malay Union and later the Federation of Malaya (*see* MALAYSIA: *History: Malaya*) mainly because Singapore's predominantly Chinese population would be a racial obstacle to a common citizenship. As a separate crown colony (from 1946) Singapore made constitutional progress despite the Communist insurrection in Malaya. Elected ministers and a Legislative Assembly with an elected majority assumed government responsibility in 1955, except for defense and foreign affairs. In 1959 the official and nominated elements were eliminated and Singapore became self-governing, although Britain still retained control of defense and foreign affairs.

Singapore joined the Federation of Malaysia on its formation in September 1963, retaining autonomy in education and labour and accepting in consequence a smaller number of seats in Parliament. The ruling People's Action Party (PAP) led by Lee Kuan Yew had refused in 1959 to form a government until eight extreme left-wing leaders who had been detained by the colonial authorities were released. This left wing opposed the concept of Malaysia, broke away to form the Barisan Sosialis Party, and was accused of being a Communist Front organization amenable to the control of the Indonesian Communist Party. The PAP faced fresh dangers of subversion when Indonesian opposition to Malaysia took the form of military and economic "confrontation" (1964). There were a number of bomb outrages and attempts to blow up military installations and public utilities; the suspension of Indonesian trade reduced the gross national product by 8%.

"Confrontation" ended in June 1966, but Singapore had left Malaysia in August 1965, invited to do so by the Malaysian government as a result of political friction between the state and central governments. This conflict had racial undertones, and it continued to affect relations between independent Singapore and Malaysia.

POPULATION

At the census in 1957 Singapore's population numbered 1,445,929, comprising 1,090,596 Chinese (75%), 197,059 Malays, 124,084 Indians and Pakistanis, 11,382 Eurasians, 10,826 Europeans (excluding servicemen), 5,426 Ceylonese, and 6,556 others. The estimated population in December 1966 was 1,945,000 (a density of 9,262 per square mile), largely a natural increase although assisted by some migration from West Malaysia. Over half the population was under 15 years of age. When Raffles founded the British settlement in 1819 fewer than 200 people lived on the island. Five years later there were over 10,000; by 1860 half the population of 82,000 was Chinese; and by 1931 the Chinese formed 75% of the population of 559,945. For many decades Singapore was a "frontier town," its population largely migratory and the women vastly outnumbered. When the first immigration quotas were imposed in 1932 there were no restrictions on women immigrants. Since World War II the policy has been increasingly selective and there is now virtually no immigration.

ADMINISTRATION AND SOCIAL CONDITIONS

When Singapore became a republic in 1965 Tun Yusuf bin Ishak, the head of state (*yang di-pertuan negara*), became president; Lee Kuan Yew remained prime minister. The president is elected by Parliament for a term of four years. The prime minister and nine other ministers form the Cabinet, which is responsible to an elected Parliament of 51 members. Voting is compulsory. There are no local government bodies. Appointments to the civil service are made by an independent Public Services Commission. Criminal and civil jurisdiction is exercised by the High Court, consisting of the chief justice and six judges. There are two civil district courts; 15 district and magistrates' criminal courts; and a juvenile court. The Shariah Court has jurisdiction in matters of Islamic law. There is a Federal Court of Appeal, and in certain cases further appeal to the Judicial Committee of the Privy Council in London is possible.

The main religions are Islam, Buddhism, Hinduism, and Christianity. Malay is the national language, but Malay, Chinese, Tamil, and English are all official languages and English is the

language of the administration.

Labour and Welfare.—Legislation governing working conditions provides in general a 44-hour week and an 8-hour day for manual workers. Employment of women and young persons is regulated, and provision is made for compensation for injury or death at work and for retirement benefits. Labour disputes which cannot be settled by agreement, with the aid if necessary of ministry officials, are adjudicated by two Industrial Arbitration Courts, which also record and supervise agreements reached by collective bargaining. There are about 100 trade unions with a total membership of 160,000. The Social Welfare Department has charge of public assistance.

The Woman's Charter, which came into effect in 1961, enforced monogamy, made marriage registration compulsory (these provisions do not apply to Muslim marriages), and raised the marriage age from 16 to 18. The validity of existing polygamous marriages was not affected. The charter also regulates divorce.

Housing.—The whole area within six miles of the Singapore River, which runs through the heart of Singapore city, is built over —a mixture of congested streets and covered five-foot ways, modern office and apartment buildings, and shopping districts and residential suburbs in which old-style colonial bungalows survive. About 500,000 people in Singapore Island live in the Housing and Development Board's apartment buildings. During 1935–59 the Singapore Improvement Trust built 23,019 homes and shops. Between 1960 and 1965 the Housing Board added 54,430 units. Another 60,000 units were being built under the 1966–70 plan. Some 2,000 ac. (809 ha.) were being reclaimed at Kallang basin and between Bedok and Tanjong ("cape") Rhu, on the south coast, for the siting of light industries and rehousing under the urban renewal plan. This plan also involved clearing the crowded and dilapidated 100-year-old shop-houses and two-storied buildings in some of the older parts of Singapore city.

Health.—In the mid-1960s the state supported 11 hospitals with 7,000 beds. They included a large general hospital, a district hospital, an orthopedic hospital, and specialist hospitals for mental and infectious diseases, including tuberculosis and leprosy, and an opium-addict treatment centre on St. John's Island, in the straits. Five private hospitals have 750 beds. The Health Institute houses the school health services and the university's Department of Social Medicine and trains health officers. Physicians, dentists, and pharmacists are trained in the University of Singapore. Between 1947 and 1965 the birth rate fell from 45.9 per 1,000 population to 29.9; the infant mortality rate from 81.3 to 26.3; and the death rate from 13.3 to 5.5, one of the lowest in the world.

Education.—Since the population of Singapore is so youthful, one in four attends school and education absorbs one quarter of the republic's annual expenditure. All children receive six years' free primary education, and the Malays enjoy free education to university standard. Half a million pupils follow a common syllabus designed to foster a national consciousness. This is also furthered by a number of integrated schools, which about 19% of the pupils attend. Another 43% are in English-language schools, 32% are taught in Chinese, 5% in Malay, and 1% in Tamil. Most children are at least bilingual. The emphasis of the second five-year plan was to ensure that after 1966 every child who qualified had the opportunity of secondary education. The University of Singapore had approximately 3,000 students in the mid-1960s; Nanyang University, founded in 1953 largely by Chinese subscriptions, had over 2,000; and the Singapore Polytechnic had 2,300, 950 of them full-time.

Defense.—Singapore's armed forces consist primarily of two battalions of infantry with an ancillary artillery unit and supporting forces. A new form of national service is being introduced. All are liable. Ten percent of the men will be selected for full-time two years' service, forming two infantry battalions which will replace the Defense Force units. The remainder will be liable for 12 years' part-time service. Singapore is also the main base for British Navy, Army, and Air Force units in the Far East. This arrangement was preserved after Singapore seceded from Malaysia, with which Singapore has a defense agreement. Over 30,000 civilian workers are employed at the British bases.

Political Parties.—In 1966 the PAP led by Lee Kuan Yew, the prime minister, held 42 out of 51 seats in Parliament. Before the end of the year the Barisan Sosialis Party, which had won 13 seats in the general election in 1963, boycotted all parliamentary proceedings. Its remaining M.P.s then resigned. In 1966 the Singapore United Malay National Organization (UMNO), a branch of the Malaysian UMNO, sought to establish itself as a separate party.

THE ECONOMY

Trade.—Singapore was founded as a free port, open to the ships and traders of every country, at a time of monopoly trading in many parts of the Far East. It remained a free port for 140 years. Until the late 1950s there were no import duties except on petroleum, tobacco, and liquors. The rapid growth of the population and the changing prospects for trade, as neighbouring countries developed their economies, then turned Singapore's attention to industry, and the first protective duties were imposed. Nearly two-thirds of the $944,000,000 spent on the 1961–65 development plan was devoted to economic development, and the $1,521,000,000 of the 1966–70 plan put the same emphasis on an industrial substructure. (Currency, quoted here and following, is in Singapore, not U.S., dollars.) The continuing importance of the entrepôt trade was recognized, however, by the provision of free trade zones and by the general absence of customs formalities in the importation of commodities which are processed for reexport. Manufactured goods, machinery, and foodstuffs form the other half of this entrepôt trade. Singapore is the world's largest rubber market and exporter, although producing practically no rubber itself.

In the 1960s Singapore's external trade averaged $7,500,000,000 yearly, with an unfavourable balance of $700,000,000. There was a setback when trade with Indonesia was broken off during "confrontation," but the figures began to rise again in 1966. The part played by the entrepôt trade in Singapore's economy is well illustrated by the fact that the gross national product, which per capita is the second highest in Asia, is less than half the value of the island's foreign trade. Singapore handles or transships a great deal of Malaysia's trade. Its principal overseas customers are the United Kingdom, the U.S., Japan, and Hong Kong; chief suppliers are the U.K., Japan, China, and Australia.

Agriculture and Fisheries.—Vegetables are cultivated intensively in parts of the island. Coconuts, fruit, tapioca, tobacco, and pepper are grown mostly on family smallholdings. Larger holdings are usually mixed farms, for raising poultry and pigs as well as vegetables. There are substantial exports of pork, poultry, and eggs. About 10,000 tons of fish are landed, from coastal waters, palisade fish traps, and fish ponds.

Industry.—Encouragement of industry is the responsibility of an Economic Development Board which manages the industrial estates, provides technical and financial assistance, and takes part in joint ventures with private investors. The largest project is the Jurong estate, of 9,000 ac. (3,642 ha.), to the west of Singapore city. It has sites for heavy and light industries, adequate power and water, access roads, a spur railway, and its own harbour. The centre of this complex is the National Iron and Steel Mill, with associated shipbuilding and structural steel interests, and a planned blast furnace capacity of 300,000 tons of steel a year. Five smaller industrial sites include a reservation for boatbuilding and marine industry. By the end of 1966 more than 150 firms had been granted pioneer certificates, entitling them to tax remissions and other privileges extended by the government to new industry. These enterprises ranged from petroleum refineries, flour mills, and chocolate and textile manufactures, to the assembly of motor vehicles and manufactures of metal, soap and toilet preparations, industrial chemicals, tires, and condensed milk.

Shipping and Transportation.—Over 23,000 ships, of a registered 100,000,000 tons total, annually enter and clear the port of Singapore. Singapore is also a bunkering port and, as a blending and distributing centre, handles 10,000,000 tons of oil a year. A great deal of the general cargo is unloaded from vessels anchored in the roads. Telok ("bay") Ayer basin is reserved for small boats engaged in the barter trade with Indonesia. The terminus

of Malayan Railways is linked to the harbour services.

Sixteen major airlines operate regular services from, or through, Paya Lebar airport, 7½ mi. [12 km.] NE of the city centre. There are 525 mi. (844 km.) of roads, and the causeway over the Johore Strait links the island with peninsular Malaysia. Island-wide motorbus services are the main form of public transport, but 3,000 trishaws survive, and the taxis are supplemented by "pirate" cars which act as minibuses and are used especially by schoolchildren and office workers.

Power.—For power the island depends on thermal-generated electricity from two power stations with an output of over 1,000,000,000 kw-hr. in 1966. Another power station comprising four 60-megawatt units was to be completed by 1970. Much of Singapore's water is drawn from Johore.

Communications.—Radio and television services in the four official languages and additional dialects are operated by the government. There are three English-language daily newspapers, three Chinese, two Malay, and two Tamil. Singapore has extensive radiotelephone communication and a public teleprinter service with most parts of the world. The SEACOM cable provides high-quality telephone circuits with Jesselton and Hong Kong.

Finance.—Singapore and Malaysia have a common banking system and until June 1967 had a common currency (*see* MALAYSIA: *The Economy: Trade and Finance*). Government revenue in 1967 was estimated at $590,500,000 (Singapore, not U.S., dollars) and expenditure at $590,200,000. Expenditure included the municipal services, which were mainly financed by a property tax (rates). Income and company tax and customs and excise duties were the principal national taxes, the protective duties, not imposed till the late 1950s, making as yet only a small contribution. Only 7% of the working population paid income tax. There were 34 commercial banks, the assets of which at the beginning of 1966 totaled $2,382,000,000. *See* also MALAYSIA.

BIBLIOGRAPHY.—R. O. Winstedt, *Malaya and Its History*, 5th ed. (1958); W. Makepeace, G. E. Brooke, and R. St. J. Braddell, *One Hundred Years of Singapore*, 2 vol. (1921); C. E. Wurtzburg, *Raffles of the Eastern Isles*, ed. by C. Whitting (1954); E. H. G. Dobby, *Senior Geography for Malayans* (1955) and *South East Asia*, 7th ed. (1960); A. E. Percival, *The War in Malaya* (1949); R. Braddell, *Lights of Singapore* (1934); M. W. F. Tweedie and J. L. Harrison, *Malayan Animal Life* (1954); Milton E. Osborne, *Singapore and Malaysia* (1964); Singapore Annual Reports. (A. KE.)

SINGAPORE (city), the capital of the Republic of Singapore (*q.v.*), lies at the southern end of the diamond-shaped island of Singapore, at the narrow gateway between the Indian Ocean and the South China Sea. Created a city by royal charter in 1951, Singapore was administered as a municipality by an elected mayor and fully elected council from 1957 to 1959. In 1959, when the colony of Singapore became internally self-governing, administrative responsibility for the city was assumed by the state.

The town, founded by Sir Thomas Stamford Raffles (*q.v.*) in 1819, was centred on the landing place he used near the mouth of the Singapore River. Within three weeks the resident in charge reported that "inhabitants are flocking in from every quarter." The commercial possibilities and the strategic value of the settlement were such that progress was uninterrupted even after the British acquisition of Hong Kong deprived it of some of its trade. The opening of the Suez Canal in 1869 increased its importance, and as steam replaced sail Singapore became a bunkering station. Keppel Harbour was developed in the early 20th century, but the old anchorage in the roads has remained in use, providing the city with one of the world's most fascinating waterfronts.

In the older part of the city are grouped the principal government and public buildings, the Supreme Court, banks and commercial offices, and St. Andrew's Cathedral, built by convict labour. The famous Botanic Gardens are three miles north of the city centre, close to the University of Singapore. The government maintains a national museum and library, a national theatre, and an aquarium all within the city area of approximately 37 sq.mi. (96 sq.km.). The estimated population in 1966 was 1,200,000, but separate city statistics had ceased to be available and an extensive building program was changing both the face of the city and the distribution of population. After World War II multistoried buildings dotted the waterfront skyline, and redevelopment after 1960 saw slum areas replaced by tall apartment buildings and modern housing developments complete with shopping arcades, schools, playgrounds, and other amenities. Much of picturesque "Chinatown" remained, but this too was changing. (A. KE.)

SINGER, ISAAC MERRIT (1811–1875), U.S. inventor who developed the first practical domestic sewing machine and brought it into general use, was born on Oct. 27, 1811, in Rensselaer County, N.Y. In 1851 he developed and patented a sewing machine which embodied a spring presser foot and feed synchronous with the needle action and thus permitted continuous and curved stitching. An overhanging arm projecting the needle bar over the work, which was spread over a horizontal table, made it possible to sew on any part of the work. Singer's machine was the first that could perform these necessary functions, and its basic design features have been followed in almost all subsequent machines. Nonetheless, Elias B. Howe (*q.v.*) must be credited with the first employment of the basic eye-pointed needle and lock stitch (in 1846), which, since Singer had embodied these in his machine, resulted in Howe's winning an infringement suit against Singer in 1854. The suit did not prevent Singer from manufacturing his machine, however, and in June 1851 he formed a partnership with Edward Clark, their company becoming, by 1860, the largest producer of sewing machines in the world. Singer left commercial development of his invention to Clark while he supervised manufacturing and the "experimental department," where he perfected other improvements. In 1863 he and Clark formed the Singer Manufacturing Company and Singer retired to England, where he died on July 23, 1875. (A. B. J.)

SINGHBHUM, the southernmost district in the Chota Nagpur division of Bihar, India. Area 5,191 sq.mi. (13,445 sq.km.). Pop. (1961) 2,049,911. Singhbhum forms part of the southern fringe of the Chota Nagpur Plateau and is a hilly upland tract composed largely of crystalline and ancient sedimentary rocks. There is a central 1,000-ft. (300-m.) gneissic and schistose peneplain drained by the Kharkai and Sanjai rivers. This fertile plain is bordered in the northwest and southwest by higher dissected and forested masses of hills with a general elevation of 2,000 ft. above sea level but occasionally rising to 3,000 ft. Toward the east there are dissected and forested hills of lower elevation (*c.* 1,000 ft.), through the centre of which runs the narrow Subarnarekha Valley. Forests occupy about one-third of the district. About one-fourth is cultivated, rice covering 80% of the sown area.

Singhbhum is one of the most important districts of India for metallic minerals. In the extreme south iron ores amounting to half the Indian output are produced from banded hematite-quartzites. Manganese (annual output *c.* 40,000 tons) and chromite (*c.* 3,500 tons) are mined near Chaibasa. Almost the whole of the country's kyanite production comes from Singhbhum, which is also the only district in India where copper (annual output *c.* 400,000 tons) is mined. Among the heavy industries are the iron and steel and allied manufactures at Jamshedpur (*q.v.*) and copper at Maubhandar. Shellac is manufactured at Chandil and Chakradharpur. The largest town is Jamshedpur but the headquarters of the administration are at Chaibasa (pop. [1961] 22,019). Singhbhum is traversed by the Bombay–Nagpur–Calcutta Railway, which passes through Jamshedpur where railway lines from the Damodar coalfields and the iron and copper regions converge. About two-fifths of the population of the district consists of tribal peoples, of whom the most numerous are the Hos and the Santals. (E. AH.)

SINGIDA REGION, TANGANYIKA, established in October 1963 following a reorganization of administrative units in Tanganyika (Tanzania), comprises Iramba, Singida, and Manyoni districts, which previously belonged to the former Central Region. It is bounded north by Shinyanga and Northern (Arusha) regions, and east by Dodoma Region, south by Iringa and Mbeya regions, and west by Tabora Region. Area, 19,050 sq.mi. Pop. (1963 est.) 393,520. The region lies on the central plateau of Tanganyika, its height ranging from 3,000 to more than 5,000 ft., and is a major watershed. Following the southeastern boundary, the Njombe River joins the Rufiji system and the Indian Ocean, whereas in

the southwest drainage is to Lake Tanganyika and the Congo River. Much of the drainage is to dischargeless basins, to the Wembere depression on the northwestern boundary (draining to Lake Eyasi), and to the Bahi Swamp. Mean annual rainfall is mostly less than 30 in., rain falling during November–April, with almost complete drought for six months.

Regional headquarters are at Singida. The main African tribes are the Iramba and Nyaturu (or Turu; in Iramba and Singida) and the Gogo (in Manyoni). The economy is based mainly on livestock, producing hides, skins, and clarified butter. Some areas are infested by the tsetse fly. Beeswax, honey, and gum arabic are collected and there is fishing and salt making in Bahi Swamp. The region is linked by road and by the Central Railway with the interior and the coast. (J. M. Ke.)

SINGING, the oldest form of music making, is an art with many associations, social and literary, as well as musical, and demands treatment, therefore, from both historical and technical viewpoints. It may best be treated in its traditional aspects in the first place, followed by stylistic and technical considerations.

Folk and Oriental Styles.—Of the vast bibliography of man as a music maker, Charles Seeger remarks: "Much of it is of music-hearing man. But little of it is of singing man; and that little is confined nearly entirely to him as a user of one single singing style—that of the *bel canto* of the opera and concert singer in the European tradition of the fine art of music" (*Western Folklore*, vol. xvii, no. 1, 1958).

In the past, differences in singing style were often remarked upon. In the 4th century, to the emperor Julian the Apostate, German singing was like the croaking of birds. A little later, Roman churchmen were complaining of the "beastly" singing of the Franks and Burgundians, who made "horrid clucking noises and crushed the melody to pieces in their throats." A 16th-century German commentator, Andreas Ornithoparcus, remarked (the translation is by the early 17th-century composer John Dowland): "The English doe carroll; the French sing; the Spaniards weepe; the Italians which dwell about the coast of *Ianua* caper with their voices; the other barke; but the Germans (which I am ashamed to utter) doe howle like Wolves." Such distinctions, whose rough justice the modern musical folklorist would in some instances admit, became ignored when, after 1600, a more or less standard manner of voice production imposed itself on the performance of Western art music.

It is only since the advent of mechanical and electronic sound recording that other singing styles than the conventional *bel canto* have come to be acknowledged, though in fact only a small proportion of the world's population practices or even admires this manner of singing. The art musics of the Orient employ quite other singing styles, as do also the various folk musics throughout the world.

The qualities of style that make a performer important in his tradition are differently assessed in different places. Notably, there are the regional variations in preferred ways of producing the voice, such as caught the ear of commentators in the past. For instance, English folk singers tend to pitch their voices high, while in Sardinia a guttural growl is practised. Pygmy peoples, wherever they occur, affect a sweet yodeling quality. In Azerbaijan and among the Kurds a strident yelping coloratura is highly prized among the professional minstrels or *ashugs*. (*See* also Folk Music.)

Various attempts have been made to classify the styles of folk singing, and Alan Lomax (*American Anthropologist*, vol. 61, no. 6, 1959) distinguishes among other families the "Eurasian," "Old European," and "Modern European" (to name only those close to the music of the Western world). We may take the Eurasian area to include Gaelic Ireland, parts of France, southern Spain and Italy, parts of Yugoslavia, southern Greece and Turkey, along with a number of Asian and north African countries. The area is characterized by solo song, unblended unison, and the use of accompanying instruments. The voice tone is high-pitched, tight-throated, often strident, with an emission suited to long-lined, decorated melodies.

The Old European tract may be said to include the Hebrides, Wales, Brittany, northern Spain, northern Italy, Switzerland, Germany, Scandinavia, Czechoslovakia, Poland, and the Ukraine. Here the song is mainly choral and collective, and accompaniment is unimportant. The voice is produced from a relaxed throat and is deeper-pitched than in the Eurasian area. The melodies are expressed in a relatively simple foursquare manner, with but sparse ornamentation.

The Modern European is described by Lomax as a hybrid style which grew up in the borderlands between Eurasian and Old European, and "most of the folk singing which western Europeans and Americans know" belongs to this style. Its area includes Lowland Scotland, parts of England, western and central France, central Spain, central Italy, and the greater part of "white" North America. It is characterized by solo songs sung in a hard, often nasal voice, with little ornament. As with the Old European—though not the Eurasian—style, the singer's interest and vocal emission are centred on text rather than tune, on sense rather than expression. With this style, timbre, dynamics, tempo, once established, remain constant throughout the song.

This classification may serve as a rough guide to the distribution of European and Western folk-singing styles, but to test its adequacy requires a more massive reservoir of knowledge than exists at present. It may be that the florid, tight-throated style of Eurasian singing—a style with common frontiers extending at least from Portugal, along both shores of the Mediterranean, and as far as India—derives from a current of musical practice that flowed out, both eastward and westward, probably from Mesopotamia in prehistoric times. The plain, relaxed style called here Old European—the frontiers of which may prove to extend from the Hebrides to the Tien Shan Mountains of Central Asia—is typical of another, even earlier, current, possibly from a Caucasian source.

It is repeatedly suggested that these various singing styles, along with those of the American Indians, Pygmoids, Bantu Africans, Australians, Melanesians, and Polynesians (to name the chief remaining families of distinguishable style), are related to institutional and emotional factors in the life of the exponents. These are considerations for the ethnographer and the psychologist. (A. L. Ll.)

From the Middle Ages to Bel Canto.—During the Middle Ages and Renaissance singers were theoretically divided into two still familiar categories: those who were technically proficient and, at the same time, were skilled musicians, and those who were merely the possessors of agile voices. A papal bull of 1324–25 likens the latter type of singer to "a drunkard, which indeed goeth home, but by which path he cannot tell." It is, in fact, possible to learn more about the abilities of the singers of this period from accounts of their performances recorded by outraged clerics and indignant musicians than from the more literary adulations of their admirers.

Two further contemporary sources of information on singing are the written music, where comparison of one manuscript with another will often show variations of vocal ornamentation that may point to a personal or a local style; and the popular music instruction books (*see* Musicology) that began to appear after 1500, many of which deal in some detail with singing. With these, however, as with their contemporary counterparts on instrumental playing, the modern reader is handicapped by the authors' assumption that the sounds they describe are familiar.

Indeed, the question of the quality of vocal (and instrumental) sound in early music—from, in fact, the Middle Ages to the advent of the phonograph—is a vexed one and is of crucial importance to any modern performance of the music. While musical styles have varied widely according to period, country, and individual, it is infinitely easier to reconstruct a past instrumental style than it is to form any idea of a past vocal style. The instrumentalist's technique is governed to a very large extent by the limitations and characteristics of his instrument. The enormous musical potentialities of the human voice, on the contrary, have remained constant for all periods; yet, while vocal styles must have varied as greatly as the instrumental style with which they coexisted, there is very little evidence to show what quality of

sound a medieval or Renaissance singer produced. It has been pointed out, however, that in many paintings of the 15th and 16th centuries singers are shown with tense facial muscles (in contrast to the relaxed expressions of the instrumentalists), which seems to point to a hard (to modern Western ears) vocal timbre and well-articulated attack. Furthermore, the vocal music of, for example, the period from the end of the 13th to the end of the 14th century requires remarkable vocal precision and agility, while the theoretical writings of that time, as well as those of later periods, suggest that an acute degree of exact intonation was demanded of both singers and instrumentalists.

While much medieval and Renaissance vocal music, in which the instrumental part often had an importance equal to that of the vocal line, required from the singer a strong sense of ensemble, the chordally accompanied song of the late 16th and early 17th centuries in Italy and France gave the solo singer a new freedom and a new dramatic role as interpreter of words and emotions. By the beginning of the 17th century a clearer view of vocal practice becomes possible, and singing teachers begin to publish books describing their methods. The old styles of vocal ornamentation, practically interchangeable with instrumental ornamentation, gave way to new, expressive, mannered, and purely vocal styles (*see* ORNAMENTATION), and the rise of opera put the virtuoso singer in a position of power over both composer and audience that he was to maintain for the next 200 years. (*See* also MUSIC, TEACHING OF; SONG.) (M. Mw.)

Bel Canto.—Singing, in its physical aspect, has a well-defined technique which depends on the use of the lungs, in the form of a bellows; on the larynx, which acts as a reed or vibrator; on the chest and head cavities, which have the function of an amplifier, as the tube in a wind instrument; and on the tongue, together with the palate, teeth, and lips, which articulate and impose consonants and vowels on the amplified sound. Though these four mechanisms function independently, they are nevertheless coordinated in the establishment of a vocal technique and are made to interact upon one another. From a musical viewpoint, the sounds produced by a singer are required to have tone quality; to be in tune; to be maintained in their tone quality; and to be capable of dynamic gradations. The technique of singing depends ultimately on the coordination of the various mechanisms designed to produce a propulsion of sound in a steady flow.

Modern styles of singing largely derive from the Italian *bel canto*, which has its origin in a style associated with the polyphonic music of the 16th century. Since this music was remarkable for its symbolism expressing the significance or the moods of the text, a great range of expression was required from the singers, who, in these polyphonic works, assumed something of the function of a vocal orchestra. The art of *bel canto* singing accordingly evolved to allow the singers the maximum power and variety of expression.

Bel canto singing was built primarily on the recognition that the intensity of vocal tone on a single note may be increased or diminished. The varying of this intensity was known as the *messa di voce*. There is, however, a difference between variation in intensity and variation in volume of vocal tone. The *bel canto* style depended on the technique of intensity; that is, tone was varied by increasing or decreasing the air pressure on the glottal lips and not by enlarging the oral chamber, which merely resulted in a larger tonal volume. The *bel canto* style was also based on the principle that the voice has two "tones," a diapason tone produced when the larynx is in a relatively low position, and a flute tone when the larynx assumes a higher position. Both singers and composers were aware of these distinctions, which, however, were largely obliterated when a broader style of singing was introduced by Wagner and later composers.

Among the different genres of the *bel canto* style was the *canto spianato*, using a great variety of colouring, as in the arias of Alessandro Scarlatti and his contemporaries; and the *canto fiorito*, or florid song, using the flute tone in agile arias and the diapason tone in dramatic arias. There was, furthermore, the *canto declamato*, or declamatory chant, of which the subdivisions were the *serioso* type, using the diapason tone, and the *buffo* type, usually sung in the flute tone. In the 18th-century *opera buffa*, accompanied recitative was sung in the full tone and *recitativo secco* in the lighter tone.

Physical aspects of the technique of *bel canto* singing demanded a stance in which the chest was raised and the stomach drawn in; the raising of the soft palate together with a corresponding lowering of the larynx; and the drawing back of the chin with the effect of opening the throat. Correct breathing was above all essential, and the Italians went so far as to declare that "he who knows how to breathe can sing." By a contraction of the upper abdominal muscles, control is achieved over the diaphragm, which thus enables the flow of air pressure from the lungs to be kept steady. This principle, which was the basis of singing in the 18th century, was later adopted by the Spanish tenor Manuel García, who declared that "the lungs are for tone emission, the glottis is for pitch, the oral cavity is for vowel and timbre, and the front of the mouth is for consonants." The function of the diaphragm is to regulate the pressure of air, while the larynx, as a nozzle in a water spray, determines the nature of the flow.

With the muscles in appropriate position and the reserve of air under proper control, accented notes in singing are given their full value not as startling percussive notes but in the manner of an accented note produced by a violinist who prepares his effect by the proper placing of his bow. An exercise known as *vibrazione* enabled the singer to control the voice at the larynx and, by attacking a note softly, to increase the volume by pressure of the larynx. These methods of tone production and control, used by García and later singers, originated from a practice known as *coup de glotte*.

Later Styles.—Later schools of singing paid much attention to the resonation of the voice in the "mask," that is, the cavities of the head, though this resonation did not affect the radiation power of the voice but only its volume. These singers, and also the still later *parlando* singers who effected a union of speech and singing, made a conscious use of resonation in this way and differed from the *bel canto* singers in that they exercised less control over physical mechanisms.

The development of the orchestra by Berlioz, Verdi, and Wagner encouraged singers to seek means of amplifying their voices by methods of resonation unknown in the *bel canto* style, and a new method was established of "singing on resonance." Jean de Reszke, who emphasized the function of the nose in resonation, was the main exponent of this school. Apart from the facial mask and the nose, other resonators were held to be the hard palate and the teeth. Followers of this school maintained that the voice should be "placed" in a particular area of resonation, according to the physique of the singer. This school also maintained that the underlying principle of the vocal art is to "sing as you speak."

Demands made on the voice by Wagner and later composers transformed the principles of the *bel canto* style, largely because the human voice would have been submerged by the vast orchestral resources drawn upon by these composers. Even so, the robust Wagnerian singers often had great difficulty in projecting their voices through the heavy orchestral texture. The technique of the *parlando* singer eventually enabled the *Sprechstimme* ("speaking voice") of Schoenberg and his followers to become established, while the use of the microphone in popular music led to an even closer approximation to the speaking voice.

In other developments of vocal music about the mid-20th century, the tendency, particularly when the singer was required to place emphasis upon an isolated word or a dramatic effect, was to associate the *parlando* and resonation styles with the remaining traditions of *bel canto* singing. *See* also VOICE.

SINGLE TAX, as the term originally developed, referred to a tax upon land values and was proposed as the sole source of government revenues, to replace all existing taxes.

The idea of a single tax on land is at least two centuries old, appearing, for example, in the writings of the 18th-century physiocrats and of the famous English economist David Ricardo. But the term itself and the modern single-tax movement originated with the publication of *Progress and Poverty*, the first systematic presentation of the plan, written by a San Francisco newspaper editor, Henry George (*q.v.*), in 1879. The proposal gained substan-

tial strength in ensuing decades, and then gradually declined in popular appeal. There were, however, in the second half of the 20th century still a number of very devoted disciples, the best known of whom was Harry Gunnison Brown, and publications in the field were sponsored by the Schalkenbach Foundation, established for this purpose.

The basic arguments for the single tax were twofold. On the one hand it was argued that land rent, unlike other income, is a product of the growth of the economy, not of individual effort, and thus society is justified in recovering it to support the costs of government. Since land rent is a surplus return, a tax on it cannot be shifted forward to consumers, and cannot reduce the supply of land available to the economy. Land rent is essentially an "unearned increment" and thus governments are justified in taxing it at a 100% annual rate. On the other hand, use of the tax would allow elimination of other forms of tax, and particularly taxes on buildings, which interfere with construction and economic development generally. Thus the tax would stimulate growth of the economy. The argument was also made that a single tax would be simple, uncomplicated, and economical to collect.

The critics of the proposal condemn the tax as contrary to usual standards of ability to pay, since there is no correlation between land ownership and total wealth and income. Portions of other incomes may be regarded as much "unearned" as land rent. If the tax were imposed today it would in large measure rest on the wrong persons, on those who had invested their capital in land rather than those who had benefited from the increase in land values. Furthermore, as a practical matter, separation of land values from those of buildings and other improvements would be extremely difficult.

While no attempt has ever been made to use the land tax as a single tax, and it is universally recognized that the levy could not meet all needs of government, several jurisdictions have applied their property taxes to land only instead of to land and buildings, or have taxed land more heavily than buildings. Examples include the states of Australia, the western provinces in Canada, the Commonwealth countries in East and Central Africa, and a few municipalities in the United States. A land tax was imposed for a short time in England before World War II, and "unearned increment" was taxed under the Labour government from 1945 to 1951. The single-tax supporters claim that the use of the plan in Australia has greatly stimulated building in the states using it, but the evidence is not conclusive.

BIBLIOGRAPHY.—Henry George, *Progress and Poverty* (1879); H. G. Brown, *The Economic Basis of Tax Reform* (1932), *Land-Value Taxation Around the World* (1955); A. N. Young, *The Single Tax Movement in the United States* (1916). (J. F. D.)

SINHA, SATYENDRA PRASSANO (1ST BARON SINHA OF RAIPUR) (1864–1928), Indian lawyer and statesman, who had an extremely successful legal career, won high esteem in nationalist circles, and was appointed to high office under the British government, was born in the village of Raipur, Birbhum district, Bengal, in June 1864. He was educated at the Presidency College, Calcutta, and was called to the bar from Lincoln's Inn. He was the first Indian to be appointed advocate-general of Bengal (1907) and the first to be appointed to the governor-general's Executive Council, in which he served as law member during 1909–10. He was knighted in 1914, presided over the Indian National Congress Party's session at Bombay in 1915, and subsequently served in the Imperial War Cabinet. In 1919 he joined the Lloyd George ministry as undersecretary for India, being raised to the peerage as Baron Sinha of Raipur, and steered through the House of Lords the bill to enact the Montagu-Chelmsford proposals for the reform of the Indian constitution. In 1920 he was appointed governor of Bihar and Orissa, being again the first Indian to hold such office under the British, but resigned on grounds of ill health in the following year. He was appointed a member of the Judicial Committee of the Privy Council in 1926. He died on March 6, 1928, at Berhampur. (KE. A. B.)

SINHAILIEN (HSIN-HAI-LIEN), an important city and rail-terminal seaport on the Yellow sea in north Kiangsu province, China. It is an amalgamation of several settlements founded at different times, bearing many individual names between 1900 and 1949. Haichow (Hai-chou) was the name once applied to an old junk port at the coastal end of a canal system. Sinpu (Hsin-p'u) was the early inland eastern terminal of the Lunghai railway. Lao-yao and Lien-yun-kang appeared on some maps to name a new deepwater port built in the 1930s south of the unusable silted estuary of the Yun-yen Ho and Shu Ho. Tung-hai and Ta-p'u both were used as names for the central portion of the urban complex. Sinhailien, applied to the whole city, and Lien-yun-kang, applied to the port, were both in use in the second half of the 20th century. Pop. (1953) 207,600. (J. E. SR.)

SINHALESE LANGUAGE. Sinhalese is an Indo-European language of the Indo-Aryan family, spoken in the island of Ceylon by about 7,000,000 people (1965). It has developed along independent lines, because of the separation of Ceylon from the mainland of India and from the main Indo-Aryan language areas of north India by the Dravidian-speaking southern regions. In the 19th century the name "Sinhalese" was usually written "Cingalese"; the people themselves call it "Sinhala" (accent on first syllable).

The language has its own syllabic script, the letters or characters being of a generally rounded appearance. For literary purposes, there is a complete set of equivalents for all Sanskrit letters, but for representing the Sinhalese sounds, 12 vowel characters and 26 consonant characters are sufficient. Orthography, however, is not very fixed. Accent is not significant. The vocabulary is basically Indo-Aryan, with a considerable number of Dravidian (mostly Tamil) loanwords. There are also some Portuguese and Dutch loanwords bequeathed by the pre-British colonizers of Ceylon, and a large number of English loanwords. English loanwords, however, are still usually avoided in written Sinhalese. Words can be classified as noun, verb, or particle. Among nouns, five case-forms can be distinguished. In literary Sinhalese, finite verbs have the archaic personal inflections, but in the spoken language these are replaced by a single invariable form. Much use is made of a verbal absolutive (having done, having gone), and of a relative participle (cow eating grass; we living house; *i.e.*, the grass which the cow is eating; the house which we are living in).

The earliest surviving specimens of the language are brief inscriptions on rock, in Brahmi letters (*see* ALPHABET: *Other Alphabets*), of which the earliest date from *c.* 200 B.C. The language of these inscriptions does not appear to be greatly different from other Indian Prakrits (*i.e.*, chronologically Middle Indo-Aryan languages) at that time. About the 5th century A.D., however, considerable developments occurred, the course of which can be traced in barest outline through inscriptions; and by the period of the earliest surviving literary works, toward the end of the 10th century, both the language and the script are much closer to their present forms than to those of the early centuries. By *c.* 1250, the literary language had attained a form which has hardly varied since, though the 20th-century spoken language (still scarcely used for literary purposes) differs from it considerably.

There is a considerable literature dating from about A.D. 1000, nearly all Buddhist in inspiration and subject matter, and much of it in prose. The best known classical poets are Śrī Rāhula (15th century) and Alagiyavanna (*c.* 1600). Modern secular literature is entirely confined to the 20th century, and by 1950 had still not attained substantial proportions. The novels of W. A. Silva (d. 1957) have had a wide popularity.

BIBLIOGRAPHY.—Introduction to W. Geiger, *Grammar of the Sinhalese Language* (1938); introduction to Sir D. B. Jayatilaka, *Dictionary of the Sinhalese Language*, vol. i, pt. i (1935); C. E. Godakumbura, *Sinhalese Literature* (1955); E. R. Sarathchandra, *The Sinhalese Novel* (1950). (C. H. B. R.)

SINING (HSI-NING), capital of Tsinghai Province, China, on the south bank of Sining River, a tributary of the Yellow River, and 110 mi. WNW of Lan-chou. Sining, which is situated on the northeast margin of the Tibetan highlands at an elevation of 7,500 ft., is the centre of an agricultural district east of Lake Koko Nor, producing spring wheat and barley. Wool, hides and salt are also marketed. Long linked only by highways with the rest of China, the city was in 1959 reached by a railroad from

Lan-chou. Sining is the gateway to the Tsaidam basin, where the Chinese have developed new petroleum fields and mining industries. Pop. (1953) 93,700. (T. Sd.)

SINKIANG (Hsin-Chiang), name of a former province of westernmost China, converted September 1955 into the Sinkiang Uighur Autonomous Region (Hsin-chiang Uighur Tzu-chih Ch'ü), a political division at the provincial level set aside for the Uighur ethnic minority. Area 635,829 sq.mi. (1,646,799 sq.km.) pop. (1957 est.) 5,640,000. The name Sinkiang means "new borderland" and is roughly coextensive with what was historically known as Chinese Turkistan.

Sinkiang is bordered on the northeast by the Mongolian People's Republic, on the southwest by Kashmir and a narrow strip of Afghanistan, and on the west and north by the Central Asian republics of the Soviet Union. Physiographically, Sinkiang consists of two mountain-ringed basins separated by the east-west-trending Tien Shan (*q.v.*). The Dzungarian basin, in the north, has an elevation of 600 to 1,500 ft., and receives summer rainfall that makes Dzungaria (*q.v.*) ideal as a grazing land. The Tarim (*q.v.*) basin, south of the Tien Shan, has an average elevation of 2,500 to 3,000 ft. It is much more arid than Dzungaria, and its centre is occupied by the sandy Takla Makan (*q.v.*) Desert. At the eastern end of the Tarim basin lies the salt lake and marshland of Lop Nor (*q.v.*).

Sinkiang is multinational. The Uighurs (*see* Uighur), a Muslim Turkic-speaking people, the most important ethnic group, numbered 3,640,125 (1953) and comprised 75% of the population. They settled primarily in irrigated agricultural oases along the northern foot of the Tien Shan, and around the edge of the Tarim basin. The second largest ethnic group is the Kazakh, also a Muslim Turkic people, related to the Kazakhs of the Soviet Union. They numbered 475,000 in Sinkiang in the 1953 census. Unlike the Uighurs, the Kazakhs are primarily nomadic herders and graze their livestock in the Dzungarian pasture lands where a separate political division (Ili Kazakh Autonomous Chou, with its capital at Kuldja) was set aside for them. The Kazakhs also have two autonomous *hsien* (counties), at Mu-lei-ho and Barkol (Chen-hsi), in eastern Sinkiang (*see* also Turkic Peoples).

There were about 300,000 Chinese in Sinkiang, residing mainly in cities and towns, where they worked as government officials, merchants, and professional people. Chinese Muslims (*Hui*), numbering 200,000 in 1953, were concentrated in the Urumchi area in the Ch'ang-chi Autonomous Chou and also in the Yen-ch'i Autonomous district, south of the Tien Shan. There were 120,000 Mongols in Sinkiang, occupied mainly as livestock herders. Mongol autonomous areas include the Boro Tala Autonomous Chou, on the south slopes of the Dzungarian Ala-Tau on the Soviet border; the Bayin Gol Autonomous Chou in the basin of the lake Baghrash Kol south of the Tien Shan; and the Kobuk-Sai-li Autonomous district, on the slopes of the Saur range on the Soviet border. The Kirgiz people (70,000 in 1953) constitute the Kizil Darya Autonomous Chou, north and west of Kashgar, adjoining the Kirgiz S.S.R. of the Soviet Union. About 20,000 Sibo people, a Tungusic group near Kuldja, and 15,000 Tadzhiks are also set up in an autonomous district. (*See* also Mongol; Tadzhik; Tungus.)

Sinkiang's economy is primarily agricultural. Its cultivated land was expanded after 1949 through addition of new irrigated areas, such as the Manas River project on the northern foot of the Tien Shan. Grain crops, mainly wheat, corn, kaoliang, and sorghum, and ginned cotton are important products of the Tarim basin oases. Sinkiang's livestock herds, estimated about 21,000,000 head in the late 1950s, included about one-quarter of China's sheep, which yielded more than 60% of the nation's wool. In the early 1960s industry was in an early stage of development, with handicrafts (cotton goods, paper, rugs, leather goods, and jade articles) accounting for two-thirds of the total value of manufacturing. The principal cities and their 1953 census populations are: the regional capital, Urumchi (140,700), Kashgar (91,000), Yarkand (80,000) (*qq.v.*), I-Ning (Kuldja) (108,200), and Aksu (90,000). The development of modern industry depends in part on the development of the region's mineral resources and the construction of modern transportation routes. Petroleum was produced on a small scale at Tu-shan-tzu, near Wu-su. A large petroleum centre at Karamai, 70 mi. N of Wu-su, was under development in the early 1960s. Mineral prospects also include tungsten, molybdenum, and other nonferrous metals, as well as coal and iron ore. A major railroad crossing Sinkiang from Kansu province to the junction of Aktogai on the Turksib Railroad in the Soviet Union was completed in the early 1960s.

South Sinkiang was controlled by China (200 B.C.–A.D. 220) during the Han dynasty. The Uighurs ruled for many centuries until the area was conquered by Genghis Khan and his successors and later by the Kalmyks. Sinkiang eventually came under Chinese control in the 17th century during the Ch'ing dynasty. It became a Chinese province in 1884. During the Nationalist period, warlords ruled Sinkiang, sometimes with Soviet support. Chinese Communists seized control in October 1949. (T. Sd.)

SINKING FUND: *see* Debt, Public.

SINO-JAPANESE WAR. For the first conflict known by this name, in which China and Japan fought (1894–95) over Korea, *see* Chinese-Japanese War. In this war, China was quickly defeated by Japan, which was then emerging as a world power. For the second Sino-Japanese War, which was part of World War II, *see* World War II: *The Chinese-Japanese War (1937–45)*. China suffered invasion on a vast scale but finally saw its enemy defeated.

SINOP (ancient Sinope), the chief town of Sinop *il* (province) on the north coast of Turkey, lies on a low isthmus joining the promontory of Karababa Tepesi (Boz Tepe) to the mainland, and 190 mi. (306 km.) NE of Ankara. Pop. (1960) 10,214. Though it possesses the only safe natural roadstead on the north coast of Asia Minor, the difficulties of communication with the interior and the rivalry of Inebolu on the west and Samsun on the east prevented Sinop from becoming a great commercial centre. The town is shut off from the plateau by forest-clad mountains. On the isthmus, toward the mainland, stands a huge but mostly ruined castle, originally Byzantine and afterward strengthened by the Seljuk sultans. Of early Roman or Greek antiquities there are only the columns, architraves, and inscribed stones built into the old walls; but the ancient local coinage furnishes a very beautiful and interesting series of types.

Sinope, whose origin was assigned by its ancient inhabitants to Autolycus, a companion of Hercules, was founded sometime in the 7th century B.C. by the Ionians of Miletus, and ultimately became the most flourishing Greek settlement on the Black Sea. This was so because it was not only a port but the terminus of a great caravan route from the Euphrates to the Black Sea, over which were brought the products of central Asia and Cappadocia (whence came the famous "Sinopic" red earth, probably red ochre). In the 5th century B.C. it received a colony of Athenians, and by the 4th it had extended its authority over a considerable tract of country. Its fleet was dominant in the Black Sea, except toward the west, where it shared the field with Byzantium. When in 220 B.C. Sinope was attacked by the king of Pontus, the Rhodians enabled it to maintain its independence. In 183 B.C. the city, taken by surprise and captured by Pharnaces I, became the capital of the Pontic monarchy (*see* Pontus). Under Mithradates VI the Great, who was born in Sinope (as was Diogenes), it had just been raised to the highest degree of prosperity, with fine buildings, naval arsenals, and well-built harbours, when it was captured by Lucullus and nearly destroyed by fire (70 B.C.). Under Julius Caesar the city became a Roman colony, but was already declining with the diversion of traffic to Ephesus, the port for the Roman Empire, and in part to Amisus (Samsun).

In the Middle Ages it became subject to the Greek Empire of Trebizond (Trabzon), and passed into the hands of the Seljuk Turks; and in 1461 was incorporated in the Ottoman Empire. In November 1853 the Russian vice-admiral P. S. Nakhimov destroyed there a division of the Turkish fleet and reduced a good part of the town to ashes; this incident helped to sway British opinion in favour of entering the Crimean War.

Sinop Il had a population in 1960 of 249,730. It is mountainous and drained by the Gok Irmak. Olives, flax, and tobacco are the chief crops. (N. Tu.; S. Er.; E. Tu.)

SINO-TIBETAN LANGUAGES: *see* SOUTHEAST ASIAN LANGUAGES.

SINS, SEVEN DEADLY, a classification that goes back to the early history of Christian monasticism. A sin was classified as deadly not merely because it was a serious offense morally but because "it gives rise to others, especially in the manner of a final cause" or motivation (Thomas Aquinas). The traditional catalogue of the seven deadly sins is: (1) vainglory or pride; (2) covetousness; (3) lust, understood as inordinate or illicit sexual desire; (4) envy; (5) gluttony, which usually included drunkenness; (6) anger; and (7) sloth, sometimes called acedia or accidia. The classical discussion is in Thomas Aquinas, *Summa theologiae*, i-ii, Q. 84, art. 4. (J. J. PN.)

SINTER, in petrology, certain mineral deposits, more or less porous or vesicular (cinderlike) in texture. At least two kinds of sinter are recognized: one siliceous, the other calcareous. Siliceous sinter is a deposit of opaline or amorphous silica from hot springs and geysers, occurring as an incrustation around the springs, and sometimes forming conical mounds or terraces. The deposition of siliceous sinter is largely due to the action of algae and other forms of vegetation in the thermal waters. Siliceous sinter has also been called geyserite and fiorite.

Calcareous sinter, sometimes called tufa, calcareous tufa, or calc-tufa, is a deposit of calcium carbonate, exemplified by travertine, which formed the principal building stone of Rome, and was named from the Italian *travertino*, a corruption of *tiburtino*, the stone of Tibur, now Tivoli (*see* STONE). So-called petrifying springs, not uncommon in limestone districts, yield calcareous waters which deposit a sintery incrustation on objects exposed to their action.

The cavities in calcareous sinter are partly due to the decay of mosses and other vegetable structures that have assisted in its precipitation. Even in thermal waters, like the hot springs of Carlsbad, Bohemia, which deposit *Sprudelstein*, the origin of the deposits is mainly due to organic agencies, as shown as far back as 1862 by the German botanist Ferdinand Julius Cohn. While calcareous deposits in the open air form sinterlike travertine, those in caves constitute stalactites and stalagmites. *See* also CALCITE; STALACTITES AND STALAGMITES.

SINT-NIKLAAS (SAINT NICOLAS), a town of Belgium in East Flanders province, lies 13½ mi. (22 km.) WSW of Antwerp. Pop. (1961) 47,819. The town is a rail junction and the centre of the Waasland market-gardening district, its market place of 7½ ac. (3 ha.) being the largest in Belgium. Notable buildings are the new town hall (1876), the old town hall (1663) and other 17th-century buildings, the Church of St. Nicholas (15th–16th century; restored 1900), and Walburg Château (16th century) in the public park. The museum contains a Mercator collection. The town received its charter in 1513 and has textile industries. (J. DR.)

SINTRA (CINTRA), a town of Lisboa district, west central Portugal, lies 17 mi. (28 km.) W.N.W. of Lisbon by rail or road. Pop. (1960) 20,321. It is picturesquely situated on the northern slope of the Serra de Sintra, a rugged mountain mass largely overgrown with pine, mimosa, cedar, eucalyptus, cork and other trees, above which rise bare and jagged gray rocks (Cruz Alta, 1,732 ft.). The beauty of Sintra is celebrated by Lord Byron in *Childe Harold* and by Luís Vaz de Camões in *Os Lusíadas*. On one of the peaks is the Palacio da Pena, a fantastic 19th-century building, partly an adaptation of a 16th-century monastery and partly an imitation of a medieval fortress. On another peak is the Castelo dos Mouros, an extensive Moorish fortification. In the town itself is a 12th–15th-century royal palace, partly Moorish, partly debased Gothic, and remarkable for the decoration of its rooms and its two immense conical chimneys. On the road to the village of Colares are the palace and park of Montserrate. The park, with its tropical luxuriance of vegetation and its variety of lake, forest and mountain scenery, is one of the finest examples of landscape gardening in the Iberian peninsula.

Sintra has given its name to two conventions, one in 1509 between Portugal and Castile settling differences concerning voyages of exploration, and one in 1808 by which the British and Portu-

BURTON HOLMES FROM EWING GALLOWAY

PALACIO DA PENA, SINTRA, CASTLE OF FANTASTIC DESIGN BUILT IN THE EARLY 19TH CENTURY

guese allowed the beaten French army to return home during the Peninsular War.

SINŬIJU (Japanese, SHINGISHU), a city and capital of North Pyongan Province, North Korea, was developed by the Japanese during their occupation (1910–45) at a site 7 mi. W of the old city of Uiju where a railroad bridge was built across the Yalu River connecting with the Manchurian city of Antung. Sinŭiju (or New Uiju) is an industrial and commercial city and wood rafted down the Yalu forms the base of a large forest products industry. Trade with Manchuria and China is funneled through the city to Korea. Pop. (1958 est.) 200,000. (S. McC.)

SINUS, in anatomy, is a space filled with blood (*e.g.*, in the dura mater covering the brain) or air (*e.g.*, in the cranial bones). The word is also used by surgeons to signify a discharging tract that will not heal and has in many cases a foreign body or dead bone at the bottom. Popularly, the term sinus is used most often to designate one of the air cavities (paranasal sinuses) connected with the nose. The largest of these cavities (maxillary sinus), is in the cheekbone. The next in size is in the forehead (frontal sinus). Smaller cavities open into the back (sphenoidal sinus) and sides (ethmoidal sinus) of the nose. The paranasal sinuses are lined with ciliated mucous membrane continuous with that of the nose.

Sinusitis is the medical term for disease of the paranasal sinuses, popularly called sinus trouble. Sinusitis may be either acute or chronic.

Acute sinusitis may occur as an aftermath of a cold, due to secondary bacterial infection, or it may be precipitated by faulty breathing habits in swimming or by sudden changes in barometric pressure in flying or diving. Allergy and conditions resulting in lowered general resistance may predispose to attacks. Infections of the upper molar and bicuspid teeth may cause acute maxillary sinusitis.

Symptoms are pain and headache, tenderness over the affected area, nasal obstruction and discharge and malaise. Conservative

treatment, bed rest, fluids, heat externally, sedation for pain, and vasoconstrictor sprays to relieve obstruction generally are effective. Persistence of symptoms may indicate the use of suction to promote drainage, and antibiotic therapy, based upon culture and sensitivity studies. X-ray examination may suggest irrigation of the maxillary or external drainage of the frontal sinus.

Chronic sinusitis may follow repeated or neglected attacks of acute sinusitis, particularly if there is impaired nasal breathing and drainage due to intranasal obstruction. These may be aggravated by poor environmental conditions, dust or excessively dry air of heated dwellings. Dental infection in the upper molar and bicuspid area may cause chronic disease in the maxillary sinus.

Symptoms are tendency to colds, purulent nasal discharge, obstructed breathing, loss of smell and sometimes headaches. Pain is rare except during acute phases. Foul-smelling discharge may indicate dental origin. While there may be increased postnasal discharge, its significance often is overstressed, as the normal drainage from the nose is posteriorly to the pharynx. Frequently symptoms attributed to sinusitis may be manifestations of general conditions (particularly allergy or disturbance of the endocrine system) or result of faulty nutritional or living conditions or of habitual use of nasal sprays or drugs. Stress and tension may be contributing factors.

Successful treatment depends upon a thorough systematic evaluation of the patient, including X-ray and other diagnostic studies, to determine not only the status of the sinuses but also the significance of any underlying conditions. Surgery may be indicated to restore ventilation and drainage, to eliminate irreversibly diseased tissues, and to avoid complications. (S. HA.; F. T. HI.; X.)

SION (SITTEN), capital of the canton of Valais, Switzerland, is at an altitude of 1,700 ft. (518 m.) in the Rhone Valley, 60 mi. (97 km.) SE of Lausanne. Pop. (1960) 16,051. It is built around two prominent hillocks: the northern has the remains of the 13th-century Tourbillon Castle and the southern, those of the castle of Valère (Valeria), now containing a historical museum and a mainly 13th-century church. Other notable buildings are the Gothic Cathedral of Notre-Dame, St. Theodule's Church, and the 17th-century town hall. The town is linked by road with the summer and winter resorts of the canton, and is on rail and road routes from Lausanne to Milan via the Simplon Pass (*q.v.*) and Tunnel. Sion (ancient Sedunum) dates from Roman times and became an episcopal see in the 6th century. From 999 the bishops of Sion held the temporal and spiritual power in Valais which they retained, at least in part, until 1798 (*see* VALAIS: *History*).

SIOUAN INDIANS, North American linguistic family that takes its name from that of the largest tribe, the Sioux or Dakota (*q.v.*). Next to Algonkians, they were perhaps the most populous stock north of Mexico. They held three territories, the largest mainly west of the Mississippi River, another east of the Appalachian Mountains in Virginia and the Carolinas; the smallest, in two fragments, in Mississippi. The last two divisions are nearly extinct. The culture was not uniform, but accorded with the region in which each tribe lived. Physical types probably varied similarly. In the Great Plains, the Siouans were the preponderant linguistic stock (*see* PLAINS INDIANS). The principal Siouan tribes were (those asterisked being separately treated): (1) in the west, *Dakota and *Assiniboin, the former really seven tribes; *Mandan, *Hidatsa, and *Crow; *Winnebago; tribes speaking the Chiwere dialect, namely the Iowa, Oto Missouri; tribes speaking Dhegiha, viz., *Omaha, Ponca, Kansa, *Osage, Quapaw or Arkansa; (2) in the south, Ofo and *Biloxi; (3) in the east, Monacan, Manahoac, Tutelo, Saponi, Occaneechi, Woccon, Catawba, Santee, Cheraw or Saraw, and probably Wateree, Congaree, Pedee, and others. In the 1960s there were about 60,000 Siouan Indians; their highest level of population was probably at least 80,000. *See* also DORSEY, JAMES OWEN; and references under "Siouan Indians" in the Index.

See W. L. Chafe, "Another Look at Siouan and Iroquoian," *American Anthropologist*, vol. 66 (Aug. 1964). (A. L. K.; X.)

SIOUX CITY, a city of northwestern Iowa, U.S., lies on the bluffs of the Missouri at the mouths of the Big Sioux and Floyd rivers, where the states of Iowa, South Dakota and Nebraska meet; a port of entry and the seat of Woodbury county.

In 1804 Captains Meriwether Lewis and William Clark visited the site of Sioux City and there buried Sgt. Charles Floyd, the only man lost on their famous journey (*see* LEWIS AND CLARK EXPEDITION). Theophile Bruguier, a French-Canadian fur trader, settled there in 1849 and married the daughter of War Eagle, a powerful chief of the Yankton Sioux. In 1854 John K. Cook platted Sioux City while surveying the region for the U.S. government. It prospered originally as a supply centre for the northern plains and as a land-office town. The first steamboat from St. Louis reached Sioux City in 1856 and in 1857 the city, with a population of 400, was incorporated. The first railroad arrived in 1868 and the first meat-packing plant was opened a few years later. During the boom of the 1880s, when settlers occupied northwestern Iowa and the Dakotas, the packing industry expanded and the population jumped from 7,366 to 37,806. Although retarded by depression in the 1890s, during which its population dropped, Sioux City grew rapidly between 1900 and 1920, when its population topped 70,000. Pop. (1960) 89,159; standard metropolitan statistical area (Woodbury county, Iowa, and Dakota county, Neb.) 120,017. (For comparative population figures *see* table in IOWA: *Population*.)

Early settlers in Woodbury county came mainly from the Ohio valley, New York, Pennsylvania and older Iowa. French Canadians, Germans and Scandinavians also settled there and a Russian group was present by 1900. A community of Negro packing-house workers developed after 1910.

Commerce is marked by heavy wholesale trade and grain exchange activity. Industries include meat packing, food processing, and feed and clothing production, as well as other diversified manufactures.

Morningside college, a Methodist liberal arts institution which maintains a conservatory of music, was founded in Sioux City as the University of the Northwest in 1889. In 1930 the Sisters of St. Francis established Briar Cliff college (Roman Catholic) for women. Sioux City has a symphony orchestra and an extensive park system in which are preserved memorials of Indian days including War Eagle's grave and the Council Oak of the Sioux. Sioux City adopted a commission form of government in 1910 and in 1953 switched to the council-manager plan. (A. G. Bo.)

SIOUX FALLS, the seat of Minnehaha County, in southeastern South Dakota, U.S., a few miles from the Minnesota border and about 75 mi. N of Sioux City, Ia., on the rolling hills and in the wooded valley bordering the Big Sioux River. Pop. (1960) 65,466; standard metropolitan statistical area (Minnehaha County) 86,575. (For comparative population figures *see* table in SOUTH DAKOTA: *Population*.) The population of the city is much like that of the Dakota farm area, heavily Scandinavian and German.

Except for an early fur-trading establishment that later became Fort Pierre, Sioux Falls is the oldest community in South Dakota. It was founded in 1857 by Iowa and Minnesota land speculators who were disappointed when in 1861 Yankton, rather than Sioux Falls, was named capital of the new Territory of Dakota. The settlement was deserted during the uprising of the Minnesota Sioux which started the following year. Permanent settlement dates from 1870; it was incorporated as a village in 1877 and as a city in 1883. With the influx of settlers during the Dakota boom of the 1880s, followed by the decline in importance of steamboat transportation which had given Yankton its earlier prominence, Sioux Falls developed into the principal community in the state. The falls of the river, from which the city takes its name, were harnessed for water power in 1873 and the shipment of "Sioux Falls granite," a hard quartzite once used for paving blocks and as a building stone, began with the arrival of the first railroad in 1878. Because of the laxity of the South Dakota divorce laws prior to 1908 Sioux Falls was for a time nationally famous as a "divorce mill."

Located in an area of diversified farming where corn is the principal crop and livestock feeding the main source of income, Sioux Falls came into its own as a result of the wide use of trucks and the increasing dependence of farmers upon specialized farm

machinery, fertilizers, and mixed feeds. A distributing centre for the neighbouring farm area of Iowa, Minnesota, and South Dakota, it has over 200 wholesale firms. Its main industries are food processing, particularly meat packing, and wood and metal fabrication. Farm prosperity of the World War II period, which lasted several years after the war, also helped to build Sioux Falls as a shopping, entertainment, and commercial centre.

Educational facilities include Augustana, a Lutheran college founded in 1860, and Sioux Falls College, a Baptist institution established in 1883. The city has several parks and two golf courses. It has superior medical facilities, including a veterans hospital and a crippled children's hospital. The state penitentiary and a school for the deaf are also located there. (E. W. St.)

SIOUX INDIANS: *see* Dakota; Siouan Indians.

SIPHNOS (modern Greek Sifnos), an island of the Cyclades, Greece, 30 mi. (48 km.) SW of Syros. Area 32 sq.mi. (82 sq.km.); pop. (1961) 2,258. Along the west slope of a limestone ridge, whose principal summits, Ayios Ilias (2,129 ft. [649 m.]) and Ayios Simeon (1,624 ft. [495 m.]), are crowned by old Byzantine churches, lie a series of villages. Apollonia, the modern capital, has the name of an ancient town; Kastro, which has medieval fortifications, represents the ancient city of Siphnos; excavations by the British have yielded finds from the 8th century B.C. to Roman times. In ancient times Siphnos was colonized by Ionians from Athens. It refused tribute to Xerxes, and sent one ship to fight on the Greek side at Salamis (480 B.C.). The great wealth of Siphnos came from its gold and silver mines, from the proceeds of which it built a treasury at Delphi about 525 B.C. (*see* Greek Art); but by the 1st century A.D. these were flooded and the island was poor. Byzantine churches and convents are scattered about the island. The School of the Holy Tomb was founded by Greek refugees from the iconoclastic persecutions at Byzantium and became a centre of Greek culture until the 19th century. After the capture of Constantinople in 1204 Siphnos became part of the Venetian duchy of Naxos from 1207, but was recovered by the Byzantines in the 1270s. Regained by the Venetians, it was ruled by the Da Corogna family (1307–1456) and after 1456 by the Gozzadini, who were not expelled by the Turks until 1617. Olive oil is a principal product. (J. Bo.)

SIPHON (Syphon), an instrument, usually in the form of a tube bent to form two legs of unequal length, for conveying liquid over the edge of a vessel and delivering it at a lower level (Lat. *sipho;* Gr. *siphon,* "a tube"). A siphon also affords a ready method of transferring liquids over a hump or elevation. The action depends upon the difference of the pressure on the liquid at the extremities of the tube, the flow being toward the lower level and ceasing when the levels coincide. Siphons are made of glass, rubber, copper, or lead, according to the liquid which is to be transferred. The simple siphon is used by filling it with the liquid to be decanted and plunging the shorter leg into the liquid to be drawn.

Innumerable forms of siphons have been devised, adapted for all purposes, and provided with arrangements for filling the tube or for keeping it full and starting it into action automatically when required. Pipes conveying the water of an aqueduct across a valley and following the contour of the sides are siphons when they depend on the principle of the above instrument. Such a large siphon must be fitted with a valve for allowing the escape of entrapped air at the top of the hump or upper elevation of the siphon pipeline. In the siphon used as a container for carbonated (aerated) waters a tube passes through the neck of the vessel, one end terminating in a curved spout, the other reaching to the bottom of the vessel. When a spring valve on the tube is opened, the water is driven out by the pressure of the gas it contains. The "Regency portable fountain," patented in 1825 by Charles Plinth, was the prototype of the modern siphon; it differed in having a stopcock in place of a spring valve. The *siphon champenois* of Deleuze and Dutillet (1829) was a hollow corkscrew, with a valve, which was passed through the cork into a bottle of effervescent liquid. The *vase siphoïde* of Antoine Perpigna, patented in 1837, was essentially the modern siphon, its head being fitted with a spring valve.

SIPHONAPTERA: *see* Flea.

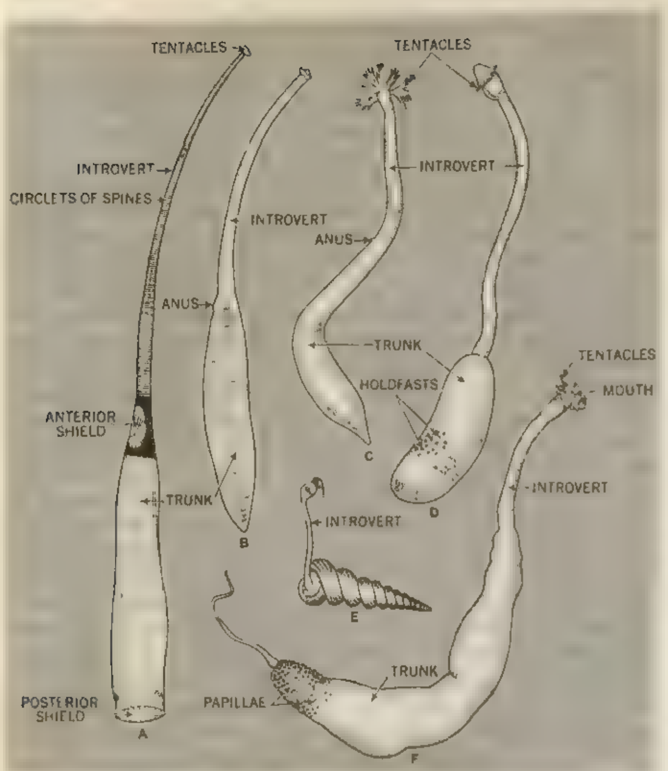
SIPPAR, an ancient city of Babylonia on the east bank of the Euphrates, identified with Abu Habbah (Tell Abu Habbah), about 20 mi. (32 km.) SW of Baghdad, Iraq. In 1881–82 H. Rassam began tracing a rectangular temenos wall which measured 1,050 × 787 ft. (320 × 240 m.); it enclosed a ziggurat with temples dedicated to the sun-god Shamash and to the goddess Anunitum. The city was subject to the 1st dynasty of Babylon. Later kings of this dynasty fostered the cult of Shamash; Hammurabi among others repaired the city walls, dug a moat, dredged the Euphrates, and built a protective quay. After the 1st dynasty of Babylon little is known about Sippar till 1174 B.C. when it was sacked and looted by the Elamite king Kutir-Nahhunte. It must have recovered, for it was captured by the Assyrian king Tiglath-pileser I, *c.* 1100 B.C. Sippar, however, survived these shocks and under the 8th dynasty of Babylon King Nabu-apal-iddina, *c.* 880 B.C., rebuilt the great temple of Shamash, which had previously been wrecked by the nomadic Sutu of the desert. This king recorded that while digging in the ruins he found the ancient image of the god, and he depicted himself on a stone memorial tablet as a suppliant before Shamash, who is enthroned beneath a canopy; in front of him the sun-disk, suspended from heaven by a rope, rests on an altar. This tablet, now in the British Museum, was found by King Nabopolassar (626–605 B.C.) when he restored the temple. Thousands of cuneiform tablets, mostly commercial, and contemporary with Nabonidus, the last king of Babylon, have come from this site. After the Battle of Opis (539 B.C.), when Nabonidus' army was defeated by the Persian king Cyrus, Sippar opened its gates and offered no resistance to the conqueror.

Bibliography.—H. Rassam, *Asshur and the Land of Nimrod* (1897); L. W. King, *A History of Babylon* (1919); W. Andrae and J. Jordan, "Abu-Habbah-Sippar" in *Iraq*, vol. i (1934); A. Parrot, *Archéologie mésopotamienne* (1946); T. Jacobsen, "The Waters of Ur" in *Iraq*, vol. xxii (1960). (M. E. L. M.)

SIPUNCULIDA, a phylum of bottom-dwelling marine worms that inhabit burrows, tubes, or borrowed shells; they are adapted to this mode of life by a forward displacement of the anus, on the dorsal surface near the base of the protrusible proboscis, or introvert. The body is thus elongated in a ventral direction almost at right angles to the anteroposterior axis, as it is in Bryozoa and Phoronida (*qq.v.*). This curious arrangement led E. R. Lankester to coin the now obsolete name Podaxonia for these three otherwise unrelated groups. The Sipunculida are distinguished from the Annelida (*q.v.*) not so much by the position of the anus—in some annelids the anus may be displaced forward dorsally—as by the total lack of segmentation. The former phylum Gephyrea, which included organisms now referred to the Sipunculida, Echiurida, and Priapulida, cannot be sustained in view of the lack of relationships between its members.

Natural History.—Although sipunculids are bottom dwellers, it is not necessary for them to inhabit U-shaped tubes with entrance and exit holes, as in the case of echiurids and many tube-dwelling annelids, since the anus is shifted forward toward the mouth. Some species burrow to considerable depths; others construct tubes by cementing sand grains together; while yet others live in abandoned tubes of annelids or shells of mollusks. Some species of *Aspidosiphon* form a symbiotic relationship with a coral that grows on the outside of the snail shell originally inhabited by the worm: as the coral grows, the worm enlarges its tube, which becomes encased in the calcareous base of the coral. In this and related genera, horny or calcareous preanal plates are used to close the mouth of the burrow.

The group is found predominantly in shallow water along shores, but representatives have been taken from great depths of the ocean: the "Galathea" expedition (1950–52) recorded sipunculids from the Kermadec Deep, off New Zealand, at 26,935 ft. (8,210 m.). Some species are almost cosmopolitan, such as *Sipunculus nudus;* others may be circumarctic or circumtropical. Each geographical area tends to have its characteristic assemblage of species. Antarctic species are frequently identical with, or closely related to, forms that occur in the far north; *e.g. Golfingia* species and *Phascolion strombi.* However, the great majority of species are found in warmer waters.



FROM L. H. HYMAN, "THE INVERTEBRATES: SMALLER COELOMATE GROUPS" (VOL. V), MCGRAW-HILL BOOK CO., INC.

FIG. 1.—TYPES OF SIPUNCULOIDS: (A) ASPIDOSIPHON SPECIOSUS; (B) GOLFINGIA MINUTA; (C) DENDROSTOMUM; (D) PHASCOLION STROMBI; (E) P. STROMBI IN A SNAIL SHELL; (F) GOLFINGIA FLAGRIFERA

Some species, like those found in the Baltic Sea, are tolerant of moderately low salinity levels. A Sumatran species, *Phascolosoma lurco,* is terrestrial, living in mangrove soil, much like an earthworm. A former supposedly planktonic genus, *Pelagosphaera,* is now known to be a larval stage: its metamorphosis has been observed and recorded.

External Characters.—The body of a sipunculid is divided into two regions, a trunk consisting chiefly of the elongated belly of the worm and a retractile introvert that bears the mouth at its anterior end. The mouth is partly or completely surrounded by tentacles or tentacle-bearing folds. The introvert was called a proboscis by earlier authors but the term, although descriptive, is misleading. Retractor muscles, typically four but sometimes only two or one, originate from the body wall of the trunk region at the base of the tentacles. Contraction of these muscles serves to withdraw the introvert into the body like the finger of a glove. Evagination depends upon hydraulic pressure developed in the coelomic fluid by contraction of the body musculature. In *Xenosiphon* eversion is effectively aided by the action of a pair of protractor muscles.

The number and arrangement of the tentacles varies widely. The mouth of *Sipunculus* is encircled by a lobed tentacular fold, frequently indented dorsally so that it assumes the shape of a horseshoe. In many genera the tentacles arise individually and in some instances they do not form a complete circle, as in species of *Aspidosiphon.* Sometimes, as in many species of *Golfingia,* the tentacles are arranged in tiers or concentric rings. Four to eight branching featherlike tentacles characterize species of the genus *Dendrostomum.* In very small species, such as those of the genus *Onchnesoma,* the tentacles may be greatly reduced or completely absent.

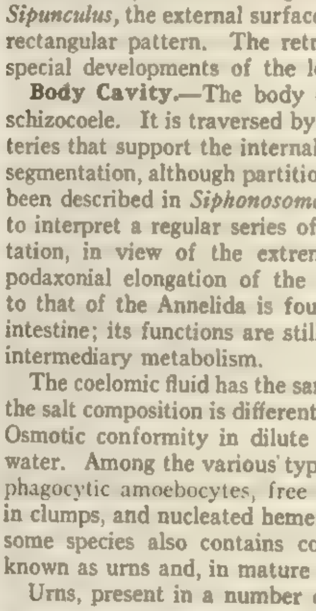
The introvert may be smooth, but, more frequently, it is armed with cuticular spines and hooks whose form and arrangement assume importance for the identification of the species. The anus, usually at the base of the introvert, is shifted far forward, near to the mouth in *Onchnesoma.*

The shape of the trunk varies greatly; it may be short and saclike, or long and vermiform. In *Phascolion strombi,* a common North Atlantic species that inhabits self-cemented tubes of sand grains or empty shells of mollusks, the shape of the body depends on that of the container in which the animal lives. Individuals that inhabit gastropod shells acquire a spiral twist that is almost completely lacking in sipunculids that dwell in straight tubes. The apex of the trunk is usually rounded, but in some deep-sea species of *Golfingia* it is drawn out to form a slender tail. In *Aspidosiphon* there is an apical caplike calcified structure resembling an acorn.

Body Wall.—A simple epithelium secretes a protective cuticle, thicker on the trunk than on the introvert. When horny or calcareous places are present, they are special secretions of the epidermis. Glandular and sensory papillae, frequently accompanied by hooks and holdfast organs, are found on the trunk and on the introvert, but some species are almost completely smooth. Beneath the epidermis is a connective tissue layer, more fully developed in some of the larger species. Coelomic diverticula may penetrate into the body wall; in *Sipunculus* they form longitudinal connecting canals between the muscle bands. The muscular coat consists of circular, oblique, and longitudinal layers; the oblique layer is sometimes lacking. The circular coat is usually divided into bundles; when the longitudinal coat is also so divided, as in *Sipunculus,* the external surface of the body reflects the underlying rectangular pattern. The retractor muscles of the introvert are special developments of the longitudinal coat.

Body Cavity.—The body cavity, or coelom, is technically a schizocoele. It is traversed by muscular strands and small mesenteries that support the internal organs but shows no trace of true segmentation, although partitions resembling segmental septa have been described in *Siphonosoma cumanense.* It would be difficult to interpret a regular series of septa of this sort as true segmentation, in view of the extreme displacement of the anus and podaxonial elongation of the body. Chloragogue tissue similar to that of the Annelida is found on the coelomic surface of the intestine; its functions are still problematical but are involved in intermediary metabolism.

The coelomic fluid has the same osmotic pressure as seawater but the salt composition is different; only traces of protein are present. Osmotic conformity in dilute media results from the uptake of water. Among the various types of corpuscles may be mentioned phagocytic amoebocytes, free chloragogue cells often aggregated in clumps, and nucleated hemerythrocytes. The coelomic fluid of some species also contains complex ciliated vaselike structures known as urns and, in mature individuals, developing sex cells.



FROM T. JEFFERY PARKER AND WILLIAM A. HASWELL, "A TEXT-BOOK OF ZOOLOGY", MACMILLAN & CO., LTD.

FIG. 2.—DISSECTION OF SIPUNCULUS NUDUS

Urns, present in a number of genera, are crowned by a single ciliated cell and are attached to the coelomic epithelium, the location depending on the species. In some forms the urns become detached and float in the coelomic fluid. The free urns of *Sipunculus* are budded off from the surface of the tentacular blind sac (*see* below), those of *Phascolosoma* are derived from the gut wall. Each urn is a closed vesicle covered by flat peritoneal cells with adherent chlorogogue tissue and containing fluid with connective tissue strands in its interior. A single ciliated cell sits like a cap on the vesicle, its circlet of cilia forming the brim, as it were, of the hat. The urns are believed, by their movement, to aid in the circulation of the coelomic fluid; the ciliary beat is such that foreign particles are collected in the central crown of the cap and eventually eliminated.

Tentacular System.—Canals lined by an endothelium similar

to that lining the coelom form a tentacular system, sometimes miscalled the vascular system. A circular lophophoral vessel, forming a ring surrounding the mouth, is closely associated above with the brain; canals extend forward from it into the tentacles and one or two backwardly projecting blind tubular sacs lie against the wall of the esophagus. These pouches have been miscalled hearts or, more often, Polian vesicles, on account of their resemblance to structures of this name in the Echinodermata (*q.v.*). Circulation of fluid, attributed to ciliary action, has repeatedly been observed within the tentacular canals: a peripheral flow in the lateral channels and a return flow in the median channel of each tentacle. In *Dendrostomum signifer* the blind sac is a pulsating organ that pumps fluid in and out of the lophophoral canals by frequent reversals of its beat. The circulation of the fluid, with its contained erythrocytes, can thus subserve a respiratory function. On the other hand, in forms like *Sipunculus*, in which the tentacles are short, the primary function may be hydrostatic. When the introvert is everted, the pressure set up in the body cavity compresses the blind sacs and fluid is thus forced into the tentacles, causing them to become extended. L. H. Hyman therefore proposed the name "compensation sacs" for the so-called Polian vesicles.

Digestive System.—The mouth lacks jaws. A ciliated esophagus, lying in the introvert, is followed by the long, coiled midgut. A true stomach is absent, although, in some species, a specialized glandular region at the junction of esophagus and midgut is usually called a stomach. The midgut passes backward to the apex of the body and then forward again to the anus. The descending and ascending limbs of the loop are usually twisted together in a spiral. This spiral is frequently supported by a spindle muscle that originates near the anus, traverses the central axis of the spindle, and is inserted apically. The midgut opens into a short hindgut that often receives a blind sac; terminally, there may be rectal glands of unknown function. A ciliated groove, externally visible as a reddish line, runs along the whole length of the alimentary canal in *Sipunculus* or is restricted, as in *Golfingia*, to the ascending limb of the intestine.

Sipunculids feed on organic matter contained in mud or sand taken up by the tentacles and swept into the mouth by ciliary action. Since the midgut is generally laden with such material, it has been suggested that these worms play an important role in the comminution of the sea bottom, similar to that played by earthworms on land. It has been suggested that the descending limb of the intestine is the main seat of digestion.

Vascular System.—A true vascular system is apparently lacking, although it may be represented, at least in *Sipunculus*, by a closed system of lacunae in the walls of the intestine.

Respiration.—Respiratory exchanges take place through the skin, and in forms like *Sipunculus*, which have subepidermal coelomic canals, this function probably is greatly facilitated. In forms like *Dendrostomum* the gill-like plumose tentacles may be the chief organ of respiration.

Hemerythrin, the respiratory pigment of the erythrocytes, is deep pink when oxidized and almost colourless when reduced. The dissociation curve of hemerythrin is such that the pigment gives up oxygen to the tissues only when there is an oxygen deficiency. Littoral species may be confined to their burrows during exposure at low tides and, as in tube-dwelling annelids and Echiurida (*q.v.*), the respiratory pigment may subserve a low tension transport function.

Excretory Organs.—A pair of excretory tubules (metanephridia) of relatively simple structure opens on the ventral surface of the body in front of the level of the anus. In *Phascolion* and *Onchnesoma* the development of one member of the pair, usually the left, is suppressed. The tube may be straight or U-shaped, and the limbs of the U may be more or less completely fused. Excretion takes place, at least in part, through the mediation of special cells lining the tube and also, apparently, by the migration of cell bodies called excretophores from the coelomic fluid into the lumen. The ciliated interior opening, or nephrostome, may also play a part in the collection of waste products, but it is believed to serve primarily for the collection of eggs or spermatozoa, which develop in the coelom. Ammonia is the chief nitrogenous excretory product.

Nervous System.—The central nervous system is of the annelidan type. Paired cerebral ganglia form a bilobed brain that is situated above the pharynx and closely associated with the special sense organs. The brain may lie superficially beneath the skin, as in *Golfingia verrilli*, or it may be more or less deeply sunken at the bottom of a cephalic tube that communicates with the exterior by a small opening above the tentacles. The development of the cephalic tube is probably an adaptation to burrowing; it reaches an extreme condition in species like *Sipunculus nudus*, which may dig themselves down to nearly a metre in loose sand. In addition to the neurons of the special sense organs, the brain receives sensory nerves from the tentacles and sends motor nerves to the retractor muscles of the introvert. Circumpharyngeal connectives unite the brain with the ventral nerve cord, which is ganglionated along its whole course and gives off mixed nerves to the body wall. There is no evidence of segmentation. The speed of conduction in the ventral cord approaches that of the somatic motor nerves of cold-blooded vertebrates, 10–20 cm. per sec.

Sense Organs.—The surface of the body responds to both tactile and chemical stimuli; sense organs vaguely resembling taste buds are located in the skin papillae and may contain both sensory neurons and gland cells.

There are three main special sense organs: (1) The cerebral or (miscalled) frontal organ, a lobe of the brain that projects forward and is united with an epithelial cushion at the base of the cerebral tube or depression. The function of the cerebral organ is unknown. (2) The nuchal organ, a ciliated paired or quadripartite raised area, resembling the organ of this name in polychaete annelids, is present in some genera, above the cerebral organ. Although well innervated, the nuchal organ serves an unknown function. (3) Paired eyes are present in sipunculid larvae and persist in the adults of various species. In the larva they are simple pigment spots associated with the future brain; in the adult they may disappear or, if present, may become deeply sunken at the bottom of ocular tubes that lie embedded in the brain and are closed to the exterior by cuticular plugs. Within the ocular tubes, a secretion forms a lenslike body. Neurons from sensory cells at the base of the eye communicate by a direct reflex pathway with the motor nerves of the retractor muscles. Retraction in response to a light stimulus is very rapid in species that have a well-developed lens.

Reproductive Organs.—The sexes are separate but externally similar. Rare instances of hermaphroditism have been described. The gonads are attached at the origin of the retractor muscles to the body wall and are of various forms. In *Sipunculus* there are paired branching glands on little stalks; in *Dendrostomum* there is a single transverse ridge. The sex cells are shed into the coelom at an early stage and grow to maturity, nourished by the coelomic fluid. The eggs are usually spherical, but in *Phascolosoma* they are flattened. The nephridia function as ducts to provide passage of the sex cells.

Development.—Segmentation is of the annelidan-molluscan type, with spiral cleavage. The blastopore forms the mouth; the anus develops later and is situated somewhat dorsally from the very beginning. As in Annelida (*q.v.*) the mesodermal bands are formed by teloblasts derived from blastomere 4d, and the coelom forms later as a schizocoele (*see* EMBRYOLOGY AND DEVELOPMENT, ANIMAL).

Fossil Remains.—It is doubtful whether any of the remarkable Middle Cambrian forms, some of which were referred by C. D. Walcott to the Gephyrea, can be regarded as sipunculids, since the intestine, when preserved, appears to be straight. An equally dubious gephyrean worm, *Epitrachys*, was described by E. Ehlers (1868) from the lithographic shales of Germany (Mesozoic). Tertiary, and even Paleozoic Devonian corals are said to have basal deformities resembling those produced by the living genus *Aspidosiphon*.

Relationships.—There can be no doubt that sipunculids have pronounced annelidan affinities, despite the lack of segmentation. Hyman regards them as protostomatous coelomates placed along the main evolutionary line of the Protostomia that leads to An-

nelida, Mollusca, and Arthropoda. Molluscan relationships have also been claimed on embryological grounds.

Classification.—W. K. Fisher (1952) recognizes 13 genera comprising, according to Hyman, some 250 valid species. Fisher's invaluable revision necessitated changes in names that had become well established in the literature. Species formerly placed in the genus *Phascolosoma* are correctly assigned to the genus *Golfingia*, whereas the name *Phascolosoma* is properly applied (for reasons of priority) to species of the obsolete genus *Physcosoma*. J. H. Gerould suggested that species of the genus *Golfingia* ("*Phascolosoma*") provide an ancestral group whose members became specialized in different directions. Three genera (*Aspidosiphon*, *Lithacrosiphon*, and *Cloeosiphon*) possess a preanal shield or cone; *Aspidosiphon* has also an apical shield. In most of the remaining genera the tentacles surround the mouth, but in *Phascolosoma* ("*Physcosoma*") they form a horseshoe enclosing the nuchal organ, dorsal to the mouth. Several genera are characterized by the division of the longitudinal musculature into separate bands and in four of these (*Sipunculus*, *Xenosiphon*, *Siphonomecus*, and *Siphonosoma*) there are integumentary coelomic canals; in *Siphonides* these bands are absent. The body musculature forms a continuous sheet in the remaining genera: these include *Golfingia* (except the subgenus *Phascolopsis*); *Dendrostomum*, with branched tentacles; and two genera (*Phascolion* and *Onchnesoma*), which, among other peculiarities, have only a single nephridium.

BIBLIOGRAPHY.—L. H. Hyman, *The Invertebrates*, vol. v (1959), is indispensable and supersedes the earlier, but still interesting, reviews by F. Baltzer in Kukenthal and Krumbach's *Handbuch der Zoologie*, vol. 2, pt. 2 (1931) and W. Harms in *Handw. Naturwiss.*, 2nd ed., vol. 9 (1934). A revision of the genera, with keys for their identification, is given by W. K. Fisher in "The Sipunculid Worms of California and Baja California," *Proc. U.S. Nat. Mus.*, vol. 102, pp. 371–450.

(G. E. P.)

SIQUEIROS, DAVID ALFARO (1896–), youngest of the "big three" of 20th-century Mexican mural painters (the others being José Clemente Orozco and Diego Rivera; *qq.v.*), was born in Chihuahua on Dec. 29, 1896. Although he signed his painting "Siqueiros," his true patronymic was Alfaro. A sublieutenant in the Constitutionalist Army at the age of 15, Siqueiros had already entered upon a lifelong dual career as artist and militant propagandist. In Paris from 1919 to 1921, he later visited the U.S.S.R., the U.S. and various Latin-American countries. In the 1920s he helped to found the Mexican magazine *Machete*. During the Spanish Civil War of 1936–39 he served in the Republican Army.

Alone or with organized groups of colleagues, Siqueiros produced thousands of square metres of vivid wall paintings, indoors and outdoors, commonly using synthetic lacquer colours sprayed from paint guns to produce his powerful, brightly coloured effects. In his studio paintings as well as on public walls, he portrayed social, political, and industrial change and ferment with a dazzling mixture of photographic realism, illusion, and fantasy. His most persuasive and permanent murals are to be seen at the headquarters of the Electrical Workers' Syndicate, the Palace of Fine Arts, the Polytechnic Institute, the National University, the nurses' home at Social Security Hospital No. 1, all in Mexico City, and at Chillán in Chile. Scores of his easel paintings—signed "*El Coronelazo*" after 1943—are in private collections and museums, notably the Museum of Modern Art in New York City.

BIBLIOGRAPHY.—MacKinley Helm, *Modern Mexican Painters* (1941); David Alfaro Siqueiros, *Siqueiros* (1951); Bernard S. Myers, *Mexican Painting in Our Time* (1956). (MACK. H.)

SIRACH: *see* ECCLESIASTICUS.

SIRAJ-UD-DAULA (d. 1757), ruler of Bengal under the nominal suzerainty of the Mughal emperor of India. The date of his birth is generally placed between 1729 and 1736. His name was Mirza Mahmud, and he succeeded his grandfather Ali Vardi Khan as nawab of Bengal on April 9, 1756. He was inexperienced, capricious, and vacillating. On hearing that the English were fortifying Calcutta against the French, he attacked and took the city after a slight resistance on June 20, 1756. The governor and many senior merchants had deserted to ships; the survivors of the garrison were secured for the night in the Fort William lockup known as the "Black Hole," where many were suffocated. There is no reason to suppose that the nawab was personally responsible for this incident, which was rather the result of ignorance and apathy than of deliberate cruelty. (*See* CALCUTTA.)

Calcutta was recaptured by Robert Clive and Adm. Charles Watson on Jan. 2, 1757. In February Siraj-ud-daula made peace with the East India Company, restoring its privileges. Cowed by an Afghan incursion into northern India, he allowed Clive to capture the French settlement of Chandernagore without interference. Thus strengthened, Clive joined a conspiracy of discontented nobles headed by Mir Jafar, Siraj's principal general. Siraj was overthrown at the Battle of Plassey on June 23. He fled up-country to Rajmahal but was captured and executed on July 2, 1757, by Mir Jafar's son Miran.

See *Cambridge History of India*, vol. v, ch. vii; Brijen K. Gupta, *Sirajuddaullah and the East India Company 1756–57* (1962).

(T. G. P. S.)

SIRENIA, an order of aquatic placental mammals, comprising the manatees (*Trichechus*), the dugongs (*Dugong*), the extinct (since the 18th century) Steller's sea cow (*Hydrodamalis*), as well as their fossil relatives of the Tertiary Period. The torpedo-shaped body ends behind in a horizontal tail fluke, as in the dolphins; but in contrast with the latter the broad muzzle is truncate and the transversely expanded lips are very mobile. The name Sirenia was given in allusion to the supposed resemblance of these animals to mermaids. A dugong as seen at a distance from the deck of a ship and especially if floating half upright, with its baby under its flipper, might well be mistaken for a mermaid; and many legends gathered around them in the early days of exploration of the Indian Ocean. For the evolutionary history and relationships of the Sirenia, *see* UNGULATE. *See* also DUGONG; MANATEE; SEA COW.

SIRENS, in Greek mythology, creatures half-bird half-woman who lured sailors to destruction by the sweetness of their song. Homer describes how Odysseus, advised by Circe, escaped the danger yet heard the songs by stopping the ears of his crew with wax so they were deaf to the Sirens and having himself tied to the mast so he could not steer the ship out of course. Apollonius Rhodius tells how when the Argonauts sailed that way Orpheus sang so divinely that none of them listened to the Sirens. Later legend relates that after one or other of these failures the Sirens committed suicide. According to Homer there were two Sirens, living on an island in the western sea between Aeaea (Circe's home) and the rocks of Scylla. Later the number was usually increased to three, and they were located on the west coast of Italy, near Naples. They are variously said to be the daughters of the sea-god Phorcys or of the river-god Achelous. Plato in the myth of Er (*Republic* 10) places a Siren in each of the eight musical spheres. In art they appear first as birds with the heads of women, later as women, sometimes winged, with bird-legs.

The Sirens seem to have evolved from two elements: a primitive tale of the perils of early exploration combined with an Oriental image of a bird-woman. Anthropologists explain the latter as a soul-bird; *i.e.*, a winged ghost which steals the living to share its fate. In this respect the Sirens have affinities with the Harpies (*q.v.*). The Sirens continued to exist in folklore throughout the medieval period and later, sometimes being thought of as fish-women rather than bird-women.

See Pauly-Wissowa, *Real-Encyclopädie der classischen Altertumswissenschaft*, 2nd series, vol. 3, col. 288–308 (1927). (H. W. PA.)

SIRICIUS, SAINT (*c.* 334–399), pope from 384 to 399, was a Roman by birth. In spite of the antipope Ursinus, Siricius was unanimously elected pope in December 384. He often followed the lead of St. Ambrose (*q.v.*), but was conscious that he bore "the burdens of all who are burdened, or rather, the Blessed Apostle Peter carries them in our person," as he wrote to Himerius bishop of Tarragona, in answer to disciplinary questions. These responses are the earliest surviving texts of papal decretals. St. Jerome's departure from Rome, while not forced by Siricius, was not unwelcome to him. Jerome's later activities in anti-Origenism displeased the pope. Ambrose and Siricius ended the Meletian controversy at Antioch by recognizing Flavian as patriarch. Al-

though he opposed Priscillianism and other errors, Siricius disapproved of Catholic extremists. Siricius was pope when the emperor rebuilt the basilica of St. Paul; a still surviving column commemorates Siricius' dedication. His feast day is Nov. 26. (W. M. K.)

SIRIONO INDIANS, a group of seminomadic, Guaraní-speaking aborigines who originally lived in isolated pockets of tropical forest land in eastern Bolivia. Of a population of about 2,000 in the 1960s less than a third still roamed the forests near the Río Blanco, the Río Grande, and the Río Piray. The remainder lived in missions or under conditions of forced labour on farms and cattle ranches in the political departments of the Beni and Santa Cruz.

Few tribes of the world are more technologically handicapped than the Siriono in their native state. Their only weapon is a cumbersome bow and arrow; their only tool, a digging stick of palm. They live almost exclusively by hunting and by gathering wild fruits and nuts, although occasionally maize (corn), cassava, and papaya are planted in natural clearings in the forest. They wear no clothing and live on the margin of subsistence the year round. They are not warlike, being preoccupied with food getting. The aged and infirm are sometimes abandoned, as is the house when a person dies.

The Siriono are organized into endogamous bands of from 60 to 100 persons. An entire band lives in a single hut, a crude, roughly rectangular palm shelter providing little resistance to wind and rain. Within the hut the matrilineal family groups cluster around a series of hammocks and hearths.

In part because of the difficult nature of supplying sufficient food for survival, ceremonial, aesthetic, and religious life is not highly elaborated among the Siriono. A form of ring dance is practised but no other forms of art are developed. Religion is highly animistic, centred on a fear of evil spirits and monsters.

See also Guaraní.

Bibliography.—A. R. Holmberg, *Nomads of the Long Bow: the Siriono of Eastern Bolivia*, Smithsonian Institution, Institute of Social Anthropology, publication no. 10 (1950); A. E. Schermair, *Vocabulario sirionó-castellano* (1958); R. Needham, "Descent, Category, and Alliance in Sirionó Society," *Southwestern Journal of Anthropology*, vol. 20 (1964). (A. R. Ho.)

SIRIUS (the Dog star) is the brightest star of the constellation Canis Major (*q.v.*); it is also visually the brightest star in the heavens. Intrinsically, it is a normal blue star 21 times as luminous as the sun; it is somewhat larger than the sun and has a considerably higher surface temperature. Its apparent brilliance in the skies is mainly because of its relative nearness to Earth. The distance of Sirius from the solar system is 8.6 light-years, which is only twice the distance of the nearest known star beyond the sun. (*See* also Star.)

Sirius was known as Sothis to the Egyptians, who early in their history were aware that this star made its first appearance of the season in the twilight before sunrise at about the time when the annual floods were beginning in the Nile delta. The Egyptians long believed that Sothis caused the Nile floods and accordingly was the "creator of all green growing things," as a Pyramid text declares. They discovered that the heliacal risings of the star occurred at intervals of 365¼ days rather than the 365 days of their calendar year, a correction incorporated by Julius Caesar (in 45 B.C.) by use of the plan of leap years in his reform of the Roman calendar. Among the Romans the hottest part of the year was associated with the heliacal rising of the Dog star, a connection that survives in the modern expression "dog days."

The fact that Sirius is a binary star was reported in 1844 by F. W. Bessel at Königsberg. He had observed that the bright star was pursuing a wavy course among its neighbours in the sky instead of going ahead uniformly in a straight line, as a single star would do. Bessel concluded that Sirius had a traveling companion, with which it revolved in a period of about 50 years. This discovery constituted the first chapter in what came to be known as the "astronomy of the invisible"; *i.e.*, the detection of unseen celestial bodies by their effects on visible bodies. The discovery of the planet Neptune two years later by its gravitational effects on Uranus was another famous example. Whereas Neptune was sighted soon after its place in the sky was predicted, the companion of Sirius waited nearly 20 years for direct telescopic verification. The companion was first seen in 1862 by Alvan Clark, a U.S. telescope maker who was using the bright star for testing the lens of the 18½-in. refractor now at the Dearborn observatory.

Sirius and its companion revolve together, once around in 49.9 years, in orbits of considerable eccentricity and with average separation of the stars equal to the distance of Uranus from the sun. Despite the glare of the bright star, the 8th-magnitude companion is readily seen with a large telescope except when the two appear closest together, as they did in 1944. The two stars are respectively 2.3 and 1.0 times as massive as the sun. Although the companion has one-third the mass of Sirius and about the same surface temperature, its light is only $\frac{1}{10,000}$ as bright as that of the brilliant star. The conclusion seemed inescapable that the companion has the mass of a star, the size of a planet and therefore a density many thousand times the density of water. Although so high a density of matter seemed surprising at that time when little was known about atomic structure, a test of the conclusion was available from the theory of general relativity. Albert Einstein showed that the lines in the spectrum of a star should be displaced to the red by an amount that is directly proportional to the cube root of the star's mean density. Small enough in the case of the sun to be masked by other factors, this displacement would be conspicuous in the spectrum of an extremely dense star.

W. S. Adams reported (1925) on his observations of the spectrum of the companion of Sirius. He had found the lines displaced in the predicted direction, but by an amount somewhat different than might be expected. It may be that the diffused light of the bright star was confused in the spectrum with that of the companion. Later, D. M. Popper studied the spectrum of another excessively dense star, the 9th-magnitude white dwarf companion of the star 40 Eridani; he reported a red shift of the lines in satisfactory agreement with the shift that would be predicted from the calculated density in this case. *See* also references under "Sirius" in the Index. (R. H. Br.)

SIRMUR (Sirmoor), a district in the southern part of Himachal Pradesh, India, situated in the lower ranges of the Himalayas between Punjab on the west and Uttar Pradesh on the east. Area 1,095 sq.mi. (2,836 sq.km.). Pop. (1961) 197,551. Sirmur is hilly except in the southeast where the Giri River, which traverses the district from northwest to southeast, has its confluence with the Jumna. Chaur Peak, on the northern border, is about 12,000 ft. (3,660 m.) above sea level. In British India Sirmur was a princely state forming part of the Punjab Hill States; it was merged with Himachal Pradesh in 1948. It was also called Nahan after its chief town (pop. [1961] 12,439), which lies about 60 mi. (95 km.) S of Simla. (S. Gl.)

SIROCCO: *see* Wind.

SIROHI, a town in the southwest of Rajasthan, India, and the administrative headquarters of the district of the same name, lies 95 mi. (150 km.) S of Jodhpur on the main Delhi-Bombay railway via Ahmedabad. Pop. (1961) 14,451. The town is noted for the manufacture of knives and sword blades.

Sirohi District has an area of 1,979 sq.mi. (5,126 sq.km.) and a population (1961) of 352,303. Formerly a princely state, it was merged with Bombay in 1949 but ethnic considerations led to its being transferred to Rajasthan in the following year. The district is much broken up by hills and rocky ranges and the Aravalli Range crosses it from northeast to southwest. About 12 mi. (19 km.) SW of Sirohi town is Mt. Abu (*q.v.*). Much of the district is covered with jungle, which is abundant in wildlife. The climate is on the whole dry and healthy, and there is a general freedom from epidemic diseases. On Mt. Abu the average annual rainfall is about 64 in. (1,626 mm.); at Erinpura, less than 50 mi. (80 km.) to the north, it is only 13 (330 mm.) in. (S. M. T. R.)

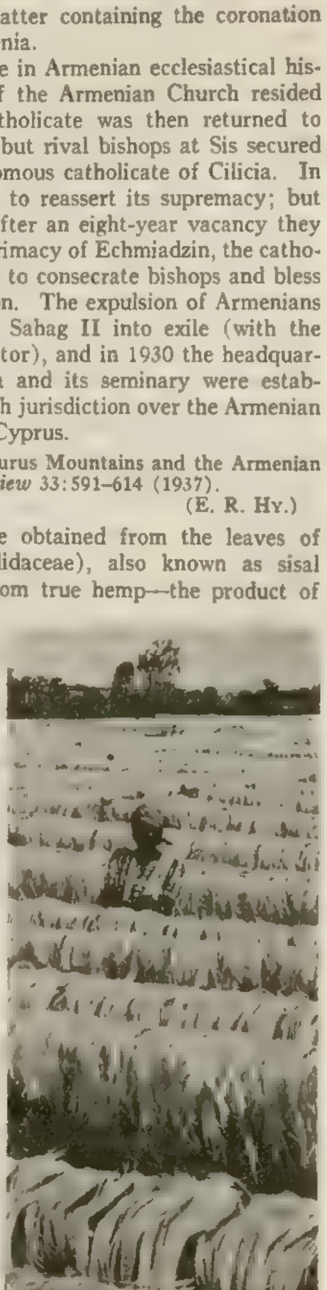
SIS (now Kozan), the former capital city of Armenian Cilicia, and now the chief town of a *kaza* in the Adana (Seyhan) *il* (province) of Turkey. Pop. (1960) 15,129. It lies on the Kirksu, a tributary of the Ceyhan River (ancient Pyramus), at the southern

end of a group of passes leading from the Anti-Taurus valleys to the Cilician plain and to Adana, which lies 40 mi. (64 km.) SW by road. First mentioned in Byzantine history as a frontier fortress unsuccessfully besieged by the Arabs in 704, Sis changed hands between Byzantines and Muslims until in the 12th century Leo II of Armenian Cilicia transferred his capital to that place from Anazarbus and erected many new buildings (*see* ARMENIA: *History*). It suffered frequent attacks from the Mamelukes of Egypt who finally occupied it in 1375. In 1488 it was taken by the Turks, and never recovered its prosperity; it is now in effect only a large village with a few thousand inhabitants. The castle and cathedral erected by Leo II survive, the latter containing the coronation chair of the kings of Lesser Armenia.

Sis has had a considerable place in Armenian ecclesiastical history. The supreme catholicos of the Armenian Church resided there from 1294–1439. The catholicate was then returned to Echmiadzin in Greater Armenia, but rival bishops at Sis secured recognition as heads of an autonomous catholicate of Cilicia. In 1885 the clergy at Sis attempted to reassert its supremacy; but when the see was filled in 1902 after an eight-year vacancy they were again obliged to accept the primacy of Echmiadzin, the catholicos of Cilicia retaining the right to consecrate bishops and bless the holy chrism for his jurisdiction. The expulsion of Armenians in 1915–20 drove the catholicos Sahag II into exile (with the relics of St. Gregory the Illuminator), and in 1930 the headquarters of the catholicate of Cilicia and its seminary were established at Antilyas, near Beirut, with jurisdiction over the Armenian Churches in Lebanon, Syria, and Cyprus.

See E. H. King, "Through the Taurus Mountains and the Armenian Cilician Kingdom," *The Asiatic Review* 33:591–614 (1937).
(E. R. HY.)

SISAL FIBRE, a hard fibre obtained from the leaves of *Agave sisalana* (family Amaryllidaceae), also known as sisal hemp. It is entirely different from true hemp—the product of *Cannabis sativa,* which is soft fibre. A closely related plant, *Agave fourcroydes,* produces the henequen fibre of Yucatán and Cuba, known in the trade as Yucatán sisal and Cuban sisal. Both sisal and henequen are indigenous to Mexico, but the East African colonies, Indonesia, and Haiti are the main sources of supply of true sisal fibre.



AUTHENTICATED NEWS

SISAL FIBRE (AGAVE SISALANA) BEING DRIED IN THE SUN, KENYA, AFRICA

The common century plant (*Agave americana*) suggests the habit of sisal and henequen. Both of these plants consist of a large rosette of rigid, straight, fleshy leaves arising from a short trunk. The sisal leaves are dark green in colour and have a terminal spine but ordinarily no marginal prickles. The life period of this plant is from 5 to 10 years. The henequen plant has grayish-green leaves which bear both a terminal spine and marginal prickles. The life period of this plant is from 15 to 25 years. Both sisal and henequen, when mature, produce a flower stalk or pole which develops from the centre of the leaf cluster and grows to a height of 15–25 ft. The light-yellow flowers are borne in dense clusters at the ends of the lateral branches.

The flowers of the sisal plant are followed by small bulbils, and those of the henequen plant by both bulbils and seed pods. The plants die after flowering. Both sisal and henequen may be propagated either from the bulbils or from suckers that grow from the rootstocks.

In harvesting, the lower mature leaves are cut off at the base, the spines and prickles are removed, and the leaves are carried in bundles to the cleaning machines. The cleaning of these fibres is a scraping process in which the leaves as they pass through the machine are carried over curved metal plates and are scraped by rapidly revolving scraping wheels that remove the pulp and waste material. The fibre is then dried either in the sun or by artificial drying methods, is sometimes brushed, and is then graded and baled for market.

These fibres are white to yellowish-white in colour, strong, flexible and in value for cordage purposes are second only to abacá. Henequen is used principally for the manufacture of binder twine, while sisal is used for binder twine and many other types of cordage. These fibres have a limited use in the manufacture of coarse fabrics. *See* also FIBRE. (H. T. ES.)

SISINNIUS, a Syrian, was pope from *c.* Jan. 15 to Feb. 4, 708.

SISKIN (ABERDEVINE), a canarylike bird, *Carduelis spinus,* which has long been known in England as a cage bird, is one of the finches (*q.v.*). It often feeds upon the catkins of alder or birch, frequently hanging upside down like a titmouse. Above, the male is olive green marked with black and yellow, and beneath, yellowish-white marked with black. The hen is grayer and streaked with white on its underparts. The male's song is a rapid musical twitter pierced with high squeaky call notes. The siskin breeds locally throughout Europe, and its range stretches across Asia to Japan.

Coniferous woods and thickets are favoured as nest sites. The nest of the siskin is like that of the goldfinch, but not so neatly built; the eggs, except in their smaller size, resemble those of the greenfinch.

A larger and more brightly coloured species, *C. spinoides,* inhabits the Himalayas. In the United States the name siskin or pine siskin is sometimes used for the pine finch (*Spinus pinus*).

SISLEY, ALFRED (1839–1899), one of the creators of French Impressionism, was born in Paris of English parentage on Oct. 30, 1839. Originally intended for commerce, Sisley began painting as an amateur, and it was in Gleyre's studio in 1862 that he began to associate with Monet, Renoir, and Bazille. The Franco-Prussian War of 1870, which brought financial ruin to the Sisley family, caused Sisley himself to flee temporarily to London, and it was at this period of crisis that he decided to make painting his full-time career. The remainder of his life was a constant struggle against poverty. He died at Moret-sur-Loing on Jan. 29, 1899, without ever having acquired the French citizenship for which he had applied some four years earlier. Shortly after his death his talent began to be widely recognized.

Sisley was essentially a landscape painter, whose early style was much influenced by Corot, and his restricted and delicate palette continued to reflect something of Corot's silvery tonalities; his snowscapes are particularly effective. Much of his best and most spontaneous work was done in the period 1872–80 in the neighbourhood of Paris, at Marly, Louveciennes, Bougival, Sèvres, Saint-Cloud, and Meudon, at a time when he was in close touch with Monet.

See J. Rewald, *The History of Impressionism,* 2nd ed. (1961); F. Daulte, *Sisley: Catalogue raisonné de l'oeuvre peint* (1959).

SISMONDI, JEAN CHARLES LÉONARD DE (1773–1842), Genevan economist, historian, and author of works representing a humanitarian protest in a materialistic age, was born in Geneva on May 9, 1773. His name was Simonde, but he italianized it, claiming Italian ancestry. Because of the revolutionary disturbances of 1793–94, his family fled to England (1793) and then to Italy (1794), settling on a little farm near Pescia, Tuscany. Sismondi's interest in agricultural problems, especially *métayage,* resulted in his descriptive *Tableau de l'agriculture toscane* (1801), published after returning to Geneva in 1800.

In 1803 he met Mme de Staël (*q.v.*); he became one of her regular guests at Coppet, and accompanied her to Italy (1805). He became the intermediary between Coppet and the *salon* of the countess of Albany (*q.v.*) at Florence, Mme de Staël and the countess sharing his aversion to Napoleon. However, when Na-

poleon proposed his *Acte additionnel* (1815), projecting a new liberal constitution for France, Sismondi, hoping that some of the achievements of the Revolution could be saved, supported him publicly. As professor of philosophy (1809) and history (1820) at the Académie de Genève, he devoted his life to research and writing. He died at Chêne, near Geneva, on June 25, 1842.

As an economist, Sismondi began by following Adam Smith (*q.v.*) with *De la richesse commerciale* (1803). Later, especially in *Nouveaux Principes d'économie politique* (1819), he was concerned less with theoretical economics than with the well-being of the people, which, in his view, should be the principal object of every government. Interest in social affairs is typical also of his immense *Histoire des Français* (31 vol., 1821–44; Eng. trans. 1849–50). It is less lively and original, however, than his *Histoire des républiques italiennes du moyen âge* (in part, 1807; in 16 vol., 1809–1818; Eng. trans. 1831), which brought him fame and many friendships. Fired with love of liberty and mankind, this work, in which medieval Italy is seen as the origin of modern Europe, helped the Italians to regain their national consciousness: such patriots as Foscolo, Mazzini, Manzoni, and Cavour regarded Sismondi with grateful admiration. *De la littérature du midi de l'Europe* (1813; Eng. trans. 1823), based on lectures at Geneva, and influenced by Mme de Staël's *De la littérature* (1800), was a useful synthesis of the Romance literatures. Sismondi also wrote on contemporary issues, such as the Negro slave trade, Italian and Greek independence, and the Algerian expedition of 1840.

BIBLIOGRAPHY.—Sismondi's journals and correspondence have been published by C. Pellegrini, 4 vol. (1933–54). The most complete biography is J. R. de Salis, *Sismondi, 1773–1842*, 2 vol. (1932). Monographs include A. Aftalion, *L'Oeuvre économique de Simonde de Sismondi* (1899); E. Halévy, *Sismondi* (1933); A. Amonn, *Simonde de Sismondi als Nationalökonom*, 2 vol. (1945–49); L. de Rosa, *Gian Carlo Sismondi e la sua opera* (1947); H. O. Pappe, *Sismondis Weggenossen* (1963). (CA. PE.)

SISTERHOODS, in the Christian Church, communities of women living and working together for specific religious purposes. See MONASTICISM; WOMEN'S RELIGIOUS ORDERS; ORDERS AND CONGREGATIONS, RELIGIOUS; *Roman Catholic Religious Orders of Women.*

SISYPHUS, in Greek mythology, the cunning king of Corinth who was punished in Hades by being made repeatedly to roll up a hill a huge stone which always rolled down again as soon as he had brought it to the summit. This fate is related in the *Odyssey*. In the *Iliad* Sisyphus, living at Ephyre (later Corinth), is the son of Aeolus (eponymous ancestor of the Aeolians), the father of Glaucus, and the grandfather of Bellerophon. In post-Homeric times he is called the father of Odysseus, cunning obviously providing the link between them. Sisyphus was reputed to have founded the Isthmian Games in honour of his relative Melicertes (*q.v.*). Later legend tells how when Death came to fetch him, Sisyphus chained him up so that no one died until Ares came to free Death and Sisyphus had to submit. But he had told his wife, Merope, not to perform for him the usual sacrifices, so when he reached the underworld he received permission to return and punish her for the omission. Once back at home, he continued to live to a ripe old age before dying a second time. Sisyphus is, in fact, like Autolycus and Prometheus (*qq.v.*), that widely popular figure of folklore, the trickster, or master thief. How he came by a royal genealogy is unknown. That he is everlastingly punished in Hades is presumably the recompense for his cheating Death, but why he is set to roll a great stone incessantly is a puzzle to which no convincing answer has yet been given. It appears to belong with other Greek imaginings of the world of the dead as the scene of fruitless labours. But Albert Camus, in *Le Mythe de Sisyphe* (2nd ed. 1949; Eng. trans. 1955), applies a contrary interpretation to that of classical antiquity. To him Sisyphus' task is typical of human endeavour, and its apparent futility should occasion the toiler happiness. (H. W. PA.)

SITAPUR, a town in the Lucknow division of Uttar Pradesh, India, and the administrative headquarters of the district of the same name, lies on the Sarayan River, 50 mi. (80 km.) NNW of Lucknow. Pop. (1961) 53,884. It has a cantonment and is a junction of the Northern and North-Eastern railways. There is considerable grain trade and the plywood manufactured in the town has an extensive market. Sitapur also has a well-equipped eye hospital.

SITAPUR DISTRICT, situated on the left bank of the Gumti (Gomati) River, has an area of 2,236 sq.mi. (5,791 sq.km.) and a population (1961) of 1,608,057. It is a vast plain, well wooded and well cultivated except in those parts where the soil is barren and cut up by erosion gullies. Its principal rivers are the Gogra and the Sarda. The main crops are wheat, rice, and barley. Khairabad, 5 mi. (8 km.) SE of Sitapur, is known for its many temples, mosques, and shrines. Biswan, 20 mi. (32 km.) E of Sitapur, has a sugar factory. (B. SI.)

SITKA (formerly NEW ARCHANGEL), historically the most notable settlement of Alaska, U.S., on the west coast of Baranof Island, in Sitka Sound. The city is on an island-studded and mountain-locked harbour, with a background of forest and snow-capped mountain cones; an extinct volcano, Mt. Edgecumbe (3,201 ft. [976 m.]), on Kruzof Island, is a conspicuous landmark in the bay. Monthly mean temperatures range from 33° F (1° C; January) to 56° F (13° C; August). Normal rainfall is 86 in. (2,185 mm.) and annual snowfall averages 26 in. (660 mm.).

Old Sitka or Ft. Archangel Gabriel, 6 mi. from the present town, was founded in May 1799 by Aleksandr Baranov, first Russian governor of Alaska. The fort was destroyed by the Tlingit Indians in 1802; the present town was founded in September 1804 when Baranov moved his headquarters there from Kodiak. Commonly known by its Indian name, Sitka, probably the name of a local tribe or village, it was the headquarters of the Russian-American Company until 1867. It was the leading trading post of Alaska and had shipyards, sawmills, flour mills, foundries and other industrial establishments. The formal transfer of Alaska to the U.S. took place at Sitka on Oct. 18, 1867. From 1867 until 1906 Sitka was the capital of Alaska. The modern industries are logging and lumber manufacture, and fishing and fish processing. There are salmon canneries, facilities for cold storage and processing of halibut, salmon, and other fishery products, and a large fishing fleet.

The public schools are attended by both whites and Indians of the Tlingit tribe. The Presbyterian-supported Sheldon Jackson High School and Junior College are accredited schools. Other institutions include the Alaska Pioneers Home for aged residents of the state; Sitka National Monument, a 54-ac. natural park; Sitka National Cemetery; Sheldon Jackson Ethnological Museum; Magnetic and Seismological Observatory of the Coast and Geodetic Survey; St. Peter's by the Sea (Episcopal, 1899); Lutheran Church (1840, the first Protestant church in Alaska); Presbyterian and Roman Catholic churches. Historic St. Michael's Cathedral (Russian Orthodox, 1848) was destroyed by fire in 1966.

On 475-ac. Japonski Island, in the harbour, the federal Alaska native service established the Mt. Edgecumbe Vocational Boarding School, a tuberculosis sanatorium, and the Mt. Edgecumbe Orthopedic Hospital, maintained jointly with the state of Alaska.

The population at the 1960 census was 3,237. (J. E. CL.)

SITTANG, a river of Burma rising in the Yamethin district and flowing south for about 350 mi. (560 km.) in a broad valley between the forested Pegu Yoma on the west and the sharp edge of the Shan Plateau on the east before emptying into the Gulf of Martaban. Geomorphologists believe that the upper Irrawaddy, north of Mandalay, originally flowed due south along the Sittang Valley but was deflected by earth movements in Pleistocene times, beheading the old lower course which became the Sittang, a misfit stream of small volume occupying a large valley. Too shallow to be important for navigation, the river is used to float down quantities of timber, especially teak. Its valley is followed by the main railway and road from Rangoon to Mandalay, and the towns of Pyinmana and Toungoo lie near its banks. (L. D. S.)

SITTARD, a town of the southeastern province of Limburg, Neth., lies 13 mi. (21 km.) NE of Maastricht. Pop. (1960) 27,548. Fertilizers, chemicals, electronic tubes, and textiles are produced and there are coal mines in the area. The town is a railway junction and on the Maastricht–Roermond motorway. Excavations at two nearby villages have shown the remains of Danubian

culture (4200–3800 B.C.). Sittard received its charter in 1243. From 1400 to 1794 it was a domain of the dukes of Jülich. After the period of French rule (1794–1814) it became part of the kingdom of the Netherlands and, apart from a period of Belgian rule in 1830–39, has since remained Dutch.

SITTER, WILLEM DE (1872–1934), Dutch astronomer and cosmologist, whose papers on relativity published in London during World War I brought Einstein's General Theory to the notice of British scientists, with important consequences for cosmology. De Sitter was born at Sneek, Friesland, Neth., on May 6, 1872. He studied at the University of Groningen to become a mathematician, but, on the invitation of Sir David Gill, her majesty's astronomer at the Cape of Good Hope, he spent two years (1897–99) at the Cape observatory, as a result devoting himself to astronomy thereafter. In 1908 he was appointed professor of astronomy at Leiden, and in 1918 he became director of the Leiden observatory. There he trained student astronomers, many of whom obtained important posts abroad, especially in the U.S. As a practical astronomer he established a station at Johannesburg to observe the southern sky and sent expeditions to stations near the equator in Kenya to determine fundamental declinations by azimuth observations of stars at low altitudes.

From the time of his visit to the Cape, De Sitter had studied the motion of Jupiter's four great satellites. Theoretical discussions published in 1918, 1919 and 1925 used observations dating from 1668 and included results obtained by him from several observatories. He died at Leiden on Nov. 20, 1934. (J. JN.)

SITTINGBOURNE AND MILTON, an urban district in the Faversham parliamentary division of Kent, England, formed in 1930 by the union of Milton Regis with Sittingbourne. Pop. (1961) 23,623. Area 7.7 sq.mi. It is on a navigable creek of the Swale, 38 mi. ESE of London by road. Sittingbourne consists principally of one long street, the Roman Watling street. Papermaking is its chief industry, brick and cement making and fruit preserving are also important, and as an agricultural market town it is the centre of the Kentish cherry-growing area. Sittingbourne is mentioned in Saxon documents in 989 and frequently in contemporary records of the 13th and 14th centuries. The first charter was obtained in 1573; a second in 1599.

From Roman until comparatively recent times Milton was a noted oyster fishery centre. It was a royal manor whose revenues formed part of the dowry of successive queens. An earthwork known as Castle Rough, in the marshes below Milton, was probably the work of Hasten the Dane in 892, and Bayford castle, 1 mi. distant, occupies the site of one said to have been built in opposition by King Alfred.

Tong castle, about 2 mi. E of Sittingbourne, consists of a high mound surrounded by a moat and is said to have been erected by Hengist (*q.v.*). The story of the founding of the castle resembles that connected with the city of Carthage. Vortigern is said to have granted Hengist as much land as an oxhide could encompass, and the hide being cut into strips the site of Tong castle was accordingly marked out. The same tradition attaches to Tong castle in Shropshire. Tradition also asserts, according to the 12th-century chronicler, Geoffrey of Monmouth, that it was in Tong castle that Vortigern met Rowena, Hengist's daughter, and became so enamoured of her as to resign his kingdom to her father.

SITTING BULL (*c.* 1831–1890), most famous chief of the Teton Dakota (Prairie Sioux) and perhaps the most famous of all American Indians, was born on Grand River, in what is now South Dakota, the son of Jumping Bull, of the Hunkpapa division of the Tetons. He joined his first war party at the age of 14 and, soon gaining a reputation for fearlessness in battle, became leader of the Strong Heart warrior society about 1856. Within five years he had extended the Sioux hunting grounds westward into what had been the territory of Shoshone, Crow, Assiniboin, and other Indian tribes. His first skirmish with white soldiers occurred in 1863, after the so-called Minnesota Massacre by Santee Sioux (*see* MINNESOTA: *History*) had involved the entire Sioux nation in trouble with the U.S. Army. For the next five years he was in frequent hostile contact with the Army, which was invading the Sioux hunting grounds and bringing ruin to an Indian economy based on the buffalo. In 1866 he became principal chief of the northern hunting Sioux, with Crazy Horse (*q.v.*), leader of the Oglala division, as his vice-chief.

In 1868 the Sioux made peace with the whites in exchange for a guaranteed reservation north of the North Platte River and the right to hunt off the reservation. The treaty was soon violated on Gen. George A. Custer's discovering gold in the Black Hills in 1874 and precipitating a rush of miners into the reservation area. Sitting Bull's continued defense of his people's hunting grounds and his refusal to be confined to the reservation led, in 1876, to a campaign against the Sioux commanded by Gen. George Crook. Summoned by Sitting Bull, the Sioux, with Cheyennes and some Arapahoes, met in a great encampment on the Rosebud River and in June engaged and defeated Crook. The camp then moved to the Little Bighorn River for game, and there on June 25 took place the battle in which Custer was killed (*see* CUSTER, GEORGE ARMSTRONG).

But though the Sioux might win battle after battle with the Army, they could never win the war. They depended on buffalo for their livelihood, and the buffalo, under the steady encroachment of the whites, were rapidly becoming extinct. Hunger led more and more Sioux to surrender, and in May 1877 Sitting Bull led his remaining followers across the border into Canada. But the Canadian government could not acknowledge responsibility for feeding a people whose reservation was south of the border, and after four years, during which his following dwindled steadily, famine forced Sitting Bull to surrender. After 1883 he lived at the Standing Rock agency, where he vainly opposed the sale of tribal lands. Hunger, disease, and, beginning about 1889, rumours of a coming Indian Messiah who would sweep away the whites (*see* GHOST DANCE) caused unrest among the Sioux. As a precaution, Indian police and soldiers were sent to arrest the chief. Seized on Grand River, Dec. 15, 1890, Sitting Bull was killed while his warriors were trying to rescue him.

Sitting Bull appears to have had qualities of leadership that were not common among Plains Indians, the power to make plans and to work steadily toward their accomplishment. Among the Sioux he was admired as a loving father of his people, a singer of songs, always affable and pleasant in manner, devoutly religious and a prophet whose prayers were strong.

For a portrait, *see* INDIAN, NORTH AMERICAN.

See Stanley Vestal (pseud.), *Sitting Bull: Champion of the Sioux*, new ed., with bibliography (1957). (W. S. C.; X.)

SITWELL, the name of an English family, three members of which achieved fame as writers in the first half of the 20th century. They are EDITH (1887–1964), (FRANCIS) OSBERT (1892–), and SACHEVERELL (1897–), the children of the antiquarian and genealogist Sir George Reresby Sitwell, 4th Bart. (1860–1943), and of Ida Emily Augusta Denison (d. 1937) daughter of Baron (later 1st Earl of) Londesborough. Edith was born at Scarborough, on Sept. 7, 1887; Osbert in London, on Dec. 6, 1892; and Sacheverell at Scarborough, on Nov. 15, 1897. They spent much of their childhood at Renishaw Hall, near Sheffield, the family seat in Derbyshire. A brilliant, eccentric, and domineering father, and a beautiful but erratic mother, who spoiled her sons but was harsh to her daughter, drove them often for understanding and affection to servants and governesses. Edith was educated privately; Osbert and Sacheverell at Eton and Osbert served in the Grenadier Guards (1912–19). All three showed a precocious interest in modern movements in art and literature, and, as soon as they could escape from parental influence, became writers. Dame Edith (she was created a Dame of the British Empire in 1954) is best known for her poetry; Sir Osbert for his prose memoirs; and Sacheverell for books on art, architecture, and travel.

Edith Sitwell's ambition to write was stimulated by reading Swinburne when she was 17 (she edited a selection of his poetry in 1960). From 1916 to 1921 she edited *Wheels*, an anthology of experimental poetry published as a counterblast to Sir Edward Marsh's *Georgian Poetry*. Her early poetry, well exemplified in her long *The Sleeping Beauty* (1924) and in the witty

series written for recitation with William (later Sir William) Walton's suite *Façade* (first performed 1922; enlarged 1926; definitive version 1942), combines a fanciful, sometimes brittle, quality with underlying melancholy, and with technical innovations and use of associative imagery derived from the French Symbolists. *Gold Coast Customs* (1929), expressing horror at the corruption of society, began a new period, and poems written during World War II—*Street Songs* (1942), *Green Song* (1944), *The Song of the Cold* (1945)—appealed to a wide public by their more sweeping rhythms and their expression of anguish at human cruelty, combined with faith in God's goodness and the holiness of Nature. This religious emphasis was continued in *Gardeners and Astronomers* (1953) and *The Outcasts* (1962). In 1955 she was received into the Roman Catholic Church. Critical and historical works include *Alexander Pope* (1930), *Aspects of Modern Poetry* (1934), *Fanfare for Elizabeth* (1946), and *The Queens and the Hive* (1962). *A Poet's Notebook* (1943) reveals wide reading and interest in technique. Her autobiography, *Taken Care Of*, was published in 1965, after her death (in London on Dec. 9, 1964). *Collected Poems* was published in 1930 and 1957; and *Façade and Other Poems, 1920–35*, in 1950.

Osbert Sitwell has written satirical and serious poetry (*Collected Satires and Poems*, 1931; *Mrs. Kimber*, 1937; *Selected Poems: Old and New*, 1943; *Wrack at Tidesend*, 1952; etc.); novels, of which the best is *Before the Bombardment* (1926), a satirical portrayal of the last phase of Victorian society in Scarborough just before World War I; short stories; and criticism, but his reputation rests on his autobiographical *Left Hand, Right Hand* (1944), *The Scarlet Tree* (1946), *Great Morning* (1947), *Laughter in the Next Room* (1948), and *Noble Essences* (1950). In these, which create with conscious nostalgia the portrait of a vanished age, the author's father is presented as a comic figure of heroic proportions, and the prose style blends delicacy with robust appreciation of the grotesque.

Sacheverell Sitwell's poetry—*The Rio Grande* (performed 1929, with music by Constant Lambert), *Dance of the Quick and the Dead* (1936), *Sacred and Profane Love* (1940), *Selected Poems* (1948), etc.—mainly in traditional metres, reveals in its mannered style the effects of his interest in the arts and music. More original are his imaginative and interpretative books on art, architecture, music, and travel, of which the first, *Southern Baroque Art* (1924), was the forerunner of much academic research. His poetic prose is seen at its best in the "autobiographical fantasia," *All Summer in a Day* (1926), and the gloomily meditative *Splendours and Miseries* (1943).

BIBLIOGRAPHY.—R. Fifoot, *A Bibliography of Edith, Osbert and Sacheverell Sitwell* (1963); R. L. Mégroz, *The Three Sitwells* (1927); M. W. Wykes-Joyce, *Triad of Genius* (1953–); R. Fulford, *Osbert Sitwell* (1951), J. Lehmann, *Edith Sitwell* (1952), both in the "Writers and Their Work Series." (G. S. F.)

SIVAJI (SIVAJI BHONSLA) (1627–1680), Maratha hero, was the son of Shahji Bhonsla, a *jagirdar* (*i.e.*, holder of the right to collect the land tax from an assigned district) of the Muslim kingdom of Bijapur in the Deccan, India. Born at Shivner, Poona district, in April or May 1627, Sivaji succeeded his father in 1655 and followed his policy of increasing his territories and achieving independence. In 1659 he defeated the Bijapuri army and himself killed its general, Afzal Khan. Sivaji then turned against the Mughal territories, plundering Surat in 1664, but was brought to bay by the Mughal general Raja Jai Singh in 1665. An attempted settlement with the emperor Aurangzeb at Agra in 1666 broke down on account of mutual suspicions; Sivaji, placed under house arrest, escaped in a fruit basket and resumed his independent career in the Deccan. He plundered Surat again in 1670 and in 1674 crowned himself at Raigarh (Rajgarh, south of Poona) as an independent sovereign with mixed Mughal and Hindu symbolism. During 1676–78 he laid the foundations of Maratha power in the south. He died at Raigarh on April 5, 1680, the master of a compact and independent kingdom in western India.

Sivaji's organization of his kingdom was based on the functional integration of the three sections of the population, Brahmans, Marathas, and Prabhus, in defense of land and religion. Working on the Maratha love of independence he evoked a fierce regional nationalism which enabled the Marathas to withstand the onslaught of the emperor Aurangzeb (*q.v.*) during 1681–1707. *See* also MARATHA.

BIBLIOGRAPHY.—J. N. Sarkar, *Shivaji and His Times*, 2nd ed. (1920); S. N. Sen, *Administrative System of the Marathas* (1923); J. Grant Duff, *History of the Mahrattas*, 2 vol., *rev. ed.* (1921). (T. G. P. S.)

SIVAS (ancient MEGALOPOLIS-SEBASTEIA), a town and capital of the *il* (province) of the same name in central Turkey, lies at an altitude of 4,420 ft. (1,347 m.) in the broad valley of the Kizil Irmak, 220 mi. (350 km.) E of Ankara. Pop. (1960) 93,368. The climate is healthful but severe in winter with average January temperature −5° C (23° F). Sivas is an active trade centre; the chief manufactures are cement, coarse cotton cloth, and woolen socks. It stands at the junction of the Kayseri-Samsun-Malatya and Kars railways.

The *medresses* (colleges) in Sivas, built in the 13th century by the Seljuk sultans of Rum, are among the finest remains of Muslim art in Asia Minor. In one of them is the tomb of its founder, Izz ud-Din Kai Kaus I (1210–19). Near the town is the Armenian monastery of the Holy Cross, in which are kept the throne of Senekherim and other relics. Mekhitar, the founder of the Mekhitarists (*q.v.*) and of the well-known monastery at Venice, was born (1676) at Sivas.

Sivas, as Sebasteia, became the capital of Armenia Minor under Diocletian at the end of the 3rd century, and in the 7th century that of the Sebasteia Theme. Justinian I (483–565) rebuilt the walls and, under the Byzantine emperors, it was second only to Caesarea (Anazarbus) in size and wealth. In 1021 Senekherim, king of the Armenian province of Vaspurakan (Van), ceded his dominions to Basil II, and became the Byzantine viceroy of Sebasteia and the surrounding country. This position was held by his successors until the town fell into the hands of the Turkmen after the defeat of Romanus IV by the Seljuks (1071).

After having been ruled for nearly a century by the Danishmend amirs, Sivas was taken (1172) by the Seljuk sultan of Rum, and in 1224 was rebuilt by Sultan Ala ud-Din Kaikobad I. In 1400, when seized and plundered by Timur, the city is said to have had 100,000 inhabitants, and to have been famous for its woolen stuffs. Mohammed II the Conqueror (1451–81) restored the citadel, and the town was for many years an important Ottoman provincial capital. Early in the 19th century, like all other Ottoman towns, it was terrorized by janizaries (*q.v.*). In 1919 a national congress was held at Sivas, which eventually led to the revolution. (*See* also TURKEY: *History*.)

SIVAS IL is one of the largest provinces in Turkey. Area 10,999 sq.mi. (28,487 sq.km.). Pop. (1960) 669,922. To the west are the Ak Mountains, and the province is drained by the Kizil Irmak, Kelkit, Calti, and Tohma rivers. It is an important cereal-producing region and also contains mineral deposits, the chief being iron ore at Divrigi, about 80 mi. SE of Sivas town. The iron ore is mainly consumed by the iron and steel mills at Karabuk in Zonguldak province on the Black Sea. (N. TU.; S. ER.; E. TU.)

SIWAH, an oasis in As Sahra' al Gharbiyah *muhafaza* (Western Desert Governorate), Egypt (United Arab Republic), near the Libyan frontier and about 350 mi. WSW of Cairo. Also known as the oasis of Amon (Jupiter Ammon or Ammonium), its ancient Egyptian name was *Sekhet-am* (Palm land). Its chief town, also called Siwah, is at latitude 29° 12′ N, longitude 25° 30′ E. The oasis is about 6 mi. long by 4 to 5 wide, with a population (1958 est.) of 5,000. Ten miles northeast is the small oasis of Zaytun (Zeitun), and westward a chain of little oases and small salty pools extends for about 50 mi. Two rock outcrops provide the sites of the old walled ruins of Siwah and Aghurmi, veritable fortresses.

The inhabitants, who live in mud-brick houses at the foot of their old strongholds, are a mixture of Berber, Bedouin, and Sudanese; they are distinct from the Egyptians, being connected more closely with North Africa than the Nile Valley. They speak their own Berber dialect, although most of the men understand Arabic, and about 10% of the people are literate. The only properly developed industry is basketmaking. The oasis is extremely fertile, with about 200 springs, of which 80 are important.

Evidence suggests that in ancient times the area of cultivated land was greater than in modern times, but attempts to enlarge the agricultural area have failed, mainly through lack of labour. Many kinds of fruit and cereals are grown, although some food is imported. The export of dates and olive oil provides the chief source of income; there are about 40,000 olive trees and 200,000 date palms. Desert caravans for trading are rare since the introduction of motor transport and the building of the road connecting with the railway at Matruh.

The oasis owes its distinction to the oracle temple of Amon (*q.v.*), already famous in the time of Herodotus and consulted by Alexander the Great. The fragmentary remains of the temple, with inscriptions dating from the 4th century B.C., lie in the ruins of Aghurmi. The oracle fell into disrepute during the Roman occupation of Egypt. Half a mile from Aghurmi is the ruined temple of Umm Beda (Um Ebeida), of the same period, with reliefs depicting the prince of the oasis making offerings to Amon. There are many remains of the Roman period. The area of the Libyan frontier is dominated by the Senusi sect who formerly prevented various explorers from penetrating westward beyond Siwah. The first European to reach Siwah since Roman times was William George Browne, who visited the oasis in 1792. In 1910 a telegraph line was built across the desert from Alexandria to the oasis.

(A. B. M.)

SIWALIK HILLS (SIWALIK RANGE), the name commonly given to the whole belt of the Himalayan foothills extending for more than 1,000 mi. from the Tista River, Sikkim, through Nepal across northwestern India into the northern part of West Pakistan. Though sometimes only about 10 mi. wide and normally reaching elevations of only 2,000–3,000 ft., the range of hills rises abruptly from the Indo-Gangetic plains and separates them from the main range of the Himalayas. The Siwaliks are built up of soft sandstones, conglomerates, and shales of Late Tertiary Age and so belong to a late stage in the Himalayan folding. In places intermediate valleys, *duns*, occur between the Siwaliks and the outer Himalayas; *e.g.* Dehra Dun (Dehra Valley) between the Siwaliks and the Mussoorie Mountains. The Siwalik Range proper, to which the name was formerly restricted, is the 200 mi. of foothills from Hardwar on the Ganges (Ganga) northwestward to the Beas (Bias) River. Everywhere in this section the poor scrub forests have long since been removed and the hills are subject to spectacular erosion. Periodic floods sweep down masses of sand and silt in the everchanging *chos* (stream beds which are dry except after rains).

(L. D. S.)

SIXTUS, the name of five popes.

ST. SIXTUS I (Xystus), pope for nine or ten years in the second and third decades of the 2nd century, was the successor of Alexander I and thus the sixth bishop of Rome after St. Peter. He is commemorated on April 6.

ST. SIXTUS II (Xystus), pope from 257 to 258, succeeded Stephen I in August 257. He restored the relations with the African and Asian churches that had been severed by his predecessor on the question of heretical baptism. He suffered martyrdom under the emperor Valerian and is commemorated on Aug. 6.

ST. SIXTUS III (Xystus), pope from 432 to 440, succeeded Celestine I on July 31, 432. He had previously been suspected of favouring the Pelagians, but on becoming pope he disappointed their expectations and repelled their attempts to enter again into communion with Rome. The dispute between Cyril of Alexandria and John of Antioch, who had been at variance since the Council of Ephesus, was settled in 433, but Sixtus himself had come into conflict with Proclus, patriarch of Constantinople. He died on Aug. 18, 440, and is commemorated on March 28.

SIXTUS IV (Francesco della Rovere) (1414–1484), pope from 1471 to 1484, was born of a poor family near Savona on July 21, 1414. He entered the Franciscan order at an early age and studied philosophy and theology at the universities of Padua and Bologna. He was chosen general of his order in 1464. Three years later, in 1467, he was, to his own surprise, made cardinal-priest of S. Pietro in Vincoli by Paul II. He was elected pope in succession to Paul on Aug. 9, 1471. The fleet that he sent, under Cardinal Oliviero Carafa, against the Turks in 1472 participated in the landing at Smyrna, but a new expedition in the following year was a failure. His relations with France were strained because of the Pragmatic Sanction of 1438 and more so after Louis XI's ordinance of 1475 requiring that no papal decree be published in France without the royal placet. Sixtus likewise continued in 1474 and in 1476 his predecessor's negotiations with the grand duke Ivan III of Muscovy for the reunion of the Russian Church with the Roman see and for support against the Turks, but without result. He was visited in 1474 by King Christian of Denmark and Norway, and in the following year he established the University of Copenhagen.

Sixtus soon abandoned universal interests in order to concentrate attention on Italian politics and showed himself a confirmed nepotist. He was cognizant of the conspiracy of the Pazzi (1478), plotted by his nephew, Girolamo Riario, against Lorenzo de' Medici (*see* PAPACY: *The Renaissance Papacy*). He entered into a fruitless and inglorious war with Florence which kept Italy for two years (1478–80) in confusion. He next incited the Venetians to attack Ferrara (1482) and then, after having been delivered by their general Roberto Malatesta from a Neapolitan invasion, turned upon them and eventually laid them under an interdict (1483) for refusing to desist from the hostilities which he had himself instigated. He relied on the cooperation of Ludovico Sforza, who speedily forsook him, whereupon peace was forced upon him by the princes and cities of Italy. Sixtus died Aug. 12, 1484.

In 1475 Sixtus instituted the office of the Immaculate Conception for Dec. 8; in 1478 he formally annulled the decrees of the Council of Constance; by a brief of January 1482 he condemned abuses in the Spanish Inquisition; and in April 1482 he canonized St. Bonaventura. He granted many privileges to the mendicant orders, particularly to his own Franciscans. The most praiseworthy side of his pontificate was his munificence as a founder or restorer of useful institutions and as a patron of letters and art. He established and richly endowed the first foundling hospital, built and repaired numerous churches, constructed the Sistine Chapel (1473–81) and the Sistine Bridge, improved church music and instituted the Sistine Choir, commissioned paintings on the largest scale, pensioned men of learning (notably Bartolomeo Platina) and, above all, immortalized himself, from 1471, as the second founder of the Vatican Library. These great works, however, were not accomplished without grievous taxation. Annates were increased and simony flourished.

(X.)

SIXTUS V (Felice Peretti) (1520–1590), pope from 1585 to 1590, was born at Grottammare in the march of Ancona, near Montalto, on Dec. 13, 1520. Brought up in poverty, he entered a Franciscan monastery in 1533. He early showed rare ability as a preacher and dialectician. Sent to Rome in 1552, he attracted the attention of the future popes Paul IV (Carafa) and Pius V (Ghislieri). He was appointed inquisitor general in the Venetian republic in 1557 but became involved in disputes and had to leave (1560). Sent to Spain with the legation under Cardinal Buoncompagni (later Gregory XIII) in 1565, he conceived a strong dislike for Buoncompagni and hurried back to Rome on the accession of Pius V, who made him vicar apostolic of the Franciscans and bishop of Sant' Agata dei Goti in 1566 and cardinal on May 17 1570. Henceforth, as cardinal, Peretti was known by the name of Montalto. During the pontificate of Gregory XIII he lived in retirement, occupied with the care of his villa and with his studies (the first volume of his edition of the works of St. Ambrose appeared in 1580). Yet he did not neglect to follow the course of affairs, though carefully avoiding every occasion of offense. This discreetness contributed not a little to his election to the papacy on April 24, 1585.

The terrible condition in which Gregory XIII had left the states of the church called for prompt and stern measures. Against the prevailing lawlessness Sixtus proceeded with an almost ferocious severity, bringing thousands of brigands to justice by proceeding with vigour against their supporters among the powerful nobility. Within a short time the country was again quiet and safe, but Sixtus had made many enemies. He turned next to financial reform. By the sale of offices, by the creation of new *monti* (loans) and by levying new taxes, he accumulated a vast reserve, which he kept against certain specified emergencies, such as a crusade or

the defense of the Holy See. He moreover pursued an active economic policy by regulating food prices and by encouraging agriculture and the production of wool and silk. Immense sums were spent on public works and buildings, yet Sixtus ended his reign as one of the richest princes in Europe, having begun it with an almost empty treasury.

It is because of his achievements in reforming the central administration of the church that Sixtus must be reckoned as one of the greatest popes. By the bull *Postquam Verus* (Dec. 3, 1586) the college of cardinals was given its definitive form, its number being limited to 70. The secretariat of state was reorganized, and in January 1588 the entire administrative system of the curia was overhauled. The number of congregations was increased to 15, of which 6 dealt with secular administration and the other 9 with spiritual matters. It was through the machinery thus established that the decrees of the Council of Trent were effectively enforced. In particular, Sixtus insisted that the decrees regulating residence and investiture be strictly observed. So he must be counted as one of the founders of the Counter-Reformation.

His intervention in doctrinal affairs was not so happy. He had good grounds for supposing that the commission appointed to prepare a revised edition of the Vulgate, which had benefited enormously by his injection of a spirit of urgency into its deliberations, had not sufficiently considered the practical problems created by their work of scholarship. His reaction was typical: he rejected the work of the commission and took the whole task into his own hands. After 18 months of intense application, he produced the *editio Sixtina* in May 1590. But his revision was far too hurried and far too conservative, and the opportunity afforded by its numerous printer's errors was taken to postpone promulgation. Sixtus' death then intervened, and his edition was eventually abandoned. By the same event, Robert Bellarmine's book on doctrinal controversies was saved from official condemnation and the Society of Jesus from a drastic revision of its constitution.

In his political relations with foreign princes, Sixtus was faced by a dilemma. Distrusting Philip II of Spain, he viewed with misgiving the extension of Philip's power but was obliged to abet him as Henry III of France showed himself ever less able to hold his own between the pro-Spanish party of the Guises and the Huguenots under Henry of Navarre. Moreover, Elizabeth of England, whose personality Sixtus greatly admired, was an obstinate heretic. Thus he excommunicated Henry of Navarre (1585), promised subsidies for a Spanish invasion of England, and denounced Henry III after the coup against the Guises. In the last months of his life, however, he saw the hope that Henry of Navarre would make good his claim to the crown of France (as Henry IV) and be converted to Catholicism. Sixtus could then begin to withstand the demands of Philip's immediate ambition, confined as this was to France since the defeat of the Armada. Negotiating still both with Philip's and with Henry's envoys, he died on Aug. 27, 1590.

Posterity ranks Sixtus one of the greatest popes. He was hasty, obstinate, severe, autocratic; but his mind was open to large ideas, and he threw himself into his undertakings with an energy and determination that often compelled success. Few popes can boast of greater enterprise or larger achievements. *See* also PAPACY.

(I. F. B.)

BIBLIOGRAPHY.—*General:* A. Fliche and V. Martin (eds.), *Histoire de l'Église, depuis les origines jusqu'à nos jours* (1935 ff.); A. Vacant et al (eds.), *Dictionnaire de théologie catholique*, 15 vol. (1903–50); *Enciclopedia Cattolica*, vol. xi, cols. 777–787 (1953).

Sixtus IV: J. Burckhardt, *The Civilization of the Renaissance in Italy*, Eng. trans., new ed. (1951); L. Pastor, *History of the Popes from the Close of the Middle Ages*, Eng. trans., vol. iv, pp. 197–471 (1894); E. Frantz, *Sixtus IV und die Republik Florenz* (1880).

Sixtus V: J. A. von Hübner, *Sixtus V*, Eng. trans., 2 vol. (1872); U. Balzani, *Sisto Quinto* (1913).

SJAELLAND (ZEALAND), the largest island of Denmark proper, with an area of 2,709 sq.mi. (7,016 sq.km.) and a population (1960) of 1,771,557. It is separated on the west from the island of Fyn (Fünen) by the Great Belt (Store Baelt) and on the east from southern Sweden (Skåne) by the busy Sound (Øresund). Extending 82 mi. (132 km.) from north to south and 68 mi. (109 km.) from west to east, the island has a basal platform of Cretaceous rocks whose raised eastern rim is exposed in the magnificent chalk and limestone cliffs at Stevns Klint. But solid geology is generally obscured by thick morainic deposits forming a gently undulating, lake-strewn landscape that attains its maximum height at Gyldenløves Høj (413 ft. [126 m.]). Coastal subsidence has produced sinuous inlets, such as Isefjord, and broad bays like Køge Bugt. Although the fertile clay loams support some of Denmark's best farmland, the urban character of the island is also pronounced. Besides Copenhagen and its suburbs, urban settlements range from the old cathedral city of Roskilde (*q.v.*) to the historic port of Elsinore (*q.v.*; Helsingør) and the ferry station of Korsør.

The island has eight major administrative units: the city of Copenhagen (*q.v.*); the adjoining borough of Frederiksberg; the four counties (*amter*) of Frederiksborg, Holbaek, Sorø, and Praestø (*qq.v.*); and the county council districts of Copenhagen and Roskilde. *See* also DENMARK. (HA. T.)

SKAGERRAK, a long, rectangular arm of the North Sea, trending southwest to northeast and separating the low-lying northern coast of the Danish peninsula of Jutland from the more mountainous fjord coast of southeastern Norway. Approximately 130 mi. long and more than 70 mi. (110 km.) wide, it is shallowest in the south off the smooth dune-fringed Danish shore, which leads to the great curving spit of the Skaw (Skagens Odde). Between the Skaw and the opposite shore of Sweden on the east the outlet of the Skagerrak narrows to 40 mi. (65 km.) before continuing southward into the Kattegat (*q.v.*) Strait and on to the Baltic Sea. The principal ports of this busy shipping lane are Oslo and Kristiansand (*qq.v.*), Nor., and Strömstad, Swed. (HA. T.)

SKAGWAY, a town of Alaska, U.S., situated at the north end of Lynn canal, a deep and narrow arm of the sea thrust far up between picturesque mountain ranges, is at the head of navigation in the waters of southeastern Alaska. Skagway owes its importance to being the seaward terminus of the White Pass and Yukon railway. This road, built in the years 1898–1900, extends up the valley of the Skagway River to the summit of White Pass, 20 mi., crossing there the international boundary and continuing thence down on the Canadian side 91 mi. to Whitehorse, head of navigation on the Yukon. Pop. (1960) 659.

Historically, it is of interest as the landing place of large quantities of supplies and many thousands of people during the Klondike gold rush of 1897–98 to reach the newly discovered deposits in the upper (Canadian) Yukon.

SKANDERBEG (GEORGE KASTRIOTI) (1405–1468), the national hero of the Albanians. He was the youngest son of John Kastrioti, head of one of the ruling families, whose possessions, extending over parts of central and northern Albania, included the famous fortress of Krujë. As John Kastrioti was defeated during the first incursions of the Ottoman Turks into Albania, he had to surrender Krujë and to give his four sons as hostages to the sultan. Thus George, a convert to Islam, was educated at the military school at Edirne (Adrianople). Sultan Murad II, impressed by his intelligence and by his ability at martial games, attached him to his personal staff. He gave him the name Iskander (Alexander), with the rank of bey—a style subsequently modified by his countrymen into Skanderbeg.

In 1443 Skanderbeg was serving on a military expedition against the Christians led by János Hunyadi (*q.v.*). During battle near Niš, in Serbia, he heard that the Albanians were rising against the Turks. Leaving the field, he hastened to Krujë, overpowered the Turkish garrison there, hoisted the Albanian flag, proclaimed himself a Christian, and declared a holy war. After reconquering the family possessions, he appealed to the Albanian leaders to unite against the foreign invader. They responded with enthusiasm, a national convention met at Lezhë (Alessio, then occupied by the Venetians), an Albanian league was formed, and Skanderbeg was appointed commander in chief.

Skanderbeg's struggle was to last for the rest of his life. Year after year Murad sent armies to subdue the rebellion; and while he fought successfully against the Turks, Skanderbeg had also to defend himself against the intrigues of the Venetians. In 1450 Murad himself, at the head of a powerful army, besieged Krujë for five months, till Skanderbeg's harassing tactics from the sur-

rounding mountains forced him to renounce the enterprise. This victory made Skanderbeg a hero throughout the Western world, but the devastation left behind by the invaders was appalling, and the Albanian league was weakened by desertions to the enemy or to the Venetians.

Going abroad in search of support, Skanderbeg concluded an alliance with King Alfonso I of Naples (Alfonso V of Aragon). Appointed captain general of the crown of Aragon, he was granted an annual pension of 1,500 ducats, while Catalan troops were sent to garrison Krujë. In 1451 Skanderbeg married Andronika, daughter of George Araniti (d. 1461), lord of Shpat and Kaninë and a member of the Albanian league.

An Albanian attempt to liberate Berat from the Turks was unsuccessful, but Skanderbeg destroyed another Turkish Army despatched against him. Pope Calixtus III named him "captain general of the Holy See." After an interval of service in Italy, on behalf of Alfonso's son Ferdinand against the Angevin rivals of the Aragonese, Skanderbeg returned to Albania to resist the continued attacks of the Turks. Sultan Mohammed II, having conquered Constantinople (1453), was eager to put an end to Christian resistance in the Balkans; but the armies which he sent every year from 1455 to 1462 were constantly defeated by Skanderbeg.

By clever diplomacy Skanderbeg obliged Venice to sign an alliance with him (1463): the Venetians sent 1,300 men to Albania for a new offensive against the Turks and granted a subsidy of 2,000 ducats. Pope Pius II's call for a crusade against the Turks fell on deaf ears, however; and when the pope died suddenly the Albanians were left to fight practically alone. For three years Balaban Pasha, an Albanian renegade in command of the Turkish Army, crossed swords with Skanderbeg. Then, in 1466 Mohammed II laid siege to Krujë in person, at the head of a strong army. Unable to take the fortress, the sultan retired, and the remaining Turkish forces were eventually decimated, Balaban Pasha himself being killed.

Skanderbeg, dressed as a simple soldier, went to Rome to seek help. At a ceremony in St. Peter's he received a sword and a blessed cap from Pope Paul II; and a subsidy of 7,000 ducats followed. Sultan Mohammed's second siege of Krujë (1467) was another disaster for the Turks, but it was Skanderbeg's last triumph. He died at Lezhë on Jan. 17, 1468, and was buried in the Cathedral of St. Nicholas, which had been rebuilt sometime earlier under his orders. His widow and his son were taken to Venice. The citadel of Krujë did not fall to the Turks till 1478.

See Athanase Gegaj, *L'Albanie et l'invasion turque au XV^e siècle* (1937); Fan S. Noli, *George Castrioti Scanderbeg* (1947). (T. Z.)

SKANDERBORG, an *amt* (county) of Denmark in east central Jutland, extending inland from Horsens Fjord. Area 664 sq.mi. (1,719 sq.km.); pop. (1960) 137,865. The fertile clay loams of the east support good mixed farming. The highest land in Jutland occurs in the hummocky moraine belt between Skanderborg and Silkeborg, the principal hills being Yding Skovhøj (568 ft. [173 m.]), Ejer Bavnehøj (561 ft.), and Himmelbjaerget (482 ft.). Broad, flat-floored valleys, like that of the Gudenaa (Denmark's longest stream), dissect the wooded upland and contain beautiful lakes, *e.g.*, Mossø. The principal towns are Horsens (pop. [1960] 37,261), Silkeborg (24,465), and Skanderborg (5,482). (Ha. T.)

SKARABORG, a *län* (county) in central Sweden between Lakes Vänern and Vättern, extending about 80 mi. (129 km.) both north-south and east-west. Pop. (1960) 250,180; area 3,262 sq.mi. (8,449 sq.km.). The county mainly drains toward Vänern and so for long has looked toward Göteborg and not to Stockholm. The higher land results from faulting and local preservation of resistant rocks; the highest summits are Mosseberg (1,072 ft. [327 m.]) and Kinnekulle (1,007 ft. [307 m.]). Limestone outcrops are extensively quarried; near Skövde are important cement works. Glacial deposits usually occur as infertile sands and gravels, but remnants of deposits from the Yoldia Sea transgression have given fertile areas. Falköping is an ancient city, and much Swedish culture and law derives from this region. Agriculture dominates the district although match making and automobile assembling are done at Tidaholm. Mariestad, on the east shore of Vänern, is the county town and a rail junction, with a 17th-century church and castle. Skara, 25 mi. (40 km.) SW of Mariestad, has a 12th-century cathedral. It became a bishopric in the 11th century and had the first Swedish high school (1641). (A. C. O'D.)

SKARA BRAE, the most perfect Stone Age village in Europe, was embalmed under a sand dune on the shore of the Bay of Skaill on the Atlantic coast of Mainland Island, largest of the Orkneys, Scot. Exposed by a great storm in 1851, four buildings were excavated during the 1860s by William Watt, laird of Skaill. After another storm in 1926 further excavations were conducted by the ancient monuments branch of the ministry of works, that had assumed guardianship of the monument, supervised from 1928 to 1931 by V. G. Childe, of the University of Edinburgh. Before 1939 the age of the site and the cultural status of its builders remained uncertain, but in that year excavation by Walter G. Grant at a contemporary village, Rinyo, on Rousay, established that the dwellings of Skara Brae were built before the first people to use bronze in Great Britain, the Beaker Folk, reached Orkney, and so belonged to the New Stone Age of Orkney. In the 1960s archaeologists estimated that Skara Brae had been built *c.* 2000–1500 B.C.

Though the dwellings at Skara Brae are built of undressed slabs of stone from the beach, put together without any mortar, the drift sand that filled them immediately after their evacuation preserved the walls in places to a height of 8 ft. (but unfortunately not the roofs). Moreover, because of the absence of trees on the island, articles of furniture that would normally be carved in perishable wood had to be made of stone and have, therefore, survived. At the same time the dwelling explored in 1928 was found exactly as it had been left by its occupants who had been forced to abandon it in precipitate haste by a violent storm. The village consisted of several one-roomed dwellings, each constructed on the same plan. Each is a rectangle with rounded corners, the largest measuring 21 ft. by 20 ft. square. Entering through a narrow doorway only 4 ft. high that could be closed by a stone slab held in place by a sliding bar, the visitor finds in the centre of the room a rectangular fireplace still filled with peat ash. Behind it against the rear wall stands an erection of two tiers of stone shelves precisely like a modern dresser. On either side of the central hearth are enclosures of stone slabs exactly similar to the fixed beds of planks built in the 19th century in Norwegian peasant houses, that on the right being always the larger. Tall uprights of stone at the corners like bed posts served to support some sort of canopy. In the wall above the bed are recesses that must have served as keeping places for personal possessions. Sunk in the floor are several cubical boxes lined with stone slabs and luted with clay along the joints, evidently designed to hold some sort of liquid. One or more small cells in the thickness of the walls

J. ALLAN CASH

EXCAVATED DWELLING AT SKARA BRAE

open off the main room. Beneath its floor a slab-lined drain runs to debouch into a main sewer.

At the moment of its abrupt desertion the village had consisted of seven or eight such huts linked together by paved alleys. By this time six huts had been put artificially underground by banking around them midden consisting of sand and peat ash stiffened with dung and other refuse, and the alleys had become tunnels through this midden roofed with slabs. On the west (the east end has been destroyed by the sea) the main tunnel gave through a doorway, identical in design with those of individual dwellings, onto an open space paved with slabs, but unroofed. A hut, entered from this space and standing free of the general midden-heap cover, differed to some extent in arrangement and furniture from the rest; it contained a kiln where the dresser should have stood and had been used for the manufacture of chert implements. The whole residential complex was drained by a built sewer into which the drains from individual huts discharged.

The inhabitants of the village lived mainly on the flesh and presumably the milk of their herds of tame cattle and sheep, on limpets and other shell fish, supplemented exceptionally by red deer, wild boar, seal and whale. There is no positive evidence that they cultivated any cereals or that they spun and wove wool or other material. Presumably they dressed in skins. Bone tools suitable for preparing and piercing such skins were very numerous, while large pins made of narwhal ivory, whalebone or the penis-bones of seals could be used for fastening them. Peat was regularly burned as fuel, while fire may have been made by striking nodular hematite against flint or chert. For their equipment the villagers relied exclusively on local materials. Axheads, or rather adz blades, were made of stone sharpened by grinding and were sometimes attached to their handles with the aid of perforated sleeves of stag's antler. Knives were made by splitting beach pebbles along their bedding planes. The bones of cattle and sheep served for the manufacture of awls, needles, fabricators and various tools of uncertain use probably employed in dressing hides. Vessels were made of pottery. Though the clay was exceedingly coarse and included large grits, and the firing was not sufficient to convert it into pottery right through, most vessels were quite richly decorated with elaborate patterns in relief formed by applying to the surface strips and pellets of clay. Vessels were also made out of the vertebrae of whales and blocks of sandstone.

No weapons at all were found at Skara Brae unless some unique objects carved out of hard volcanic stone were so used. These include balls covered with pyramidal spikes and other spikey forms, all carved with great accuracy. As ornaments the villagers wore pendants and beads made out of the marrow bones of sheep, the roots of cows' teeth (the nerve canal serving as a string hole), the teeth of killer whales and boars' tusks.

Nearly 1,000 finished beads were found together in one cell, and unfinished beads in all stages of manufacture lay about the hut floors. Red, yellow and white pigments contained in little paint pots of stone or cetacean bone were presumably used in painting the person. Games were played with dice of walrus ivory and with knucklebones.

The skeletons of two old women, buried together doubled up, were found in a cist under the wall of the hut discovered in 1928 and have been interpreted as a foundation sacrifice. Watt had previously discovered a skeleton in a niche beside the main alley, but the village cemetery has not been identified. The two old women must have stood about 5 ft. 2 in. in height, not much, if anything, below the average of the female stature prevailing in the mid-20th century. Both had suffered from osteoarthritis. The leg bones show peculiarities due to habitual squatting. The skulls are just dolichocranial (index 74.5).

A number of stones in the walls of the huts and alleys have been roughly scratched with lozenges and similar rectilinear patterns.

Beneath the walls of the huts just described the foundations of older huts were discovered in 1930. In plan and furniture these agreed precisely with the former and the relics on their floors were of the same kind as those described above. The pottery from the lower levels was adorned with incised as well as relief designs. Among these was the true spiral represented on one sherd —the only example of this pattern in pottery known in prehistoric Britain.

See V. G. Childe, *Skara Brae: Pictish Village in Orkney* (1931), *Scotland Before the Scots* (1946); G. Bibby, *Testimony of the Spade* (1956). (V. G. C.)

SKARZYSKO KAMIENNA, an industrial town of Poland in Kielce *wojewodztwo* (province). Pop. (1960) 35,000. It is situated in the Little Polish Highlands, in the wide valley of the Kamienna River, a tributary of the Vistula, and is surrounded by extensive woods. The development of the town has been associated with the iron industry established there in the 19th century, using local ores from the Old Polish Basin. The old plants of Kamienna were closed in 1878 as unprofitable, but new and continued development started in 1881, after the building of blast furnaces. The town is now an important centre of the metal and mineral industries and is a railway junction on the Warsaw–Cracow line. (T. K. W.)

SKAT, a game of cards, much played in Germany and by people of German origin throughout the world. It is an elaboration of *Schafkopf* (sheepshead), a Wendish game believed to date at least to the middle 1700s. In 1811 a resident of Altenburg who had newly learned *Schafkopf* carried it into the local tarok club, where it was modified by the addition of features from tarok and kalabrias. The very name of the game is a tarok term, derived from the Italian *scartare*, "to discard," or *scatola*, a "place of safe-keeping." In the earliest writings, the game is spelled "scat." Much of the early development of skat is credited to F. F. Hempel, an advocate in Altenburg. The first book on it was published by his cousin, J. F. L. Hempel, in 1848. After a long period of experimentation, the game was codified by a congress of more than 1,000 skat players, convened at Altenburg on Aug. 7, 1886. Subsequent modifications chiefly concerned the use of the skat (widow) and elimination of the game *frage*.

Skat is very popular in some sections of the United States and large tournaments are played for prizes. Laws of the game are promulgated by the North American Skat league and do not in all respects conform to the procedure of play in Germany.

Skat is played by three players with a pack of 32 cards. Each receives ten cards and two are dealt face down for a skat (widow). The players bid by numbers, in multiples of two, and the high bidder then names his game. In the game tournee the trump suit is fixed by a card turned from the skat—it may be a grand (in which only knaves are trumps) if a knave is turned. Solo is played without help of the skat, and the player may name a suit or grand. Gucki is a grand using the skat. Null is a bid to take no tricks, at no trumps. Ramsch is played only if all three players pass, the object being to win fewest points in tricks, with knaves trumps.

The bidding and scoring values of the games are determined by their base values times the sum of all due multipliers. The multipliers arise from various sources—points won in play, prediction or announcement as to this outcome, and matadors, which are top trumps in unbroken sequence.

To make his bid, the player must take a majority of the 120 card points: each ace 11, ten 10, king 4, queen 3, knave 2. Except in ramsch, the other two players combine in temporary partnership to try to defeat him.

A modern variant, *Räuber Skat*, popular in club play but not in tournaments, eliminates the game tournee. Instead, the high bidder always has the option of taking the skat or of playing without it. In the latter case, handplay, he earns an extra multiplier.

See Walter J. Zarse, *Sheepshead and Skat* (1935); "The Official Rules of the North American Skat League," in *The Official Rules of Card Games* (1946). (G. Mh.)

SKATE SAILING: *see* Ice Skating.

SKATING: *see* Ice Skating; Roller Skating.

SKEET SHOOTING: *see* Trapshooting and Skeet Shooting.

SKEGNESS, an urban district and seaside resort in Lincolnshire, Eng., lies 42 mi. E of Lincoln by road. Pop. (1961) 12,847. Since 1873 the fine sand dunes, beach, and bracing air have drawn tourists in large numbers from the Midlands and farther afield.

There are a pier and picturesque gardens and waterways along the front. The adjoining Seacroft and North Shore golf courses attract national competitions, and the Skegness and Gibraltar Point Nature Reserve, with its bird observatory and field research station, is now a centre for natural history research with a full-time warden and residential accommodation for students. The wold village of Somersby, birthplace of Alfred, Lord Tennyson, is about 12 mi. NW. (F. T. BA.)

SKELETON, INVERTEBRATE. The function of the skeleton is to provide a more or less rigid framework upon which the muscles can react. In many invertebrate animals no such framework exists; the necessary rigidity is provided by the hydrostatic pressure of the body fluids acting upon a soft and pliable skin. This is seen in caterpillars and, to a lesser extent, in soft worms and in mollusks, where the necessary pressure is maintained by the steady contraction of an interlacing sheet of superficial muscles, the so-called turgor muscles. The muscles that move the body are attached to the skin, which is made rigid by this internal pressure. The firmness of some invertebrate animals is maintained by a mechanism similar to that in many plants; the individual cells that form the body wall are more or less distended by water to give some degree of rigidity. But in many invertebrates, rigid skeletal materials are deposited either on the outer surface of the body or within the cells. In some animals these rigid structures do not afford attachment for the muscles but merely provide a protective armour.

It is usual to classify skeletal structures as external (exoskeleton) or internal (endoskeleton). But the one type merges with the other, even within the same animal; and in many examples of what appears to be an exoskeleton, the hard material is laid down inside the cells that form the outer covering.

The invertebrate skeleton is exceedingly varied in composition and extent. These variations cut across the ordinary classification of animals. In the Protozoa, for example, there are all grades between soft forms (*e.g., Amoeba*) and forms with a cuticle, which may be proteinaceous (in *Monocystis*) or composed of cellulose (in the plantlike flagellates). Other protozoa have definite shells, composed of protein with various foreign bodies incorporated in it, siliceous plates, lime (in most Foraminifera) or cellulose (in the resting stages of slime molds). The Radiolaria have an internal lattice of silica laid down inside the cell, a kind of endoskeleton.

Other groups show a similar variation. Some sponges deposit needlelike spicules of calcium carbonate in the jelly (mesogloea) beneath the outer epithelium; others produce spicules of silica held together by a horny substance termed spongin; and yet others have a skeleton of spongin alone and no spicules. Similarly, among coelenterates, there are Hydrozoa that have a horny covering protecting the polyps, and others with an external calcareous skeleton. Anthozoa show the same diversity. In *Alcyonium* (dead men's fingers) the skeleton is mainly internal, consisting of spicules of calcium carbonate in the mesogloea. In the red coral of commerce (*Corallium*) the spicule-bearing cells migrate inward and build up a central skeleton. In the common reef-building corals the calcareous skeleton is secreted by that part of the ectoderm that forms the basal disc. This secretory process is continuous and the polyp raises itself progressively upon a growing stem of lime.

In the echinoderms the calcareous deposits are laid down in the mesoderm. They may be so scattered that they merely serve to impart a leathery consistency to the skin (as in holothurians); or they may form well-articulated ossicles united by muscles and thus serve as a typical internal skeleton (as in the arms of brittle stars); or they may be set closely together to form a rigid armour (as in the sea urchins). Many sea urchins have projecting spines on which the epidermis has been worn away so that the calcareous substance is exposed.

The shell of mollusks is an example of an external skeleton. It is secreted by the ectodermal epithelium of the mantle, and consists of an outer layer of the horny substance conchiolin, an intermediate prismatic layer composed of calcite, and a smooth inner layer (the nacreous layer), also composed mainly of calcium carbonate. The first two layers are secreted by a marginal band of cells, so that the shell grows at its outer margin. The nacreous layer is secreted by the general surface of the mantle and is the material of which pearls are formed around foreign bodies introduced into the mantle cavity.

The exoskeleton attains its most elaborate forms in the arthropods (crustaceans, insects, etc.). Many soft invertebrates, such as worms and mollusks, secrete a protective layer of slime or mucus over the surface of the integument. In the arthropods a solid substance termed chitin forms the basis of the exoskeleton. Chitin is related chemically to mucus. It consists of a nitrogen-containing carbohydrate (a polysaccharide, polyacetyl glucosamine) intimately associated with protein. If the protein component is removed from natural chitin, there remains a material, often called pure chitin, the physical properties of which are very similar to those of the cellulose of plants.

Natural chitin forms a cuticle that is tough but flexible. It exists in this state in the integument of caterpillars, which, as noted above, have a hydraulic type of skeleton dependent on internal turgor. It exists in this soft state also in the joints of the limbs or between one segment of the body and the next. But in most arthropods the segments of the body or of the limbs are in the form of rigid plates that form a true exoskeleton linked to adjacent segments by flexible membranes. The hard material responsible for this change in the chitin is termed sclerotin, and the hardening process is called sclerotization.

The chemical nature of sclerotization is not fully understood, but it is certainly closely related to the hardening of proteins by tanning. Substances capable of causing tanning (notably certain quinones) are almost certainly produced in the cuticle of arthropods during the hardening process. The resulting product (sclerotin) is a kind of natural plastic. In its horny consistency it closely resembles keratin; both are cross-linked or polymerized proteins, but the chemical nature of the linkage is different in the two substances. It is probable that other skeletal proteins in invertebrates, such as the spongin of sponges and the conchiolin of mollusks, are also tanned proteins allied to sclerotin.

In many crustaceans (crabs, lobsters, etc.), much of the cuticle is rendered hard by the incorporation of lime in the form of aragonite or calcite. But sclerotin is actually harder than calcite, and those parts of crustaceans that need to be of maximum hardness, such as the mandibles and the tips of the claws, are in fact composed of sclerotin. The hardest parts of insects are often the darkest; the hardening and darkening processes are closely associated chemically.

Besides functioning as a skeleton, the cuticle of terrestrial arthropods must act as a waterproof covering in order to prevent these small animals from drying up. This waterproofing is effected by the secretion of a layer of crystalline wax on the surface layer of the cuticle. Such a wax layer, if freely exposed, would be excessively fragile. It is commonly protected by a thin layer of cement substance poured over its surface by small dermal glands. Other glands may discharge quantities of wax, lac and other products upon the surface of the insect.

The cuticle of arthropods is the product of a single layer of epidermal cells; indeed cytoplasmic processes from these cells (so-called pore canals) penetrate far into the cuticle, so that it is really a living structure. It can undergo endless modifications of form to produce sensory hairs, pigment-bearing scales, claws, wings and other kinds of tools. In some insects it shows brilliant metallic colours that result from the presence of some fine structure in the cuticle (multiple thin plates or ridges) that gives rise to interference. In order that the arthropod may grow, the old cuticle is shed from time to time, after a new and larger cuticle has been laid down beneath it. This process is termed molting or ecdysis. Until the new cuticle hardens, the arthropod is in a very vulnerable condition.

See also ARTHROPODA.

BIBLIOGRAPHY.—L. A. Borradaile and F. A. Potts, *The Invertebrata*, 4th ed., (1961); A. G. Richards, *The Integument of Arthropods* (1951); V. B. Wigglesworth, *The Principles of Insect Physiology*, 5th ed. (1953), "The Physiology of Insect Cuticle," *Annual Review of Entomology*, 2:37–54 (1957). (V. B. W.)

SKELETON, VERTEBRATE. In many animals and in all vertebrates, the general form of the body is maintained by means of certain elements that are hardened to form a structural framework called the skeleton. This persists after death for a variable length of time after the rest of the body has disintegrated. In higher vertebrates, the skeleton is rendered more rigid and permanent by the formation of bone. In many animals the shape and relative positions of the various parts of the skeleton determine the general body form.

In most invertebrates, the skeleton is on the surface to form a protection as well as a supporting framework, and is called an exoskeleton (*see* SKELETON, INVERTEBRATE). In the vertebrates there is an internal skeleton, called an endoskeleton; the exoskeleton is either greatly modified or completely suppressed. In the following account the human skeleton, consisting of 206 distinct bones plus a few others called sesamoid bones, is taken as an example of the mammalian skeleton; the bones of the head are discussed separately in the article SKULL. For the relationship of the skeleton to other anatomical structures *see* ANATOMY, GROSS (plates). This article is divided into the following sections:

I. VERTEBRAL COLUMN

The vertebral column (also known as the backbone, spine, or spinal column) consists of a series of vertebrae connected by thick elastic intervertebral discs and other ligaments (*see* fig. 1). The column lies in the middle of the back of the neck and trunk and is surmounted by the cranium; the column and cranium together comprise the axial skeleton. Each vertebra, with certain exceptions, consists of a thick anterior part, the body, and a posterior part, the neural arch (*see* fig. 2).

The paired nerves connected with the spinal cord (*q.v.*) emerge on each side of the vertebral column between successive vertebral bodies and neural arches. The bodies, with the intervertebral discs, form the weight-bearing portion of the vertebral column. The neural arches form partial protection for the spinal cord and for the proximal parts of the lower spinal nerves. Each neural arch consists primarily of two cylindrical pedicles, projecting backward from the body, and two flattened laminae, which meet in the midline posteriorly. At the junction of each pedicle and lamina, there is a laterally directed transverse process, and vertically disposed superior and inferior articular processes (pre- and postzygapophyses). The superior articular processes of one neural arch articulate with the inferior articular processes of the arch above at a synovial joint. A spinous process projects backward from the fused laminae and can be felt under the skin.

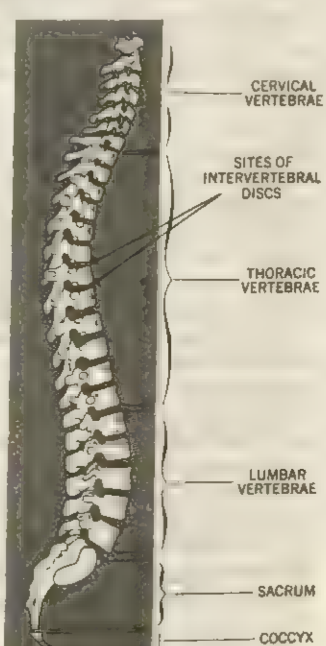


FIG. 1.—RIGHT SIDE VIEW OF THE ADULT VERTEBRAL COLUMN

In the young child, the vertebral column consists of 33 separate vertebrae. In the adult, some of these are fused together, so that there are only 26 separate bones. The vertebrae are grouped from above downward according to position into cervical, thoracic, lumbar, sacral, and coccygeal. For the human infant, the number of vertebrae in each group is expressed in the formula C_7, Th_{12}, L_5, S_5, Coc_4, but in the adult the five sacral vertebrae have fused to form a single triangular bone, the sacrum, and the four coccygeal to form the coccyx.

Vertebrae.—The cervical vertebrae are distinguished by the presence of a foramen (hole) in the transverse process. These foramina, excepting that of the 7th vertebra, are in succession traversed by the vertebral artery carrying the blood to part of the brain. The 1st cervical vertebra, or atlas, is atypical in that it has no body but is in the form of a ring. On each side of it there is a thick mass of bone by which it articulates above with the occipital bone of the skull and below with the 2nd vertebra, the axis. The body of the latter is surmounted by a stout, peglike odontoid process, which is regarded as the body of the atlas. This process forms a pivot round which the atlas moves, with the skull, when the head is turned from side to side.

Each of the typical thoracic vertebrae (1st to 9th) is distinguished by the presence, on each side of its body, of a small, smooth facet for articulation with the head of a rib. In addition, these vertebrae have a facet on each transverse process for articulation with the tubercle of a rib. The body of the 10th thoracic vertebra, however, articulates only with one rib. In the case of the 11th and 12th vertebrae, there are articular facets on the body but not on the transverse processes.

The lumbar vertebrae have neither foramina in the transverse processes nor articular facets on the bodies. The bodies are massive in order to transmit the weight of the trunk.

The sacrum is a large triangular bone formed by the fusion of five vertebrae; with the two hipbones, it forms the pelvic girdle. The anterior surface of the sacrum is concave and is marked by four transverse lines which indicate the sites of fusion of the original five sacral elements (*see* fig. 5). On each side of the fused bodies there are foramina (holes) for transmission of the anterior divisions of the upper four sacral nerves. The convex posterior surface shows four rudimentary spinous processes, and on each side of these are two rows of tubercles representing the articular processes and the transverse processes. Between the rows of tubercles four pairs of foramina transmit the posterior divisions of the sacral nerves. The part of the sacrum lateral to the foramina is known as the lateral mass. The upper lateral aspect of this mass presents an L-shaped articular surface for articulation with the corresponding hipbone. Posterior to the articular surface, the lateral mass is rough, for the ligaments that attach the sacrum to the hipbone. The human sacrum is broader in proportion to its length than that of other mammals, so that great solidity is given to the lower part of the spine; this breadth is probably correlated with the firmness of the sacrum's articulation with the hipbones, as an adaptation of the spine to man's erect posture. The sacrum is broader in women than in men and hence the whole pelvis is wider.

The coccyx consists of four rudimentary vertebral bodies fused together and, in the adult, commonly joined to the sacrum. The first body has a pair of small projections which are the vestiges of the neural arch.

Articulated Vertebral Column.—The vertebrae are held firmly together by a number of strong ligaments. The main ligaments are the intervertebral discs, the anterior and posterior longitudinal ligaments between the bodies, and the ligamentum flavum between the laminae. The intervertebral discs form flexible connections between the bodies of adjacent vertebrae from the axis above to the sacrum below. Each disc consists of fibrocartilage which encloses a central mass, the nucleus pulposus. The fibrocartilage is firmly attached to the adjacent vertebrae, whereas the nucleus pulposus is a semifluid structure acting as a shock absorber and can be deformed by pressure. Occasionally, the pressure developed in the nucleus pulposus is so great that the surrounding fibrocartilage ruptures and the nucleus protrudes ("slipped disc"; for this and other spinal ailments *see* SPINE, DISEASES AND DISABILITIES OF).

The articulated vertebral column forms the central axis of the skeleton and supports the weight not only of the head and trunk but also of the upper limbs. It gives attachment to the muscles

that move the column itself and to other muscles that move the head and limbs.

The vertebral column of an adult in the ordinary standing position exhibits curves in the sagittal (anteroposterior) plane of the body (*see* fig. 1), and minor curves in the coronal (transverse) plane. In the infant at birth, there are two primary sagittal curves concave forward which meet at the sacrovertebral junction. The lower curve involves the sacral and coccygeal vertebrae, the upper affecting the remainder of the column. When the infant begins to lift up its head, a secondary sagittal curve, convex forward, develops in the neck region, and when the child begins to walk, a further secondary sagittal curve convex forward is developed in the lumbar region. In the adult spine, therefore, there is a series of alternating convex and concave curves formed in adaptation to the erect posture of man. As a result of these curves the weight of the body is transmitted to the sacrum with the least expenditure of muscular effort. The curves in the coronal plane are not constant; the frequent convexity to the right has been attributed to the greater use of the right arm.

Movements of the vertebral column are: (1) forward bending, known as flexion, and the opposite movement, extension; (2) lateral bending, known as lateral flexion; and (3) rotation, which occurs about an axis passing vertically through the bodies of the vertebrae concerned. The range and kind of movements show great variations in the different parts of the column. They are greatest in the cervical region, at the thoracolumbar junction and at the lumbosacral junction. The limits of movement are determined by the planes of the joints between the articular processes, by the flexibility of the intervertebral discs, and by the extensibility of the muscles and ligaments. In the thoracic region movements of the vertebral column are more restricted because the ribs and costal cartilages resist distortion.

II. THORACIC SKELETON

The thoracic skeleton consists of the thoracic part of the vertebral column, the ribs with their costal cartilages, and the sternum. The bodies of the vertebrae project into the thoracic cavity from behind so that there is a recess of the cavity on each side of them.

The sternum, or breastbone, consists of three parts: an upper, the manubrium; a middle, the body; and a lower, the xiphoid process (*see* fig. 3). The anterior and posterior surfaces of the body are marked by transverse lines, indicating its division into four originally distinct segments. Each lateral margin of the sternum presents seven notches for articulation with the seven upper costal cartilages; at each side of the upper margin of the manubrium is a depression where the clavicle articulates; the xiphoid process remains cartilaginous up to a late period in life.

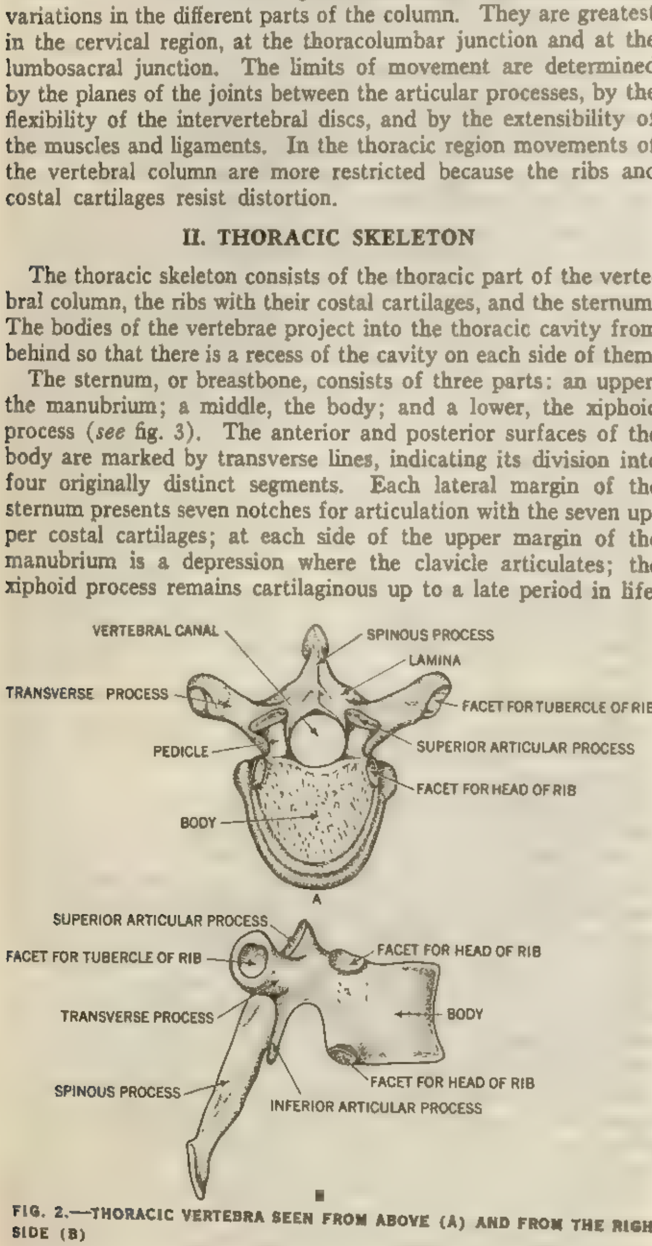


FIG. 2.—THORACIC VERTEBRA SEEN FROM ABOVE (A) AND FROM THE RIGHT SIDE (B)

Most skeletons have 12 pairs of ribs, which articulate behind with the thoracic region of the vertebral column. Occasionally cervical or lumbar ribs are present. From the vertebral column the ribs are directed first backward, then downward and forward to end at a lower level by joining the costal cartilages. The upper seven ribs are attached to the sternum by their costal cartilages and are called true ribs. The lower five ribs are called false ribs. The cartilages of the 8th, 9th, and 10th ribs do not reach the sternum directly but join the cartilages of the ribs immediately above. The cartilages of the 11th and 12th ribs have no anterior attachments, and these ribs are known as floating ribs.

A typical rib (from 3rd to 9th) consists of head, neck, and shaft. The head articulates with two adjacent vertebral bodies. A short neck separates the head from a tubercle, articulating with a transverse process. Beyond that the shaft curves round the side of the thorax to the costal cartilage.

III. APPENDICULAR SKELETON

The upper and lower limbs in man resemble one another in their general plan, but differ greatly in their detailed structure and in their mode of attachment to the trunk. Their functions also have undergone pronounced modifications; the upper limbs have been freed from supporting the body weight and have become grasping organs, whereas the lower limbs have to be used not only to propel the body but also to act as struts to prevent it from falling forward during locomotion. Each limb is attached to the axial skeleton by means of a bony girdle; that of the upper limb is called the shoulder (pectoral) girdle, and that of the lower limb the pelvic girdle. The shoulder girdle, consisting of the clavicle and scapula (*see* fig. 3, 4), has considerable mobility on the trunk; its sole skeletal connection with the trunk is through the articulation of the clavicle with the upper end of the sternum at the sternoclavicular joint. There is a wide range of free movement of the upper limb on the shoulder girdle, and this is supplemented by movements of the girdle on the trunk. The movements at the shoulder joint can take place about an infinite number of axes. Specific names are given for these movements, such as flexion, extension, abduction, adduction, rotation, and circumduction.

The lower limb articulates with the hipbone. This is firmly connected with the sacrum to form the pelvic girdle (*see* fig. 3, 4), which acts as a single skeletal unit in all ordinary body movements. It is through the hipbone that the weight of the trunk is transferred from the sacrum to the head of the femur. Conversely, the pelvic girdle transmits to the trunk the forces arising from action of the lower limb. Movements at the hip joint, like those at the shoulder joint, take place about a number of axes; the movements, however, are more limited. Certain of them are distinguished by names. Forward movement of the thigh relative to the pelvic region is called flexion, the backward movement extension. Movement of the limb toward the midline of the trunk is called adduction, the movement away from it abduction. Medial and lateral rotation of the limb can also occur, and the combination of all the movements is called circumduction.

Each of the limbs consists of three segments: proximal, intermediate, and distal.

The proximal segment in both arm and leg contains one long bone, the humerus in the arm and the femur in the leg.

The intermediate segment consists of two bones, the radius and ulna in the forearm and the tibia and fibula in the leg (*see* fig 5). The articulation of the intermediate with the proximal segment in each limb is hingelike; the joint is termed the elbow in the upper limb and the knee in the lower limb. Axial rotation (rotary movement) is fairly extensive at the knee except when the leg is completely straightened (extended) but is negligible at the elbow. The movements possible in the intermediate parts of the two limbs differ considerably; rotation, known as pronation and supination, can take place within the forearm by a movement between the radius and the ulna, but in the lower limb no comparable movement is possible.

The distal segments of the upper and lower limbs consist respectively of the hand and foot. These comprise: (1) a series of small bones, collectively called the carpus in the hand and the tarsus

in the foot; (2) five elongated bones, metacarpals or metatarsals, each of which supports (3) a finger or toe, known generally as a digit. Most of the digits consist of three bones called phalanges; the thumb and great toe are exceptional in having only two.

The articulation of the carpus with the radius and ulna, known as the wrist, allows movements around a transverse axis (*i.e.*, movements producing flexion and extension of the hand) and around an anteroposterior axis permitting abduction and adduction of the hand. Movements at the articulation of the tarsus with the tibia and fibula, known as the ankle, are almost restricted to a hingelike flexion and extension. In the foot there is a considerable range of movement between certain of the tarsal bones. The movement that turns the sole of the foot to face inward is known as inversion, while the movement by which it can be turned outward is known as eversion. These movements permit the sole of the foot to adapt itself to the surface contour on which it rests.

There is a general resemblance between toe and finger movements, but the joint at the base of the thumb between the first metacarpal and the adjacent carpal bone (trapezium) permits a variety of movements which is lacking in the big toe. At the metacarpophalangeal, metatarsophalangeal, and interphalangeal joints, hingelike flexion and extension movements take place. The movements are freer in the hand than in the foot.

A. Upper Limb

The bones of the upper limb are the clavicle and scapula of the shoulder region, the humerus of the upper arm, the radius and ulna of the forearm, and the numerous small bones of the hand (*see* fig. 3, 4).

Clavicle.—The clavicle, or collarbone, is an elongated bone which extends horizontally from the upper end of the sternum outward, to articulate with the acromion, the outermost process of the scapula (point of the shoulder). It presents a strong sigmoidal

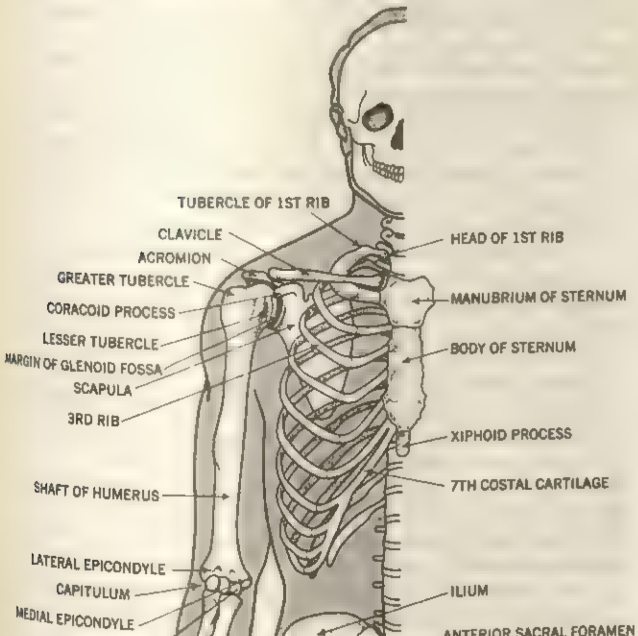


FIG. 3.—FRONT VIEW SHOWING RIGHT HALF OF SKELETON OF THE TRUNK AND UPPER LIMB

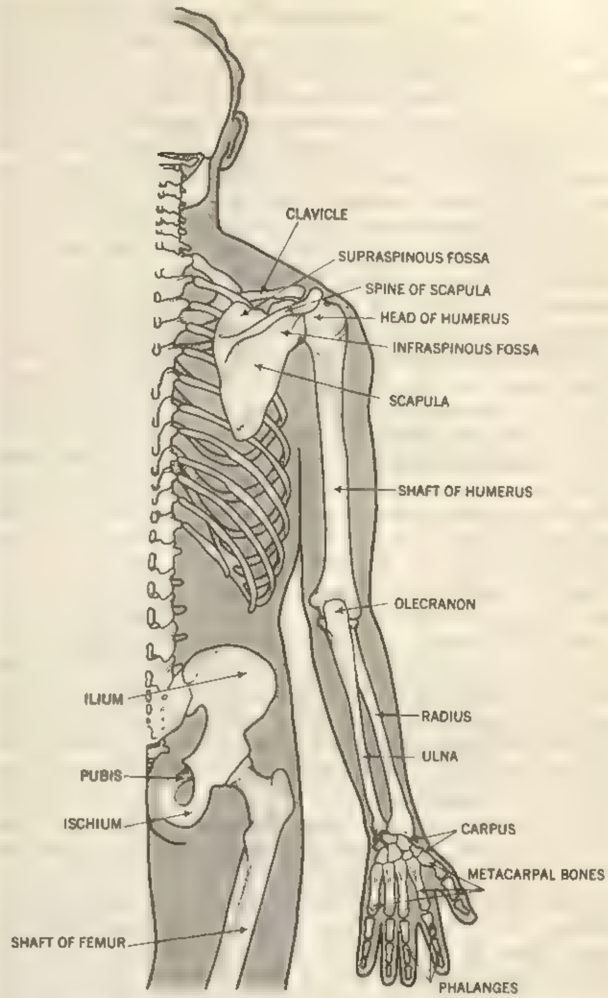


FIG. 4.—REAR VIEW SHOWING RIGHT HALF OF SKELETON OF THE TRUNK AND UPPER LIMB

curve, and is slender in the female but robust in muscular males. Its sternal end is thick; its acromial end, flattened from above downward, has an articular surface that meets a corresponding surface on the acromion. Strong ligaments connect the ends of the clavicle with the sternum and 1st rib and with the acromion and coracoid process of the scapula. The shaft gives attachment to muscles of the shoulder girdle.

Scapula.—The scapula, or shoulder blade, lies upon the upper part of the back of the chest, between the levels of the 2nd and 8th ribs. It is triangular in outline. Its anterior surface, which is concave, faces the ribs. Its posterior surface is crossed obliquely by a prominent ridge, the spine, which subdivides this aspect of the bone into a supraspinous and an infraspinous fossa (concavity). The spine and these two fossae give attachment to muscles which act in rotating the arm. The spine arches forward to end in the broad flat acromion by which the scapula articulates with the clavicle. To the margins of the scapula are attached muscles which participate in moving or fixing the shoulder as demanded by the movements of the upper limb. Of the three angles of the scapula, the outer is thickened and presents the shallow glenoid fossa, which articulates with the head of the humerus to form the shoulder joint. Overhanging the glenoid fossa is a beaklike projection, the coracoid process.

Humerus.—The humerus is a typical long bone, and as such presents a shaft and two expanded ends. The upper end possesses a hemispherical head with a smooth surface which articulates with the glenoid fossa of the scapula; where head and shaft merge, two tubercles are found, to which are attached rotator muscles which arise from the scapular fossae. Between the tubercles is a groove which carries the long tendon of the biceps muscle. The shaft of the humerus is three-sided in section above, but flattened and ex-

panded below. A broad, shallow groove winds round the back of the bone, lodging the radial nerve. The lower end of the bone consists of articular and nonarticular regions. The lateral articular region, the capitulum, articulates with the head of the radius, and the medial, the trochlea, articulates with the ulna. The nonarticular regions have projections, the medial and lateral epicondyles; each is surmounted by a supracondylar ridge. The medial epicondyle and ridge give attachment to the muscles that pronate the forearm and flex it and the fingers; the lateral epicondyle gives origin to muscles which participate in supination and in extension of the wrist and fingers.

Radius.—The radius, the lateral bone of the forearm, conforms to the type of a long bone. The upper extremity, or head, which is disc-shaped, articulates above with the humerus; the margin of the disc articulates with the ulna and with the annular ligament. The head of the bone is joined to the shaft by a neck, below which is a tuberosity into which the biceps muscle is inserted. The shaft of the radius affords attachment for muscles which act upon the fingers and the wrist. Its sharp inner margin gives attachment to the interosseous membrane, which forms a partition between the structures of the front and back of the forearm, affording surfaces for muscle attachments. The lower, or distal, end of the radius is much broader than the proximal end, and is marked posteriorly by grooves that lodge the muscle tendons passing to the back of the hand. From its outer margin a pointed styloid process projects downward. The inner margin of the distal extremity has a smooth concave surface for articulation with the lower end or head of the ulna, and its broad concave lower surface articulates with the scaphoid and lunate bones of the wrist.

Ulna.—The ulna is the long bone on the medial side of the forearm. Its proximal end is subdivided into two strong processes by the trochlear notch, the smooth surface of which articulates with the humerus. The lower, or coronoid, process is rough, receiving the insertion of the chief flexor of the forearm, the brachialis muscle, whereas the upper and posterior process, the olecranon, gives insertion to the powerful extensor of the forearm, the triceps muscle from the back of the arm. Immediately behind the trochlear notch on the lateral side of the ulna is the radial notch for articulation with the margin of the head of the radius. The shaft of the bone has three surfaces giving attachment to muscles which act upon the wrist and fingers and in the movements of both pronation and supination. The sharp lateral margin of the shaft gives attachment to the interosseous membrane. The lower end, or head, of the ulna, much smaller than the upper, has a styloid process and an articular surface, the lateral surface of which meets the lower end of the radius. The lower part is in contact with a fibrocartilage which binds the ulna to the radius and separates the distal end of the former from the wrist joint.

The ulna, much more than the radius, establishes the hinge joint at the elbow. The radius, although entering into this joint, is involved, functionally, less with the arm than with the hand with which it alone articulates. By muscular action the radius revolves about the stationary ulna, carrying the hand in the movement called twisting the wrist. When the palm faces forward in supination, the radius and ulna are parallel. When it faces backward in pronation, the radius crosses obliquely over the ulna.

Hand.—The hand consists of the carpus, or wrist; the metacarpus, or palm; and the five digits individually named. Anatomists describe it with the palm turned to the front, and with its long axis in line with the long axis of the forearm.

The eight carpal bones are arranged in two rows. Named in order from the radial to the ulnar side, the proximal row comprises the scaphoid, lunate, triquetral, and pisiform bones, and the distal row the trapezium, trapezoid, capitate, and hamate bones.

The five metacarpal bones are miniature long bones. The metacarpal of the thumb is the shortest, and diverges outward from the rest; its carpal extremity is saddle-shaped for articulation with the trapezium. Its shaft is somewhat flattened, and its distal end is smooth and rounded, articulating with the first phalanx of the thumb. The bases of the other metacarpal bones articulate with the trapezoid, capitate, and hamate bones; the distal end or head of each articulates with the proximal phalanx of a finger.

The five digits of the hand are called, by anatomists, the pollex, index, medius, annularis, and minimus—popularly, the thumb and the index, middle, ring, and little finger. The skeleton of the thumb consists of two phalanges while that of each finger has three. Each phalanx is a miniature long bone.

The carpal bones although firmly joined together, are nevertheless capable of slight movements that are essential in the mechanism of the hand as a grasping organ. The palmar contour of the wrist skeleton is like a gutter, bridged over by a dense ligament that makes it into a tunnel through which pass blood vessels, nerves and tendons for the fingers. The thumb has a great range and variety of movement, a human characteristic made possible in part by the conformation of the base of its metacarpal with the trapezium. The other carpometacarpal joints permit only slight motion, in contrast with the free movement of the fingers.

B. Lower Limb

The bones of the lower limb are the hipbone, the femur of the thigh, the patella of the knee, the tibia and fibula of the lower leg, and the numerous bones of the foot.

Hipbone.—The hipbone (also called the innominate bone or os coxae) is a large irregular plate which forms the lateral and anterior wall of the pelvis (*see* below). In early life it is made up of three separate elements—ilium, ischium, and pubis—which unite about the 18th to the 20th year into a single bone. At the site of junction of the three elements there is a cup-shaped depression, the acetabulum, for the head of the femur (*see* fig. 3, 4). From the acetabulum the ilium extends upward and laterally, the ischium downward, the pubis forward and medially. Below the acetabulum is a large opening, the obturator foramen, bounded by the pubis, the ischium, and the ilium.

Ilium.—The ilium is a broad, bony plate, the lower, narrower part of which enters into the formation of the acetabulum, whereas the upper, expanded part presents an elongated free margin, the sinuous crest of the ilium. This crest affords attachment to the broad muscles that form the wall of the abdominal cavity. The external or lateral surface of the ilium gives origin to the muscles of the buttock. The internal surface is hollowed out anteriorly and gives origin to the iliacus muscle, which flexes the thigh. Posteriorly the ilium has an L-shaped surface articulating with the side of the sacrum in the sacroiliac joint.

Pubis.—The pubis consists of a body (main portion) and two rami (secondary structures arising from the body) and is therefore somewhat U-shaped. The body is the flattened portion adjacent to the symphysis, and has an outer and an inner surface. The upper ramus passes backward to become continuous with the ilium; the lower ramus is continuous with the ramus of the ischium.

Ischium.—The ischium also is somewhat U-shaped. The upper extremity completes the acetabulum, whereas the lower is united with the inferior ramus of the pubis, and together these form the lower boundary of the obturator foramen. The conjoined rami afford origin to the adductor muscles that act in drawing the thigh toward the middle line. The upper, posterior branch is a stout bone which gives origin to the powerful hamstring muscles of the back of the thigh. The spine of the ischium separates the greater from the lesser sciatic notch. In the sitting position the body rests on a tuberosity on the back of the ischium.

Pelvic Girdle.—Articulation of the hipbones with each other at the pubic symphysis, and with the sides of the sacrum behind, forms the pelvic girdle—the bony walls of the lower body cavity known as the pelvis. (This term is popularly but incorrectly applied to the hipbone or to the pelvic girdle, rather than to the cavity which the bones define.) The pelvis is subdivided into a major (false) and a minor (true) pelvis. The true pelvis encloses a short, wide canal and possesses a superior opening, or inlet, a cavity and an inferior opening, or outlet. The false pelvis consists of the iliac fossa (concavity) of each side above the level of the inlet of the true pelvis. Because of the inclination of the pelvis, the upper part of the sacrum is nearly 10 cm. (4 in.) higher than the upper margin of the pubic symphysis. The female pelvis is distinguished from the male by certain sexual characteristics

The bones are more delicate and the ridges and processes for muscular attachment are less prominent. The depth is less, the inlet more nearly circular, the pubic arch wider, the distance between the ischial tuberosities greater and the acetabulum smaller in the female than in the male. In addition, the pelvis is broader and its capacity greater, and this contributes to the greater breadth of the hips in women. The greater capacity of the female pelvis affords room for expansion of the uterus during pregnancy and for the passage of the child at the time of birth.

Femur.—The femur is the longest bone in the body. At its upper extremity is the smooth hemispherical head, articulating with the acetabulum. Supporting the head is a strong, elongated neck which extends downward and outward to join the shaft. At the junction of neck and shaft two processes, the trochanters, are situated. To the lateral, or greater, trochanter are attached muscles that rotate the thigh; into the lesser trochanter is inserted the iliopsoas muscle, a flexor of the thigh. The neck of the femur in the adult forms an angle of about 125° with the shaft. The shaft is almost cylindrical at its centre, but it is expanded above and below. Its front and sides give origin to the extensor muscles of the leg; behind, there is a long ridge, the linea aspera, into which are inserted the adductor muscles of the thigh. The lower end of the femur is adapted to the upper surface of the tibia and presents a large, smooth, articular surface, the anterior portion forming a trochlea (pulley) for the patella. The lower and posterior part, subdivided into two condyles (rounded projections), moves upon the condyles of the head of the tibia. In the deep fossa between the femoral condyles are attached the cruciate ligaments that control and limit movements at the knee joint.

Patella.—The patella, or kneecap, is a triangular, flattened bone of the sesamoid type (*see* below) developed in the tendon of the great extensor muscles of the leg. Its anterior surface and sides are rough, giving attachment to tendinous fibres; its posterior surface is smooth, and enters into the formation of the knee joint by articulating with the anterior part of the lower end of the femur.

Tibia.—The tibia, or shin bone, is the medial and larger of the two bones of the lower leg. The femur moves and rests upon its upper end, and through it the weight of the body in the erect position is transmitted to the foot. Its upper extremity (head) is broad and expanded into two condyles, the lateral one bearing inferiorly a small articular facet for the head of the fibula. Superiorly, the two condyles have nearly flat surfaces, articulating with the condyles of the femur. The two condyles are separated by an intermediate rough surface, from which an intercondyloid eminence projects; standing opposite the intercondylar fossa of the femur, this eminence affords attachments to the interarticular cruciate ligaments and the meniscoid cartilages of the knee. These cartilages are movable rings which help to compensate for the flatness of the condylar articular surfaces by making shallow sockets for the femoral condyles. The shaft of the tibia is three-sided. Its medial surface is subcutaneous and forms the shin. Its lateral and posterior surfaces give origin to muscles which move the foot. The anterior margin is the sharp ridge of the shin, and terminates superiorly in a tubercle into which the great tendon of the extensor muscles of the leg is inserted. The lateral margin of the bone gives attachment to the interosseous membrane of the leg. The lower end of the tibia, smaller than the upper, is prolonged into a broad process, the medial malleolus, which forms the inner prominence of the ankle. The lower end of the tibia articulates with the talus.

Fibula.—The fibula is the long, slender bone buried in the muscles of the outer side of the leg. The upper end articulates with the lateral condyle of the tibia. The shaft is four-sided and gives attachment to many muscles. Separating the anterior from the medial surface is a slender ridge giving attachment to the interosseous membrane. The lower end expands into a strong process, the lateral malleolus, forming the lateral prominence of the ankle and presenting a smooth medial surface articulating with the talus. Behind this is a rough surface where the strong ligaments that bind the tibia and fibula together are attached.

Foot.—The foot consists of the tarsus, the metatarsus, and the five free digits, or toes. The human foot is placed in the prone position, with the back of the foot directed upward. The axis of the foot is at about a right angle to the axis of the leg, and the great toe, or hallux, corresponds to the thumb.

The seven bones of the tarsus, or ankle, are arranged in three transverse rows: a proximal, consisting of the talus and calcaneus; a middle, consisting of the navicular; and a distal, composed of the first, second, and third cuneiform bones and the cuboid bone. The tarsal bones like the carpal bones are short and, with the exception of the wedge-shaped cuneiforms, are irregularly cuboidal. The dorsal and plantar surfaces of most of these bones are rough for the attachment of ligaments; the exception is the talus which, since it is mortised between the bones of the leg and the calcaneus, presents smooth articular surfaces on its dorsal and plantar surfaces as well as on its sides.

The posterior portion of the calcaneus projects backward to form the prominence of the heel. With this exception the tarsal bones have articular surfaces upon their anterior and posterior surfaces. Their lateral and medial surfaces are also articular, excepting the lateral surfaces of the calcaneus and cuboid, which

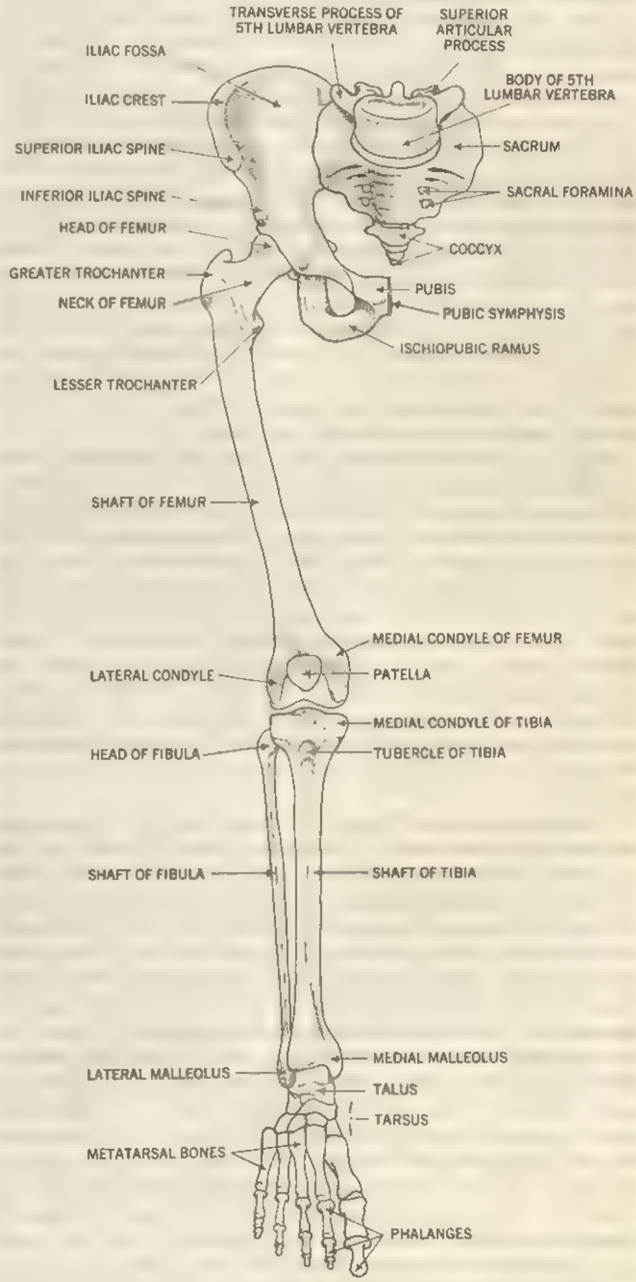


FIG. 5.—FRONT VIEW OF SKELETON SHOWING RIGHT HALF OF THE PELVIC GIRDLE AND RIGHT LOWER LIMB

form the outer margin of the tarsus, and the medial surfaces of the calcaneus, navicular, and first cuneiform, which form the inner margin.

The metatarsal bones and phalanges agree in number and in general form with the metacarpal bones and the phalanges in the hand. The bones of the great toe are more massive than those of the other digits, and this digit, unlike the thumb, does not diverge from the other digits but lies almost parallel with them.

C. Sesamoid Bones

In addition to the bones described above, certain supernumerary bones develop in certain tendons of the muscles of the hand and foot; these are known as sesamoid bones. The patella also originates in tendinous tissue, and is therefore sesamoid in character; it begins to ossify about the third year.

IV. EMBRYOLOGY

The skeleton is developed from the embryonic mesoderm, so it is necessary to refer briefly to the origin of this tissue. At about the 15th day after fertilization of the ovum, when the embryo consists of a bilaminar disc with layers called ectoderm and endoderm, a longitudinal thickening appears in the posterior part of the embryonic disc as the result of multiplication of the ectodermal cells. This thickening, called the primitive streak, gives rise to the notochord and to the mesoderm. The primitive streak is comparable to the blastopore of lower vertebrate embryos. The axis of the embryo is first laid down by the formation of a cylindrical mass of cells, the notochord, proliferated from the primitive (Hensen's) node at the anterior end of the primitive streak. The notochord lies ventral to the developing central nervous system and forms the first supporting structure for the developing embryo.

In fishes such as the shark, cartilaginous vertebrae are formed round the notochord and to some extent compress it. Nevertheless, it persists as a continuous structure through the length of the vertebral column. In the higher vertebrates, including man, the notochord is a temporary structure which persists only as a minute canal in the bodies of the vertebrae and in the central part of the nucleus pulposus of the intervertebral discs.

While the notochord is being laid down, cells proliferate from the sides of the primitive streak. These form the mesoderm, which spreads out as a sheet on each side and, as a result of migration and multiplication of cells, soon comes to occupy most of the space between the ectoderm and the endoderm on each side of the notochord. The mesodermal sheets soon become differentiated into: (1) a mass lying on each side of the notochord (paraxial mass) which undergoes segmentation into hollow blocks, the mesodermal somites; (2) a lateral plate which becomes separated into an outer layer, the somatopleuric mesoderm, against the future body wall, and an inner layer, the splanchnopleuric mesoderm, against the endoderm of the future gut; (3) an intermediate mass, the nephrogenic cord, which gives rise mainly to the urinary apparatus.

The segmentation of the paraxial mesoderm is a fundamental feature of the development of the vertebrates. The axial skeleton and associated structures develop from part of the somite, while the appendicular skeleton arises from the somatopleuric mesoderm of the lateral plate. Each somite differentiates into: (1) a lateral and superficial plaque, the dermatome, giving rise to the integumentary tissue; (2) a deeper lateral mass, the myotome, which gives origin to the muscles; and (3) a medial ventral mass, the sclerotome. The sclerotomic cells from each pair of somites migrate medially until they meet in the middle line round the notochord to separate it from the neural tube dorsally and from the aorta ventrally. The sclerotomic tissue retains its original segmentation and becomes condensed to form the forerunner or blastema of the body of the future vertebra. From each posterolateral half of the condensation extensions pass backward and eventually meet posteriorly round the neural tube to form the blastema of the neural arch of the vertebra. In the interspaces between adjacent myotomes of each side, an extension from each sclerotomic mass passes laterally and forward to form the costal or rib element. It is only in the thoracic region that the costal elements develop into the definitive ribs. In the other regions the costal elements remain rudimentary. As early as the seventh week of fetal life the mesenchymal blastema of the future vertebra becomes chondrified; *i.e.*, the mesenchymal cells are converted into cartilage cells. In this cartilaginous vertebra, ossification centres appear about two weeks later, and the cartilage is gradually replaced by bone. The mesenchyme of the primordia of the ribs also undergoes chondrification and later ossification. In the thoracic region, where costal elements are best developed, a cartilaginous sternal bar is formed in relationship with their anterior or growing ends.

The appendicular skeleton begins to develop in the primitive limb bud at the fifth week of fetal life in the core of mesenchyme derived directly from the unsegmented somatopleuric mesoderm. This mesenchyme becomes condensed to form the blastemal masses of the future limb bones. By the seventh week of fetal life, the mesenchyme has become transformed into the cartilaginous precursors of the individual bones (except in the case of the clavicle). The cartilaginous models determine the general shape and relative size of the bones, and there is convincing evidence from experimental investigations that the shape of the bones of higher vertebrates is determined by intrinsic factors inherent in the tissues and that extrinsic influences are concerned with providing the proper conditions for maintaining the normal structure once development has begun.

The first mesenchymal condensations of the appendicular skeleton are in the region of the future girdles, and those for the shoulder girdle appear a little earlier than those for the pelvic girdle. The mesenchymal condensations for the other bones of the limbs appear in a proximodistal sequence.

Ossification of Bone.—The ossification of bone is usually described as occurring by two different methods: endochondral (in cartilage) and intramembranous (in membrane). However, the fundamental process is the same for both, since the cells that are responsible are identical.

All the bones of the skeleton (excluding certain of the skull bones and the clavicle) are developed in cartilage; *i.e.*, by endochondral ossification. It must be emphasized that there is no direct transformation of the cartilage into bone. The cartilage of the model becomes calcified and is then removed by a process of erosion and replaced by bone (*see* Bone; Connective and Supporting Tissues: *Anatomical Types of Connective Tissues: Bone*).

V. COMPARATIVE ANATOMY

A. Vertebral Column and Thoracic Skeleton

Just as in the embryonic development of man the notochord forms the earliest structure that stiffens the embryo, so in the animal kingdom it appears before the true vertebral column is evolved.

Lower Chordates and Fishes.—Possession of the notochord distinguishes members of the most advanced phylum, Chordata. In the ascidians or sea squirts (subphylum Urochordata) the notochord is present in the larval stage, in the tail region only, disappearing after the animal's metamorphosis to the adult form. In amphioxus (subphylum Cephalochordata) the notochord is permanent and extends the whole length of the animal. In the cyclostome fishes (class Agnatha, the most primitive group within the subphylum Vertebrata) the notochord and its sheath persist throughout life, and in the adult lamprey rudimentary cartilaginous neural arches are found. Among the sharks (class Chondrichthyes) modern representatives possess a vertebral column composed of cartilaginous, partly calcified, centra which have their origin within the sheath of the notochord, causing its partial absorption. Among the bony fishes (class Osteichthyes) the sturgeon has a persistent notochord with fibrous sheath upon which appear paired cartilaginous arches: dorsally, the neural arches and, ventrally, the hemal arches. The vertebrae of the more advanced bony fishes, such as the salmon and cod, are completely ossified; each centrum develops in the sclerotomic mesoderm outside the notochordal sheath (perichordal development).

Amphibians.—A vertebra of a modern tailed amphibian is

in one piece, and the neural and hemal arches are in continuity with the centrum. In general, a vertebra is formed from the sclerotomic tissue of two somites, the tissue from the posterior part of one somite joining that from the anterior part of an adjacent one. Articulation between neural arches of adjacent vertebrae is by special articular processes (zygapophyses). The first vertebra of the amphibian vertebral column, the atlas, presents a pair of sockets on the anterior surface of the centrum. These sockets articulate with the two condyles of the skull. The second vertebra does not differ from those following.

Reptiles.—In living reptiles the vertebrae are completely ossified. The neural arch has a spinous process and pre- and postzygapophyses; at the junction of arch and centrum there is a facet for articulation of the head of a rib. Groups of vertebrae can be distinguished; *e.g.*, the cervical vertebrae can be recognized, as the neck is differentiated from the body. The atlas, in the form of a ring, lacks a centrum, and articulates with the single condyle of the skull by a facet on the anterior surface of the ventral bar of the ring. The second vertebra, the axis, possesses a centrum which articulates anteriorly with the odontoid process; this process is the dissociated centrum of the atlas and lies within the ring of the latter in contact with its ventral bar. The costal elements of two sacral vertebrae unite the pelvic girdle with the axial skeleton. In the tail region hemal arches are found enclosing the caudal artery and vein. In some species the hemal arches are separate and are called chevron bones. A special feature of the caudal vertebrae in the lizard is the transverse division of the centra, adapted to the shedding of the tail.

The fibrocartilaginous intervertebral discs uniting the centra of crocodiles have been identified as representing intercentra (interpreted as ossifications in the hypochordal bars). Ribs are present in the cervical, thoracic, and lumbar regions, and in the tortoises they fuse with the cervical vertebrae. The more cranial thoracic ribs are attached to the sternum through cartilaginous sternal ribs; the other ribs do not reach the sternum. In the crocodiles and the lizardlike *Sphenodon*, a spur from each thoracic rib overlaps the next rib behind and is known as an uncinate process. These processes are developed in connection with the origin of the external oblique muscle of the abdomen. The ventral elements of some of the hinder ribs of crocodiles lie unattached in the intersections of the rectus and internal oblique muscles and are known as abdominal ribs.

The sternum may be calcified but is seldom ossified in the reptiles. In the lizards it is a cartilaginous plate articulated with the coracoid processes of the pectoral girdle and with the anterior thoracic ribs. The sternum is absent in the turtles and in the snakes; in the crocodiles it is a wide plate joined by the coracoids and by two pairs of ribs.

Birds.—The skeletons of modern birds show reptilian features with some specialized adaptations to their bipedal locomotion and power of flight. Between the bodies of the vertebrae are intervertebral discs. The neck is very movable, and with its variation in length the number of cervical vertebrae ranges from 25 in the swan to only 9 in some of the small birds. The tendency for the vertebrae to fuse in certain regions is a characteristic of birds. The two true sacral vertebrae become secondarily fused with adjacent lumbar vertebrae, and indirectly even with thoracic, to constitute a synsacrum. This fuses with the ilium of the pelvic girdle, making a rigid support for the posterior part of the bird's body. The caudal vertebrae are fused into a single bone, the pygostyle, that supports the tail. The sternum is a very large bone standing like a shield in the front of the chest. In flying birds a median keel, the carina, projects ventrally, providing additional surface for the attachment of the pectoral muscles that move the wings. The flightless birds, such as the ostrich, have a keelless, raftlike sternum.

Mammals.—In mammals, the vertebral centra articulate by means of intervertebral discs of fibrocartilage. Epiphyseal bony plates formed on the generally flat ends of the centra are characteristic of mammals; they are rudimentary in the duck-billed platypus and in the sea cows. Regional differentiation in the mammalian backbone is marked. The number of vertebrae in each group, excepting the caudal, is moderately consistent with few exceptions. Some exceptions to the group averages may be cited. Whereas 7 cervical vertebrae are the rule, there are 9 or 10 in the three-toed sloth and only 6 in the two-toed sloth and the manatee. The thoracic vertebrae commonly number 13 or 14, although they vary from 9 in some whales to 24 in the two-toed sloth. The average number of lumbar vertebrae may be taken as 6; there are but 2 in the duck-billed platypus and 21 in the dolphin. Rib elements are fused to the transverse processes of the cervical vertebrae, and in the lumbar vertebrae they form the so-called transverse processes.

There is an increase in the number of vertebrae that compose the sacrum. In the early developmental stages of human embryos the beginnings of the hipbones lie opposite those segments of the spinal column that go to form the lower lumbar and upper sacral vertebrae. As development proceeds, the sacroiliac joints become established between the hipbones and the upper sacral vertebrae; the sacrum, derived from the 25th to the 29th vertebrae inclusive, becomes a single bone by their fusion. The whales and sea cows lack a sacrum, although vestiges of a pelvis occur. In some anteaters the posterior sacral vertebrae are fused with the ischium through ossification of the sacrotuberous ligament. The sacrum of some armadillos consists of 13 vertebrae, caudal vertebrae having become fused with it. The cervical vertebrae of some whales are fused together. The centrum of the atlas of most mammals fuses with that of the axis and projects from it as the dens, but in the duck-billed platypus it is a separate bone as it is in reptiles.

The spinous processes of the thoracic vertebrae, excepting the last, point caudally and those of the lumbars generally point cranially; at the transitional zone between these groups the spines of one or two thoracic vertebrae are upright and the name anticlinal is given to them. Lying ventral to the intervertebral discs in some mammals (whales, pangolin) are paired ossicles, the intercentra, homologous with the anterior arch of the atlas. The tail vertebrae vary in number from none in the bat to 49 in the pangolin.

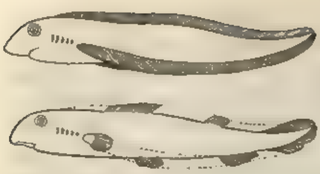
The ribs in mammals correspond in number of pairs to the number of thoracic vertebrae. The tubercle articulates with the transverse process; the head of the rib with the capitular facets of two vertebrae, that with which the tubercle articulates and the one anterior to it. The ribs of some whales articulate only with the transverse processes of the vertebrae. The ventral ends of the ribs join the costal cartilages, the relations of which follow, with minor variations, the pattern described for the human skeleton. Sternal ribs, connecting the more anterior vertebral ribs with the sternum, may be cartilaginous, calcified, or ossified. The mammalian sternum is composed of several pieces: the presternum anteriorly, followed by the mesosternum, made up of a number of segments, and a terminal single xiphisternum. The sternum of the rorqual whale and of the manatee is a single bone.

B. Appendicular Skeleton

The paired appendages are not found in ancestral vertebrates and are not yet present in the living cyclostomes. Appendages make their first appearance during the early evolution of the fishes. Usually two pairs of appendages are present and consist of fins in the fishes and limbs in land vertebrates. Each appendage includes not only the skeletal elements within the free portion of the limb but also the basal supporting structure, the limb girdle. This portion of the appendage lies partly or wholly within the trunk and forms a stable base for the fin or limb. Each girdle consists of ventral and dorsal masses. In lower fishes these are composed of cartilage; in bony fishes and in land vertebrates the masses become partly or completely ossified.

The anterior appendages, the pectoral fins or forelimbs, articulate with the pectoral girdle, which is situated just behind the gill region in fishes and in a comparable position at the junction of the neck and thorax in land vertebrates.

The posterior appendages, called pelvic fins or hindlimbs, articulate with the pelvic girdle, which is situated in the trunk region usually just in front of the anus or cloaca. It is by way of the girdles that the weight of the body in land vertebrates is trans-



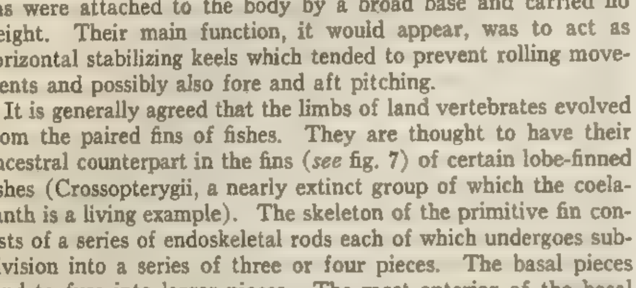
FROM H. H. WILDER, "HISTORY OF THE HUMAN BODY"; HOLT, RINEHART & WINSTON, INC.

FIG. 6.—THE THEORY OF ORIGIN OF MEDIAN AND PAIRED FINS OF THE FISH AS THE RESULT OF PERSISTENCE OF CERTAIN PARTS OF THE MEDIAN AND LATERAL FIN FOLDS

mitted to the limbs. As the hind limb is usually of greater importance in weight bearing, especially in bipedal vertebrates, it articulates with the vertebral column by means of the costal elements of the sacral vertebrae. The vertebrae to which the pelvic girdle is attached usually fuse to form the sacrum. In fishes, however, since the posterior appendages usually do not support the body weight but are used only in locomotion, a sacrum as such does not develop.

The origin of paired fins has been much debated and many theories have been put forward in explanation. According to the widely accepted fin-fold theory, the paired limbs are derived from the local persistence of parts of a continuous fold which, in ancestral vertebrates, passed along each side of the trunk and fused behind the anus into a single fin (*see* fig. 6). The primitive paired fins were attached to the body by a broad base and carried no weight. Their main function, it would appear, was to act as horizontal stabilizing keels which tended to prevent rolling movements and possibly also fore and aft pitching.

It is generally agreed that the limbs of land vertebrates evolved from the paired fins of fishes. They are thought to have their ancestral counterpart in the fins (*see* fig. 7) of certain lobe-finned fishes (Crossopterygii, a nearly extinct group of which the coelacanth is a living example). The skeleton of the primitive fin consists of a series of endoskeletal rods each of which undergoes subdivision into a series of three or four pieces. The basal pieces tend to fuse into larger pieces. The most anterior of the basal pieces fuses across the middle line with its fellow of the opposite



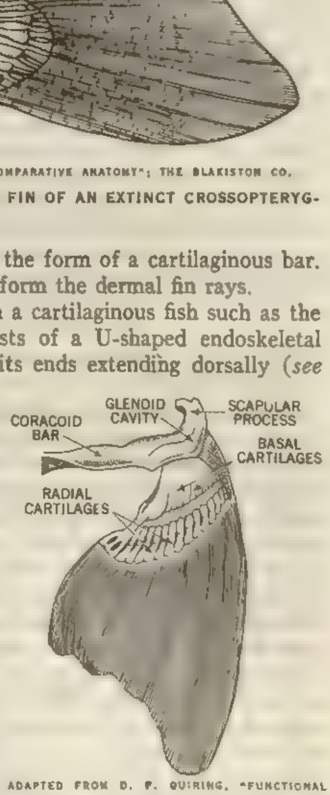
ADAPTED FROM H. V. NEAL AND H. W. RAND, "COMPARATIVE ANATOMY"; THE BLAKISTON CO.

FIG. 7.—RIGHT PECTORAL GIRDLE AND FIN OF AN EXTINCT CROSSOPTERYGIAN FISH (SAURIPTERUS)

side to form a primitive girdle in the form of a cartilaginous bar. The more distal pieces persist to form the dermal fin rays.

Pectoral Girdle.—*Fishes.*—In a cartilaginous fish such as the dogfish the pectoral girdle consists of a U-shaped endoskeletal cartilaginous inverted arch with its ends extending dorsally (*see* fig. 8). The crosspiece of the "U" forms the coracoid bar, the ends passing dorsally on each side to form the scapular processes. At the junction of the coracoid with the scapular process is an articular region, the glenoid cavity, for articulation with the basal cartilages of the fin.

In all other major groups of vertebrates the pectoral girdle is a composite structure consisting of the endoskeletal structures to which secondary dermal components are added as the result of ossification of dermal elements. These become ossified to form dermal bones. In primitive bony fishes such as the lungfishes, sturgeon, and coelacanths, the main element added is a vertically placed structure, the cleithrum, which supports the scapula (*see* fig. 9). The cleithrum may be joined by a supracleithrum, which in turn is surmounted by a posttemporal element, so that the shoulder girdle is connected with the rearward sides of the skull. The most ventral of the added dermal bones are the clavicles, which unite below the gill chamber with each other or with the sternum. In the holostean fishes (*e.g.*, gar pike) and in the teleosts (*e.g.*, salmon and codfish) the clavicle is lost, leaving only the cleithrum.



ADAPTED FROM D. P. QUIRING, "FUNCTIONAL ANATOMY OF THE VERTEBRATES"; REPRODUCED BY PERMISSION OF THE MCGRAW-HILL BOOK CO.

FIG. 8.—LEFT PECTORAL GIRDLE AND FIN OF A CARTILAGINOUS FISH

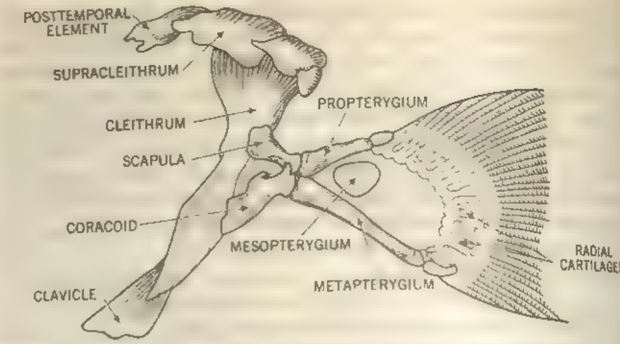


FIG. 9.—RIGHT PECTORAL GIRDLE OF A TELEOST FISH

Amphibians.—In tailed amphibians such as newts and salamanders the dermal elements of the pectoral girdle have been completely lost and only the endoskeletal parts remain; they have reverted mainly to being cartilaginous bars. This retrogression is probably the result of their adaptation chiefly to an aquatic mode of life where less support by the girdles is required. The ventral part of the girdle forms the coracoid and the dorsal part the scapula; the latter is the only part which becomes ossified, and only a rudimentary sternum is developed.

In the frog, the pectoral girdle consists of a dorsally placed, ossified scapula, to the end of which is joined an incompletely ossified plate, the suprascapula. The precoracoid and coracoid are quite distinct, the former being cartilaginous and overlaid by an investing dermal bone, the clavicle.

Reptiles.—In most reptiles there is a primary girdle for the forelimb, consisting of a scapula and a single coracoid. The pectoral girdle of the lizard consists of bones formed in cartilage the scapula and the large coracoid, participating in the glenoid cavity; and the dermal bones, the clavicle and interclavicle the latter being a single T-shaped bone, with the stem in the midline and in contact with the sternum. The curved clavicles articulate with each other at their medial ends. The cartilaginous suprascapula is present.

In the alligator the pectoral girdle consists of a stout dorsal scapula and a well-developed precoracoid. Clavicles are absent, but there is a well-developed midventral dagger-shaped interclavicle.

Birds.—In birds the pectoral girdle is essentially similar to that found in the reptiles. The precoracoid forms a stout bar which reaches to the sternum. The wishbone, or furcula, is formed from the dermal part of the girdle and consists of two clavicles united together in the middle line by the interclavicle. Carinate birds (those with a keeled sternum) have a sabre-shaped scapula and a stout coracoid joined together by ligaments, where the glenoid cavity for articulation with the humerus is located. The coracoid is jointed to the sternum; at its dorsal end is the acrocoracoid process. The furcula stands in front of the coracoids, its ends connected by ligaments with the acrocoracoid and with the rudimentary acromion process of the scapula. The girdle of the flightless ratite birds (those with a flat sternum) is little developed, being represented by an ankylosed scapula and coracoid.

Mammals.—Among mammals, the monotremes have two coracoids, articulating medially with the presternum and laterally with the scapula, which enter into the formation of the glenoid cavity. An interclavicle (episternum) and an investing clavicle resembling these bones in reptiles are also present. The clavicle articulates with the acromion of the scapula. In the opossum the scapula has a spine ending in the acromion, with which the clavicle articulates

A much-reduced coracoid fuses with the scapula and does not meet the sternum. The scapula of the placental mammals has a spine ending, generally, in an acromion, and the body of the bone is triangular. In mammals which use the forelimb for support in standing, the vertebral margin is the shortest and the long axis of the scapula runs from it to the glenoid cavity; but in those whose forelimb is used for prehension, such as the primates, or for flight, such as the bats, the vertebral margin is elongated and the distance from it to the glenoid cavity is decreased, so that the long axis is parallel with that of the body instead of being transverse. In the placental mammals, the coracoid, although developing independently, has dwindled to a beaklike process, and fuses with and becomes part of the scapula and does not articulate with the sternum. There has occurred a change in the structure of the pectoral girdle, from its form in animals (*e.g.*, reptiles) in which the forepart of the body is propped up on the forelimbs and in which the coracoid is functional, to its form in pronograde mammals (those which carry the body horizontally) in which the foreparts are suspended between the two scapulas by the serratus anterior muscles. These muscles spring from the ribs and are inserted into the vertebral margin of each scapula.

The clavicle is present generally in those placental mammals that have prehensile forelimbs (primates, many rodents and marsupials, and others), or whose forelimbs are adapted for flying (bats). In many mammals it is suppressed, as in cats, or absent, as in whales, sea cows, and hoofed animals.

Pelvic Girdle.—*Fishes.*—The pelvic girdle of the elasmobranch fishes consists of a curved cartilaginous puboischial bar or pair of bars lying transversely in the ventral part of the body anterior to the cloaca; projecting dorsally on each side is an iliac process. Connected with it is a basal cartilage carrying a series of radialia, the skeleton of the paired pelvic fins. The pelvic girdles of many bony fishes are situated far forward near the gills.

Amphibians.—There are marked variations in the form of the pelvic girdle in the amphibians. In the frog the three parts of the hipbone (ilium, ischium, and pubis) are present. The public elements, however, remain wholly cartilaginous. The hipbone is characterized by the great length and forward extension of the ilium. The girdle is connected with the costal element of one vertebra, thus establishing a sacral region of the vertebral column. The acetabulum is situated at the junction of the three elements.

Reptiles.—The pelvic girdle of some reptiles has a loose connection with the spine. In most reptiles, the ilium is joined to two sacral vertebrae. Both the pubic and the ischial parts usually meet in the ventral symphysis, from which a cartilage or bone, the hypoischium, projects backward to support the margin of the cloacal orifice, and another, the epipubis, projects forward. A few snakes (*e.g.*, boas) retain vestiges of a pelvic girdle and limb skeleton.

Birds.—In most birds the ilium extends forward and backward and is fused with the many vertebrae which form a synsacrum. The slender ischia and pubes do not form symphyses except in the ostrich.

Mammals.—In most mammals the ilium articulates with the sacrum and the pubes meet in a symphysis anteriorly. A cotyloid bone, formed in the cartilage in the bottom of the acetabulum, is usually found. The symphysis pubis is not present in certain mammals (*e.g.*, moles). In monotremes and marsupials, the marsupial bones that support the pouch have been regarded as part of the epipubis.

Limbs.—*Fishes.*—The pectoral fin of the elasmobranchs has basal cartilages articulating with the pectoral girdle and carrying a number of radial cartilages consisting of varying numbers of short segments beyond which are delicate fin rays (*see* fig. 8).

The proximal segment of the pelvic fin of sharks is supported by a single basal cartilage and by one or two radialia. In the pectoral fin of the primitive ray-finned fish *Polypterus*, three elements constitute the proximal segment of the fin, bony rods named propterygium and metapterygium on the margins and an intermediate cartilage partly ossified, the mesopterygium (*see* fig. 9).

Other Vertebrates.—In the land vertebrates, many modifications have occurred in the appendages as a result of the transformation of the fins from balancing organs to weight-bearing propulsive levers. Limbs with digits replace the paired fins. The position of the limbs undergoes progressive changes in adaptation to new conditions. The skeleton of the free limb of the land vertebrate is divisible into three segments: proximal, medial, and distal.

The proximal segment consists of a single bone, known as the humerus in the forelimb and as the femur in the hindlimb. The humerus articulates by its rounded head with the glenoid cavity of the scapula and by condyles with the bones of the forearm. Its shaft is usually twisted and presents ridges and tuberosities for the attachment of muscles. In many species the distal expanded end is perforated by epicondylar foramina. Foramina are characteristic of the humerus of most reptiles; an internal (entepicondylar) foramen is characteristic of some. These foramina are absent in crocodiles. The humerus of flying (carinate) birds is, like many of their bones, hollow and contains air. It is short and has, next to the head, a large tuberosity for the insertion of wing muscles. Epicondylar foramina are absent. In nonflying (ratite) birds, the skeleton of the wing is greatly reduced. The entepicondylar foramen is widely distributed among mammalian orders and seems to be linked, in part, with the prehensile limb, as is suggested by its presence in cats and absence in dogs and in the hoofed animals. The entepicondylar foramen transmits the median nerve and brachial artery. The supracondylar process sometimes found in man is a vestige of the boundary of the foramen.

The femur is essentially a cylindrical structure with expanded ends. At the proximal end, for articulation with the acetabulum, is the rounded head, and near it there are usually two elevations (trochanters) for the attachment of muscles. The occurrence of three trochanters is characteristic of certain mammals; *e.g.*, horse and rhinoceros. Distally, the femur expands into two condyles for articulation with the tibia. In many types there is an articular facet on the lateral surface for the head of the fibula.

The medial segment of the limb typically contains two bones, radius and ulna in the forelimb and tibia and fibula in the hindlimb. In the forelimb, the radius is anterior or preaxial, and the ulna postaxial in the adjustment of the limb for support and locomotion on land. The radius and ulna are fused in the tailless amphibians (*e.g.*, frogs) and are separate in those with tails (*e.g.*, salamanders), as they are in all of the reptiles. The ulna is the stouter bone in the bird's wing and supports the secondary feathers. In pronograde mammals the radius is fixed in pronation; *i.e.*, the forelimb is rotated so that the shaft of the radius crosses in front of that of the ulna. The radius transmits the weight of the forepart of the body to the forefeet; whereas it is the ulna that makes the elbow joint with the humerus, and into its proximal end are inserted the flexor and extensor muscles of the forelimb.

Tibia and fibula are separate in urodeles, united in tailless amphibians. In land reptiles the tibia articulates with both condyles of the femur and with the tritibiale of the ankle; the fibula articulates with the postaxial femoral condyle and with the tritibiale and fibulare. The tibia of birds is long, the fibula reduced. In mammals the fibula is generally reduced, and it may be fused with the tibia and excluded from the knee joint.

The distal segment of the limb comprises the carpus, metacarpus, and phalanges in the forelimb, and the tarsus, metatarsus, and phalanges in the hind limb. In a typical limb there are five digits (fingers or toes) which contain the phalanges.

The carpus and tarsus of the higher vertebrates are probably derived from a primitive type by the fusion or suppression of certain of its elements (*see* fig. 10). The bones of a generalized carpus (or tarsus), as that of a tortoise and certain other reptiles, are in three transverse rows: a proximal row of three, named radiale (or tibiale), intermedium, and ulnare (or fibulare); a distal row of five carpalia (or tarsalia), numbered one to five from the radial (or tibial) margin; and an intermediate row of one or two centralia.

In many of the urodele amphibians (*e.g.*, salamanders) the carpus is generalized. In the frogs and toads, however, it is more specialized, only six carpals being present, the third, fourth, and fifth carpalia having probably fused with either or both centralia. In birds the radiale and ulnare are distinct, but the distal bones are fused with the metacarpus to form a carpometacarpus. In mam-

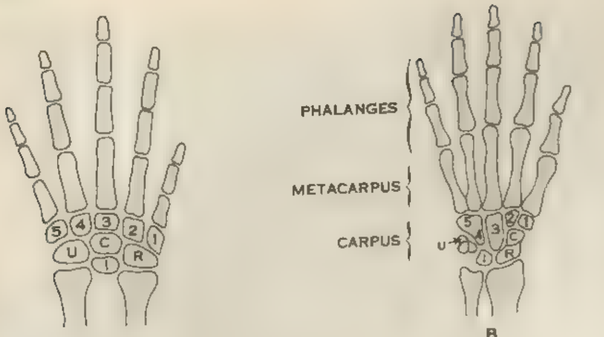


FIG. 10.—DISTAL SEGMENT OF FORELIMB SHOWING (A) PRIMITIVE CARPUS AND (B) HUMAN CARPUS

Key: R=radiale, I=intermedium, U=ulnare, C=centralis, 1–5=carpalia

mals various examples of fusion and suppression occur. In man the radiale forms the scaphoid; the intermedium, the lunate; the ulnare, the triquetral; the pisiform, an ulnar marginale, represents the remains of an extra digit, which may, however, be simply a sesamoid bone. The trapezium and trapezoid are carpalia 1 and 2; the capitate is derived from carpal 3, whereas carpalia 4 and 5 have fused to form the hamate. An os centrale is present in the carpus of many monkeys. In mammals the number of digits varies but the number of phalanges in each digit present usually corresponds with that of man. In some species, however, the phalanges are more numerous, as when the limb is modified to form a paddle (*e.g.*, in the whales).

The tarsus of urodele amphibians has the typical composition. In the frogs and toads the intermedium is absent; two long bones are identified as tibiale and fibulare. Among the reptiles there is much variation in the composition of the tarsus. Generally the hinge of the ankle is intratarsal, the row of tarsalia being distal to the hinge. In most living reptiles the tibiale and intermedium fuse to form the talus. In birds the ankle hinge is of the reptilian pattern in being intratarsal. The three tarsal cartilages of the embryo fuse to form the talus, which fuses with the tibia to form the tibiotarsus. The tarsalia fuse with the ends of the united metatarsals to make a tarsometatarsus. In the mammalian tarsus the talus is generally composed of the fused tibiale and intermedium, but in some a centrale is included to form a tritibiale. The ankle hinge is not intratarsal, but located between the bones of the leg and the first row of tarsal bones—tibia and talus usually.

Suppression of digits in hoofed mammals (*see* fig. 11) frequently has occurred in the following sequence: the pollex (first digit) is the first to go, then the minimus (fifth digit), index (second digit), and annularis (fourth digit). Thus, among the even-toed ungulates (artiodactyls; *e.g.*, pig and hippopotamus) the pollex has disappeared and the other four digits are present, although the second and fifth are much reduced. In the camel only the third and fourth digits persist and are of equal importance. Among the odd-toed ungulates (perissodactyls; *e.g.*, horse) the right digit is dominant and the others reduced to mere rudiments or splints.

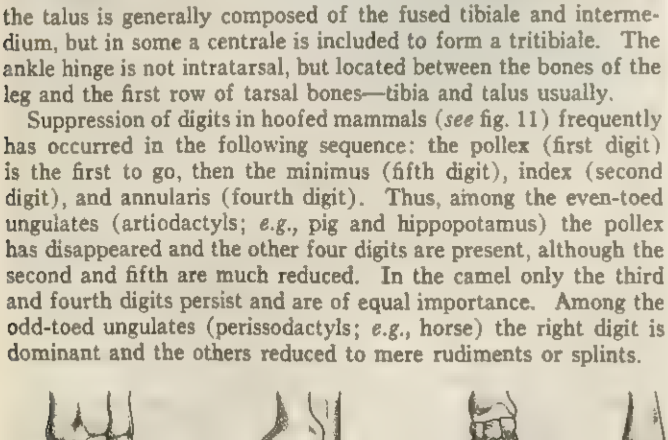


FIG. 11.—LEFT FOREFEET OF UNGULATES: (A) PIG, (B) RHINOCEROS, (C) CAMEL, (D) HORSE

See also BONE; CARTILAGE; CONNECTIVE AND SUPPORTING TISSUES; JOINTS AND LIGAMENTS; MUSCLE AND MUSCULAR SYSTEM; and references under "Skeleton" in the Index.

BIBLIOGRAPHY.—T. H. Bryce, "Osteology and Arthrology" in Sir R. Quain, *Elements of Anatomy* (1915); D. J. Cunningham, *Text-Book of Anatomy*, 9th ed. by J. C. Brash (1951); *Gray's Anatomy* 33rd ed. by D. V. Davies and F. Davies (1962); W. J. Hamilton *et al.* (ed.), *Textbook of Human Anatomy* (1956); L. H. Hyman, *Comparative Vertebrate Anatomy* (1942); D. P. Quiring, *Functional Anatomy of the Vertebrates* (1950); A. S. Romer, *The Vertebrate Body* 3rd ed. (1962); J. Z. Young, *The Life of Vertebrates* 2nd ed. (1962).

(WM. J. H.; X.)

SKELTON, JOHN (*c.* 1460–1529), the only English poet of major importance between Chaucer and Wyatt, whose reputation, long misunderstood and misrepresented, has been restored by modern scholarship. Nothing is known of his parentage or place of birth. He was educated at Cambridge, and later attained the status of "poet laureate" (*i.e.*, a degree in rhetoric) at Oxford (*c.* 1488), at Louvain (between 1488 and 1493), and at Cambridge (1493). William Caxton, in 1490, in the preface to *Eneydos*, paid tribute to him for translating classical authors, notably Cicero's letters and Diodorus Siculus, into "polished and ornate terms." From all this early work, only his translation of Diodorus has survived; when first published, in 1956, it proved to be a notable piece of Tudor ornate prose, marked by considerable linguistic inventiveness.

JOHN SKELTON, FRONTISPIECE TO THE GARLANDE OF LAURELL, PRINTED BY RICHARD FAUKES, 1523

Skelton's success as rhetorician and translator led to royal patronage. From *c.* 1489 he was associated with Henry VII's court at Westminster, first as a court poet and later, in addition, as "scolemaster" to the duke of York (later Henry VIII), and possibly to his elder brother Arthur. When Skelton took holy orders in 1498, Henry VII made an offering of 20*s.* "at master Skelton's mass." Skelton took his tutorial duties seriously. In 1501 he composed a moral treatise for his pupils, the *Speculum Principis*, long accounted "lost," and unprinted till 1934. To this period must also belong his (still undiscovered) *New Gramer in Englysshe*. His reputation by the end of the 15th century was such that Erasmus, visiting England in 1499, referred to him as "the incomparable light and glory of English letters," and addressed him in a Latin panegyric as "Skelton, immortal bard, worthy of highest fame"—impressive tributes, even discounting the fashionable Humanist adulation.

The office of schoolmaster ceased in 1502, when Arthur's death made Henry heir to the throne, and his household was reorganized. Soon afterward Skelton became rector of Diss, in the diocese of Norwich, where his name appears in the records from 1504. Up to this point he was better known as translator and scholar than as poet. Of his extant works, a handful of lyrics probably belongs to this period, a group of court poems certainly (including an elegy on Henry Percy, earl of Northumberland). His only notable poem from this time at court was his *Bowge of Courte* (1498). This is medieval in form, a dream allegory in rime royal (*q.v.*). In content it is an exposure of the disheartening experience of life at court, and is Skelton's first considerable satire.

Skelton remained rector of Diss till his death, but was resident in London from 1512. Tradition and legend, perpetuated in the anonymous jestbook *Merie Tales Made by Master Skelton* (1567), portray him while in Diss as a lively buffoon. Modern research has dispelled this image. He figures in the Norwich records as an orthodox and responsible cleric, trusted by his superiors. Further

confirmation of his status came in 1504, when Cambridge followed Oxford in conferring on him the right to wear the habit granted by Henry VII.

The years at Diss produced a group of minor satires and two longer poems which established him as a new and, to some, a disconcerting poet. Both *Phillip Sparow* (written 1508) and *Ware the Hauke* (1504–12) broke new ground. In them he began the series of poems in short rhyming lines based on natural speech-rhythms, to which the name "Skeltonic" has been given. The contents demonstrate both his orthodoxy and his audacity:

Was never byrde in cage
More gentle of corage
In doing his homage
Unto his soverayne.
Alas, I say agayne,
Death has departed us twayne!
The false cat hath thee slayne:
Farewell, Phyllyp, adew!
Oure Lorde thy soule reskew!
Farewell without restore,
Farewell for evermore!

Phillip Sparow is a lament for the death of a young lady's pet bird. It is also a mockery of the Office for the Dead, and the "Commendations" for the soul of the dead in the Office have become complimentary commendations on the young lady. *Ware the Hauke* is an angry attack on an irreverent hunting priest who had flown his hawk into Skelton's church. In these poems the poet shows that the range of the new Skeltonic can include both tenderness and vituperation.

The accession of Henry VIII in 1509 turned Skelton's mind from satire back to compliment. He wrote complimentary verses both in Latin and English. He presented the king with a copy of the *Speculum Principis*, written while he was Henry's schoolmaster. He added a volume of the *Chronique de Rains* (a French chronicle of the life of Richard Coeur de Lion), with a dedication (from "Skelton Loyall") in his own hand. A note to the former complained that he was a man completely forgotten. Henry eventually responded. By *c.* 1512 he had conferred a new title on his old tutor, that of Orator Regius; the patent conferring it was seen by a French scholar in the 18th century, but (like so much documentation for Skelton) is now lost. But most of Skelton's poems from this time forward carry the new title as well as that of "poet laureate": he was letting no one forget either of his honours.

Some time after his arrival in London he acquired a house (*tenementum*) within the sanctuary of Westminster. The Westminster Abbey Register Book notes that he was living there in 1518; and, since in 1512, shortly after he resettled in London, he wrote a Latin epitaph on Henry VII at the request of the abbot of Westminster, it is likely that his occupation of the *tenementum* began in that year. He died in Westminster on June 21, 1529, and was buried in St. Margaret's. This continuous residence within the abbey precincts is of importance for the understanding of some of the later satires in which he attacked Cardinal Wolsey. One of the most persistent "legends" concerning Skelton is that Wolsey in revenge drove him to take sanctuary. Anthony à Wood, in his *Athenae Oxonienses* (1691), adds picturesque details, including a pursuit by Wolsey's officers. Like the *Merie Tales*, this has proved to be fiction: Skelton was, understandably, unpopular for a time with the cardinal, but he made his peace, and at no time had to flee to the sanctuary in which he had lived for many years.

The London years were Skelton's most productive. As Orator Regius he considered he had full licence to praise the king, attack the king's enemies, entertain him, and on occasion admonish him. The defeat of the Scots (whom Skelton loathed) at Flodden produced one of the few poems printed in his lifetime, the *Ballade of the Scottyshe kynge* (1513). This savage satire was revised and sharpened the following year as *Against the Scottes*. Skelton turned his considerable powers of poetical invective against the king's enemies. In the same year (1514) he entertained the court with a series of "flyting" poems of mock abuse, *Against Garnesche*, written "Be the kynges most noble commandement," so continuing a court-poet's exercise that goes back to the troubadours.

In 1516 Skelton wrote his longest contribution as court poet, a "goodly interlude and a mery" entitled *Magnyfycence*. This play (there is no record of performance) retains the old form of the morality, with a cast of allegorical abstractions (*see* DRAMA: *Medieval Drama*). It is the first secular morality in English. It is political satire and admonition in a new shape. Magnyfycence represents the king, and the Vices and Virtues are the two factions among his counselors. The extravagant claims of Wolsey are covertly attacked. As a drama, it must be presumed a failure, since it has apparently never been produced. But some of the serious scenes (*e.g.*, the fall of Magnyfycence) are impressive; and as a creator of stage comedy Skelton was ahead of his time.

In 1517, probably as court entertainment, he wrote *The Tunnynge of Elynour Rummyng*, a brilliant portrayal of drunken women in an alehouse:

But let us turne playne
There we lefte agayne.
For, as yll a patch as that,
The hennes run in the mashfat;
For they go to roust
Straight over the ale joust,
And donge, when it commes,
In the ale tunnes.
Then Elynour taketh
The mashe bolle, and shaketh
The hennes donge away
And skommeth it into a tray
Whereas the yeest is,
With her maungy fystis:
And somtyme she blennes
The donge of her hennes
And the ale together;
And sayeth, Gossyp, come hyther,
This ale shall be thycker.

Though it was written in the tradition of the Gluttony scenes from *Piers Plowman*, and long remained in popular demand, it contributed largely to Skelton's later reputation as a "beastly" poet.

The years 1521–22 produced his three major political and clerical satires, *Speke, Parrot* (1521), *Colin Clout* (1522), and *Why Come Ye Nat to Court?* (1522). They are all poems of admonition to the king, directed against abuses in the church and the mounting power of Wolsey. In them Skelton formulates the position of the conservative clergy. He attacks the New Learning, rightly seeing in it the seeds of what was to be the Reformation. Wolsey's Humanism was as much a menace as his ambition. The church must be its own reformer. *Speke, Parrot*, one of the most allusive poems in English, demanding a knowledge of the learning and temper of the time and even of the day-by-day movement of events, returned to the medieval rime royal. *Colin Clout* and *Why Come Ye Nat to Court?*, in contrast, are directed toward their targets in vituperative Skeltonics, with precision of attack and deadly linguistic control.

Wolsey, however, was too strong an opponent. What reconciliation took place can only be a matter of inference. All of Skelton's major poems written later bear elaborate dedications to Wolsey. In his remaining years Skelton turned to other themes. It is significant that his next poem, the *Garlande of Laurell* (1523), is an elaborate *apologia pro vita sua*. Written while he was the honoured guest of the ladies of the Howard family, it contains (as a compliment to them) a group of lyrics Elizabethan in grace and felicity. The main theme of the poem (in rime royal) stems from Chaucerian dream allegory and *The Hous of Fame*. Dame Pallas and the Queen of Fame debate his entry to the Court of Fame, and Skelton stoutly defends his claim, listing his works, to hear in the end the assembled poets of antiquity and his English predecessors, Chaucer, Lydgate, and Gower, declare that he has triumphed. It is egotistical, jaunty, and completely successful.

His remaining poems, in Skeltonics, return to public issues. *The Duke of Albany* (1523) is an attack on the French and the Scots; *A Replycacion agaynst Certayne Yong Scolers* (1528), on Lutheran heresies within the church. In this latter work, the contrast between the sophisticated aureate prose of the preamble and the hard-hitting colloquial language of the Skeltonics is a reminder of Skelton's skill in Renaissance decorum at different levels.

Reputation.—Skelton fell rapidly out of fashion in the 16th century. His combination of religious orthodoxy in thought and

radical originality in presentation offended a predominantly Protestant England that looked to the Italy of the high Renaissance for its literary models. His use of colloquial speech-rhythm, with its echoes of Anglo-Saxon and alliterative Middle English verse, found no response among poets who preferred the sonnet and the canzone, and the "drumming decasyllabon." The emergence of the buffoon of the legend completed his eclipse as a poet who merited serious attention. Poets of the 20th century, reintroduced to speech-rhythm in verse by Gerald Manley Hopkins, have found in Skelton a source of new techniques. W. H. Auden and Robert Graves have both acknowledged their debt. Graves wrote a poem on him in Skeltonics, and has even claimed Skelton as "one of the three or four outstanding English poets."

Modern study of Skelton began with Alexander Dyce's edition in 1843. A group of studies by German scholars appeared toward the end of the 19th and in the early 20th century. Serious study began in England and the U.S. in the 1930s, and it is to a small group of scholars of that period that more precise knowledge of him is due. For a poet so original and so intensely personal as Skelton, much of whose poetry was written for a specific occasion, good critical work has had to wait till biographical facts and a firm chronology of the poems and the occasion of each were established, and is still hampered by lack of a full modern edition.

Bibliography.—*Bibliographies and Editions:* There are bibliographies, full up to dates of publication, in the books by W. Nelson, I. A. Gordon, and M. Pollet; and an excellent select bibliography in P. Green's (see *Biography and Criticism* below). The best edition of the poems, although inevitably dated, is still that by A. Dyce, 2 vol. (1843; U.S. reprints, 1855, 1856, 1887). *The Complete Poems*, ed. by P. Henderson (1931; 3rd ed., 1959), has a modernized text. There is a useful volume of selections, ed. by V. de Sola Pinto (1950). *Magnyfycence* was ed. by R. L. Ramsay for the Early English Text Society (1908; reprinted 1958). The prose works were first published as follows: *Speculum Principis*, ed. by F. M. Salter, in *Speculum*, vol. 9 (1934); *The Bibliotheca Historica of Diodorus Siculus*, ed. by F. M. Salter and H. L. R. Edwards, for the Early English Text Society (1956–57).

Biography and Criticism: F. Brie, "Skelton-Studien," in *Englische Studien*, vol. 37 (1907); W. Nelson, *John Skelton, Laureate* (1939); I. A. Gordon, *John Skelton, Poet Laureate* (1943); H. L. R. Edwards, *Skelton: the Life and Times of an Early Tudor Poet* (1949); M. Pollet, *John Skelton* (1962). Useful critical studies include W. H. Auden, "John Skelton," in *The Great Tudors*, ed. by K. Garvin (1935); E. M. Forster, in *Two Cheers for Democracy* (1951); J. Holloway, in *The Charted Mirror* (1960); P. Green, *John Skelton* (1960); A. R. Heiserman, *Skelton and Satire* (1961). (I. A. G.)

SKEPTICISM. Extreme skeptics deny that the human mind can attain knowledge. Total skeptics extend this denial to all kinds of claims to knowledge. But even the original Greek school of skeptics included few, if any, who were not partial or moderate: the former allowed that certain kinds of claims to knowledge were valid; the latter refrained from dogmatic denials of the possibility of knowledge and merely advised suspense of judgment. This restraint was more in harmony with the etymology of the word "skeptic," which is derived from the Greek word *skeptesthai*, "to examine." Dogmatic denials of the possibility of knowledge are always difficult to substantiate, and sometimes they seem to imply knowledge of the existence of the very things which are said to be unknowable.

Differences within the Greek school of professed skeptics have caused uncertainty about the classification of philosophers outside the school. Evidently a skeptic need not be dogmatic. But if a philosopher shows moderate skepticism it may not be a systematic doctrine, but perhaps only an attitude of mind. It may even be unconscious, and possibly the philosopher himself would repudiate the suggestion that he was a skeptic. For if partial skepticism is allowed it is by no means clear what kind of claims to knowledge must be admitted by everyone who is not a skeptic. Consequently the title "skeptic" has no agreed application, and its use reveals as much about the person who applies it as about the person to whom it is applied. For instance, skepticism about ordinary claims to knowledge is quite a different thing from skepticism about the claims of speculative philosophers, so that ordinary people and speculative philosophers are likely to disagree in their diagnoses of skepticism. Also, the same people who apply the word pejoratively to the rejection of their own claims to knowledge at other times may use it as a term of praise to point the contrast with credulity or superstition. However, most systematic skeptics, who stand at the centre of these extensions of their title, do have one tendency in common: they are reluctant to allow claims to knowledge about anything beyond the claimant's own immediate reactions; and consequently systematic skepticism has often been closely associated with some radical form of empiricism.

Greek skepticism passed through three stages. The school was founded by Pyrrho toward the end of the 4th century B.C. Next, in the 3rd and 2nd centuries, its ideas were absorbed, extended and modified by the philosophers of the Middle Academy and of the New Academy. Finally, it was revived as an independent school by Aenesidemus in the 1st century B.C. and carried on by Menodotus, Sextus Empiricus and others in the next two centuries.

Pyrrho's chief interest was ethics. He taught that, since knowledge is unattainable, the only way to achieve happiness is to practise suspense of judgment. He appears to have formulated this moral precept without any reservations. So he was neither a total nor an extreme skeptic. Perhaps it was the definiteness and simplicity of his moral teaching which enabled his follower Timon to claim that his philosophy did not conflict with common sense. Pyrrho's skepticism was directed chiefly against the dogmatic speculative theories on which the Epicureans and Stoics based their moral doctrines. However, skeptics who begin by attacking the recondite theories of philosophers often end by colliding with common sense, and Pyrrho's skepticism was not confined to philosophers' claims to knowledge.

The skeptical teaching of Pyrrho's school was taken over by Arcesilaus, the founder of the Middle Academy, and worked out in a more systematic way. This development was continued by Carneades, who founded the New Academy. Both were interested in epistemology for its own sake, unlike Pyrrho, who took it up because of its bearing on ethics. They produced a sustained and thorough criticism of the theory of knowledge which had been elaborated by the Stoics. But their work was not entirely negative. For though Carneades denied the possibility of knowledge, he offered a substitute: he admitted that judgments may have different degrees of probability.

The new skeptical school founded by Aenesidemus owed much to the work of the Middle and New Academies. It was systematic and self-conscious and classified the arguments which it inherited under ten headings (*tropoi*). Aenesidemus also took over the criticism of the powers of reason, which had been begun by his predecessors, and carried it much further. This inevitably raised a methodological problem: if reason is unreliable, the philosophical arguments which support this conclusion are themselves unreliable. This difficulty was usually evaded by a simile: the skeptic's arguments would be compared, perhaps, to fire, which consumes both its fuel and itself. Such comparisons are hardly adequate to remove the difficulty; what was needed was some way of distinguishing, as Kant did, between legitimate and illegitimate uses of reason. But the lack of any such distinction did not halt the attack on reason. Deductive reasoning was declared invalid, and inductive reasoning was submitted to a criticism which on many points anticipated Hume's views on causation. The conclusions of inductive arguments were rejected if they could not in principle be checked by observation. Even those which merely moved from phenomena to phenomena were said to yield something less than certainty. But at this point skepticism was tempered by the school's interest in medicine. This they regarded not as a science but as an art (*techne*), based on repeated and careful observation of the phenomena, never claiming absolute certainty and never postulating unobservable causes. In an instructive passage Sextus commends the grammar which children learn but condemns abstruse theories about the nature of the letters and their origin. Evidently his views on empirical matters were practical, cautious and positivistic, but not obviously skeptical. However, some philosophers would detect skepticism even here, on the ground that inductive arguments involve more than Sextus allows.

Outside the school of professed skeptics the attribution of the title is highly debatable, except in the case of the Middle and New Academies, which stood consciously within the skeptical tradition. But when the historian of ideas considers philosophers outside this

continuous Greek tradition, it is as if he had moved out of harbour into the open sea. However, it is just this move which is most likely to lead to some understanding of the nature of skepticism.

Michel de Montaigne is often classified as a skeptic because, in his *Essais,* he discussed every question in a detached and critical way. He certainly was not an extreme skeptic. But almost all philosophers have something of Montaigne's spirit. So, if Montaigne counts as a moderate skeptic, as no doubt he should, few philosophers have entirely lacked moderate skepticism, even in the Middle Ages, which constituted in some ways the least skeptical of all periods in the history of thought.

However, the rather more extreme forms of skepticism are perhaps worth diagnosing. But here too a great range of variation is encountered. Skepticism about a particular human faculty can be used by two different philosophers to support two entirely different conclusions. This is not altogether surprising, since the same doubt can be associated and contrasted with different certitudes. Plato impugned the reliability of the senses and claimed that the intellect could apprehend a transcendent world of "forms." Sextus rejected all speculation about transcendent reality, took over the Greek corpus of arguments against the senses and accepted medicine as something less than science. Sextus' combination of skepticism about the senses with empiricism is not as inconsistent as it might appear to be. For it is possible to try to allow for sources of error in perception and so to claim that it provides an adequate basis for probable conclusions.

Even Berkeley (*Principles of Human Knowledge,* 1710), who maintained that perception yields information only about the sensations of the perceiver, sincerely believed that this did not make him a skeptic. In this he was not entirely unjustified. For, if sensations are the only direct objects of perception (which not everyone would admit), it is not clear how the existence of anything beyond them could be validly inferred. Moreover, Berkeley claimed that material objects were merely collections of sensations in the minds of human beings and God, and this theory might well seem to be less skeptical than any theory which treats sensations as clues to the existence and nature of objects which are never directly perceived.

Berkeley's investigation of perception raises two connected questions about the nature of skepticism. When a philosopher says, not that a certain kind of thing exists but is unknowable, but that it does not exist, is he a skeptic? If, for instance, God's existence is denied, there is no doubt that this is true skepticism. But if a philosopher can plausibly claim that ordinary empirical judgments, when they are correctly construed, do not commit the speaker to belief in the existence of a certain type of thing, then the denial that these things exist is not obviously skeptical. This raises the second question, which is a question about the relation between ordinary judgments and philosophical doctrines: how is it possible to decide whether a philosopher's account of what ordinary judgments imply is correct? Many would say that Berkeley does not give a correct account of what ordinary perceptual judgments imply. But it is not easy to adjudicate in cases like this, if only because ordinary people do not often consider such matters deeply. Yet, until this question is decided, it is impossible to say whether a philosophical doctrine like Berkeley's is skeptical or not; unless, of course, the analysis of ordinary judgments is abandoned and the question is decided by appeal to some speculative theory.

Skepticism about the powers of reason can also lead in diverse directions. It led Sextus to pare down all ambitious claims to knowledge, leaving only the probable results of strictly empirical inquiries, shorn of all theoretical superstructure. It took Pascal, in his *Pensées,* to an entirely different conclusion: the impotence of reason seemed to him to leave all science open to doubt and to show up by contrast the reliability of religious faith. The difference between the sciences which these two had in mind probably explains their divergent estimates of the bearing of skepticism about reason on the claims of scientists. Clinical medicine is largely concerned with symptoms, probabilities and happy diagnoses, whereas applied mathematics is rigorous and theoretical.

Hume, who was also a skeptic about the powers of reason, gave a much more subtle account of science (*Treatise of Human Nature,* 1739; and *Dialogues Concerning Natural Religion,* 1779). Like Sextus, he condemned all theories which postulate unobservable causes behind the phenomena. Like Sextus too, he believed that reason, working on the results of observation, could not possibly establish any contingent general statements. For reason can establish only a priori statements (on this point he was more liberal than Sextus). Contingent general statements are believed not on rational grounds but only because continued observation of a conjunction of events has produced a habit of expectation. Such expectations are based on feeling and not on reason, and no justification of feeling is possible. It is merely an ultimate fact that human beings have this feeling. So there are two central propositions in Hume's philosophy: matters of fact cannot be established purely by a priori reasoning; they can be established only in a precarious way by nonrational inference; or else, perhaps, by direct observation. But even direct observation was not left unscathed by Hume. By an elaborate argument he sought to prove that no rational account of perceptual judgments is possible, and here too he fell back on feeling. Human beings, he said, are so constituted that they must believe in the existence of material objects beyond their sensations even though they can give no rational account of this belief.

Hume's account of inference inevitably seems skeptical to speculative philosophers. For it confines a posteriori inference to phenomena, and it declares a priori inference incapable of yielding new factual conclusions. However, when it is judged at the bar of common sense, it is not so easy to see how it should fare. Hume himself often said that, judged by these standards, it was skeptical, but it is not clear that he was right. Admittedly his refusal to allow that a posteriori inference is rational conflicts with ordinary ideas and linguistic usage. But on the other hand ordinary people might well rest fairly content with his account of the evidence which supports such inferences. For contingent general statements are based ultimately on the observed conjunctions of events, and it is not obvious that they assert anything more than that these conjunctions have always held and will always hold. No doubt the theory needs qualifying and modifying in various ways if it is to fit the facts exactly. But it is doubtful if ordinary people would agree with the much larger complaint of Hume's philosophical critics, that his theory altogether fails to account for the necessity of contingent general statements. However, Hume's theory of inference, like Berkeley's theory of perception, goes far beyond the point to which ordinary people carry the analysis of their judgments. Consequently it is not easy to estimate their reactions to the theory, or, in particular, whether they would call it skeptical.

However, it is possible to detect a pattern in the development of such difficult cases and so to understand why they are difficult. The ordinary person makes judgments which he does not analyze. Some philosophers hold that these judgments commit him to hypotheses which transcend the available evidence. Then other philosophers, allegedly skeptics, react by saying that they commit him to nothing beyond the bare phenomena. The controversy proceeds by a kind of diastole and systole. Perhaps neither the expanding nor the reducing theories are correct, but rather something intermediate. In any case this pattern of philosophical controversy explains why it is so difficult to decide whether reducing theories like Hume's are, from the point of view of the ordinary person, skeptical.

Kant thought that Hume's theory was skeptical, and he tried, in his *Critique of Pure Reason,* to work out a theory of knowledge which would be acceptable to an empiricist and yet free from what he thought was skepticism. At the same time he wished to curtail the speculative flights of reason. What he had in mind was not so much the speculative hypotheses which philosophers sometimes reach in the course of analyzing ordinary judgments, but rather the elaborate deductive systems of the metaphysicians of the 17th century. He called his own philosophy critical, to distinguish it both from what he regarded as skepticism and from what he regarded as dogmatism. This word "critical" is close to the etymology of the word "skeptic," in fact closer than the popular connotation of skeptic. The whole field would be easier to survey if Kant's distinction between criticism and skepticism had always been ob-

served. As it is, the word "skeptic" usually covers both. However, even Kant's use of the word is not consistent. It can mean one who denies that something exists, or one who allows that it exists but maintains that its nature cannot be known, or one who claims that a whole area of so-called knowledge is vitiated by inconsistencies. Hume's theory of perception belongs to this third type.

Kant believed that human reason imposes certain categories on the raw material of experience which somehow comes from things in themselves. Phenomena are the result of this operation. Causality is one of the categories. So, if a philosopher doubted the validity of causal inferences, as Kant thought that Hume did, he could be answered by the contention that causality is already an ingredient in phenomena as we know them. This illustrates the positive part of Kant's philosophy: reason is rehabilitated in the sphere of phenomena. Kant's negative teaching was that reason cannot extend its operation beyond phenomena. For instance, any attempt to use the category of causality to reach some speculative conclusion would inevitably lead to contradictions; the apparatus of reason, having no material to work on, would cease to function properly.

In spite of Kant's intention there is an element of skepticism in his system. Ostensibly it is critical and only seeks the limits of the powers of reason. But it gives these limits in terms which imply awareness of the existence of things in themselves. It is hard to see how Kant can consistently allow this awareness of the existence of things in themselves. However, since he does allow it and since at the same time he denies the possibility of any knowledge of the nature of things in themselves, he is on this point a skeptic. On the other hand, he was not in the least skeptical about the possibility of knowledge of what he called phenomena.

Kant used the term "phenomenon" in a neutral way that did not imply an unfavourable contrast with things in themselves. Most of his idealist followers thought that this was an inconsistency and rejected the hypothesis of things in themselves. Some rejected even the claim that phenomena are knowable, supporting this rejection with arguments very like those which Kant had used to show that things in themselves are unknowable. They tried to establish that there are latent contradictions even in quite ordinary claims to knowledge. This kind of dialectical argument, which they inherited from Kant and adapted to a more subversive use, had existed in antiquity and had often been used to undermine the beliefs of ordinary people in just this way. Hume's treatment of perception is an example of this kind of reasoning, but it is as old as Zeno. It produces a special kind of skepticism, which is conspicuous in the writings of 19th-century idealists and was taken to the extreme limit by F. H. Bradley (*Appearance and Reality*, 1893), who found that all claims to knowledge are self-contradictory.

This kind of skeptic does not deny that things of a certain kind can be known to exist; nor does he admit that they exist but deny that their nature can be known: instead he maintains that parts, or perhaps even the whole, of what is alleged to be knowledge cannot be knowledge since it contains hidden contradictions. The first two of these three kinds of philosophical skepticism rely on analogies with ordinary situations in which people refrain from claims to knowledge, perhaps misleading analogies, but at least intelligible ones. This third kind of skepticism is more difficult to understand. A pragmatist would ask how anything which works so well as science could possibly be self-contradictory. An idealist such as Bradley would reply by introducing the doctrine of degrees of reality. What the scientist discovers cannot be perfectly real, since it contains contradictions, but nevertheless it is not sheer illusion, since successes can be achieved even on lower levels of reality. But the scientist must appreciate that he is on a lower level and must not make any claims about ultimate reality, which is the domain of the metaphysician. This kind of system is, from the point of view of the ordinary person, skeptical, and skeptical in a very peculiar way. But the philosophers who build such systems would say that they give every claim to knowledge its due and would confine the title "skeptic" to those who reject transcendent reality.

These speculative adventures inevitably produced reactions in the 19th century. Both pragmatists and positivists defended science and common sense against these dialectical attacks and professed themselves at least agnostic about transcendent reality. However, their defense of science was often associated with very reductive theories, similar to Berkeley's and Hume's, and this controversy provides yet another example of the perennial pattern of extreme contraction alternating with extreme expansion. Also agnosticism about transcendent reality is a difficult position to maintain because, if it is really true that ordinary judgments commit people to belief in transcendent reality, agnosticism about transcendent reality is strictly inconsistent with the claim to understand and accept ordinary judgments.

These two weaknesses in the empiricist reaction were not irremediable. The pragmatists investigated the meaning of scientific statements in a sympathetic way and claimed that they contained no contradictions and did not point to any transcendent reality. They also investigated the meaning of other kinds of empirical statements and tried to give a comprehensive criterion of meaningfulness. Their work was carried on by the logico-analytic school in the 20th century; and, though it sometimes produced reductive theories, this did not always happen. When it did happen, it was not because there is any good reason why theories about the meaning of empirical statements should be reductive, but rather because the logico-analytic school inherited the tradition of Berkeley and Hume and because in any case extreme reaction is natural.

G. E. Moore, Bertrand Russell and Ludwig Wittgenstein were the three most important philosophers of this school. Its attitude to speculative philosophy was critical rather than skeptical because when its members rejected speculative theses, it was on the ground not that they were false but that they were meaningless, or at least involved misuse of language. However, if the idealist attack on empirical statements is to be classified as a kind of skepticism, then it might be correct to call the logico-analytic philosophers skeptics because of their attack on speculative philosophy. For, though the objects of the two attacks were different, their methods, the detection of contradictions and the diagnosis of meaninglessness, were very similar. But, if the word "skeptic" is used pejoratively, it is important to remember that there is a certain presumption in favour of common sense. The logico-analytic school did not merely rest on this presumption: it tried to show that both the apparent contradictions in ordinary empirical statements and the apparent meaningfulness of speculative theses are the results of illusions; *cf.* particularly Moore's "Defence of Common Sense," in *Contemporary British Philosophy*, 2nd series (1925); and Russell's *Mysticism and Logic* (1918).

Two different tendencies can be detected in the writers of this school. Some of its adherents tried to give a single criterion of meaningfulness which would cover all ordinary statements and exclude those of speculative philosophers. In this way they hoped to evade the difficulty which Kant had failed to evade, the difficulty of limiting knowledge without implying some awareness of what lies beyond the limit. For if statements about what lies beyond the limit could be shown to be meaningless, the inconsistency would be avoided. This systematic attempt to delimit the sphere of meaning was made in Wittgenstein's *Tractatus Logico-Philosophicus* (1921), which was closely related to Russell's work in logic. Its weakness was that it set up scientific statements as the type to which all meaningful statements must conform. This places it in the tradition of the 19th-century pragmatists and positivists. In fact some of the philosophers who tried to work out these ideas of Wittgenstein in more detail called themselves logical positivists. The other tendency within the school was a reaction against any attempt to give a systematic account of meaning. Every type of ordinary statement should be examined separately and judged by its own standards. This tendency begins in the writings of G. E. Moore. It also appears in the later writings of Wittgenstein, who added the corollary that, if speculative philosophers are wrong in thinking that their statements mean something, this too should be established by a detailed examination of each individual case. This tendency, judged by ordinary standards, was obviously not skeptical; it was a vindication of common sense in which many speculative philosophers would detect credulity.

There are many reasons why the diagnosis of philosophical skepticism will always be debatable. Greek skepticism had many fac-

ets, and outside the Greek school different philosophers show different combinations of these facets. However, when the word "skeptic" is used about the rejection of particular claims to knowledge on ordinary grounds the difficulties are less formidable. In such cases both the claimant and the rejector will usually be appealing to similar standards: they will differ in their application of these standards; and they may also differ about the force of the word "skeptic," whether it is pejorative, commendatory or neutral. These complications need to be unraveled, but the task of unraveling them is comparatively easy. But when philosophers question whole departments of knowledge, they are really questioning the standards which are ordinarily taken to be reliable within these departments. At the same time they frequently offer different, stricter standards and sometimes make strange new claims to knowledge of their own. This enormously widens the scope of the debate and makes it very hard to decide which philosophers are skeptics, which dogmatists and which merely critical. It might appear that this question could be safely settled only by a philosopher who adopted the standpoint of common sense. But even here the difficulties do not end. In fact the greatest difficulty of all is to understand the intricate relations between ordinary beliefs and philosophical doctrines. For few philosophers entirely abandon ordinary beliefs. Even those who do not analyze them sympathetically usually recast them and let them stand in an altered form. And there is no simple way of deciding when justice has been done.

For theological skepticism *see* AGNOSTICISM.

See also references under "Skepticism" in the Index.

BIBLIOGRAPHY.—On the Greek skeptics *see* E. Bevan, *Stoics and Sceptics* (1913); M. M. Patrick, *The Greek Sceptics* (1929). *See* further the bibliographies of the separate articles in this encyclopaedia on the philosophers mentioned above. (D. F. P.)

SKIING, a way of moving over snow wearing a pair of long flat runners called skis attached to the shoes or boots. It is practised for recreational, competitive, or utilitarian purposes. On the level or on slight grades, the skier uses a gliding gait; downhill he slides effortlessly over the snow, turning to avoid obstacles, to slacken speed, or to change direction; special steps are used in climbing steeper slopes. Ski jumping is usually performed on artificial inclines of varying size especially designed for the sport.

Skis were first used for utilitarian purposes, for transportation, hunting, and in warfare, as they are still. Recreational skiing was a natural outgrowth of such uses. In the Americas and the Alps the sport developed chiefly as downhill skiing.

At first, only the mountain railways and aerial cable cars, or funiculars, of the Alps were available for the uphill transportation of skiers. In the early 1930s, many new devices were designed and built for this purpose, including chair lifts, numerous types of bar lifts such as the T-bar tramway, gondola lifts, and in the U.S. and Canada, rope tows. These greatly promoted the popularity of skiing, for they made possible four or five times as much skiing in a day as when the skier had to climb uphill; and they made the sport more enjoyable for all participants. In the Scandinavian countries and Eastern Europe touring and cross-country skiing developed as the more popular sports.

Competitive skiing comprises the Nordic (or classic) events—cross-country racing and jumping; and the Alpine (or downhill) events—downhill (or straight) racing, slalom (zigzag downhill) racing, and giant slalom (see *Competitive Skiing*, below).

Following a description of the equipment used in skiing, this article discusses the history of skiing and the development of modern Alpine skiing, competitive skiing, and the organization of the sport; selected skiing terms are given in the glossary. For Olympic championships *see* OLYMPIC GAMES.

EQUIPMENT

The Ski.—The three chief types of skis, downhill, jumping, and cross-country, have certain characteristics in common. They are pointed, turned up, and usually slightly wider at the front (the tips or shovel) and are squared at the rear (or heel). They are thickest in the midsection, under the foot, thinnest just before the ends, and are built with a slight arch, or camber, so as to distribute the skier's weight evenly along the length and thus

BY COURTESY OF (ABOVE) THE TRUSTEES OF THE BRITISH MUSEUM; (LEFT) THE NORWEGIAN SKI MUSEUM

(Above) Lapp hunters on skis, from a map of 1539 shown in the "Historia" of Olaus Magnus, 1555; (left) the earliest known pictorial representation of skiing, a rock carving dating from 2000 B.C. from Rødøy, Norway

impress a maximum running surface, with maximum edge-length, on the snow. Over the years skis have been refined in their shape and camber and in precision of stiffness.

Skis vary in length, the downhill ski for the average adult man being about 7 ft. (213 cm.) long, although much shorter skis, slower but more maneuverable, have enjoyed some vogue. They were formerly fashioned from one piece of wood, usually hickory, but from the 1930s laminated constructions also were used. After 1950, plastic running surfaces were introduced to increase speed and durability, and skis of metal, usually with a wood or plastic core, became increasingly popular. The downhill ski, which is about 3 in. (7.6 cm.) in width, typically has a shallow groove running longitudinally along the centre of the bottom (or sole); this imparts directional stability. In addition, skis have sharp edges of steel along the undersurface, in order to bite into hard snow or ice. Jumping skis are longer (about 8 ft. [244 cm.]), wider, thicker, and heavier and ordinarily have three grooves in the bottom. They have no steel edges. Cross-country skis are narrower and lighter in weight than downhill skis and often are laminated. They have one groove but no metal edges.

Other Equipment.—For downhill skiing, the skier wears close-fitting, heavy leather boots with flat and stiff soles in order to exercise precise control over the skis. The boots are firmly attached to the skis by bindings, often with safety or release bindings, developed to free the skier's foot in case of falls. For jumping and cross-country a lighter and more flexible boot is used, with a binding which allows the heel to be raised.

For downhill skiing a light pole, or stick, of metal or bamboo, about four feet long, is carried in each hand. For cross-country skiing the pole varies in length. These aid the skier in pushing himself along on the level, and in climbing. They also are used in maintaining balance when running downhill and to assist in turning. Each stick has a wrist strap at the top and a ring, or wheel, near the bottom, which prevents the point from sinking too deeply into the snow. Various waxes, depending on temperature and snow conditions, are used on the bottoms of skis to prevent snow from sticking to them and to reduce friction. To aid in climbing, strips of sealskin or plush are sometimes strapped to the bottoms of the skis to grip the snow.

The refinements in ski equipment after World War II were accompanied by a great increase in the proficiency of the average recreational or competitive skier. As the parallel (skidded) technique supplanted the earlier stemmed (steered) turns, speed and maneuverability were greatly increased. (R. Po.)

HISTORY

The oldest ski that has been found, in a bog at Hoting in Sweden, is believed to be at least 4,500 years old, and the oldest pictorial representation dates from about 2000 B.C. Some sort of skis or ski-like snowshoes were probably used in many regions where men found them useful to travel or hunt in snow-covered country. Procopius, the Byzantine historian, *c.* A.D. 550, mentions "Skridfinns," that is, gliding Finns, in apparent contrast to the Finns who did not glide.

Military Skiing.—Skis were early used for military purposes; according to Norwegian historians, the Vikings used skis on raids in Norway in the 10th and 11th centuries. One record shows that in 1200 King Sverre of Norway sent out men on skis to reconnoitre before battle. In the 15th, 16th, and 17th centuries skis were used in warfare in Norway, and in Russia, Sweden, Finland, and Poland. Subsequently, skis have been used in warfare in many places, as along the Austrian-Italian front running through the Alps in World War I. During the Russian invasion of Finland in 1939–40 much use was made of ski-equipped troops.

Ski Mountaineering.—Skis have been used in mountaineering since the 1880s, in Norway, and in the Alps since the mid-1890s. The first important expedition in central Europe was the incomplete traverse of the main massif of the Bernese Oberland by a German party, including W. Paulcke. From the beginning of the 20th century the French, Italians, and British also took part.

Development of Skiing as a Sport.—There are records of military ski competitions in Norway in 1767 and civilian competitions at Tromsö in 1843. But skiing as a competitive sport really began about 1860 when Sondre Nordheim, from Morgedal in Telemark, invented bindings around the heels fastening the boots to the skis, thus making jumps and turns possible. Competitive skiing also started in California in the 1860s, but because there the skiers had only toestrap bindings it was impossible for them to make turns and they ran only straight downhill courses.

The first big ski contest was held at Huseby Hill, Oslo, in 1879, with 10,000 spectators in the presence of the king. This contest, held yearly, was moved to the Holmenkoll Hill in 1892. In 1888 Arctic explorer Fridtjof Nansen led a party across Greenland on skis, dragging sledges. His book, *Paa Ski Over Grönland*, was translated into English, German, and French in 1891 and exerted a great influence on the development of skiing in central Europe, particularly in the Alps and the Black Forest of Germany. The first competitions were in jumping and slalom without gates—slalom gates were invented by the British in 1922.

As the sport spread, both proficiency and equipment developed rapidly. Because the terrain in the Alps was steeper and more difficult, a new skiing technique was needed. The erect, Norwegian Telemark and Christiania style was gradually supplanted by an Alpine way of skiing downhill and turning. (J. VA.)

MODERN, ALPINE SKIING

As Alpine skiing developed, the Nordic manner of skiing was modified and eventually basically changed to allow the skier to cope with the long, steep runs in the high mountains, and the obstruction-filled glaciers and forested hillsides. Outgrowths of these changes were the ski school where the new method was taught, and the downhill and slalom competitions that tested the relevant skills.

A pioneer in adapting skiing technique and ski bindings to the requirements of Alpine terrain was Mathias Zdarsky of Austria. He was probably the first regular ski teacher in that country, having registered over 1,000 pupils of his "Lilienfeld technique" as early as 1904. Later he extended his courses to Germany and Switzerland.

Viktor Sohm of Bregenz, another Austrian of the 1890s who conducted courses, adhered to the Norwegian style. A professional ski teacher was first heard of in 1906, Reinhard Spielman, who taught on the Semmering, near Vienna.

Col. Georg Bilgeri of Austria, W. Paulcke and Henry Hoek of Germany, Col. Christopher Islin of Switzerland, and Sir Arnold Lunn and Vivien Caulfeild of Great Britain should also be mentioned as pioneering contributors to modern skiing.

Ski Schools.—To Hannes Schneider is attributed the credit for laying the foundation both of the modern ski technique and the thoroughly systematized ski school. He started teaching in St. Anton in Austria in 1907, developing the crouched position and the stem Christiania, as well as the position of "vorlage," or putting the weight forward. He taught in a methodical progression from snowplow, through snowplow turn, stem turn, and stem Christiania, to parallel swing initiated with only a slight, quick stem. He soon found graded classes and additional instructors necessary. During World War I, he further developed his methods while training Austrian mountain troops.

Since World War I instruction has been systematized, and, as more efficient techniques have been developed, largely in competition, these have been taught. Concurrently, more and more people attended the ski schools, and the sport spread to all parts of the world where there was snow, for the satisfaction of advancing in skill, and the ability to ski faster and in control add greatly to the pleasure derived. The first ski teachers examination was held in 1925 in Austria. This has become increasingly exacting over the years, and other countries, including Switzerland, France, Italy, Germany, and the United States, have adopted similar tests, with certification considered a mark of professional competence. By 1930 there were several well-organized ski schools in Switzerland, and shortly thereafter courses for teachers were instituted. Under the leadership of Christian Rubi, the Swiss Ski Schools promoted a uniform method of instruction at the various important resorts.

While Schneider's Arlberg ski school put great emphasis on the stem and stem turns as the basis of its technique, the Swiss avoided what they considered excessive attention to these positions, and later the French schools, too, took a more direct route to "parallel" skiing. Émile Allais and Paul Gignoux brought the French theories to the forefront when they declared war on the basic stem and rotation. Ahead of their time were Eugen Matthias, physiologist of the University of Munich, and Giovanni Testa, head of the ski school in St. Moritz, who first taught the reverse (or inside) shoulder technique and turning with knees together in the mid-1930s. Their book, *Natürliches Skilaufen* (1936), attracted considerable attention, but was not at the time recognized as prophetic of what was to come.

Since World War II, the techniques taught and the sequence of instruction in the ski schools of the various countries have tended to become more similar. In particular, less emphasis is placed on a thorough and almost exclusive grounding in the stemmed turns. Instead, a parallel position of the skis is inculcated; and to make this habitual, early stress is placed on sideslipping, traversing, edge control and straight running of slopes. The chief differences are between the Austrian system of rhythmic short-swinging (*wedeln*) turns, distinguished by counterrotation and leading with the inside shoulder in turning; and the French turning style of *projection circulaire*, in which the shoulders are kept almost square to the direction of movement during a turn.

Among the important factors leading to these developments were the introduction of the steel edge, in the 1930s, and the great increase in the number of ski lifts, beginning in the late 1940s. The former formed the basis for parallel, narrow-tracked, and precise running and turning; the latter—which greatly increased the numbers of skiers and multiplied the number of runs each one could make downhill, with the consequence that nearly all skiing must be done on well-broken and well-worn trails rather than on virgin snow—made necessary the development of sideslipping and short-swinging (short-turning) techniques.

The Basic Movements.—Brief identifications of the basic skiing movements referred to above and commonly taught in most schools follow. The best way to learn these movements is with a certified instructor.

Stemmed Turns.—Formerly the focus of instruction, these are based on placing one or both skis at an angle to the direction of movement. The chief turning force is shifting of weight. The snow-plow turn is made from a snowplow, or double stem (skis in a "V"), position by shifting the weight to one ski and exerting outward-turning pressure with that heel to make a steered turn.

The stem turn is made from the traversing position (with skis parallel) by stemming (angling out) the upper ski and turning as in the snowplow. The stem Christiania is similar except for the manner of shifting the weight and moving the body, while the Christiania is a swing from a slight initial stem that continues across the fall-line. The pure Christiania is a series of sweeping swings back and forth across the fall-line.

Sideslipping.—This maneuver became necessary to avoid ruts and bumps as hard-packed trails became common, and as it also inculcates the parallel ski position and teaches edge control, has become the basis of modern instruction. The skis are skidded sideways by flattening the skis to disengage the edges and pushing the heels downhill and the skis out from under the body.

Traversing.—This is running downhill diagonally across the fall-line with the skis parallel, and close together for greatest control.

Straight Running.—In the downhill position for running in the fall-line, the skis are flat, parallel, close together, and equally weighted. The goal is a graceful, relaxed position, slightly forward-leaning (vorlage), with smooth movements to adjust the skier's centre of gravity to changing conditions of snow and slope.

Short-Swinging and Wedeln.—These consist of consecutive short, parallel swings across the fall-line, one after the other. In the short swing there is alternating pole planting; in wedeln there is practically no edging. The legs and hips are used a great deal in the sequence of swing and counterswing.

COMPETITIVE AND RECREATIONAL SKIERS

Three Olympic gold medal winners: (top) Finnish ski-jumper Veikko Kankkonen leans far forward to reduce wind resistance and gain lift; (centre) Marielle Goitschel of France negotiates gates during woman's giant slalom race; (left) Austrian skier Egon Zimmermann leaps during men's downhill race; (right) recreational skiers on a slope at St. Moritz, Switz.

(TOP) CENTRAL PRESS PHOTOS (BOTTOM LEFT) PICTORIAL PARADE, (RIGHT) HANS STEINER—PHOTO RESEARCHERS, INC.

Safety and Courtesy.—With large numbers of new skiers appearing on the slopes each year, a code of safety and courtesy has been suggested for the protection both of the skier himself and of the other persons on the hill.

"Ski in Control"—skiing in such a way that the skier at all times can turn or stop—is the basic objective of the following "rules of the road" that are followed by most experienced skiers.

1. Skiers approaching each other on opposite traverses pass to the right.
2. An overtaking skier warns the skier ahead with the cry, "On your right," or "On your left." The call, "Track!" is a warning that a fast skier is coming down a trail or course, or a loose ski or other object is sliding downhill.
3. An uphill skier is responsible for avoiding a skier below him.
4. A skier entering a trail yields right-of-way to skiers already on the trail.
5. A moving skier avoids a stationary skier, and passes at an ample distance.
6. A skier does not stop at any spot that will obstruct a trail, or impede the passage of following skiers, or where he is not plainly visible from a safe distance above. A skier who has fallen in such a spot moves to a safer location as soon as possible.

Ski patrols composed of volunteers or professionals who are specially trained for first aid and rescue work are available at all times in most ski areas to assist skiers who are lost, ill, or injured, or otherwise are in need of help.

COMPETITIVE SKIING

Downhill races are run on defined courses of a length, steepness, and difficulty appropriate to the skill and endurance of the entrants. For men's Olympic and world's championship events, a vertical descent of at least 2,625 ft. (800 m.) is prescribed, with sufficient steep and difficult terrain. Such courses may be between 1½ and 3 mi. long (2.4 and 5 km.), although some races are considerably longer. Average speed of the winners is usually between 40 and 50 mph (64 and 80 km.ph). Speeds of more than 100 mph (161 km.ph) have been recorded on special short courses.

For a slalom race the course is defined by gates consisting of pairs of poles with flags, between which the runner must pass. The course is carefully set so as to test the skill and judgment of the competitor. For Olympic and world championship events the men's course must have a vertical drop between 650 and 975 ft. (198 and 297 m.) and for other international races between 390 and 650 ft. (119 and 198 m.). Fifty to seventy-five gates are required for men, forty to fifty-five for women. For less skilled groups, courses are shorter and less difficult. The Arlberg-Kan-

dahar race rotates between Mürren, Switz.; St. Anton, Aus.; Chamonix, France; Sestriere, Italy; and Garmisch-Partenkirchen, Ger., and is the oldest open international event decided on the results of a downhill and a slalom, having been first run in 1928.

The giant slalom has characteristics of both the downhill and the slalom. The gates are set wider and farther apart, the length of the course approaches that of the downhill and the average speeds achieved are between those of the two other Alpine events. Because of the persistent efforts of Sir Arnold Lunn, who devised the modern slalom race, the International Ski Federation recognized downhill racing in 1930, and the first downhill and slalom world's championships were held in Mürren in 1931. The best Alpine competitors have come from Austria, France, and the other western European countries.

Cross-country races are held over rolling terrain. The standard distances in international competition are 15, 30, and 50 km. for men, and 10 km. for women, although the Vasa race in Sweden is 56 mi. long. In the relay race four men each run a 10 km. course, or three women a 5 km. course.

Jumping competitions are held on carefully graded and prepared hills, classed according to size as 100-m., 80-m., 70-m., etc. The approach, or inrun, often starts on a scaffold, or tower; down it the jumper in a crouched position accumulates speed until he reaches the takeoff, where he leaps outward, upward, and forward. The best competitors lean far over toward the points of their skis in order to minimize wind resistance and to get an aerodynamic lifting effect to increase the length of their jump. The landing is made on the steep landing hill in a more upright position with the shock of contact taken up by the knees. After the slope levels off, the jumper stops by turning on the outrun. The performance is decided half by the distance covered and half on form, as marked by the judges, on the basis of style marks.

The best jumpers and cross-country runners have come from the Scandinavian countries, Finland, and eastern Europe.

A combination of the results of the downhill and slalom is called the Alpine combined event, and the jump and cross-country together are called the Nordic combined. The biathlon, an Olympic Games event combining cross-country racing and rifle shooting, is a modern offshoot of Nordic and military skiing. Skijoring, in which the skier is pulled by a horse, car, or plane, is little practised.

ORGANIZATION OF THE SPORT

The International Ski Federation (Fédération Internationale de Ski, or FIS) was founded, and the first winter Olympic games were held, in Chamonix in 1924, but with Nordic ski events only. The Alpine events were included for the first time in 1936, at Garmisch-Partenkirchen in Germany.

The world's first ski club was the Trysil, in Norway, formed in 1861. The first in the United States was organized at Berlin, N.H., in 1872. The Ski Club of Great Britain, founded in 1903, was the sport's first national administrative body. It was superseded by the National Ski Federation of Great Britain (1964). The first intercollegiate ski meet, between McGill, Dartmouth, and Montreal, was held in Quebec, at Shawbridge, in 1914.

The basic skiing organization is the local club, which is composed of individuals. In the United States, most clubs are members of one of the regional associations (or divisions) which compose the U.S. Ski Association, formerly the National Ski Association of America (1904) founded at Ishpeming, Mich. The latter, along with the national associations of the other skiing nations throughout the world, compose the FIS, which sets the rules for and sanctions international competitions in all forms of the sport, and engages in other coordinating activities.

GLOSSARY

Alpine (events, races).—Downhill and slalom competitions.

Arlberg strap.—A leather strap which is attached to the cable of a binding and passes over the instep and around the ankle (see *Safety Strap*).

Basket.—See *Ring*.

Bob-run.—A hard-surfaced channel that has been hollowed out by many skiers running in the same track.

Bow.—A forward inclination from the waist, as in bowing.

Brink.—Point or line at which a sharp increase of grade occurs.

Camber.—The arch of the ski, from front to rear.

Change edges.—Shift the lateral cant of the skis so that the previously disengaged edge, or edges, are weighted and canted.

Christiania.—A ski turn executed entirely, or largely, with skis parallel

Classic.—See *Nordic*.

Climbers.—Attachments to the bottoms of skis to facilitate climbing by preventing backsliding.

Comma.—Body position in traversing or turning in which the upper body is angled away from the slope and the knees, held together, are bent toward the slope.

Control gate.—A pair of flags on a downhill slalom course through which a racer must pass.

Corn snow.—Granular snow created by alternating thawing and freezing.

Cornice.—Overhanging drift of snow on the lee side of a ridge or summit.

Counterrotation.—Turning the body in the direction opposite to the direction of the turn.

Creepers.—See *Climbers*.

Critical point.—The place or line on the landing hill of a ski jump beyond which the gradient decreases and landings should not be made.

Double stem.—Position of the skis making an inverted "V," with tips together.

Dry course.—Ski instruction given indoors, or on ground without snow.

Edging.—Canting the ski so that one edge bears more weight than the other and thus "bites" the snow.

Egg Position.—Compact, crouching downhill running position with elbows close to body and hands together in front.

Fall-line.—Direction of steepest descent at any point.

FIS—Fédération Internationale de Ski, or International Ski Federation, governing body of the sport.

Flying mile (or kilometre).—A measured straight downhill course, often shorter than a mile or kilometre, on which speeds are timed.

Gate.—Two poles between which competitors must pass.

Geländesprung (Ger.).—A leap from a bump or brink, during which the knees are retracted.

Giant slalom.—A race between gates which are wider and further apart than in a slalom.

Godille (Fr.).—See *Wedeln*.

Grat (Ger.).—A ridge or col.

Hairpin.—A sharp turn in a slalom course forced by two gates

Heel-turning pressure (*Heel thrust*).—Force exerted in the horizontal plane by the heel pivoting on the ball of the foot.

Herringbone.—Method of climbing by alternately placing one ski above the other in a diagonal position, with the tips out.

Inrun.—The track from the top of a jumping hill to the takeoff.

Inside (ski, foot, hip, shoulder).—The one toward the inside of the turn.

Jump turn.—A turn executed while in motion by lifting the skis from the snow with the aid of one or both poles, and turning them in the air.

Kick turn.—A turn made from a stationary position by placing first one ski and then the other in the desired direction.

Knoll.—The part of a ski jump beyond the takeoff where the landing hill begins.

Landing hill.—That part of a ski jump where the landing should take place.

Lip.—The furthest edge of the takeoff of a jumping hill.

Long-thong (or *longlanière*).—A long strap for attaching the boot to the ski.

Mogl (or *mogul*).—A pronounced bump created by skiers wearing grooves in the snow.

Nordic (events).—Jumping and cross-country competitions.

Outside (ski, edge, foot, hip, shoulder).—The one toward the outside of the turn.

Outward lean.—Bending sideways from the waist toward the outside of the turn.

Piste (Fr.).—A prepared snow track for skiing.

Powder.—A light, dry, uncompacted snow.

Prejumping.—Making a jump just before a large bump or brink in order to clear it and land on its down side.

Ring (wheel, or basket).—Attachment to a ski pole just above its point to prevent its penetrating into the snow excessively.

Ruade (Fr.).—A turn made by lifting the heels of the skis clear of the snow.

Rücklage (Ger.).—Position with body back, weight on the heels, useful in skiing in deep powder snow.

Safety strap.—A small strap connecting ski to boot, to prevent the ski from running away if it comes loose.

Schuss.—A steep hill that can be run straight at high speed; running such a hill.

Schussboomer.—A person who skis out of control or at speeds beyond his ability.

Scissors.—Ski position with tips diverging, back ends together.

Sealskins.—Strips of sealskin attached to the bottom of skis to aid in climbing.

Seelos.—A famous slalom runner, after whom was named a com-

bination of three gates in a slalom course with the second gate between and at right angles to the first and third.

Shovel.—The upturn at the forward end of the ski.

Sideslip.—Sideways or diagonal sliding of the skis.

Sidestep.—Moving uphill by alternately placing the skis across the hill higher and parallel to each other.

Single stem.—Position with one ski at an angle to the other and to the direction of movement.

Sitzmark.—Depression in the snow made by a skier sitting down when falling.

Skating.—Moving forward on the level or a gentle slope by skating steps.

Skins.—See *Sealskins;* also applied to such devices made of plush, canvas, etc.

Slalom.—A race down a course controlled by flags.

Snowplow.—Descent in the fall-line in a double stem.

Spring snow.—Granular or corn snow (*q.v.*).

Stem.—A position of a ski or the skis divergent to the direction of movement.

Stembogen (Ger.).—A turn made by use of a stem.

Swing.—A turn made at high speed, largely with skis parallel.

Téléférique (Fr.).—A large car, usually suspended from a stationary track cable and pulled by a traction cable, for carrying passengers uphill.

Telemark.—A steered turn in which the outside ski is weighted and advanced about a half ski-length beyond the other.

Tempo.—A high speed turn made with skis parallel.

Tramway.—Any device other than a rope tow for hauling skiers uphill

Transition.—Area or line where the grade of a slope lessens or flattens out.

Traverse.—A course diagonally across and down a slope.

Trestle.—A girder structure for all or part of the inrun or landing hill of a ski jump.

Unweighting.—Removing the body weight from one or both skis.

Vorlage.—Position with body, knees and weight forward.

Wedeln.—Making a series of short connected parallel turns.

Wheel.—See *Ring.*

Windslab.—A crust formed on top of snow by wind action.

(R. Po.)

Bibliography.—Ski Club of Great Britain, ***British Ski Year Books*** (1903–); Sir Arnold Lunn, *A History of Ski-ing* (1927), *The Story of Ski-ing* (1952), *The Englishman on Ski* (1963); Charles Dudley, *60 Centuries of Skiing* (1935); National Ski Association, *American Ski Annuals* (1936–); Emile Allais and Paul Gignoux, *Ski Français* (1937); Walter Prager, *Skiing* (1939); Benno Rybizka, *The Hannes Schneider Ski Technique*, new ed. (1946); Tyler Micoleau, *Power Skiing Illustrated* (1949); Robert Bourdon, *Modern Skiing* (1953); Roland Palmedo, *Ski New Horizons* (1956); *The New Official Austrian Ski System;* ed. by the Austrian Association of Professional Ski Teachers, *et al.*, tr. by Roland Palmedo (1958); Tomm Murstad, *Look and Learn* (1957); James Riddell, *The Ski Runs of Switzerland* (1957), and Jeanette Oddie, *The Ski Runs of Austria* (1958); Howard Bass, *The Magic of Ski-ing* (1959); Jean Vuarnet and Georges Joubert, *Ski ABC* (1957), Francesco Freund and Fulvio Campiotti, *The New Skier* 1960), *Deluxe Book of Skiing* (1963); Clemens M. Hutter, *Wedeln* (1960); Jean Franco and Marius Mora, *Ski de France* (1962); Curtis Casewit, *Ski Racing* (1963); J. Kerr Hunter and Odd Gulbrandsen, *Ski-ing in Britain* (1963); Walter Pause, *Salute The Skier* (1963); Johnny Caldwell, *The Cross-Country Ski Book* (1964); *The American Ski Technique* (1964).

(R. Po.; H. B.)

SKIKDA (formerly Philippeville), a town and Mediterranean seaport of northeastern Algeria, in the *département* of Constantine, lies on the Gulf of Stora, about 55 mi. (89 km.) NNE of Constantine by rail and road. Pop. (1960) 80,281. Formerly the population was predominantly European, partly of Italian and Maltese origin, but during the war that preceded and followed Algerian independence (1962) there was an influx of Muslim refugees and an exodus of Europeans. Founded by the French in 1838 to serve as the port of Constantine, Skikda is rectilinear in plan and occupies the site of ancient Rusicade, which was the port of Cirta, as Constantine was called until the 4th century. It lies near the mouth of the Safsaf River, escaping the humidity of the alluvial plain by its situation on a saddle between two hills. The road and railway to Constantine follow the course of the former Roman road. The Safsaf Valley is rich in vines, citrus fruit trees, and early vegetables; Skikda exports its products as well as those of the upper plains of the Constantine region (chiefly cereals, wool, and mutton), dates from the oases of the Biskra region, and iron pyrites from the surrounding district. It imports combustibles and construction materials, and also a variety of manufactures and sugar and coffee, which are redistributed through Constantine to the whole of eastern Algeria. There is considerable passenger traffic through the port, which, with the village of Stora, is a centre for sardine fishing and canning. (J.-J. Ds.)

SKIN, the covering, or integument, of an animal, as distinguished from the internal parts. In animals with backbones (vertebrates) the skin is complex in structure and is a vital organ in that some of its functions are essential for the maintenance of life. In all vertebrates the skin consists of two major layers. The richly cellular, relatively thin outer layer, derived from the embryonic ectoderm, is called the epidermis; the thicker and tougher inner part, of mesodermal origin, is largely composed of fibrous connective tissue and is called the dermis, cutis, or corium.

Skin in Lower Vertebrates.—Before considering the skin of man, the integuments of various classes of vertebrates will be surveyed in the order of their position on the evolutionary scale.

Cyclostomes.—In the lamprey the surface of the skin is smooth, with no scales. Gland cells producing slime are mixed with the epithelial cells, as in most aquatic vertebrates.

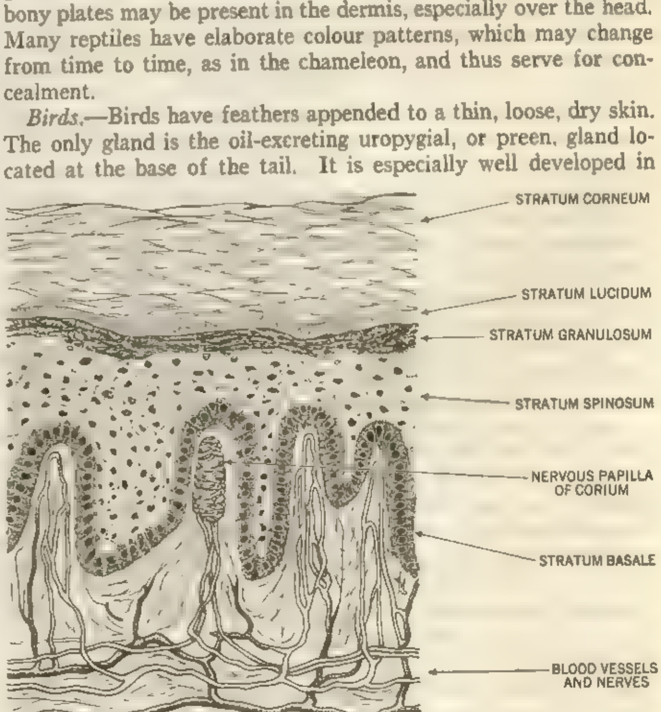
Elasmobranchs.—Among the elasmobranchs, sharks have a very tough skin. Over it are scattered denticles or placoid scales, each with a pulp cavity, around the edge of which is a layer of odontoblasts. These cells secrete the dentine or calcareous material of the scale. Outside the dentine is the enamel, secreted by the overlying ectoderm. Usually the denticles pierce through the ectoderm, after which no more enamel can be added.

Fishes.—The dominant fishes of the modern era are teleosts, characterized by bony scales covered with skin. The epithelium of a trout's epidermis provides the animal with an inert covering of keratin. The scales lie in the dermis as thin, overlapping plates with the exposed part bearing the pigment cells. The scale is deposited in a series of annual rings, since its growth occurs rapidly in spring and summer and hardly at all in winter.

Amphibia.—All adult amphibians have skin that is moist and used for respiration. Its outer layers are horny—a character typical of land vertebrates. In the adult frog the epidermis is renewed at intervals by a process of molting, which is under control of the pituitary and thyroid glands. The wartiness of the skin of toads is due to local thickenings (of interest in considering the origin of feathers and hair). Both mucous and poison glands derived from the epidermis are present in the skin of amphibians.

Reptiles.—The skin of reptiles is dry and contains few or no glands. The outer division is horny or scaly and is shed in flakes; bony plates may be present in the dermis, especially over the head. Many reptiles have elaborate colour patterns, which may change from time to time, as in the chameleon, and thus serve for concealment.

Birds.—Birds have feathers appended to a thin, loose, dry skin. The only gland is the oil-excreting uropygial, or preen, gland located at the base of the tail. It is especially well developed in



FROM CUNNINGHAM TEXT-BOOK OF ANATOMY, OXFORD MEDICAL PUBLICATIONS

FIG. 1.—VERTICAL SECTION OF EPIDERMIS AND PAPILLAE OF DERMIS

aquatic birds. Feathers are the major keratinous structures, but scales like those of reptiles are present on the legs and feet and sometimes elsewhere. The bill and claws are also specialized scale-like structures and may be molted. Nerve endings are present throughout the skin. (*See* also FEATHER.)

Mammals.—The presence of hair appended to the skin (even though the area may be extremely small) is a requisite for classification within the Mammalia. In addition to hairs—both the long tactile bristles and the smaller hairs more generally distributed over the body—the skin appendages may include horns, claws, hoofs, nails, and glands. The armadillo is the only living mammal that has a true bony exoskeleton. (*See* also HAIR; HORN.)

Skin in Man.—The skin of man, from the surface to the point where the dermis merges with the usually fatty subcutaneous tissue layers, varies in thickness from less than 1 mm. ($\frac{1}{25}$ in.) on the eyelids to 3 mm. ($\frac{1}{8}$ in.) or more in the region between the shoulder blades, and on the palms and soles. It is elastic and presents a smooth, soft, pliable surface broken by intersecting lines, folds, ridges, and tiny pores. The more or less parallel, closely set, fine ridge patterns of palmar and plantar skin are characteristic for each individual and form the basis for the widely used fingerprint (*q.v.*) system of personal identification.

Epidermis.—The average thickness of the epidermis is only about 0.1 mm. ($\frac{1}{250}$ in.), although on the palms and soles, or where subjected to rubbing or pressure, it may be 10 to 15 times thicker. The epidermis can be differentiated into several distinct layers (*see* fig. 1). The deepest layer, which rests on the dermis, is called the stratum basale, or basal cell layer, and consists of a single layer of columnar epithelial cells arranged in regular palisade fashion. Interspersed between and slightly below the cells of this layer, along the junction between it and the dermis, is a network of melanin pigment-forming cells (melanocytes) having many slender, branchlike, cytoplasmic extensions; these extensions communicate with one another as well as extend between the cells of the deeper portions of the epidermis, to which they can transfer melanin pigment. There are about 1,000 to 3,000 melanocytes in each square millimetre ($\frac{3}{2,000}$ sq.in.) of skin. (*See* also COLORATION, BIOLOGICAL.)

The layer above the basal cell layer is called the stratum spinosum. This layer is several cells thick; its cells have irregular polyhedral shapes, and they appear to be connected with one another by numerous microscopic intercellular bridges, spines, or prickles. The stratum spinosum and the stratum basale together are sometimes called the stratum germinativum, or Malpighian layer.

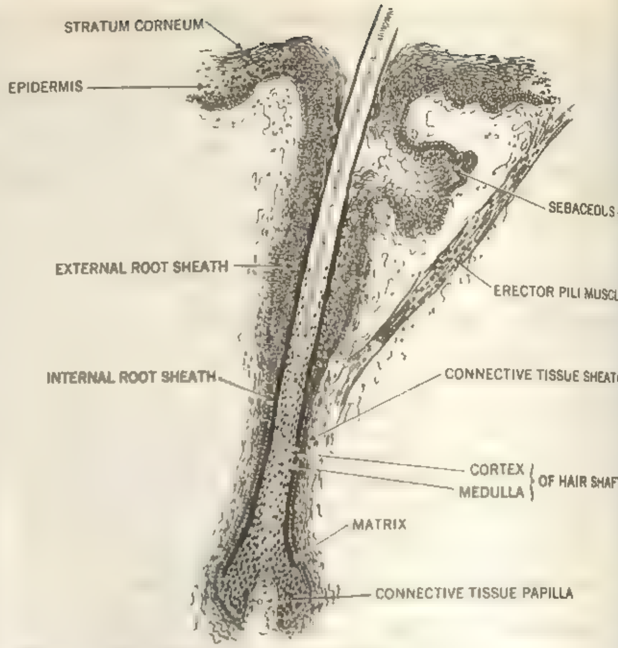
In the next layer, the stratum granulosum, which is from two to four cells thick, the cells become flattened and diamond shaped, and a granular material, keratohyalin, appears in their cytoplasm.

The next layer, the stratum lucidum, is not always apparent. Where well developed it appears as a clear, homogeneous line in microscopic cross-sectional preparations. It contains a lipid-resembling substance, eleidin. The major barrier against the passage of water and salts through the skin lies at the stratum lucidum.

The most superficial layer, the stratum corneum, consists usually of 10 to 20 layers of very flat, dead, dry, keratinized cells, which adhere to one another tightly except nearest the surface, where they are continually shed as microscopic flakes.

Cell division in the epidermis takes place in the basal cell layer and in the lower layers of the prickle-cell layer. As a result of such proliferation, cells are continuously pushed toward the surface. As they move outward they undergo a sequence of changes in which they die, and their proteins become transformed into the tough, chemically inert, fibrous protein called keratin. There are no blood vessels in the epidermis, which therefore has to receive its nutrition from tissue fluids that diffuse through the spaces between the intercellular spines of the stratum spinosum.

Dermis.—The dermis, or corium, has an outer thin, relatively loose, papillary layer adjoining the epidermis and an inner thick, dense, reticular layer that blends into the subcutaneous connective tissue. Papillae of the outer layer (from about 40 to 140 per sq.mm., depending on the region) are tiny, fingerlike processes that fit into sheathlike sockets in the overlying epidermis. The



AFTER A. W. HAM, "HISTOLOGY"; J. B. LIPPINCOTT CO

FIG. 2.—LONGITUDINAL SECTION OF HAIR FOLLICLE

papillae carry capillary vascular loops and also may contain specialized sensory nerve endings. A thin layer of mucoprotein cement substance lies between the corium and the epidermis.

Most of the structural material of the corium consists of the tough, elastic fibrous proteins collagen and elastin. Electron microscopic examination of collagen fibres reveals them to be composed of bundles of very fine fibrils showing a characteristic regular pattern of cross striations. The fibrous components of the corium are imbedded in a complex matrix of ground substance consisting of mucopolysaccharides, proteins, water, and miscellaneous dissolved substances. This ground substance has important functions in regulating diffusion of material through the corium.

The fibres of the corium run in all directions, but predominantly in the plane of the body surface and parallel to the lines of tension and compression of the skin in ordinary movements. (A round puncture wound becomes elliptical after withdrawal of the instrument, and a linear incision gapes more widely if it is transverse to the prevailing fibres.) The corium is well supplied with blood vessels and nerves. Striated muscle extends into the dermis of the face and neck to permit facial movements. Smooth muscle is widely distributed in the corium in the form of thin bundles attached to hair follicles. Their contraction elevates hairs to a more erect position and produces "goose flesh."

Cutaneous Appendages.—The cutaneous appendages in man consist of pilosebaceous units (hair follicles and oil glands), sweat glands, and nails.

The nails are platelike, translucent, keratinous structures that grow over the ends of the fingers and toes from specialized, proximal epidermal matrices called nail roots; these extend to the distal edge of the lunula, a moon-shaped whiter area often visible at the base of the nail plate. Nails consist of densely compacted, highly cornified, dead epithelial cells containing only remnants of degenerated nuclei. The distal edges of the nails are free, and on the fingers serve to assist in the manipulation of small objects as well as for scratching. Six to seven months are required for a newly generated portion of nail to reach the distal free edge.

Eccrine sweat glands occur throughout the skin of man in large numbers, although in many other mammals they occur only on the foot pads. Each gland is an independent, slender, blind tubule about 0.02 mm. ($\frac{1}{1,250}$ in.) in diameter and has two cell layers, an inner epithelial lining (either secretory or ductal in type), and an outer myoepithelial layer of more or less contractile cells. The secretory portion of the gland lies coiled in a ball-like structure in the lower dermis or upper subcutaneous tissue. From there the duct follows a slightly spiral course through the dermis, and then

a much more tightly spiral course through the epidermis and especially the stratum corneum to open in a tiny sweat pore. The total number of sweat glands in man is over 2,000,000, with an estimated tubular length of about 13 km. (8 mi.). These glands actively secrete a watery fluid containing salts. The chief function of sweat secretion in man is the regulation of body temperature by evaporative cooling. Sweat also helps maintain normal hydration of the outer layers of the skin, and on the palms its moisture enhances the grip. (*See* also PERSPIRATION.)

Pilosebaceous units are densely distributed in almost all parts of the skin (fig. 2). Each unit consists of a hair follicle with its hair and its appended sebaceous (oil) gland and erector muscle. In the armpits, perineal and pubic regions, and about the nipples, apocrine sweat glands are also appended to the pilosebaceous units. These glands, which open into the pilosebaceous canal, have a complex, branched, tubular structure; the tubes are about ten times larger than those of eccrine sweat glands. Their mode of secretion also differs: part of the cytoplasm of the secretory cells is pushed off (decapitation secretion). Apocrine secretion is an opalescent fluid that under bacterial action gives rise to substances partly responsible for unpleasant body odours.

Sebaceous glands are multilobulated structures 0.2 to 2 mm. ($\frac{1}{125}$ to $\frac{2}{25}$ in.) in diameter with short ducts that open into the necks of the hair follicles. These glands excrete a highly complex mixture of lipids, called sebum, which diffuses upward in the follicle and impregnates the hair and surrounding horny layers of the skin to participate in forming the greasy skin surface film. This film helps maintain normal pliability of the skin by retarding loss of water from the horny layer, delays absorption of many foreign substances, and contains substances that are protective against exogenous infections. In sebum formation, germinative cells in the periphery of the gland proliferate and, as the maturing cells move toward the centre of the gland, discrete droplets of lipid progressively accumulate in their cytoplasm. Finally these cells burst and liberate the lipid material.

Hair, in man, has little functional value when compared with its role in fur-bearing animals. Nevertheless, human hair has considerable cosmetic and biological interest. Hair filaments are dead, thin, flexible shafts of highly keratinized epithelial cells developing slanting tubular invaginations of the epidermis—the hair follicles. The deepest portion of each follicle is enlarged to enclose a vascular papilla, projecting from the dermis into the bulb. The epithelium immediately above and around this papilla is the germinative matrix for the hair. There are great individual and regional variations in types and distribution of hair. On the scalp, where hair is usually densest and longest, the average total number of hairs is between 100,000 and 150,000. Hair is continually shed and renewed by the operation of alternating cycles of growth, rest, dedifferentiation, and renewal of growth of the hair follicle and its associated structures. This cyclic activity is intrinsic in the individual pilosebaceous units. The average life of different varieties of hair varies from 4½ months for downy hairs to three to five years for long scalp hairs. Scalp hair grows approximately 1 cm. ($\frac{2}{5}$ in.) per month; in persons of European origin it reaches an average length of 60 to 75 cm. (24 to 30 in.). (*See* also HAIR: *Anthropology*.)

Skin Functions.—The skin is a remarkably effective protective barrier against a wide range of injurious chemical, physical, and biological factors. Since it is nearly waterproof it enables the relatively fluid body to exist in dry air, and to be immersed in fresh water without becoming swollen and in salt water without becoming shrunken. The keratinized layers of the skin, which are practically inert chemically and are poor conductors of heat and electricity, also serve, with melanin pigment, as an effective screen against potentially harmful amounts of ultraviolet radiation. Vitamin D is produced in the epidermis when the skin is exposed to sunshine. Body temperature regulation is achieved by the extensive vascular plexuses distributed in the skin and by the cooling effect of sweat evaporation from its surface. Sensory functions are carried out through a rich variety of nerve endings, which bring the primary sensory modalities of touch, heat, cold, and pain into consciousness from discrete receptor points in the skin. Total body respiration, as well as excretion of metabolic wastes, is aided by the skin only slightly. Most substances are very poorly absorbed through intact skin except in the gaseous state. Materials readily soluble in both fatty substances and water are generally better absorbed than those soluble in only one or the other.

See also SKIN, DISEASES OF; SKIN, SENSORY FUNCTIONS OF; and references under "Skin" in the Index.

BIBLIOGRAPHY.—Stephen Rothman, *Physiology and Biochemistry of the Skin* (1954); William Montagna, *The Structure and Function of Skin*, 2nd ed. (1962); J. Z. Young, *The Life of Vertebrates*, 2nd ed. (1962). (A. L. Lz.)

SKIN, DISEASES OF. The human skin is subject to many diseases, some of them inconsequential and passing, some chronic and productive of prolonged partial or total disability, and others associated with congenital or acquired systemic diseases. Although the cause of many of these conditions is definitely known and the treatment reasonably precise and effective, a large number have no cause definitely established and treatment remains empiric and sometimes unsatisfactory. The diagnosis and treatment of skin diseases is called dermatology, or cutaneous medicine.

Incidence.—Accurate figures on the comparative incidence of various skin diseases were difficult to obtain until World War II, when the medical records of soldiers and sailors of many nations showed that skin diseases were a significant source of disability. From 4 to 20% of patients in U.S. and British military hospitals throughout the world were admitted because of diseases affecting the skin. The percentage was highest in the tropics, where the climate greatly increases the incidence of skin diseases.

Under conditions of dislocation of the population, crowding, and inadequate hygienic conditions, parasitic infestations such as scabies show a sharp rise, as do venereal diseases affecting the skin. Drugs account for a considerable number of skin maladies. Penicillin is a major cause of reactions, usually hives. Sensitiveness to sulfonamide compounds is not uncommon. Many drugs render the skin abnormally sensitive to sunlight (photosensitivity).

Types of Skin Lesion.—A localized, abnormal change in living tissue due to injury (trauma) or disease is called a lesion. The principal types of lesion seen in skin diseases are the following: Maculae are areas of altered skin colour that do not involve swelling or other distortion of tissue. They vary in size from small spots, *e.g.*, freckles, to fairly extensive areas called patches or blotches. Papules are small, solid lesions raised above the skin surface. They are usually conical or rounded in shape and about 1 mm. to 1 cm. ($\frac{1}{25}$ to $\frac{2}{5}$ in.) in diameter. Hives, or wheals, are a kind of papule, usually white or pink in colour, that soon disappears; *e.g.*, mosquito bites. Nodules are rounded or irregularly shaped swellings that are larger than papules and extend well below the skin surface. Tumours are swellings that resemble nodules but are substantially larger and deeper. Any cancerous growth is likely to be called a tumour—though not all tumours are cancerous. Vesicles are lesions filled with clear serous fluid; large vesicles are called bullae (blisters). Pustules (pimples) are papule-like lesions containing pus; they often occur in association with vesicles. Ulcers are open sores resulting from the destruction of tissue. They may extend deep into the skin, and they usually exude pus and other matter which, on drying, forms a scab. (*See* also TUMOUR; ULCER.) An extensive but temporary skin eruption of any kind is popularly called a rash.

Classification.—Diseases of the skin may be classified in several ways. From a systemic point of view, for example, there are (1) those which arise solely within the skin and ordinarily remain confined to it, such as superficial fungal infections and acne; (2) those in which the spread of infection or absorption of toxic agents from the skin may produce disease of internal organs, as in impetigo; and (3) those in which skin changes are an expression of internal disease, as in chicken pox and measles.

In this article, common diseases of the skin are described in groups according to their predominant causative factors. The groups are as follows: (1) Disorders of structure (*e.g.*, moles) and function (*e.g.*, excessive sweating). (2) Acne—the most common disease of the skin—and other conditions arising from endocrine and metabolic disturbances. (3) Noninfectious inflam-

mations: the various forms of dermatitis, or eczema. (4) Infections: those caused by fungi (*e.g.*, ringworm of the scalp, athlete's foot), by bacteria (*e.g.*, impetigo, boils), by viruses (*e.g.*, warts, cold sores, shingles), and by parasites (*e.g.*, scabies). (5) Miscellaneous maladies: psoriasis, hives, photosensitivity, skin cancer, seborrheic keratosis, pemphigus, and congenital conditions.

For details of anatomical structures mentioned in this article *see* SKIN: *Skin in Man*.

Disorders of Skin Structure or Function.—Several skin disorders may involve the pigment cells (melanocytes), located in the basal portion of the epidermis, or outermost layer of the skin. On exposure to sunlight the pigment cells are stimulated to greater output. In brunettes the increase is considerable, and the pigment is laid down evenly, resulting in a deep tan.

Freckles.—In blondes and redheads the pigment cells respond only slightly, and rather unevenly, to sunlight. As a result there is little or no tanning, and the pigment may be laid down irregularly, as freckles. (*See* also ALBINISM.)

Vitiligo (*Leukoderma*) is the loss of ability to form pigment. It is of cosmetic significance only, but may be quite disfiguring in Negroes and other dark-skinned persons. The tendency to it is inherited. The normal skin colour rarely returns, and there is no satisfactory treatment.

Moles.—The pigment cells give rise to ordinary moles, or pigmented nevi (sing., nevus). These contain masses of melanocytes, with nerve elements and connective tissue in thicker lesions. Moles may be flat or fleshy. Their colour varies from almost colourless to bluish, dark brown, or black. They sometimes contain stiff hairs. Moles may be present at birth, but more frequently they appear during childhood. Pregnancy often promotes enlargement of pigmented moles or causes new ones to appear. Moles sometimes disappear in old age. Everyone has some moles, which may number in the hundreds. Occasionally a single mole may cover a considerable area of the body.

Melanomas.—Pigmented nevi are almost always harmless, but they sometimes give rise to tumours known as melanomas. This condition most often arises from a flat, or only slightly raised, and hairless mole. The earliest warning is a change in pigmentation, which tends to become irregular. The development of a zone of pigment around the base, or slow enlargement, is cause for investigation. The entire mole must then be excised surgically, and examined microscopically for evidence of cancer. Malignant melanomas may sometimes develop without any evidence of a preceding mole. (*See* also below, *Skin Cancer.*)

Prickly Heat.—The skin regulates body temperature in part by the evaporation of sweat, produced by glands in the underlying layer of the skin. When ducts are blocked, as in sunburn or eczema, the sweat may be trapped. If the blockage is superficial, tiny blisters appear, but are of no consequence. Deeper blockage results in an inflammation called prickly heat (miliaria). This produces itching, which can be continuous and intense in a warm climate. The itching usually ceases promptly in cooler surroundings.

Excessive Sweating.—Sweating is greatly influenced by emotional stimuli. Under severe stress, almost anyone will react with increased sweating of the hands, feet, and armpits. In some persons the reaction occurs from the ordinary stresses of living, and this excessive sweating (hyperidrosis) may be almost continuous and very annoying. (*See* also PERSPIRATION.)

Seborrhea and Wens.—Hyperactivity of the sebaceous (oil) glands produces excessive oiliness of the face, scalp, centre of the chest, and genitalia, where these glands are most numerous. Often there is yellowish scaling (*see* below, *Seborrheic Dermatitis*). Occasionally a saclike growth called a wen develops from a sebaceous gland; it is filled with sebum and keratin, and is unsightly but benign (although subject to bacterial infection). Draining the contents is useless, since the wen soon fills again; surgical removal may be required.

Chapping.—The skin surface is kept supple by the oily secretion of the sebaceous glands and by fatty substances supplied from the epidermis. The stratum corneum (outermost, horny layer of the skin) also contains a fairly constant water content. Under very dry conditions, as in overheated houses in cold weather, the amount of water in the stratum corneum may drop below a critical level, with loss of elasticity and resultant chapping and fissuring. (*See* also below, *Ichthyosis.*)

Bedsores are ulcerations of the skin over bony prominences, *e.g.*, the hips or elbows, of bedridden persons. They may be caused by interrupted circulation due to body pressure upon the mattress, or by friction against the bedclothes. The condition is aggravated by uncleanliness, which also increases the danger of secondary infection. Prevention and treatment of bedsores includes frequent changes of position, daily bathing, and the application of lotions.

Chilblains are red or purplish patches that itch and burn. They usually occur on the hands, feet, and legs, and are commonest in children and young women. Chilblains result from exposure to alternate extremes of heat and cold.

Calluses and Corns.—The skin responds promptly to physical or chemical injury. With chronic rubbing, as from a tight shoe, the epidermal cells become much more active and horny tissue increases at the skin surface. This localized thickening is known as a callus, or callosity. If it has a conical shape, penetrating the dermis and causing pain when pressed, it is called a corn.

Acne Vulgaris.—The most common of all chronic skin diseases, acne is so frequent during adolescence as to constitute almost a normal physiological response to puberty. With the normal rise in circulating hormones of the male type, with development of secondary sex characteristics such as the growth of pubic and axillary hair, the size and activity of the sebaceous glands also increases. At the opening of the follicle (pore) through which hair grows and the sebaceous gland empties, a plug known as a blackhead appears, and is followed by the typical acneous pustule.

Mild acne frequently clears up spontaneously, with little or no residual pore enlargement or scarring. The reasons for this spontaneous cure are uncertain. More severe cases involve inflammation, with repeated infection or chemical irritation and disturbance of the precise cellular arrangement of the epidermis. In such cases acne may continue into the twenties or even longer, and produce irreversible scarring. The psychic implications of prolonged severe acne are often considerable, and young individuals so afflicted may become depressed and withdrawn. Treatment must include some psychotherapy in addition to other measures.

The treatment of acne is far from satisfactory in terms of producing prompt, permanent subsidence of all signs of the disease. In mild cases, where spontaneous involution may be anticipated, drying lotions containing sulfur, resorcinol, and other agents are ordinarily helpful, along with careful drainage of superficial pustules and the gentle expression of blackheads. Picking or pinching the lesions can be very harmful, and may become an almost uncontrollable habit. Regular washing of the face is advisable, though this should not be too frequent or strenuous. Reasonable use of cosmetics is not harmful. The relation of flareups to ingestion of particular foods should be watched for. Dietary restrictions usually include the avoidance of chocolate, nuts, and cola drinks.

Sunlight ordinarily has a beneficial effect on acne. If the weather is warm and muggy, severe exacerbation of acne may occur (so-called tropical acne). In females, exacerbation of acne is frequently noted before or during menstruation.

Antibiotic and sulfonamide therapy is widely used in severe acne in which infection is obviously present. However, approximately one person in three does not show any lasting response to such treatment. Therapy must be long continued and, in the case of the antibiotics, is expensive.

Since the basic reason for the development of acne is the presence of androgenic hormones, many kinds of hormonal treatment have been tried. None is entirely satisfactory and free of side effects. In males the administration of hormones of the female type (estrogens) is helpful; however, the doses are necessarily large and will produce undesirable feminizing effects in time. In females the administration of hormonal contraceptive compounds has a good effect on acne, though this effect is not ordinarily apparent until treatment has been given through three or four menstrual cycles.

Deficiency Diseases.—Improper diet causes a number of diseases involving the skin. One of the most serious of these is pellagra (*q.v.*). For deficiency diseases generally, *see* MALNUTRITION.

Dermatitis and Eczema.—A dermatitis is an inflammation of the skin not caused by an infectious agent. It may be caused by many substances. Dermatitis is characterized by redness, swelling, blister formation, oozing, and (almost always) itching. Eczema is, essentially, a chronic dermatitis. Dermatitis and eczema may be classified in the following principal types:

Contact Dermatitis results from contact of the skin with an irritating substance or a substance to which the person is allergic (*see* ALLERGY AND ANAPHYLAXIS). Inflammation resulting from contact with a strong chemical, such as a concentrated acid or alkali, is called a primary irritant dermatitis. There is no allergic component in this. Other substances causing contact dermatitis are not primarily irritant in themselves but produce a reaction in persons sensitized by previous exposure to the substance.

The most common cause of allergic contact dermatitis in the United States is poison ivy (*q.v.*) and its relatives. Individuals are not born with the sensitivity, but acquire it by repeated exposure to the plant substance. Many other plants are capable of causing sensitization; *e.g.*, the European primrose. (*See* POISONOUS PLANTS: *Plants Poisonous on Contact.*) Chemical compounds that may cause allergic contact dermatitis include penicillin, sulfonamides, paraphenylenediamine dyes, and chromates.

Atopic Dermatitis, also called allergic infantile eczema and disseminated neurodermatitis, is a highly characteristic, itching type of chronic dermatitis. It appears most commonly in tense, nervous persons, especially those with dry skin. The tendency is inherited. Evidence of related allergic disturbances such as hives, hay fever, and asthma is found in the family members or in the afflicted person. Severe cases are difficult to control medically.

Seborrheic Dermatitis (*Dandruff*) manifests itself as scaling of the scalp, which is often oily. Inflammation may appear in other areas of the body; *e.g.*, behind the ears, in the armpits, and in folds of the skin. The term dandruff is a rather inexact one, simply indicating excessive scaling of the scalp. Although the most characteristic type is seborrheic dermatitis, dandruff also occurs in persons with dry skin. Some dandruff is caused by psoriasis (*see* below) which has not been recognized. Infections, particularly of the ringworm variety, also may produce scaling.

Localized Neurodermatitis (*Scratch Dermatitis*).—Occasionally, localized areas of skin become itchy although there is no apparent source of irritation. The process is perpetuated by scratching. The scratching substitutes pain for the itch, but the itching soon returns and the scratching is repeated.

Chronic Dermatitis of the Hands and/or Feet.—The hands are susceptible to inflammatory disease because of chemicals with which they come in contact; the feet, because of shoes, which harbour fungi and bacteria causing superficial infection.

Stasis Dermatitis is a characteristic inflammation of the ankles and lower legs in persons suffering from interference of return circulation, usually due to varicose veins (*q.v.*). The dermatitis is chronic, and ulceration may occur. An essential part of treatment is to relieve the stasis (stoppage of blood flow) by supportive stockings or by surgery on the varicose veins.

Dermatitis Medicamentosa (*Drug Eruption*).—Almost any drug is capable of producing a sensitization reaction. The reaction may be a dermatitis, hives (*see* below), or a mixture of various skin lesions with large blisters. It is estimated that 8 to 10% of the U.S. population is sensitive to penicillin in varying degrees. The most common reaction to penicillin is the development of hives, sometimes of large size.

Fungus Infections.—Fungi, a class of primitive plants, include microscopic organisms capable of producing infections, either confined to the skin surface (superficial infections) or involving internal organs as well as the skin (deep infections).

Ringworm (*Tinea*) is the inclusive name for a number of superficial infections caused by the fungi *Trichophyton* (the most common genus), *Microsporum*, and *Epidermophyton*. These organisms feed on keratin, the protein building-material of the nails, hair, and stratum corneum. Most ringworm infections respond well to treatment with griseofulvin, a fungicidal antibiotic.

Ringworm of the Scalp (*Tinea Capitis*) is highly contagious among children. Within the round, spreading lesions the hairs become brittle and break off at the skin surface. The disease has declined in Western communities; it formerly was epidemic in institutions such as orphanages. The common types of ringworm of the scalp are easily detected by a characteristic fluorescence when examined under ultraviolet light with a special filter.

Favus (*Tinea Favosa*) is another type of ringworm, now rare. The cuplike, yellow-crusted lesions occur most commonly on the scalp, and may itch intensely. The condition is chronic and contagious. Favus sometimes results in permanent, patchy baldness.

Pityriasis (*Tinea Versicolor*) is a fungus infection that produces, mainly on the upper trunk, numerous blotchy macules that vary in colour from white to pinkish-buff to brown. The disease is chronic, and of cosmetic significance only; it is not contagious.

Athlete's Foot.—The most common site of this type of ringworm infection is the feet. The incidence is higher in males than in females. Because infection ordinarily begins in early adult life, and because it was erroneously thought to be transmitted in gymnasiums and swimming pools, the term athlete's foot became popular. However, irritation of the skin of the foot can be caused by many agents other than fungi, and the diagnosis must be proved by the demonstration of fungi in scales scraped from the skin surface or the toenails. Inflammation between the toes may be produced by moisture alone, through excessive sweating or failure to dry thoroughly after bathing; or by the superficial bacterial infection known as erythrasma (*see* below).

Jockstrap Itch is a fairly common superficial fungus infection in the groin in males. It is characterized by round, scaling patches. Diagnosis must be confirmed by demonstration of the organism, since other maladies tend to involve this area as well. If a ringworm infection is present it is ordinarily quickly controlled by griseofulvin, scrupulous hygiene, and the prevention of irritation, as from the wearing of a jockstrap. Women occasionally have a ringworm infection in the groin, but the inflammation is more frequently candidiasis or erythrasma.

Candidiasis (*Moniliasis*) is caused by the same yeastlike organism that produces thrush (*q.v.*) in the mouth. It occasionally occurs in the vagina, with accompanying inflammation of the vulva and perianal region. The organism also attacks the webs of the fingers and the nails of housewives, bartenders, and others whose hands are often wet. As a skin disease it is most frequent in obese persons and diabetics. It occurs with some frequency in persons receiving broad-spectrum antibiotics over an extended period of time. Candidiasis is treated with the fungicide nystatin.

Deep Fungus Infections.—Certain fungi are capable of invading internal organs as well as the skin and mucous membranes. Blastomycosis, found mainly in the midwestern U.S., produces skin ulcers. Many persons are infected and recover spontaneously without being aware that they have had the disease. Coccidioidomycosis (valley fever) is found mainly in the San Joaquin Valley of California, though cases have been reported elsewhere. Infection begins in the lungs when the spores of the fungus are inhaled. A characteristic feature is the development of inflamed nodules on the lower legs. A high proportion of residents of the endemic areas acquire the infection, and all but a few recover spontaneously. *See* also FUNGUS INFECTIONS.

Bacterial Infections constitute another major group of skin diseases. They may occur suddenly and acutely on normal skin.

Impetigo, usually caused by staphylococci, is marked by purulent lesions that become thickly crusted. In some cases there are severe systemic complications. The disease is extremely contagious among infants, but not particularly so among older children and adults. Infections in nurseries may often be traced to an adult carrier. The lesions are treated by cleansing with soap and water, followed by application of a broad-spectrum antibiotic ointment. Systemic antibiotic therapy is sometimes necessary.

Boils and Carbuncles (*Folliculitis*).—When bacterial infection is mainly around hairs, superficial folliculitis or a deeper boil occurs. If several boils coalesce a carbuncle results, forming a dangerous lesion in old or debilitated persons. Some individuals, often

young and in otherwise excellent health, are susceptible to recurrences of boils, for reasons not fully understood.

Erythrasma is a bacterial infection resembling a dark, blotchy type of ringworm, with which it was long confused. Most prevalent in warm climates, it is treated with oral antibiotics.

Erysipelas is a form of cellulitis (disease of the subcutaneous tissue) caused by streptococci. The red, plaquelike swellings are seen most commonly on the face and scalp, and may recur at the same site. The disease is accompanied by lassitude and fever. It is treated with bed rest, cool compresses, and antibiotics.

Leprosy.—*See* LEPROSY.

Venereal Diseases (*q.v.*) are prominent among the many systemic bacterial infections which are characterized, at some stage, by skin lesions. Pinta, a disease encountered in the American tropics and caused by a spirochete, is of special interest. It involves bizarre changes in the colour of the skin, which may become white or a shade of blue, brown, or red.

Virus Infections.—Many virus infections produce lesions of the skin. Some of these are systemic infections; *e.g.*, measles and chicken pox (*qq.v.*). This section discusses virus infections that remain entirely or almost entirely within the skin.

Warts are extremely common. They can be spread from one part of the skin to another. In moist parts of the body, as around the anus or genitalia, warts may be large and exuberant (venereal warts—a term that does not imply venereal disease). Warts on the sole of the foot (plantar warts) may cause considerable pain and disability in walking. Warts are definitely contagious, not infrequently occurring in several members of the family or on the feet of persons who walk barefoot on contaminated floors. Warts may appear or disappear for no apparent reason. In children especially they may be successfully treated by psychotherapy or by folk remedies of the "hexing" kind. No known compound has a specific effect on the virus. Methods of treatment include application of a cauterant compound; freezing with liquid nitrogen; and electrosurgery.

Herpes Simplex (*Cold Sore, Fever Blister*) occurs most frequently on or near the lips but may be seen on the genitalia and in other places. It may recur repeatedly at the same site. The initial infection apparently occurs sometime during childhood; among infants it may produce marked systemic illness, with extensive lesions of the skin and mucous membranes, but among older children the first infection may go undetected. Recurrence is triggered by a variety of factors, such as an upper respiratory infection, excessive exposure to sunlight, the menstrual period, emotional tension, and possibly a food or a drug. There is often tingling or burning in the affected site for a few hours to a day or two before the onset of redness and a round cluster of blisters. The lesion is usually single, but occasionally there may be more. The blisters gradually dry up in about a week and are replaced by a scab. Ordinary cases are treated with drying lotions or solutions. Supportive therapy is given in cases of severe initial infection in infants, and for persons with eczema, particularly infants or children, who sometimes develop very widespread cold-sore lesions that look much like chicken pox and may be accompanied by high fever. Herpes simplex about the eye may sometimes involve the eye itself (herpetic keratitis).

Herpes Zoster (*Shingles*) is caused by the same virus as that of chicken pox. The virus affects the skin and nerves. Herpes zoster has a characteristic linear distribution following a nerve segment, along which are seen round groups of small blisters similar to, but usually much larger than, those in herpes simplex. Most common on the back, lesions extend from the spine toward the chest. However, any nerve segment may be involved; *e.g.*, the lesions may extend down a leg or an arm. The most serious type of herpes zoster involves the first branch of the fifth cranial (trigeminal) nerve, which supplies the scalp, forehead, and the region of the eye; the lesions may cause severe eye damage. The symptoms and the severity of lesions vary with age. Before any lesions appear, there is frequently a dull ache for a day or so in the affected site. In most young persons the lesions are hardly more severe than those of a cold sore; in aged persons or those debilitated by some systemic disease, the pain and inflammation may be severe. Occasionally the pain persists long after evidence of inflammation has subsided (postherpetic neuralgia). Herpes zoster is treated with soothing powders or lotions, and pain-relieving drugs.

Parasitic Infestations.—Many insects regard man as an attractive food or habitat. Biting insects such as mosquitoes, flies, and ticks produce local reactions in the skin varying from a small hive to a large blister. Bees, wasps, and hornets inject a venom, producing reactions of a severity that depends on the sensitivity of the person bitten. (*See* INSECT BITES AND STINGS.)

Pediculosis is the name of skin disorders caused by various species of lice that infest the scalp, groin, and body. Good hygienic practices and the availability of miticides such as DDT have made louse infestations rare in civilized communities.

Scabies is caused by the itch mite, *Sarcoptes scabiei*. The female burrows into the stratum corneum, laying her eggs in a tunnel that can be seen as a dark wavy line. The initial lesion is soon obscured by the secondary lesion, which usually takes the form of hives or scratch dermatitis (as the result of itching). Common sites of infection are between the fingers, in the genital region in men and the breasts in women, along the belt line, and about the elbows and wrists. Under conditions of poor hygiene and crowding, scabies may become epidemic. Transmission occurs almost entirely by skin-to-skin contact and not to any appreciable extent through wearing apparel and bedding.

Psoriasis is a fairly common and extremely persistent skin disease of unknown origin. The age of onset varies from infancy or early childhood to age 65 or later; usually, however, the disease begins in early adult life. A tendency to psoriasis is inherited.

The typical lesion of psoriasis is a red, sharply circumscribed patch with silvery scales. The most common sites are the scalp, the elbows, and the knees. The disease may remain in rather diminutive form for many years; at times it may become more extensive, with numerous lesions on the trunk and the extremities. A characteristic involvement of the nails is seen; this may be confused with ringworm infection.

Treatment of psoriasis is often discouraging, and no method will effect a permanent cure. Exposure to sunlight often is helpful. Medications usually contain mercury, tar, salicylic acid, sulfur, and anthralin, singly or in combinations. The best single local medication is a cream or liquid containing a potent corticosteroid.

Hives (Urticaria).—This is a local and usually transitory edema of the skin produced by the release of histamine (*q.v.*) and, probably, other substances. The condition may be short-lived or chronic. A simple example of a hive is the reaction resulting from an ordinary mosquito bite. The most common cause of extensive hives is penicillin or a compound chemically related to this antibiotic. Other drugs may give rise to urticaria, depending on the person's sensitivity. Foods may also produce these lesions. In some persons, hives persist or recur for months or years. The cause in such cases is difficult to determine. Psychological factors may have a dominant role. Some persons tend to develop hives wherever trauma is exerted on the skin, such as by rubbing the finger firmly over it; this is called dermographism.

Photosensitivity.—Many persons are intolerant to sunlight. This is especially true of blondes and redheads, whose skin is reddened by relatively brief exposure to the sun, and in whom the protective responses to sunlight, such as thickening of the skin and tanning, do not come into play. In rare instances infants are born with a very marked intolerance to sunlight. Persons with liver disease, particularly that caused by alcoholism, and the biochemical disturbance known as porphyria (*q.v.*), may become intolerant of sunlight. An increasing number of drugs are known to produce photosensitivity.

Lupus Erythematosus is a disease that is exacerbated by sunlight. The cause is not known. Chronic discoid lupus erythematosus is characterized by red patches that appear chiefly on the bridge of the nose, the cheeks, and the scalp. The borders of the lesions have a gray scale which extends into the dilated follicles. There may be residual scarring. Antimalarial drugs are effective in treatment. A far more serious form is systemic lupus erythematosus, a collagen disease (one involving connective tissue in blood vessels throughout the body). Symptoms include fever, pain in

the joints, and distress caused by lesions in the kidneys and elsewhere. Red, patchy skin lesions occur in about half the cases; they are commonest on the face but also may involve the hands and chest. The prognosis is unfavourable, because of complications. Treatment consists largely of the administration of corticosteroids to abate the symptoms. (Lupus vulgaris, or tuberculosis of the skin, is not related to the two diseases just discussed. It is marked by nodular lesions that produce scarring.)

Skin Cancer.—The possibility of curing most types of skin cancer approaches 100% if the precancerous or cancerous condition of the skin is recognized early and the tumour is promptly eradicated. The only cancerous lesions to which this statement does not fully apply are the malignant melanomas (*see* above) arising from moles.

There are many types of skin cancer. Most of them arise from the skin itself, but occasionally they result from cancer of internal organs or from leukemia. Most skin cancers are of two types: basal-cell epithelioma (rodent ulcer) and squamous-cell (prickle-cell) epithelioma. Basal-cell epithelioma lesions are ordinarily slow-growing, and may require years to achieve a size of 1 in. ($2\frac{1}{2}$ cm.) in diameter. They have a characteristic appearance to the practised eye, and may be recognized very early in their course. They occur most frequently on exposed areas, especially where the sebaceous glands are most numerous. Fortunately these tumours invade only by local extension, and do not spread through the lymphatics or bloodstream to other parts of the body. Squamous-cell epithelioma is a more malignant type, which may spread to neighbouring lymph nodes and other structures, particularly in the case of tumours on the lips. This type of cancer is the most common one occurring within the mouth. Squamous-cell epithelioma is almost always preceded by a rather long-standing precancerous lesion, usually a senile (actinic) keratosis (a small, wart-like lesion) seen most commonly in areas which have had chronic exposure to sunlight.

Certain chronic stimuli are known to promote degenerative changes in the skin which may result in cancer. By far the most important of these is excessive exposure to sunlight. The incidence of skin cancer is highest in sunny regions such as the southwestern U.S., and the disease most often afflicts persons whose work or recreation keeps them in sunlight a great deal; *e.g.*, farmers, sailors, fishermen, and sunbathers. The more common types of skin cancer are practically unknown in dark-skinned Negroes, whose pigmentation is a defense against sunlight.

Seborrheic Keratosis.—This very common skin lesion is frequently misinterpreted as a skin cancer or mole. The lesions begin as light-coloured, rough-surfaced, slightly raised papules which gradually darken and extend in size. They are seen most frequently on the trunk, especially on the shoulders and back; some persons have a great many of them. Women frequently develop the lesions on the sides of the neck. These keratoses are of cosmetic significance only, and are easily removed by surgery.

Pemphigus is a disease of unknown cause which involves degeneration of tissue. Blisters that form in the mouth and throat leave raw areas when they rupture; those that form on the skin (which acquires a distinctive odour) are followed by the exfoliation of large scales. Pemphigus may be fatal unless treated intensively with corticosteroid drugs.

Congenital Conditions.—The general characteristics of the skin are, of course, genetically determined. The number of congenital defects is very large, but fortunately most of them are rare. However, two are so common, and at times so disabling or disfiguring, as to deserve mention.

Ichthyosis (*Fishscale Skin*).—Some individuals are born with very dry skin. The mechanism involved is a greatly increased adherence of the dead cells of the outermost, horny layer. The surface is dry and rough, is intolerant of even the mildest irritants, and has a tendency to severe chapping and fissuring, especially in cold weather. These persons may also have a deficiency of sweat glands and occasional associated disturbances of the growth of hair, teeth, and nails.

Hemangiomas (*Vascular Birthmarks*) are malformations of the blood vessels of the skin. The most frequent type is immature (strawberry) hemangioma, which may be present at birth or appear shortly thereafter. It is a bright-red nubbin composed principally of small capillary blood vessels. The lesion ordinarily enlarges for several months or a year, and occasionally becomes ulcerated. Strawberry hemangioma usually recedes after the first year of life, and no treatment may be necessary. A second, very common type is flat hemangioma (port wine stain, nevus flammeus). The most frequent site is the back of the neck; less frequently it is on or near the eyelids. Sometimes the lesion may be large and prominent, and can be very disfiguring if on the face. There is no satisfactory treatment, but the area can be satisfactorily obscured by heavy make-up preparations. A third, much less common type is mature (cavernous) hemangioma. This is present full-blown at birth, is composed of rather large blood vessels and other tissue elements, and does not change in size except in relation to the growth of the part involved.

See also BALDNESS; BURNS; FROSTBITE; GANGRENE; INFLAMMATION; WOUND; and references under "Skin, Diseases of" in the Index.

BIBLIOGRAPHY.—S. M. Bluefarb, *Cutaneous Manifestations of the Malignant Lymphomas* (1959); G. Hildick-Smith, H. Blank, and I. Sarkany, *Fungus Diseases and Their Treatment* (1964); W. F. Lever, *Histopathology of the Skin*, 3rd ed. (1961); R. M. B. MacKenna, *Modern Trends in Dermatology*, 2nd series (1954); W. Montagna, *The Structure and Function of Skin* (1962); D. M. Pillsbury, W. B. Shelley, and A. M. Kligman, *A Manual of Cutaneous Medicine* (1961); R. L. Sutton, Jr., *Diseases of the Skin*, 11th ed. (1956); R. L. Baer and V. H. Witten (eds.), *Year Book of Dermatology and Syphilology* (1956–63); D. M. Pillsbury and C. S. Livingood (eds.), *Proceedings of the XII International Congress of Dermatology* (1962). (D. M. P.)

SKIN, SENSORY FUNCTIONS OF. Sensory nerve impulses from the skin are employed either in keeping the organism informed of its relationship with its immediate surroundings, or in the control of mechanisms for counteracting the effect of changes in the physical conditions within or at the surface of the skin. Thus, the sensory functions of the skin are not concerned solely with conscious experience.

This complex role can be demonstrated in many ways. For instance, if a beam of infrared radiation (heat energy) is transferred to the skin at a suitable rate, a report of warmth is evoked; the temperature of the skin surface starts to rise, at first rapidly, then more slowly until a steady state is reached, despite the fact that heat is still being transferred to it at virtually the same rate. The new state of equilibrium is brought about by reflex action. The sensory impulses evoke a discharge of motor impulses which result in a local increase in the blood flow, thus enabling the heat to be removed as fast as it is received. As soon as the surface temperature of the skin stops rising, the sensation of warmth is no longer perceived despite the fact that the skin is still being irradiated.

By contrast, if the experiment is repeated, using a much greater rate of heat transfer, the subject will almost instantaneously withdraw from the path of the beam and a report of pain will be made even though the actual rise in temperature of the skin surface has not been particularly large.

These observations make it clear, among other things, that the conscious experience evoked by stimulation of the skin is conditional both upon the physical characteristics of the stimulus and upon the conditions prevailing within the skin itself at the time of stimulation. In other words, in respect of heat transfer, the skin does not behave like a thermometer, for the number of impulses reaching the central nervous system from the skin does not merely reflect alterations in the physical state of the environment but also depends upon the changes brought about in the skin itself.

Further, even if the nerve impulses reaching the central nervous system from the skin do not vary, the conscious experience evoked depends upon the precise conditions prevailing in the central nervous system at the time of stimulation. Stimuli which consistently evoke reports of pain in a subject at rest may fail to reach consciousness when the same person is absorbed in a game of chess.

Although the sensory functions of the skin are not limited to arousing conscious experiences, these are of the more immediate interest, primarily because of their well-known relationship to dis-

ease and also because more is known about them than about most of the other sensory functions of the skin.

Physical Background of Common Sensibility.—It was formerly believed that the skin contained a mosaic of discrete nerve terminations divided into four species, each of which was specialized structurally and functionally to serve one of the primary modalities of common sensation; *i.e.*, touch, warmth, cold, and pain.

This concept probably arose directly from the necessary codification of common experience into convenient abstract terms. Moreover, it proved to be a useful working hypothesis, for it led to the recognition that impulses which evoke pain travel up to the brain in a different part of the spinal cord from those transmitting detailed tactile information. However, more critical and refined work (discussed below) has yielded observations which cannot be reconciled with the earlier view.

Anatomy of the Cutaneous Nerve Terminals.—The greater part of the human body in both sexes is covered with hairy skin, although the hairs in certain zones are so fine that they cannot be seen without the aid of a lens. Indeed, the only zones where hairs are entirely absent are the palms of the hands, the soles of the feet, and certain exposed mucous membranes, such as those of the lips, the anus, and parts of the external genital organs. In hairy skin there are only two kinds of nerve termination: (1) compact basketlike networks surrounding the necks of the follicles through which the hair shafts pass and (2) diffuse arborizations of fine filaments which intermingle extensively with one another and end freely and at random among the various tissue elements in the skin; all of these arborizations are anatomically alike. In hairless zones, compact nerve entanglements of diverse shapes and sizes, often enclosed within cellular capsules, take the place of the basketlike networks around the hairs, but it is not possible to classify them into sharply demarcated groups.

These observations strike at the roots of the hypothesis upon which the mechanism of cutaneous sensibility was originally erected. For it is clear that over the greater part of the body there are only two types of nerve terminal to subserve four sensory modes. Further, since it is unlikely that the stimulation of hairs arouses in conscious experience sensations unrelated to touch, the diffuse arborizations must be related, among other things, to the conscious experiences reported as warmth, cold, and pain.

Pattern of Innervation of the Skin.—It has been demonstrated that individual nerve terminals do not have private lines of communication passing directly to the central nervous system. On the contrary, each parent nerve subserves many terminals which are evenly scattered over a relatively large area by means of a mazelike network of nerve fibre bundles in the skin itself. Indeed, the arrangements are such that random destruction of up to one-quarter of the parent nerve fibres in a sensory nerve trunk (a bundle into which parent nerves are gathered as they pass from the skin to the spinal cord) results in degeneration of a small proportion of nerve terminals evenly scattered throughout the area of skin subserved by the whole trunk. Moreover, the subject is usually quite unaware of any alteration in the sensory capacity of the skin involved; certainly none can be detected by the use of such test objects as needles and wisps of cotton wool.

Although little quantitative information concerning the pattern of innervation of human skin had been obtained by the mid-1960s, the innervation of the hairs in the skin of the rabbit ear had been analyzed in some detail, and comparative studies leave no doubt that hairs are innervated on the same principle in man. Briefly, the number of hairs in the rabbit ear is about five to six times greater than the number of parent nerve fibres leaving the ear. In the ear itself each parent nerve is served by numerous branches which come from hairs scattered over areas ranging in size from a few millimetres square to as much as half the ear. Moreover, the area served by one parent fibre always overlaps that supplied by others. Thus even the smallest hairs are served by branches from at least two parent nerves, large so-called guard hairs by as many as 20, and medium-sized hairs by about 5 or 6. It has also been found that individual hairs are connected with parent nerve fibres of slightly different thickness which conduct impulses at different speeds so that they reach the spinal cord at different times. In addition to this, the hairs themselves are so grouped that it is virtually impossible to stimulate a single hair in isolation, unless it be a large guard hair (which is served by many parent axons), by any form of naturally occurring stimulus. Thus, any significant stimulus must result in a highly complex spatiotemporal pattern of impulses reaching the cord. The sensory experience evoked must therefore be the result of the integration of this pattern of impulses with other incoming patterns, the transfer of the new pattern which emerges to higher centres where further integration probably takes place, and the final transfer of resulting activity to circuits responsible for perception.

In the case of the rabbit, it has been calculated that these peripheral arrangements require fewer parent nerve fibres to carry an equivalent amount of information as compared with the system envisaged in the original hypothesis, which assumed that each terminal in the skin had a private line of communication with the central nervous system. A corollary of this is the negligible effect on sensory acuity caused by destruction of up to one-quarter of the parent fibres in any part of a sensory nerve trunk passing between the skin and the spinal cord. Focal disturbances of sensation within the area supplied are not encountered.

Density of Innervation.—This varies considerably from place to place over the body surface. The skin of the back, for example, is not subserved by so many parent nerve fibres as the skin of the face and of the extremities, particularly that of the palms and soles. However, the nerve terminals which end freely within such sparsely innervated areas of skin are almost as numerous as (though they are finer than) those in more densely innervated zones.

Skin which is supplied by a large number of parent nerve fibres is generally capable of resolving fine details in the configuration of the environment. For instance, the fingertips can be used by blind people for reading by the Braille System. Pain evoked from lavishly innervated areas, such as the face, is always reported as being more intense, and it certainly is more likely to cause fainting than that evoked by the same stimulus from less densely innervated areas of skin.

The effect of stimulation on conscious experience is not, however, always directly proportional to the density of innervation of the area concerned. The foreskin and mucous membranes of the glans penis and clitoris are among the most lavishly innervated surfaces of the body, in terms of both parent nerves and the density and size of their terminal ramifications. Yet these regions are virtually "silent" as compared with those covering the trunk when the same stimulus is used. Moreover, the sensation evoked is peculiarly uninformative, as regards both the position and clarity of image which is perceived. Despite this apparent insensitivity it is known that a large outburst of activity follows stimulation of the monkey's glans penis with a wisp of cotton wool even when the animal is fully anesthetized. Division of cutaneous nerve fibres in this region (*i.e.*, circumcision) evokes pain of greater intensity than that which follows an incision of comparable length in skin elsewhere.

It seems, then, that the nature of the conscious perception evoked by naturally occurring stimuli (short of severe injury) in these zones is not dependent upon the sensory impulses evoked but solely upon the state of the nerve cells to which the impulses are conveyed, in the spinal cord or elsewhere. The mechanism apparently is such as to prevent the majority of impulses reaching the level of consciousness except upon specific occasions.

Specificity of Nerve Terminations in the Skin.—It is accepted that the nerve terminals related to hairs and compact nerve terminals help to keep the organism informed about its environment. It is, then, of more than passing interest to find that such endings, as well as the spinal tracts associated with them, are absent in aquatic animals, whose watery environment is relatively homogeneous, and where such information is presumably less necessary to survival. It is not known whether hairs can be stimulated in such a way that nerve impulses specifically related to the terminals evoke reports of pain. It is not possible to cause pain simply by vibrating hair shafts mechanically at

speeds which are either about the same as or greater than the number of impulses the nerves serving them could signal.

Vibration in a single plane, however, is unlikely to stimulate maximally and simultaneously all the nerves which serve as hair. Pulling hairs out of the skin is certainly painful, but this not only damages the nerves of hair shafts but also damages the nerves serving neighbouring structures. The latter nerve endings are of the diffuse type and the ones most likely to be responsible for the pain. However, in certain pathological conditions in which the normal pattern of nervous impulses from hairs in a given area is disturbed, the sensations evoked by their stimulation are referred to as "unpleasant" and at times "painful," but the subject always adds that sensations are not of the kind he experiences in everyday life and he is unable to describe them adequately. In this connection, it is perhaps significant that tactile stimulation in man under local anesthesia of the spinal pathways conducting the impulses from hairs (and other proprioceptive organs; *i.e.*, impulses from deeper structures) gives rise to a report of intolerable pain.

More is known concerning the specificity of diffuse nerve endings, although the evidence concerning their precise function in human skin is indirect. It is possible that, despite their close interrelationship, complex patterned arrangement, and lack of anatomical differentiation, they may be highly specific from a functional point of view. Unfortunately, there is no known area of human skin subserved solely by nerves which end diffusely. They are invariably accompanied by nerves which form compact terminals of one kind or another, which may affect the sensory experience they evoke. However, the transparent covering of the eye, the cornea (*see* EYE, HUMAN: *Anatomy and Physiology of the Eye*), does contain only diffuse nerve endings. It is true that it is structurally unlike skin, but it is highly significant that when it is suitably stimulated, reports in each of the four primary modalities are invariably evoked. Fortunately, the microscopic anatomy of the cornea varies little among the higher animals, including man. Moreover, it is a relatively uncomplicated tissue in that it contains neither blood vessels nor nerves having other than sensory functions. In the cat it has been shown that one and the same nerve ending discharges impulses along its parent nerve when the cornea is stimulated by objects which in man evoke reports related to each of the four sensory modes. The amount of activity evoked by a particular stimulus varies from ending to ending. Moreover, different stimuli evoke different patterns of activity from the same ending. For instance, light touch evokes a short, high-frequency outburst in the case of all active terminals. Injury (pain) evokes a prolonged outburst, the frequency of which depends upon the degree of damage which has been inflicted. Thermal stimuli usually evoke a discharge with a sharp rise and fall in frequency. In the case of some terminals, the activity evoked by warmth is far greater than that evoked by an equivalent amount of cold and vice versa; in the case of others, the responses evoked by either warmth or cold are indistinguishable. More rarely, while tactile and injurious stimuli evoke activity from the terminal, thermal stimuli are ineffective. These observations strongly suggest that stimuli having different physical characteristics evoke different patterns of activity from nerve endings, the specificity being due to their varied environmental conditions.

Conclusions.—The anatomical configuration and arrangement within the skin of the cutaneous nerves and their terminals are such that it is probable that the great majority of stimuli encountered in everyday life evoke a unique space-time pattern of activity which is projected onto the central nervous system. Here the impulse pattern is first integrated with other incoming impulse patterns. The resulting pattern either increases, decreases, or has no effect on the excitatory state of the neurons to which these impulses are conveyed. The result is usually to set up yet another impulse pattern, which is projected to yet another set of sensory cells where further analysis ensues until at last the impulses are perceived as a sensation localized in the skin. The pattern on its way to the brain may become dissipated or transformed at each relay station, the degree of transformation depending upon the excitatory state of the station and the number of sensory impulses which converge upon it at a given time. There are no groups of specially modified terminals in the skin reporting by private lines the local state in respect of four clear-cut modalities. On the contrary, the amount of information available from the periphery is far more complex and varied. The particular information which dictates activity can be divided into four sensory modalities, but this selection is made in the central nervous system as the result of learning. As everyone knows, pain is not a single entity, but acts as a warning signal. It ordinarily follows when an unusually large and prolonged series of impulses from the skin or sensory relay station converge onto the circuits responsible for conscious experience. Warmth and cold can be regarded as preliminary warnings, informing the organism of impending danger. Touch evoked from diffuse endings, such as that produced by an insect crawling over the skin, can also be regarded as a warning signal which alerts the organism against a possible bite.

The arrangement of the nerves and their terminations in the skin of amphioxus resembles that in the human cornea. Moreover, the sensory functions of the skin in these animals appear analogous to that of the cornea. Touch, warmth, cold, and injurious stimuli all cause different reactions, but each can be interpreted as a response designed to avoid danger. *See* also NERVOUS SYSTEM; PAIN; SENSATION; TOUCH. (A. G. M. W.)

SKIN DIVING, diving and swimming underwater at considerable depth without the use of diving apparatus, is the most ancient diving technique and undoubtedly has been practised since the time man learned to swim. Although by strict definition the term skin diving applies only to "natural," or "free," diving, that is, without the use of any underwater breathing apparatus, in general use it also encompasses the closely related sports of diving with self-contained underwater breathing apparatus (SCUBA) and spearfishing. Skin diving is employed by pearl and sponge divers in many parts of the world and skin divers and SCUBA divers are engaged in commercial enterprises, such as underwater construction and salvage, and in scientific exploration of all kinds including oceanography, ichthyology, and archaeology. But the greatest popular interest is in diving as a sport.

The sport was first widely publicized in the 1930s with the publication of Guy Gilpatric's humorous book, *The Compleat Goggler*. In the warm waters off the coasts of southern California and in the Caribbean and the Mediterranean, pioneer skin divers, equipped only with waterproof goggles and a crude hand spear, and holding their breath while swimming underwater, began hunting fish and searching for mollusks and crustaceans. Later, a face mask, giving clearer underwater vision, replaced the goggles. Next came the snorkel, a short J-shaped tube, equipped with a mouthpiece, permitting a diver to swim or float on the surface and breathe while his face was in the water. Swimmers confined to the surface and using their masks for observation are called "snorklers." At the same time swim fins, or "flippers," shaped like duck feet, enabled even mediocre swimmers to swim effortlessly on the surface, and the dive became easier with quicker propulsion down from and back to the surface. Later still 1–2 lb. (450–900 g.) of lead attached to the belt helped journeys into depths of over 30 ft. (9 m.) by canceling the buoyancy of the human body without in any way impeding the rise to the surface. These aids brought skin diving within the ability of more people, and its popularity was greatly increased. But the lack of adequate underwater breathing equipment, or any protection against cold water, limited the sport to a handful of hardy swimmers, and confined its practice to warm southern waters.

With this simple equipment, the early skin divers learned to catch lobsters, shellfish, and octopus with their bare hands, and to stalk and spear fish in their own element. They also learned to explore underwater reefs and coral formations, archaeological remains and the hulks of sunken ships.

The sport continued to attract adherents in the 1940s but the mass invasion of the underwater world came after World War II in the 1950s. Publicity given to frogmen (*q.v.*) of underwater demolition teams, plus tremendous advances in skin-diving equipment, brought thousands of new converts to the sport, and its popularity spread around the world. Also aiding the rapid spread

of the sport were motion pictures such as *The Silent World* and television programs such as *Sea Hunt*.

A World Underwater Federation was formed at Monaco in 1959; the national diving federations represented were Belgium, Brazil, France, Great Britain, Greece, Italy, Malta, Monaco, the Netherlands, Poland, Portugal, Spain, Switzerland, the United States, West Germany, and Yugoslavia. Skin diving clubs were formed in Europe in the 1930s and the first clubs in the U.S., such as the San Diego Bottom Scratchers and the Los Angeles Neptunes, were formed in the 1940s. By the 1960s there were 1,250 organized diving clubs in the U.S. and Canada, including the Dusty Divers in New Mexico, and there were 50 or more clubs in Great Britain. There is no settlement or town of any size on the Mediterranean or Australian coasts without a club and the sport can be considered worldwide.

Development of SCUBA and Other Equipment.—With self-contained equipment the breathing gas is carried in tanks strapped to the diver's body, so the diver is independent of a supply of air or oxygen from the surface and is free to move about without impeding hoselines. The equipment falls into two categories: open-circuit and closed-circuit. While elementary versions of both systems were developed before 1880, the closed-circuit was refined for military applications during World War II. The first popular open-circuit SCUBA was developed by the French Navy, and in 1942 Comdr. J. Y. Cousteau and E. Gagnan took out the first patent for the "Aqua-Lung," a trade name that has often been applied erroneously to all SCUBA. Open-circuit units consist of an air reservoir of from one to three cylinders charged with the breathing mixture (usually compressed air) to a pressure of from 120 to 200 atm., and a demand regulator which feeds air through a hose and mouthpiece to the diver at ambient pressure when he inhales. While SCUBA generally is inadvisable for dives deeper than 130 ft. (40 m.) descents to much greater depths have been made by free divers. Since a human being labours under several physiological handicaps at depth, SCUBA divers should be trained and should not dive alone.

RUSS KINNE—PHOTO RESEARCHERS, INC

FULLY EQUIPPED OPEN-CIRCUIT SCUBA DIVER

The closed-circuit SCUBA (also called oxygen rebreathers or regenerative systems) consist essentially of three parts: (1) a relatively small steel cylinder charged with the breathing gas (usually pure oxygen) to a pressure of from 120 to 200 atm.; (2) a canister of soda lime or other carbon dioxide absorbent into which exhaled gas passes for removal of carbon dioxide; and (3) a breathing bag which acts as a reservoir for oxygen at ambient pressure between the cylinder and inhaling hose. Closed-circuit systems using pure oxygen are limited in depth to 30 ft. (9 m.). Prolonged respiration of a pure oxygen atmosphere at greater pressure ultimately results in convulsion and unconsciousness. Closed-circuit units are suitable for use only by specially trained and experienced divers. Closed-circuit divers have military uses, since they cannot be traced from the surface. They are used for prolonged work only in shallow water, to avoid subjecting the diver to "the bends" (*see* DIVING APPARATUS: *Effects of Air Pressure on the Diver*).

Other than the breathing apparatus, equipment used in SCUBA diving is identical to that required for skin diving. Further accessories include a weight-belt, a depth gauge and wrist watch (both mandatory in deep SCUBA diving), knife, compass, and self-inflating rescue device. (*See* also DIVING APPARATUS.)

At about the same time SCUBA was introduced, permitting extended submersions at greater depths and tremendously increasing the scope of diving, an adaptation of the rubber cold water suits used by navy frogmen was perfected, which permitted skin divers to stay warm and dry while submerged. These "dry" suits permitted skin diving in northern waters and beneath the ice of the Arctic. Later they were replaced by the more practical "wet" type foam rubber suits, which were less cumbersome and gave fewer problems to a diver. The diver's time in the water is still limited, however, by chill of the water.

Underwater Activities.—The mobility and simplicity of SCUBA opened a whole new world of underwater occupations, avocations, sports, and hobbies. SCUBA equipment can easily be transported into remote areas out of reach of conventional deep sea divers. Free swimming underwater, with no connecting lines to the surface, permits divers to work from shore, in shallow turbulent water and over extremely rough terrain.

One of the results was a rapid development of underwater photography. Equipped with pressure-proof still and motion-picture cameras, skin and SCUBA divers record the mysteries of underwater jungles, the colourful beauty of submarine coral gardens, and the excitement of exotic, little-known marine life. Conrad Limbaugh, skin diving scientist for Scripps Institution of Oceanography at La Jolla, Calif., used his camera to record damage to marine habitat from pollution. He also recorded the existence of fish cleaning stations, which large fish visit to be picked clean of skin parasites by smaller fish and other animals in a process called cleaning symbiosis. Cousteau used artificial lighting at great depths to record on film the brilliant colours of deep water marine life which cannot be seen with natural light. Dimitri Rebikoff of France synchronized powerful stroboscopic lights with an underwater motion-picture camera, which enabled him to film the living, moving colour of the depths. Water filters the colour out of sunlight, and below 52 ft. (16 m.) only blues and greens can normally be seen. In the early 1960s Cousteau pioneered the building of underwater stations to which divers can return between dives. By undergoing the decompression process only once just before the final return to the surface (sometimes after several weeks) Cousteau's divers increased their useful time under water from a maximum of about one hour a day to many hours.

SCUBA divers are engaged in underwater gold mining, tapping deep water rivers and streams in the Mother Lode country of the famous California gold rush of 1849. They are searching the hulks of ancient wrecks for sunken treasure and artifacts in all of the oceans and inland waters of the world. Few of their treasure hunts have resulted in quantities of gold, silver, or jewelry, but their discoveries of archaeological and historical significance have been of immeasurable value. Artifacts have been recovered from ancient wrecks in all waters of the world. Bronze Age artifacts recovered from a wreck off Cape Gelidonya, Turk., placed its age in the neighbourhood of 3,300 years, for example, and skin divers probing the bottom of Guatemala's Lake Anatitlán discovered sacrificial offerings of an ancient race of Maya Indians. In other areas of interest, skin diving scientists have discovered rivers of flowing sand on the ocean floor off Cape San Lucas, Mex., and stalagmites and the fossils of prehistoric human and animal bones at the bottom of Florida's Silver Springs.

The impact of skin and SCUBA diving upon the ecology of marine life has yet to be determined, but in some areas conservationists are taking precautions to preserve the natural beauty of submarine gardens by setting aside areas as marine reserves from which no plant or animal life may be taken. Such an area is the Pennekamp Coral Reef preserve, off Florida, which was set aside in 1960 to protect the natural beauty of the coral formations there. (P. R. GL.; L. V. B.)

Spearfishing.—The greatest single underwater sports attraction remains underwater hunting or spearfishing. The sport is of relatively recent origin. It required the invention of a device to extend the diver's range of vision underwater, and waterproof goggles (later replaced by face masks) did not appear until the early 1930s. As mentioned above, the sport was publicized by Guy Gilpatric. From this late start it spread rapidly throughout the world.

The arsenal of underwater weapons ranges from the simple

band spear to guns capable of penetrating the largest fish. The simplest weapon is the Hawaiian sling, a hollow wooden tube with an elastic loop at one end. The shaft, which is tipped by one of a variety of spearheads, is drawn through the tube and pulled back, stretching the loop. When released, the shaft is propelled forward. In the mid 1930s, Alec Kramarenko was granted a patent on an underwater gun in which the spear was propelled by a compressed spring. Shortly thereafter appeared a propulsion type gun invented by a Frenchman, Maxime Forlot and the popular spear gun propelled by a rubber elastic band designed by his compatriot, Georges Beuchat.

The foregoing are rifle-type weapons in which the spear either travels through a long barrel or is guided along its upper surface. In all cases, the spear shaft is released by a trigger mechanism. Early guns and carbon dioxide and compressed air types have a line attached to the shaft making it easier to retrieve. On strike, the line holds the quarry securely. A pneumatic air gun was invented in 1956 by Juan Vilarrubis of Spain and became popular because of its accuracy, power, and simplicity of operation. The targets include man-eating sharks, vicious barracuda and huge fish weighing several hundred pounds, as well as the ordinary game prized by fishermen everywhere. In most cases, stalking and shooting his quarry underwater is only the beginning of a skin diver's battle. After a fish is hit, it must be held on a harpoon line and landed. With large fish this may mean an underwater "ride" as the diver is towed through the water.

Underwater spearfishing is so popular among diving clubs that local, national, and international competitions are held each year. Local champions compete for the national championship and the privilege of representing their country in international competition, where world championships are determined. The competitions do not permit the use of SCUBA, and the contestants dive while holding their breath. Champions such as Jim Christiansen and Terry Lentz of the United States, Claudio Ripa of Italy, and Bruno Hermany of Brazil proved divers could hold their breath more than three minutes and work at depths exceeding 100 ft. In 1960 Enzo Majorca of Italy set a record depth for free diving. Holding his breath, and diving in the Mediterranean, he reached the depth of 51 m. (approximately 170 ft.).

(P. R. Gl.; L. V. B.; V. El.)

Bibliography.—L. Bridges and W. Barada, *Mask and Flippers* (1960); J. Y. Cousteau *et al*, *Complete Manual of Free Diving* (1957); J. Y. Cousteau and F. Dumas, *The Silent World* (1953); Guy Gilpatric, *The Compleat Goggler* (1938, rev. ed. 1957); I. S. Ivanović, *Modern Spearfishing* (1954). (P. R. Gl.; L. V B.)

SKIN GRAFTING: *see* Burns; Plastic Surgery; Transplants, Tissue and Organ.

SKINNER, OTIS (1858–1942), U.S. actor, who played hundreds of roles in all parts of the world in a career extending over more than 60 years, was born in Cambridge, Mass., June 28, 1858, the son of a minister. His first stage appearance was at the Philadelphia Museum in 1877 in *Woodleigh*. After two years in the stock company at the Walnut Street Theatre in Philadelphia, Skinner made his New York debut at Niblo's Gardens, Sept. 4, 1879, in *The Enchantment*. During the next five years he developed a classical repertory and a successful acting style, first with Edwin Booth, at Booth's theatre, and then, for three years, with Lawrence Barrett. In 1884 he joined Augustin Daly's company at Daly's theatre, remaining with it for four years. He made his London debut in 1886 with Daly's company, playing at the Strand Theatre. After two years with the Booth-Modjeska company, he became, in 1892, leading man opposite Helena Modjeska. In 1903 he starred with Ada Rehan. In addition to his Shakespearean roles, Skinner's chief successes were in *Kismet*, which he played between 1911 and 1914, and *Blood and Sand* (1921), in which he played the matador, Juan Gallardo. Skinner was active in the theatre until his death on Jan. 4, 1942. He was the author of *Footlights and Spotlights* (1924) and *Mad Folk of the Theatre* (1928).

Skinner's daughter, Cornelia Otis Skinner (1901–), actress and monologist, was also the author of many sketches, a play, *Captain Fury*, verse, and light essays. (S. W. H.)

SKINNER'S CASE, the name usually given to the celebrated dispute between the House of Lords and the House of Commons over the question of the original jurisdiction of the former house in civil suits.

In 1668 a London merchant named Thomas Skinner presented a petition to Charles II asserting that he could not obtain any redress against the East India Company, which, he asserted, had injured his property. The case was referred to the House of Lords, and Skinner obtained a verdict for £5,000. The company complained to the House of Commons which declared that the proceedings in the other house were illegal. The Lords defended their action, and after two conferences between the houses had produced no result the Commons ordered Skinner to be put in prison on a charge of breach of privilege; to this the Lords replied by fining and imprisoning Sir Samuel Barnardiston, the chairman of the company. Then for about a year the dispute slumbered, but it was renewed in 1669, when Charles II advised the two houses to stop all proceedings and to erase all mention of the case from their records. This was done and since this time the House of Lords has tacitly abandoned all claim to original jurisdiction in civil suits.

SKIPTON, a market town and urban district in the Skipton parliamentary division of the West Riding of Yorkshire, Eng., on the Leeds and Liverpool canal, 43 mi. E. of York by road. Pop. (1961) 13,008.

Lying in the hilly district of the upper valley of the Aire where it is joined by the Eller Beck from the north, it is the nodal town of the Craven gaps, the lowest ways through the Pennines between Trent and Tyne, and so has become an important route centre and the principal market town of Craven. Weekly sales of cattle and sheep take place.

During the Middle Ages the Craven gaps were the chief routes through the Pennines along which the Angles, and later the Danes, moved westward, and Skipton was the capital of Craven. At the Norman Conquest, it became part of the possessions of Earl Edwin and was granted to Robert de Romille, who built the castle in the 11th century. It was taken by the parliamentary forces in 1645 after a three-year siege, partly demolished in 1648, restored by Lady Anne Clifford in 1657–58, and now consists chiefly of the gateway with its two round towers and an original Norman arch to the inner court. In the grounds are the remains of the castle chapel. The parish church of the Holy Trinity, mainly Perpendicular, was also partly destroyed during the Great Rebellion but has been restored. The grammar school was founded in 1548.

There are rayon and silk factories in the town and a large limestone quarry nearby.

SKOBELEV, MIKHAIL DIMITRIEVICH (1843–1882), Russian army officer, one of the conquerors of Turkistan and a prominent commander also in the Russo-Turkish War of 1877–78, was born in St. Petersburg, on Sept. 29 (new style; 17, old style), 1843. After graduating from the General Staff Academy in St. Petersburg, he was sent to Tashkent in 1868. With the exception of two years' service in the Caucasus, he remained in Turkistan until 1877. In 1873 he took part in the expedition against Khiva, commanding the advance guard of Col. N. P. Lomakin's column, which marched on the khanate from the Mangyshlak Peninsula on the Caspian Sea. In 1875–76 he held a command in Gen. K. P. Kaufmann's expedition against Kokand, earning promotion to the rank of major general and appointment as the first Russian governor of Fergana.

When war broke out between Russia and Turkey, Skobelev was transferred to the European front. He captured the bridge over the Seret at Barboşi in April 1877 and crossed the Danube in June. He commanded the Caucasian Cossack brigade in the attack of the Green Hills at the Second Battle of Pleven (Plevna). He captured Lovcha on Sept. 3, and distinguished himself again in the Third Battle of Pleven. In command of the 16th division, he took part in the investment of Pleven and also in the fight of Dec. 10, when Osman Pasha surrendered. In January 1878 he crossed the Balkans in a severe snowstorm, defeating the Turks near Shipka Pass and capturing 30,000 men and 90 guns. Later he captured Edirne and San Stefano. Dressed in a white uniform and mounted on a white horse, and always in the midst of the battle he was known by his soldiers as the "White General."

Skobelev subsequently returned to Turkistan. He captured Göktepe on Jan. 24, 1881, and, after slaughtering about 8,000 people, reduced the Akhal-Tekke country to submission. He was advancing on Ashkhabad when he was recalled and given the command of the Minsk Army Corps. In the last period of his short life he was active in politics. He made speeches in Paris and in Moscow at the beginning of 1882 in favour of a militant Pan-Slavism, predicting inevitable conflict between German and Slav. He was at once recalled to St. Petersburg. He died suddenly of heart disease in Moscow on July 7 (N.S.; June 25, O.S.), 1882.

ŠKODA, EMIL VON (1839–1900), Czech engineer and industrialist who founded one of the most important industrial complexes of Europe, including the Škoda armament works, was born at Cheb (Eger) on Nov. 19, 1839. He was the nephew of Josef Škoda, a celebrated professor of medicine in the University of Vienna. Škoda studied engineering in Germany and in 1866 accepted the position of chief engineer in a small machine factory in Plzeň (Pilsen) that belonged to Count Ernst von Waldstein. Three years later Škoda himself bought this factory, which at the time employed only 33 people. He expanded the undertaking at a great pace in the 1880s, building a railway connection with the main Vienna–Cheb line in 1886, and finally adding an armaments factory in 1890. In 1899 he founded a Škoda works company, under his own direction. At the time of his death on Aug. 8, 1900, the Škoda factories employed some 4,000 workers. (E. Wi.)

SKOKIE, a residential and light industrial village of Cook County, in northeastern Illinois, U.S., is on the northern edge of Chicago and west of Evanston. Called Niles Center until 1940, Skokie was first settled in 1834 and attracted a colony of settlers from Luxembourg. By the time of its incorporation in 1888 the trading centre was known primarily for its greenhouse industry, started during the previous decade.

In the late 1920s a development boom occurred in the community following the extension of service by the Chicago and North Western and the Chicago, North Shore and Milwaukee railroads. The depression of the 1930s halted the growth of the village but World War II began a great influx of industry. The community nevertheless retained much of its earlier residential character, and residential and commercial building (including extensive shopping facilities) increased sharply in the postwar years. Between 1950 and 1964 the population increased from 14,832 to 67,865. (For comparative population figures *see* table in ILLINOIS: *Population.*)

The village adopted a council-manager form of government in 1957. (Ar. C. H.)

SKOPJE (Serbo-Croatian, SKOPLJE; Turkish, USKUB; Albanian, SHKUP; ancient, SCUPI and JUSTINIA PRIMA), the capital city of the Socialist Republic of Macedonia, Yugos.; an episcopal see for both the Serbian Orthodox and Roman Catholic churches; and one of two centres of the Muslim administrative and supervisory body for Yugoslavia.

The city was almost entirely destroyed by an earthquake on July 26, 1963. It stood on the Vardar river in the midst of mountainous country, 210 mi. S.E. of Belgrade. Pop. (1961) 171,893, of mixed origin. The old part of the city was on a slope on the left bank overlooked by an ancient fortress with, north of it, a Roman aqueduct. Other old buildings were the medieval Turkish Kursumlihan (caravansary) and the mosques of Sultan Murad II (1430) and the ghazi ("warrior") Isay Bey (1476).

Modern Skopje, begun in the early 19th century, developed with the construction of the Belgrade-Salonika railway through the town; but most building was done after World War II. The old city was linked with the new part on the right bank, across the old bridge. Outstanding of the modern buildings were the Macedonian parliament, the university (founded 1946) and the bishops' palaces. Skopje's position made it an important road junction for the whole Balkan peninsula. Industries included textiles, steel, ceramics, chemicals and glass; local metalliferous ores are mined. It was long a centre of silver filigree work and other handicrafts. Three miles southwest is the monastery of St. Panteleimon (1164) with fine frescoes.

Ancient Scupi, a city of Moesia, became the capital of Dardania under the Roman emperor Diocletian in the 4th century A.D. Justinian I rebuilt it after its destruction by an earthquake in 518. The Serbs first seized Skopje in 1189; their self-proclaimed emperor Stephen Dushan was crowned there in 1346. In 1392 the Turks made it their capital. It declined during the 18th century but revived during the 19th, especially with the coming of the railway at mid-century. By the London and Bucharest treaties of 1913 it was incorporated into Serbia; and in 1918 became part of the new state of Yugoslavia. During World War II the Germans occupied it in April 1941 and it was then garrisoned by Bulgarian troops. In 1945 it was made capital of Macedonia in the Federal People's Republic of Yugoslavia.

The earthquake of 1963 killed more than 1,000 of Skopje's inhabitants and rendered 120,000 of them homeless. Relief in cash and in kind, including medical, engineering and building teams and supplies, was immediately dispatched from many countries. (V. Dr.)

SKUA, a rapacious sea bird, *Catharacta skua*, of the family Stercoraridae, related to gulls. (Closely related are the three species of jaegers [*q.v.*], sometimes grouped with the skua in a common genus, *Stercorarius*, in which event all are referred to as skuas.)

The great skua, or bonxie (*C. s. skua*), nests on the ground of tundra from the Faeroes, Shetland and Orkney Islands north-

ERIC HOSKING

SKUA (CATHARACTA SKUA) AT NEST

ward to Iceland. In the southern hemisphere distinct geographic races inhabit the tip of South America, various subantarctic islands and the Antarctic continent. The species is thus "bipolar," a unique phenomenon among birds. The southernmost form *C. s. mccormicki*, is pale-coloured and strongly distinct from the others. It has been encountered far in the interior of the icy continent, closer to the pole than any other animal.

Skuas lay two eggs but usually devour one of the chicks soon after it hatches. They have a hawklike plumage, with a conspicuous whitish patch at the base of the wing quills, and talons on their webbed feet. In flight they suggest a small eagle rather than a gull. Their beating flight is surprisingly strong, a fact hardly evident until they are seen to overtake longer-winged and supposedly swifter species.

On breeding grounds skuas are fearless of man. They are the scourge of neighbouring birds and (in the north) also of lemmings. The southern races prey upon petrels and the eggs and chicks of penguins. At sea, in the nonbreeding period, they capture living organisms for themselves but also pursue and harry other sea fowl, forcing them to disgorge their catches. (R. C. Mu.)

SKULL. The skull comprises the bony structures of the head. Its main parts are the cranium, or brain case (eight bones),

and the facial skeleton (14 bones, including the upper and lower jaws). The function of the skull, apart from providing for the suspension of the jaws, is to protect the brain and the organs of special sense: the nose, eyes, and ears. The skull is poised on the first, or atlas, vertebra in a manner which permits nodding movements of the head. Rotation of the head is carried out by movement of the ringlike atlas around the peglike odontoid process of the second, or axis, vertebra.

The following account of the human skull considers its structure, the infant skull, evolution of the skull, and the skull in anthropology.

Structure.—The adult skull has a dome-shaped brain case composed of the frontal bone over the forehead; the parietals at the top and upper sides; the temporals and sphenoids at the lower sides; and the occipital at the back (*see* fig. 1 and 2). The bones of the cranial vault are joined to each other at the sutures, which are actually fibrous joints and are shown in the figures as serrated lines between the bones. They serve in the young skull to join the bones together while allowing a slight degree of movement between them—an arrangement conferring some elasticity on the skull. This elasticity, together with the generally convex shape of the bones, enables the cranium to withstand blows of considerable force. The sutures become progressively and slowly obliterated by the formation of a bony union between the individual bones, this process beginning in the young adult skull and being advanced by the time of middle age.

Frontal Aspect.—The human forehead (*see* fig. 1) is smooth and almost vertical; it lacks the well-developed brow ridge of the great apes. The orbits, or eye sockets, roughly quadrilateral in shape, face forward and slightly outward; this position of the eyes in man is of great importance since it permits an overlap of the visual fields that makes true stereoscopic vision possible. This overlap is also favoured by the great reduction of the nose area (corresponding to the projecting snout of other animals, which causes their eyes to be placed far over on the sides of the head). The orbits of the human skull are bounded at the outer edge by the zygomatic bone. The orbital cavities are conical in shape and taper as they are traced deep into the skull. At the apex of the orbit are several holes through which the nerves to the eye and to the eye muscles gain entrance to the orbit. The bones forming the inner walls of the orbit are thin and delicate; those of the roof and floor are more massive.

In the centre of the facial skeleton is the external orifice of the nasal cavities. The bridge of the nose is made by the flat nasal bones and the rest of the nasal orifice is bounded by the maxillary bones (*see* below). The external projection of the nose on the face of a living person is due almost entirely to cartilaginous structures that are not properly a part of the skull. The greater part of the nasal cavities lies deeply buried within the skull. On the outer wall of each nasal cavity are two scrolls of bone, the turbinate bones; in life the scrolls are covered by a thick mucous membrane which is believed to assist in warming the incoming air before it passes into the lungs. In the midline between the nasal cavities is the bony nasal septum (partition).

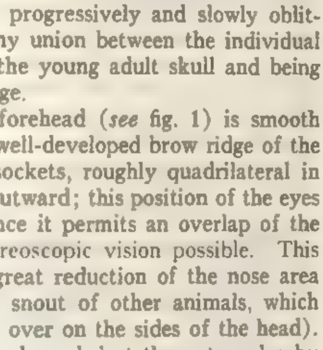


FIG. 1.—HUMAN SKULL FROM FRONT

The maxillary bones (one maxilla on either side) make up the main mass of the upper jaw and carry 16 teeth in the adult. Internally the maxillary bones are hollowed out to form the maxillary sinuses, which are lined with mucous membrane continuous with that of the nose.

The mandible is the single large bone making up the lower jaw; like the maxillaries it bears 16 teeth. Holes in the frontal bone, maxillary bones, and mandible provide for the exit of branches of the trigeminal nerve passing to the skin of the face.

Lateral Aspect.—At the outer margin of the orbit the zygomatic bone sends backward a prolongation which meets a similar process of the temporal bone to form the zygomatic arch (*see* fig. 2). The space between the zygomatic arch and the side of the brain case allows the powerful temporalis muscle to gain attachment to the mandible; its function is to close the jaws. At the posterior end of the zygomatic arch is the external auditory canal, or ear hole; its walls are made up chiefly of a tubular bone, the tympanic. A slender, pointed bony structure projecting down from the base of the skull from a point just behind the tympanic bone is the styloid process. A fibrous band, the stylohyoid ligament (indicated by dotted lines in fig. 2), is found in the living person attached above to the tip of the styloid process and below to the hyoid bone (*see* below, *Evolution of the Skull*). Behind the external auditory canal is a bulging area of bone called the mastoid process. It is a typical feature of the human skull and contains many small cavities, or air cells, continuous with the cavity of the middle ear. These cells are readily infected and may give rise to mastoiditis (*q.v.*).

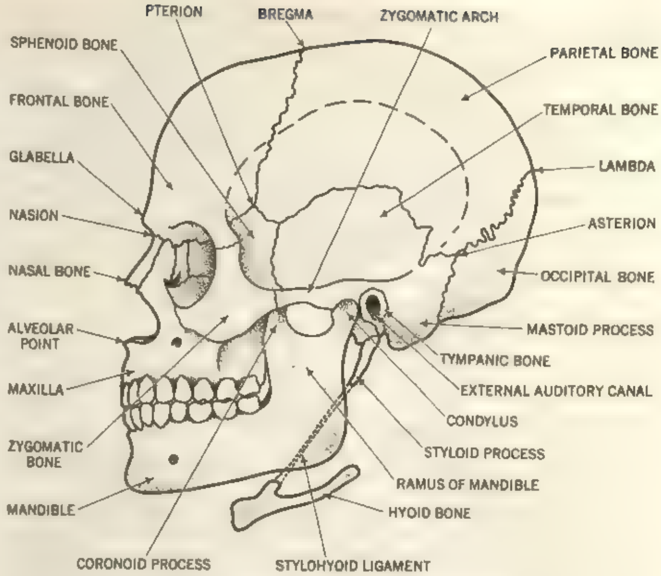


FIG. 2.—HUMAN SKULL FROM LEFT SIDE

The shape of the mandible can be recognized more clearly from the side. It consists of a horizontal body which carries the teeth on its upper surface and, posteriorly, an ascending portion, the ramus. At the upper margin of the ramus are two prominences, one in front called the coronoid process and one farther back called the condylus. The condylus is rounded and coated with cartilage in life. It fits into a depression in the undersurface of the temporal bone to form with it a movable joint, the temporomandibular joint, which permits opening and closing of the mouth and also side to side movement as in chewing.

Of special note is the pterion, the point of juncture of the frontal, parietal, and sphenoid bones. The conformation of the sutures at this point is very distinctive in the human skull and is of interest when comparing the human skull with that of closely related species such as the anthropoid apes. From the surgeon's standpoint the pterion is an important locus, since it overlies the large middle meningeal artery inside the skull; this artery is frequently torn as a result of a blow on the side of the head.

Sagittal Aspect.—A vertical section through the middle of the skull (fig. 3) shows the relative thinness of the bones of the vault of the skull overlying the brain. In a young person these bones consist of three layers, an inner and outer table of compact bone with a middle layer of bone marrow. The latter layer, called the diploe, disappears with advancing age. When a blow is administered to the skull the inner table is apt to be more severely damaged than the outer table and sharp slivers of bone may be driven into the brain. Moreover, since the dome of the skull is somewhat elastic, at least in the young person, the force of the blow may be transmitted around to the base of the skull, producing a fracture there.

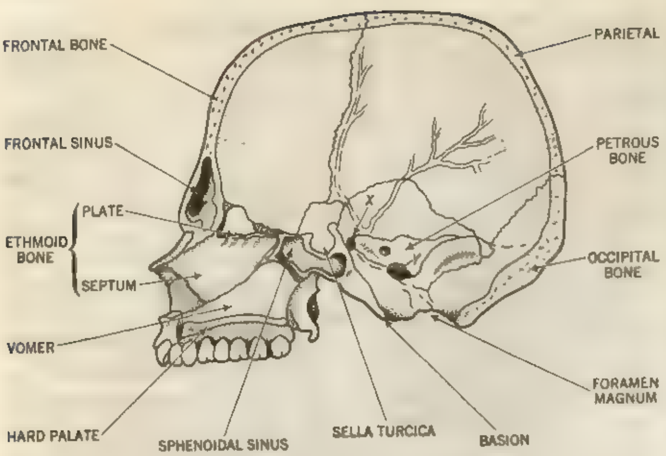


FIG. 3.—INTERIOR OF HUMAN SKULL IN VERTICAL SECTION

The base of the skull, from front to back, is made up of basal portions of the frontal, ethmoid, sphenoid, temporal, and occipital bones. The base supports the weight of the brain, which is protected by fluid-filled spaces between it and the base of the skull. The sectional view also shows the large air sinuses within the frontal bone above the nasal cavity and within the body of the sphenoid. These sinuses are continuous with the nasal cavity. Behind the frontal bone is the ethmoid bone. In the floor of the skull the ethmoid forms a perforated plate through which pass the nerves to the nose. It also has an upward projection, the crista galli, and a downward projection which forms the upper part of the nasal septum. The upper surface of the sphenoid bone is hollowed out to form the sella turcica in which the pituitary gland lies. This important gland, on which the proper functioning of all other endocrine glands depends, is thus situated in an invulnerable place in the skull; its inaccessibility makes a surgical approach very difficult.

A massive portion of the temporal bone in the floor of the skull, called the petrous bone, is the hardest bone in the body. It houses the internal mechanisms of the ear. On the inner side of the petrous bone is a hole for the passage of the auditory nerve. The medulla oblongata (basal portion of the brain) and the spinal cord are continuous with each other through the foramen magnum, the large hole in the base of the skull. The great internal carotid artery gains access to the skull at the point indicated in fig. 3 by X and the great internal jugular vein leaves the skull at the point indicated by Y.

It has been noted that the ethmoid bone takes part in the formation of the nasal septum. The rest of the septum consists of the vomer and the maxillary bones. The hard palate in the roof of the mouth is made up of the maxillary bones and the palatine bones.

The Infant Skull.—The skull of a newborn baby (*see* fig. 4) shows clearly the multiple bones in the brain case. At this stage

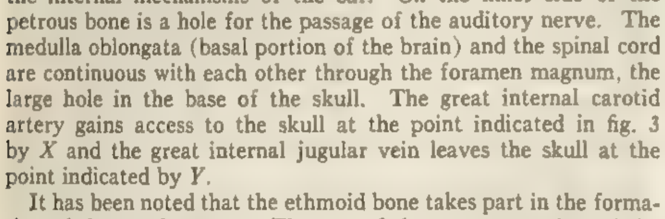


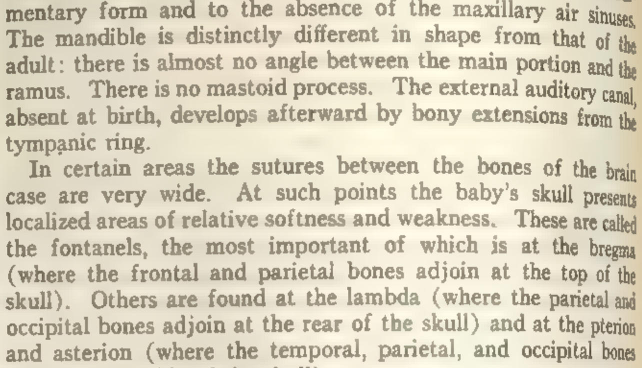
FIG. 4.—SKULL OF NEWBORN INFANT FROM LEFT SIDE

the bones are widely separated by sutures which thus confer great flexibility on the skull and permit overriding of the bones upon each other during the process of birth. The facial skeleton is quite small by comparison with the brain case. The maxillary bones are very small, due to the absence of teeth except in rudimentary form and to the absence of the maxillary air sinuses. The mandible is distinctly different in shape from that of the adult: there is almost no angle between the main portion and the ramus. There is no mastoid process. The external auditory canal, absent at birth, develops afterward by bony extensions from the tympanic ring.

In certain areas the sutures between the bones of the brain case are very wide. At such points the baby's skull presents localized areas of relative softness and weakness. These are called the fontanels, the most important of which is at the bregma (where the frontal and parietal bones adjoin at the top of the skull). Others are found at the lambda (where the parietal and occipital bones adjoin at the rear of the skull) and at the pterion and asterion (where the temporal, parietal, and occipital bones adjoin on the side of the skull).

Evolution of the Skull.—The head of a hypothetical animal which may have preceded the vertebrates is illustrated in fig. 5A. The head already possesses a well-developed brain, two nasal pits, two eyes, and two ears, the latter being concerned only with balance. Instead of a bony vertebral column there is a semirigid rod called the notochord which traverses the full length of the animal below the spinal cord and ends in front just behind the pituitary gland (fig. 5A).

At a later stage in the evolution of the skull (fig. 5B) certain



AFTER "HISTORY OF THE HUMAN BODY" BY H. H. WILDER, HENRY HOLT & CO., N.Y.

FIG. 5.—DIAGRAMS SHOWING THREE STAGES IN THE EVOLUTION OF THE SKULL: (A) HEAD OF A HYPOTHETICAL PREVERTEBRATE ANIMAL; (B) LATER STAGE OF ANIMAL WITH VERTEBRAL COLUMN; AND (C) VERTEBRATE WITH PRIMITIVE SKULL

additions are made to the head. Cartilaginous capsules partially invest the nasal pits and the eyes and completely invest the ears. The need for added rigidity is met by the development of two rods of cartilage on either side, the cranial bars (trabeculae) in front and the parachordal cartilages at the back of the head.

At a still later stage (fig. 5C) a general consolidation of the cartilaginous elements of the primitive skull takes place by the fusion of the cranial bars and the parachordal cartilages into a single cartilaginous cranial base and the fusion of this in turn with each of the sense-organ capsules. Further extensions of the cartilaginous skull also take place above the brain.

This sequence of events in the evolution of the skull can be deduced with a fair degree of confidence since it is essentially what takes place in the development of the embryos of all vertebrates, including man. Speculation about the skull becomes fact with the conditions present in the skulls of a large group of fish having cartilaginous skeletons, the elasmobranchs. These primitive fish, which include the sharks and rays, have skulls essentially similar to that depicted in fig. 5C. The next stage in the evolution of the skull is the appearance of added protection for the head in the form of bony plates. The living primitive fish of the ganoid group, which includes the sturgeons and gars, have this armour plating over the head and some of the body. The bony plates of the skull are actually developed from the connective tissues of the skin and are called dermal bones. These bones be-

come extremely numerous in the bony fish known as teleosts, the group which includes most of the ordinary fish. A striking feature of the dermal bones in all vertebrates above the fish is their development not on the surface of the head but deep in the connective tissues under the skin. This is the way they develop in the human embryo.

The same evolutionary sequence is also marked by important changes in the cartilaginous skull lying inside the outer skull of dermal bones. Centres of bone formation appear within the cranial base and the sense-organ capsules and convert them into solid bone. This change is foreshadowed in the elasmobranchs and is well advanced in the teleosts. At the same time there are varying degrees of fusion between the dermal bones and between these and the underlying cranial base and the sense-organ capsules. In essence this completes the story of the evolution of the skull. There is, however, an infinite variety in the detailed arrangements of the elements of the skull in different groups of animals.

There remains to be considered the so-called visceral skeleton. This includes the jaws and the hyoid apparatus (*see* fig. 2). In the fish the gills are supported by cartilaginous bars, the branchial, or visceral, arches. The upper jaw consists of a plate of cartilage, the palatoquadrate bar; the lower jaw consists of Meckel's cartilage. The jaw joint is formed between the quadrate and the articular cartilages, which are merely the hind ends respectively of the palatoquadrate and Meckel's cartilages.

In the ganoids and teleosts new dermal bones are laid down around the cartilages of the jaws; the articular and quadrate cartilages continue to form the jaw joint. In higher vertebrates there are profound modifications in the visceral skeleton of the skull. A new joint is developed between the dermal lower jaw and the temporal bone. The quadrate and articular cartilages are thus no longer necessary. They do not disappear, however, but are apparently transformed into the small bones of the ear, the quadrate becoming the anvil bone (incus) and the articular becoming the hammer bone (malleus).

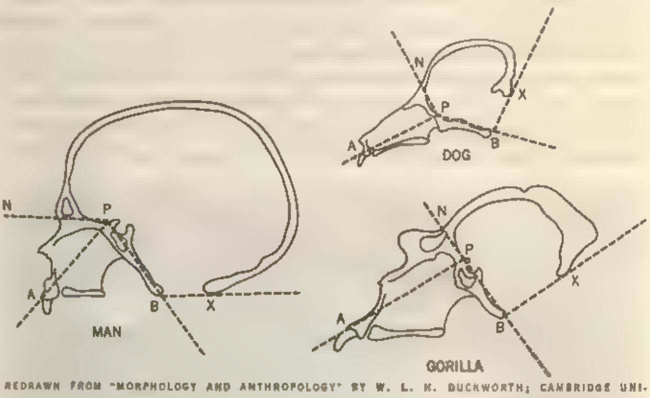
The remaining gill arches of the fish also become greatly modified in higher vertebrates with the disappearance of gills and the development of lungs. The second and third gill arches become transformed into the hyoid bone together with the styloid process and the stylohyoid ligament (*see* fig. 2). The stirrup bone (stapes) of the ear appears to be developed from the second arch. The remaining arches become modified in man as the cartilaginous framework of the larynx, or voice box.

The human skull, then, may be regarded as consisting essentially of an inner skull, very ancient from the evolutionary point of view, which includes the base of the skull and the sense-organ capsules; and an outer, more recent skull, made up of the so-called dermal bones. The dermal bones are typically found in the dome of the skull, the face, and the mandible. They have proved extremely plastic in adapting themselves to the enormous development of the brain that characterizes man.

The Skull in Anthropology.—The student of fossilized remains who wishes to establish their possible affinities with man must keep clearly in mind exactly what criteria make a skull typically human. Doing so necessitates a thorough knowledge of the many natural variations of the skull in different groups and races of mankind and in modern man's progenitors. (*See* RACES OF MANKIND; ANTHROPOLOGY, *Physical Anthropology.*) Some of the distinctly human features are reviewed below.

The capacity of the brain case is relatively and absolutely large in all known races. If the capacity is less than 1,350 cc. the skull is called microcephalic; if 1,350 to 1,450 cc., mesocephalic; and if over 1,450 cc., macrocephalic. Another characteristic feature of the human skull is the smooth, almost vertical forehead, lacking the pronounced brow ridges found in many apes and certain extinct forms of man. There is a marked concavity above the canine (cuspid) tooth in the upper jaw, the canine fossa, which is absent or is replaced by a convexity in apes and in extinct man. The mandible is characterized by a definite angle between the ramus and the horizontal portion. The condylus of the ramus is also typical. The bony prominence of the chin is a distinctive human trait, as is the absence of a shelf of bone at the back of the chin known as the simian shelf, found in all apes but generally absent in the mandible of early man. The mastoid and tympanic region and certain sutural areas, such as the pterion, also serve to distinguish the human from the anthropoid skull.

The human skull differs markedly from those of other animals in the way it is poised upon the vertebral column and in the proportions of the face. This is illustrated in fig. 6, where the skulls of man, gorilla, and dog are compared (cross-section views), certain planes of reference being indicated by dotted lines. The plane of the foramen magnum (BX) is horizontal in the human skull, slightly inclined to the horizontal in the gorilla, and almost vertical in the dog. This indicates that in man, who is truly an erect animal, the skull is perfectly balanced on the top of the spine. In the gorilla this is not so and powerful muscles are developed at the back of the neck to counterpoise the weight of the face with its massive jaws. In the dog, which walks on all fours, the skull is held at right angles to the axis of the spine (BX). Other angles, indicated in fig. 6, may be of value in establishing that a skull is not anthropoidal. The sphenethmoidal angle (NPB) is drawn by projecting a line (PN) forward in the plane of the ethmoid plate from a point just behind this plate known as the prosphenion (P) and another line (PB) from the latter point passing through the axis of the base of the skull to the front of the foramen magnum (B). The difference in this angle in the skulls of man, gorilla, and dog is striking. In man the plane of the ethmoid plate has been rotated downward and forward, a change associated with the enormous development of the frontal lobes of the brain. The sphenomaxillary angle (APB), determined by adding a line (PA) from the prosphenion to the alveolar point (A) just above the upper incisor teeth, indicates that there has been a striking reduction in the projection of the face, or snout area, in man together with some rotation of the maxillary area downward.



REDRAWN FROM "MORPHOLOGY AND ANTHROPOLOGY" BY W. L. H. DUCKWORTH; CAMBRIDGE UNIVERSITY PRESS

FIG. 6.—FACIAL ANGLES IN THE SKULLS OF MAN, DOG, AND GORILLA

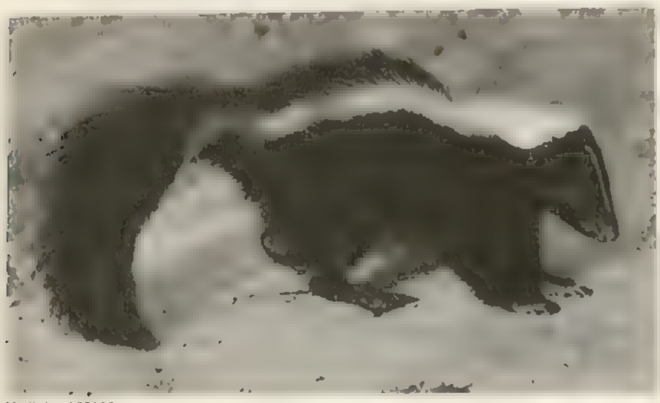
Certain cranial indexes can be calculated from measurement of the diameters of the skull. Large calipers are used for this purpose. The length of the skull is the distance from the glabella (the midpoint between the brows; *see* fig. 2) and the most projecting point at the back of the head. The breadth of the skull is the distance between the most projecting points at the sides of the head, usually a little above and behind the ears. The cephalic index is the breadth × 100 divided by the length. An index of less than 75 means that the skull is long and narrow when seen from the top; such skulls are called dolichocephalic, and are typical of native Australians and the Kaffirs. An index of 75 to 80 means that the skull is nearly oval; such skulls are called mesaticephalic and are typical of Europeans and the Chinese. A skull having an index of over 80 is broad and short, and is called brachycephalic; such skulls are common among Mongolians and the Andaman Islanders. Other cranial indexes are the index of height and the gnathic, or alveolar, index. The former is calculated as the height of the skull from the basion (front of the foramen magnum; *see* fig. 3) to the bregma (*see* fig. 2) × 100 divided by the breadth. The latter is reckoned as the basialveolar length (from basion to alveolar point; *see* fig. 2 and point marked *A* in fig. 6) × 100 di-

vided by the basinasal length (from basion to nasion [root of the nose; *see* fig. 2]) and gives a measure of the degree of snout projection (prognathism).

See also references under "Skull" in the Index.

BIBLIOGRAPHY.—Sir Henry Morris, *Human Anatomy*, 11th ed. (1953); A. S. Romer, *The Vertebrate Body*, 2nd ed. (1955); L. S. B. Leakey, *Adam's Ancestors*, 4th ed. (1960); L. B. Arey, *Developmental Anatomy*, 7th ed. (1965). (J. Ds.)

SKUNK, any of the several species of the weasel family (Mustelidae) noted for the offensive odour produced by glands on either side of the anal opening. The animal can eject the yellow, odoriferous liquid as far as 12 ft. Skunks, often called "polecats," belong to the subfamily Mephitinae and are exclusively American. The typical striped skunk (*Mephitis mephitis*) is cat-sized, squat, and bushy tailed; it has a shaggy, glossy black coat



JOHN H. GERARD

COMMON STRIPED SKUNK (MEPHITIS MEPHITIS)

marked by a white stripe (sometimes paired) extending from the base of the head to, and sometimes including, the tail. The striped and closely related hooded (*M. macroura*) skunks are found from central Canada to northern Mexico; other *Mephitis* species occur southward to Nicaragua. The smaller spotted skunk (*Spilogale putorius*), unlike other skunks, climbs well. Its black coat is marked with four to six white, sometimes broken stripes. Spotted skunks (including the *S. gracilis*, *S. pygmaea*, and *S. angustifrons*) occur over most of the United States southward to Costa Rica. The hog-nosed skunks (*Conepatus mesoleucus*, *C. leuconotus*, and others) occur from the southern United States into South America. Their pelage is black except for the white back and tail.

Skunks are omnivorous. They feed on small arthropods, mice, rats, eggs, birds, and plants; the several species show somewhat different food preferences. Skunks are calm, deliberate, amiable creatures seldom molested by other animals because of their repulsive scent. They usually amble along, rarely proceeding at a fast gait. Most skunks accept man's presence and will establish a den under a building as well as in natural sites like tree stumps, caves, rock outcroppings. In the striped skunk four to seven young are produced in the spring, about two months after mating.

When deodorized, skunks make delightful and affectionate pets. The pelt is of value commercially and is usually plucked and dyed to simulate more precious furs; hence skunks are both trapped and raised on fur farms. The scent is persistent and when treated is useful as a perfume base. (K. R. KN.)

SKUNK CABBAGE (*Symplocarpus foetidus*), a fleshy herbaceous plant of the arum family (Araceae), so called because of its fetid odour and large leaves, native to eastern North America and northeastern Asia.

The plant grows in swampy places and in very early spring (March or sometimes February) sends up from thick rootstocks grotesque, swollen, shell-like, purple-brown spathes, each enclosing many small flowers borne in a short thick cluster. These are soon followed by numerous ovate leaves, one to three feet long, and later by large globular masses of fleshy berries. The similar western, or yellow, skunk cabbage (*Lysichitum*) occurs in western North America.

SKY, surrounding space as seen above the horizon. From the surface of the earth it has the appearance of a great dome, low and rather flat. When clouds fully obscure the space beyond, the impression of flatness is accentuated. When unclouded, the dome seems closer to the observer at a point directly overhead. This article is concerned with the physics of colour and brightness in the sunlit sky of the earth. In the 1960s it was expected that the first extraterrestrial sky to be observed would be that seen from the surface of earth's moon (almost devoid of atmosphere, and differing substantially in appearance). For additional aspects of sky phenomena, *see* ASTRONOMY; AURORA POLARIS; HALO; METEOROLOGY; MIRAGE; RAINBOW; SPACE EXPLORATION; UNIDENTIFIED FLYING OBJECT (UFO); VAN ALLEN RADIATION BELTS.

Although subject to such variations as spectacular sunrises and sunsets, the characteristic colour of a cloudless sky on earth is dark blue at high elevations and pale blue at lower levels. Even when viewed from high altitudes, earth's sky appears brightest and whitest near the horizon (*see* ATMOSPHERE; CLOUD; HORIZON; TWILIGHT).

The colour of the sky is closely associated with polarization of light, arising from the electromagnetic nature of light waves. If oscillations in the electromagnetic wave are largest in one direction normal to the axis along which the light is propagating, the light is said to be polarized. Since people normally cannot distinguish such light from ordinary light in which the oscillations are equal in all transverse directions, special optical analyzers, usually incorporating sheet polarizers (*e.g.* Polaroid plates), are used to make the polarization visible. A polarizer blocks the transmission of oscillations in all directions except one. When a polarizer is rotated between the observer and the bluest part of the sky, the brightness of the sky as viewed through the polarizer will change. With brightness at minimum, rotation through 90° will increase intensity to a maximum that may be as much as four times brighter (*see* LIGHT: *Polarization and Electromagnetic Theory*).

Polarization of skylight was discovered by D. F. J. Arago (*q.v.*; 1809), who found that maximum polarization appears at right angles to the sun in the sun's vertical (the plane through the zenith and the sun). He also found in the vertical a point with vanishing polarization (the Arago neutral point) at an angular distance of 18° from the antisolar point. Another neutral point was discovered (1840) by J. Babinet in the vertical at an angular distance of about 18° above the sun. With a special arrangement to diminish glare near the sun, Sir David Brewster (*q.v.*) discovered a third neutral point in the vertical 18° below the sun.

The observed colour and polarization of light from a cloudless sky were once attributed to reflection of sunlight on hypothetical thin bubbles of water or other small particles (R. Clausius), or to transmission through a turbid medium (Tyndall effect). However, the first effective effort to relate and explain colour and polarization was by Lord Rayleigh (1881) in his mathematical theory of light scattering by small dielectric particles (*see* LIGHT: *The Atomic Theory of Refraction*).

Rayleigh Theory of Primary Scattering.—Using J. C. Maxwell's equations of the electromagnetic field (*see* ELECTROMAGNETIC WAVES), Lord Rayleigh showed that a volume of matter (with dimensions negligible as compared with the wavelength of incident light, and with a dielectric constant slightly different from its environment) propagates its own electromagnetic waves. In other words, light falling on the small volume is reemitted (scattered) in all directions.

The amount of scattered light is larger for shorter (blue) wavelengths than for longer (red) ones. The intensity I of the scattered light is inversely proportional to the fourth power of the wavelength; $I \sim \lambda^{-4}$.

When the volume is illuminated by unpolarized light, the scattered light is polarized. The intensity $I\perp$ of the oscillations vibrating perpendicularly to the scattering plane (*i.e.*, to the plane containing the direction of the incident and of the scattered beam) is the same for all values of the scattering angle θ between the direction of the incident and scattered beam. The intensity $I\parallel$

of the oscillation vibrating in the scattering plane varies as $\cos^2 \theta$. Hence the degree of polarization P is given by

$$P = (I_\perp - I_\parallel)/(I_\perp + I_\parallel) = \sin^2 \theta/(1 + \cos^2 \theta) \quad (1)$$

Therefore in the forward and backward direction ($\theta = 0°, 180°$) the scattered light is unpolarized ($I_\perp = I_\parallel$), while at right angles to the incident light, the scattered light is completely polarized ($I_\parallel = 0$). Since the intensity $I = I_\perp + I_\parallel$, it varies with the scattering angle as $1 + \cos^2 \theta$. The distribution of the intensity is thus symmetrical to the principal directions $\theta = 0°, 180°$ and $\theta = 90°, 270°$.

After scattering, the incident parallel beam has lost all the energy scattered in all directions around the scattering volume. This attenuation is inversely proportional to the fourth power of the wavelength (λ^{-4}) of the incident light; blue light thus loses more energy than does red.

Since air molecules have negligible size with respect to the wavelengths of visible light, Rayleigh theory can be applied to a pure, molecular atmosphere. If sunlight were scattered by air molecules only once (primary scattering), skylight would exhibit the properties noted above. In locations where the atmosphere is relatively free of haze, dust, or industrial pollution, skylight measurements tend to approach those predicted by Rayleigh theory. For example, observations of the remarkably clear and dry atmosphere above Mt. Wilson in California led F. E. Fowle (1914) to conclude that scattering there is almost entirely the result of air molecules themselves. The deep blue colour is a function of the λ^{-4} dependence of skylight intensity on wavelength. The skylight is highly polarized, with maximum intensity perpendicular to the sun in the sun's vertical plane (as discovered by Arago).

Even for very clear air, however, the colour of the sky becomes whitish toward the horizon. The light at right angles to the sun at the sun's vertical is only partially polarized ($P = 75\%$). Furthermore, degree of polarization varies with wavelength; according to Rayleigh theory, as indicated in equation (1), polarization should be equal for all wavelengths and depend only on scattering angle θ. The original theory also deviates from the observation that the three neutral points in the sun's vertical have angular distances from the sun (or from the antisolar point) that vary with wavelength. These discrepancies increase in magnitude and in day-to-day variation with increasing atmospheric turbidity.

Lord Rayleigh himself questioned the assumption of primary scattering as being too limited. Light scattered once by an air molecule may be scattered by other molecules several times (multiple scattering) before it enters the observer's eye or the aperture of a measuring instrument. J. L. Soret later (1888) showed in a very crude quantitative way that partial polarization at 90° from the sun and the existence of neutral points were the result of multiple scattering.

Variations in colour and polarization after large volcanic eruptions suggest another reason for deviations from original Rayleigh theory. Estimates of the diameters of volcanic dust particles in the atmosphere, based on the position of Bishop's ring (a brownish ring around the sun observed after the Krakatau [*q.v.*] eruption of 1883), were comparable with or larger than the wavelengths of light. Direct measurements of haze or dust particles confirmed this estimate (*see* DUST), emphasizing that Rayleigh theory failed to include a general law of scattering for such large particles.

Multiple Scattering.—Inclusion of multiple scattering in the theory was considered even as late as 1942 as a "dream of the future" (C. Jensen). However, S. Chandrasekhar (1948) ingeniously solved the problem of radiative transfer in a plane-parallel atmosphere, taking account of primary scattering, multiple scattering, and polarization of scattered light. The mathematical complexity of the solution required high-speed computers; intensity and polarization of skylight for different wavelengths were computed and summarized in extensive tables.

Multiple scattering explained such deviations from the Rayleigh theory of primary scattering as partial polarization at 90° from the sun and its increasing degree of polarization with increasing wavelength. The modified theory also predicted three neutral points at angular distances (from the sun or the antisolar point) that vary with sun elevation and that decrease with increasing wavelength.

Quantitative empirical verification of the modified theory required that the previous visual measurements in broad spectral regions be replaced by accurate measurements in narrow spectral bands. A photoelectric polarimeter, constructed by Z. Sekera (1951), provided measurements that agree remarkably with theoretical values for the very short wavelength $\lambda = 3{,}650$ Å, and for very low atmospheric turbidity. Increasing turbidity and longer wavelengths yielded deviations (and daily variations) that increased noticeably. Discrepancies from theoretical values were noticed primarily in a decrease in degree of polarization and in a shift of the neutral points.

Large-Particle Scattering.—A more general theory of scattering, developed by G. Mie (1908), considers the measured size distribution of large haze or dust particles in the atmosphere, and is in very good agreement with measurements: the volume scattering coefficient varies with the wavelength at a much slower rate ($\sim \lambda^{-b}$, where $b \leqq 1.5$) or is independent of the wavelength. The intensity of light scattered by large particles shows a similar dependence on wavelength; when the atmospheric content of such particles is high, the normally dark-blue sky has a white, milky appearance. Large-particle scattering is highly unsymmetrical, showing up to a thousand times more forward than backward scattering. Thus scattering by haze or dust particles is predominant close to the position of the sun in the sky; there is always a white area of sky around the sun (the sun's aureole), the width of which can be used as a measure of atmospheric turbidity. Since light scattered by such particles tends to be only slightly polarized, increased atmospheric turbidity can be detected as a decrease in polarization of the clear sky. This decrease is proportional to the ratio of the volume scattering coefficient of the haze or dust particles to that of the air molecules; the ratio increases with wavelength. Hence deviations from modified Rayleigh theory increase from blue to red; since dust or haze content is variable, the deviations show similar daily fluctuations.

Twilight Colours.—The twilight sky (on the side of the rising or setting sun) at times shows a brilliant display of colour. Red prevails close to the horizon and gradually blends through orange, yellow, and purple into blue around the zenith. As the sun approaches the horizon, sunlight traverses relatively long paths through the atmosphere to reach the observer, and undergoes appropriately increased attenuation. If turbidity is low, blue wavelengths tend to be attenuated most, and their diminution or disappearance is sensed by the observer as the yellow, orange, or red of the sun or sky. In a hazy atmosphere attenuation tends to be the same for all wavelengths, and the observer senses decreased brightness of the sun, but little, if any, change in colour, even when the sun is at the horizon.

BIBLIOGRAPHY.—S. Chandrasekhar and D. D. Elbert, *Illumination and Polarization of the Sunlit Sky on Rayleigh Scattering* (1954); Z. Sekera, "Polarization of Skylight," in S. Flügge (ed.), *Handbuch der Physik*, vol. 48 (1957); H. C. van de Hulst, *Light Scattering by Small Particles* (1957); K. L. Coulson, J. V. Dave, and Z. Sekera, *Tables Related to Radiation Emerging from a Planetary Atmosphere with Rayleigh Scattering* (1960); W. A. Shurcliff and S. S. Ballard, *Polarized Light* (1964). (Z. S.)

SKY DIVING: *see* PARACHUTE: *Sport Parachuting*.

SKYE, the largest island of the Inner Hebrides, Inverness county, Scot. From the mainland it is separated by the Sound of Sleat, Kyle Rhea, Loch Alsh and the Inner sound, and from the Outer Hebrides by the Minch and Little Minch. At Kyleakin, on the western end of Loch Alsh, the channel is only about ¼ mi. wide, and there is a ferry. The length of the island from southeast to northwest is 48½ mi., but its coast is deeply indented, so that no part of the interior is more than 5 mi. from the sea. It has a total area of 670.3 sq.mi. The population was 23,082 in 1841, 9,908 in 1931 (or 14.8 to the sq.mi.) and 7,478 in 1961. The chief arms of the sea are Lochs Snizort and Dunvegan in the north, Loch Bracadale in the west, Lochs Scavaig and Eishort in the south and Loch Sligachan in the east. The jagged mass of the Cuillin hills (Coolins) dominates the view. Their highest point is Sgurr Alasdair (3,309 ft.) and at least six other peaks exceed 3,000 ft.

To the north of Loch Slapin stands the group of Red hills of which the highest points are Ben Caillich and Beinn Dearg Mor and near Loch Ainort rises Ben Glamaig. About 8 mi. N. of Portree is the curious basaltic group of the Storr, consisting of pinnacles and towers, the most remarkable of which, the Old Man, forms a landmark for sailors.

Most of the land is moor and hill pasture but during the first half of the 20th century there was a considerable development in agriculture. The crofting system was still general at mid-20th century and while more crops were grown, the climate appeared to be better adapted for sheep and cattle. In 1953 there were approximately 104,284 sheep and 11,000 cattle on the island. The farms and small holdings were stocked principally with Blackface sheep although there were also Cheviot stocks. The condition of the crofters, which was pitiable in the extreme, was greatly improved by the Small Landholders (Scotland) acts 1886–1911, and in later years by the introduction of the government subsidies for cattle, potatoes and, until mid-20th century, sheep. The introduction of grants for various purposes enabled the black houses to be replaced principally by houses built of brick or concrete. With the opening in 1952 of the Storr Lochs hydroelectric power scheme, approximately 95% of the houses were supplied with electricity. The many ejections between 1840 and 1880 and the emigration that followed were mainly responsible for the serious decline of the population.

While the railheads at Kyle of Loch Alsh and Mallaig make the mainland markets accessible, road transport is extensively used. The fishing industry, which was at one time the mainstay of the population, declined during the first half of the 20th century. Whisky is distilled at Carbost. At Loch Cuithir, 4 mi. N. of Storr, work on the extensive diatomite deposits was restarted and in 1953 a factory, in which the raw material is processed, was opened at Uig. An afforestation scheme led to the reopening in 1953 of one of the 18 forestry schools closed as a result of the depopulation of the island.

The inhabited isles off the coast of Skye are mainly situated near the eastern shore. Of these the principal is Raasay, 13 mi. long by about 3½ mi. at its widest. Off its northwestern shore lies the isle of Fladda.

To the north of Raasay is Rona (seal island, from the Gaelic *ron*, a seal), 4½ mi. long with a maximum breadth of 1½ mi. The island is now unoccupied except for the lighthouse keepers whose families were removed to Portree in 1952. Scalpay, immediately south of Raasay, has a hill of 1,298 ft. The other isles are Pabay, Ornsay and Soay.

The islanders of Soay were evacuated in 1953 to Mull where they were established on new holdings by the department of agriculture for Scotland.

Portree (pop. [1961] 2,025), the capital, lies at the head of a fine harbour about the middle of the eastern seaboard. Steamers run daily to and from Mallaig and Kyle of Loch Alsh. There is a factory for tweeds, tartans and other woolens. During the cattle sales, held twice a year, large numbers of sheep and cattle are transported to the railhead at Kyle of Loch Alsh. The name Portree was derived from the fact that James V landed there during his tour of the western Highlands. The place thus became, in Gaelic, Port-an-Righ (the king's harbour).

It was to Portree that Flora Macdonald (1722–90) conducted Prince Charles Edward when he escaped from Benbecula. The sheriff's courthouse is situated in Portree.

Among other places in Skye associated with the Young Pretender are Prince Charles's point near Monkstadt, where he landed with Flora Macdonald, and Kingsburgh, on the eastern shore of Loch Snizort. The castle of the Macleods of Macleod, on a rocky promontory at Dunvegan, was erected in the 9th century and has the reputation of being the oldest continually inhabited house in Scotland.

The MacCrimmons, the famous race of hereditary pipers, hailed from this quarter of Skye and were attached to the Macleods of Dunvegan. At Duntulm is the ruined castle of the Macdonalds, another of the great clans of Skye chieftains. There is also much of archaeological interest in Skye. (J. Pp.)

SKYROS (Scyros; modern Greek Skiros), a rocky island of the Northern Sporades, Greece, in the Aegean Sea, 22 mi. (35 km.) NE of Euboea. Area 77 sq.mi. (199 sq.km.); pop. (1961) 2,882. The earlier inhabitants of Skyros were non-Greek Dolopes. There was a sanctuary of Achilles (*q.v.*) on the island; the legend was that he conquered Skyros, or that his mother Thetis disguised him as a woman in the palace of Lycomedes, to keep him from the Trojan War, but he was discovered by Odysseus, and accompanied him to Troy. Lycomedes was also said to have killed Theseus (*q.v.*) on Skyros. The island was seized *c.* 475 by Cimon (who recovered the "bones of Theseus") and settled by Athenians. It was under Macedonian rule in 196 B.C., when the Romans restored it to Athens. It was sacked by Goths, Heruli, and Peucini in A.D. 269. The ancient city was on a rocky peak, on the northeastern coast, where the modern capital Skiros stands (pop. [1961] 2,411). The English poet Rupert Brooke was buried on the island. (J. Bo.)

SKYSCRAPER. Until the middle of the 19th century, a commercial building more than four or five stories in height was rare, because tenants and their clients were reluctant to climb higher. However, desirable sites in business districts were expensive, and as business organizations multiplied, became larger and therefore demanded more space, the logical way to make economical use of those sites was to build upward. The key to the problem, a safe and efficient means of vertical transportation, was provided with the invention of the elevator safety device by Elisha Graves Otis in the 1850s (*see* Elevator).

The first safe passenger elevator was installed in the H. V. Haughwout store in New York city in 1857. Thereafter the development of the skyscraper as a unique architectural type was fairly rapid. The first taller-than-average structure (130 ft., compared with 60 ft.) to exploit the potentialities of the elevator was the Equitable Life Assurance Society building, New York city, erected between 1868 and 1870 by Arthur Gilman, Edward Kendall and George B. Post. In 1873 Post planned the headquarters for the Western Union company that rose 230 ft. and in the same year Richard M. Hunt designed the *New York Tribune* building to the height of 260 ft. Both buildings, in New York city, were completed in 1875.

Earlier Philadelphia architects had shown an interest in a solution that was more expressive of structure. In 1849 William L. Johnston introduced in the eight-story Jayne building a skeletal granite construction with columns that rose six stories without interruption, creating an illusion of unusual height.

In Chicago, during the last quarter of the 19th century, height and structural design were amalgamated in a distinctive manner. William Le Baron Jenney's Home Life Insurance Company building (1885) in particular marked a turning point in the construction of tall buildings: an internal metal frame (part cast iron, part wrought iron) supported the dead weight of floors and roof and the major load of the ten-story building, eliminating the necessity for massive foundations and thick bearing walls that had made structures of this height uneconomical. Jenney did not create skeleton construction without assistance from earlier architects and engineers.

Iron had been used in architecture for some time before the mid-19th century when European and U.S. architects and engineers exploited its structural advantages in a wide variety of buildings. In 1849 James Bogardus erected the Harper Brothers building, New York city, in which a rigid frame of iron columns and beams was the main support for the floor and roof loads; this represented a transition between the old masonry method and the system used by Jenney, in which the interior frame carried not only the roof and floor loads but the weight of the side walls as well.

During the 1870s in the American east and midwest, opposition developed to the rich external decorative treatment of the typical Victorian building. American architects began to employ surfaces relatively free of ornament, accenting the "skin and bones" of their buildings as could be seen in the Shillito Warehouse in Cincinnati (James McLaughlin, 1877). The Montauk block (1882) in Chicago, by Daniel Burnham and John W. Root,

often called Chicago's first skyscraper (130 ft.), was constructed of pressed brick and was nearly devoid of ornament. This approach to architectural design won many adherents in Chicago during the 1880s, giving rise to the term "commercial style." It appeared in the east concurrently but was used primarily for warehouses as distinguished from office buildings. By 1890 Chicago skyscrapers had developed a distinctive quality evolving from the first Leiter building (Jenney, 1879), the Borden block (1879–80) and the Troescher building (1884), the latter two by Louis Sullivan in partnership with Dankmar Adler. The 13-story Chamber of Commerce building (1889–90), by Edward Baumann and Harris Heuhl, with its division of the principal façade by columns into a series of bays with triple windows, its flat roof and its minimal use of ornament, was characteristic.

The buildings of Sullivan in particular expressed not only their business function and their structural character but their loftiness. One of his finest examples is the Wainwright building in St. Louis (1890), in which the skeleton frame of Jenney was combined with the kind of vertical articulation employed in Philadelphia. Sullivan divided the edifice into three main parts, a base, an upper section and an attic story, capped by a flat, projecting cornice. The window treatment expresses the fact that the main section is composed of identical offices. For vertical emphasis he accented columns, a number of which are not structural, that rise without interruption from base to cornice. This formula was repeated in his Prudential building (Buffalo, 1895) and Bayard building (New York, 1899).

Architects along the eastern seaboard were more derivative, basing their work on older styles of architecture. In the 1880s a Romanesque manner, popularized by H. H. Richardson, enjoyed a vogue that spread westward throughout the country. The 14-story Ames building (1889–91) by George F. Shepley, Charles A. Rutan and Charles A. Coolidge, one of Boston's first skyscrapers, is typical.

The next decade saw a shift to the Neoclassical style, made fashionable by the architects Charles Follen McKim, William Rutherford Mead and Stanford White, founders of the firm of McKim, Mead and White. Skyscrapers were modeled after the Greek column with its tripartite division of base, shaft and capital. The Graham building (1898) in New York city, by Charles W. Clinton and William H. Russell, which was typical, had a base of lower floors containing a lobby and banking space, a shaft composed of identical offices and a capital of several attic floors topped by a projecting cornice.

Some architects, such as Bruce Price, conceived of the skyscraper as a tall, free-standing, vertical shaft. The American Surety building (1895) in New York exemplifies this concept. Others found a precedent in the medieval bell towers of Europe as in the case of the Metropolitan Life Insurance building (1909), New York, modeled by Napoleon Le Brun after the campanile in Venice, and the Woolworth tower (1913), New York city, by Cass Gilbert, in the Gothic style.

Many speculators took advantage of the absence of building codes to raise gigantic structures without concern for aesthetic expression or for the many civic problems such buildings created. The result was the adoption in many cities of zoning laws to prevent the formation of dark, canyonlike streets, and thus the setback skyscraper became commonplace in U.S. cities. One of the finest of this type was the *Daily News* building, New York city (Raymond M. Hood and John M. Howells, 1930). It was at this time that the International Style with its emphasis on the horizontal rather than vertical began to make its influence felt on skyscraper design. Hood and Howells' striped and glassy McGraw Hill building (1931), New York city, and George Howe and William Lescaze's Philadelphia Saving Fund Society building (1932), Philadelphia, reflect the thinking of Walter Gropius, Ludwig Mies Van der Rohe and other European architects during the early 1920s.

During the 1920s New York skyscrapers reached their greatest heights: the Chrysler building (William van Alen, 1929), 77 stories (1,046 ft.), and the Empire State building (Shreve, Lamb and Harmon, 1930), 102 stories (1,250 ft.). Thereafter architects tended to think in terms of 30- or 40-story structures, recommending buildings of smaller size but better design as safer enterprises economically.

One of the first projects illustrating this changed attitude was Rockefeller Centre in New York by Reinhard & Hofmeister; Hood and Fouilhoux; Corbett, Harrison and MacMurray (1929–40). On its three-block site, the associated architects could have raised the highest structure anywhere; instead they chose to build a major skyscraper of 66 stories surrounded by structures of about half that height. Consideration was given to light, air and open space as well as to fine-quality materials, landscaping and ornamental works of art. The financial success of this project and the prestige it acquired led to similar planning in the United States and elsewhere.

The most notable examples of the World War II period were the glass-sheathed Equitable Life Insurance building in Portland, Ore. (Pietro Belluschi, 1948), Lever House (Skidmore, Owings and Merrill, 1952), and the *Time* and *Life* building (Wallace K. Harrison and Max Abramovitz, 1959), both in New York city. All of these reflect the influence of Mies Van der Rohe, who along with Philip Johnson designed the Seagram building, New York city (1957). The aluminum-clad Alcoa building in Pittsburgh by Harrison and Abramovitz (1952) shows the trend toward metal curtain wall construction. Less influential but nevertheless significant is Frank Lloyd Wright's Price tower at Bartlesville, Okla. (1955–56). Noteworthy too are the sunscreen-covered structures of South America deriving from the work of Le Corbusier. Typical is the Caramuro building by Paulo Autunes Ribeiro at Salvador, Bahia (1946). Le Corbusier was also instrumental in promoting the use of wedge and claw-shaped forms into skyscraper design as a means of achieving maximum amounts of well-lighted working space.

See also MODERN ARCHITECTURE and biographies of architects.
(W. R. WE.)

SLANDER: *see* DEFAMATION.

SLANE (SLÁINE), a picturesque little village situated in the Boyne Valley 8 mi. (13 km.) NE of Navan in County Meath, Republic of Ireland. Pop. (1961) 421. It was the birthplace (1891) of the poet Francis Ledwidge, who was killed in World War I. Mervyn Archdall (1723–91), the author of *Monasticum Hibernicum* and rector of Slane, is buried in the Protestant churchyard in the village. The old village square is now surrounded by shops and houses, many of them Georgian, and forms the hub of social and commercial life. Slane Castle, seat of the marquis of Conyngham, stands by the river approximately 1 mi. W of Slane, and beyond it are the ruins of Castle Dexter. Rosnaree, the legendary burial place of the 3rd-century King Cormac Mac Art, lies 2 mi. (3 km.) S.

Just north of the village is the Hill of Slane where traditionally St. Patrick challenged the Druids' power by lighting the Paschal fire himself in place of the king. He later built a monastery there but no trace of it remains. (W. M. PO.)

SLANEY, a river of the Republic of Ireland, 73 mi. (117 km.) long, rises in Lugnaquilla (3,039 ft. [926 m.]), County Wicklow, the highest mountain of southeast Ireland. The Slaney turns westward in a steep torrent course to the Glen of Imail, a basin cut in granite at 500–700 ft., and then assumes a north-south line and wanders through the farmed lowlands of County Carlow and through Tullow to Bunclody just inside County Wexford. There it is joined by the Clody and continues southeast through the Slaney Gap, the only lowland route through the mountain chain that dominates the southeast of Ireland. In its journey from the gap across the fertile lowland of Wexford, the river is between 50 and 100 ft. below the general surface which stands at about 200 ft. above sea level. It receives two major tributaries, the Derry just above Bunclody, and the Bann north of Enniscorthy, both of which flow from the northeast in wide, well-farmed corridors of lowland. At Enniscorthy (*q.v.*) the river becomes tidal, and quays were built in the 19th century for barge traffic to Wexford (*q.v.*). The Slaney finally enters Wexford Harbour, a sheltered inlet of much more importance for shipping in the 19th century than at present. (T. W. FR.)

SLANG, an informal and colloquial variation of or addition to standard speech. It may be a variation in any element of a language—in its sounds ("atta dirl"), in its stress (posi*tiv*ely"), in its intonation ("Is that so?"), in its syntax ("We was robbed"); but it is in vocabulary ("spiv") and meaning ("She's a peach") that slang is commonest.

Slang is formed by normal linguistic processes found in all living languages, processes such as compounding ("low-down," "sob stuff"), word clipping ("pro," "mike," "pix"), abbreviation ("OK," "VIP," "snafu," "q.t."), onomatopoeia ("boom," "whiz," "bang"), generalization of proper names ("bobby," "guy," "real McCoy"), borrowing from dialects and foreign languages ("vamoose," "savvy," "pronto," "loco"), and extension of meaning by analogy ("Park your hat," "He got pickled").

ORIGINS AND USES OF SLANG

Slang and Society.—People belonging to the same social group —of the same trade, profession, hobby, age, or social position— tend to behave in the same way. This behaviour influences not only the clothes they wear but also the language they use. The language of a social group, particularly its slang, is one of several forms of behaviour that keep the group distinct from other groups. As a verse writer has aptly put it, "The chief use of slang/Is to show that you're one of the gang" (R.D.C. in the *New Statesman and Nation*, London, Nov. 16, 1946).

One of the commonest social groupings is that of people who work together. Students have their "lab" and their "gym." Secondhand car dealers have their "cream puffs" (excellent cars) and their "dogs" (dilapidated ones). For an order of poached eggs on toast a waitress may tell the cook to "drop two on." To a television producer a poor show may be a "wart." A circus "geek" (conditioned freak) can use a "stick" (decoy) to attract a "tip" (prospective customer); but this is not the same sort of tip that a hotel porter might expect from a "front" (inexperienced traveler). A professional flyer may wear a "brain bucket" (crash helmet) in case he has to "buy a farm" (crash); but his "soup" (horsepower) has nothing to do with the "soup" (nitroglycerin) of the professional safecracker. "Cat" does not refer to the same thing in the underworld as it does in the jazz world. Yet both underworld and jazz slang form the basis of much "teen talk."

Some social conditions are more favourable than others to the creation of slang. Excitement, crowding, and the sudden regrouping of people for a particular purpose, as in wartime, may result in the formation of a large number of slang words. This goes on from person to person, from ally to ally and from one side to the other. All the participants learn things about the others, and all come out with larger vocabularies. Forceful expressions by the thousand are created and used by members of the armed forces. "Big Bertha" is remembered as the nickname of the long-range German cannon in World War I, immortalizing Bertha Krupp of the munitions family. The innocent-sounding "pineapple" of the same war was a hand grenade with markings suggestive of the fruit. World War II added another type of grenade, the "Molotov cocktail," and a vast new vocabulary of slang including such words as "blitz," "doodlebug" and "walkie-talkie." The Korean War also had its slang; Korean soldiers were known as "gooks" and as "ROKs." A few slang expressions, such as "OK," the international sign of approval or assent (*see* O.K.), have spread far beyond their original linguistic boundaries.

How Slang Develops.—A witticism of a single person may be enough to launch a new slang form. The novelty may produce a response so effective that it is taken up by others and soon becomes widespread. The more it is used, the more popular it becomes, especially if it is used by associates or admired persons. But after much use the novelty wears off, and the expression, having no special effect and therefore no further function, dies out. If, however, the older form or its central meaning has in the meantime gone out of fashion, the slang expression may remain in use as the normal form. This was the case when the Latin word *testa* ("earthen pot"), used jocularly to describe the head of a baldheaded man, supplanted the normal word *caput* and later became the usual form for head in Italian (*testa*) and in French (*tête*). Similarly, *gamba* ("hoof"), the Latin slang word for leg, became the standard word for leg in both French (*jambe*) and Italian (*gamba*), the latter being the word from which "gam" in modern American slang was most probably derived.

When literary men in the 17th century wrote about the "common people" they often used a Latin expression, *mobile vulgus*, meaning the fickle or movable crowd. Sometimes in writing this was abbreviated to *mob. vulg*. Later the *vulg*. was dropped, and careless writers began to use *mob*. alone, still pronouncing it in full, *mobile*. The next step was to say simply *mob*, which was considered shockingly low and regarded as the crudest type of slang. The usage persisted, and in the 18th century "mob" became tolerated in conversation, although not in writing. Early in the 19th century came another promotion, and the word, at long last, entered the ranks of standard English. A similar process was responsible for about 2% of conventional English words; such words as "trip," "bet," "donkey," "shabby," "chap," "bore," "cab" and "kidnap" were once condemned as slang.

Some words may remain as slang for centuries, neither disappearing nor becoming part of the standard vocabulary. Such is the case of the medieval "booze," the Elizabethan "rook" (verb), and the 18th-century "swop." The same applies to certain expressions such as "dead as a doornail," which was used in the 14th century and may have been old even then. In the 16th century Shakespeare, in *Twelfth Night*, wrote "laugh yourselves into stitches." In the 17th century George Fox, founder of the Quakers, wrote about a professional man who had "done his stuff," proving early currency for another expression still in use.

Yet for every slang expression that has lived, hundreds of others have died. "Gamp" was once the usual slang name for an umbrella. It was named after a large, bulging umbrella carried by Sairey Gamp, in Charles Dickens' *Martin Chuzzlewit*, published in 1844. In the latter part of the 19th century it became a well-established colloquialism, but it was hardly important enough to last.

It is in the area of fundamental emotions that slang is most changeable and most productive. The slang of approval ("swell," "nifty," "bong," "dilly," "the end," "smashing") and disapproval ("phooey," "nuts," "baloney," "hogwash," "flapdoodle," "poppycock") may change within a generation. So may the words for woman and girl ("skirt," "biscuit," "swan"), and these are usually numerous in any generation. When a young man refers to his sweetheart as his "soul mate," his "steady," his "main dish," his "heavy" or his "onliest" he has hardly made a beginning; he could use a different expression every day for three years and still have words left over for variety. It is the same with money, which has unconventional synonyms by the bookful, including such picturesque terms as "berries," "cabbage," "chink," "ducats," "jack," "mazuma," "moola," "potatoes," "simoleons," "spondulics" and "wampum." The expression "to die" seems much too commonplace; a person is said to "drop off," "fade out," "go to grass," "go under," "go up," "go west," "kick the bucket," "pop off the hooks," "shuffle off" or "push the clouds," with probably a thousand variations. A person suffering from the effects of drinking to excess is said to be "afloat," "blowed," "boozy," "cockeyed," "corned," "disguised," "fractured," "groggy," "high as a kite," "ratty," "stewed," "vulcanized" or "smashed." The idea of stealing is expressed in slang in many forms, some of them with a disarming suggestion of innocence, such as "promote," "salvage," "rustle" and even "liberate." Other synonyms, ranging all the way from "annex" and "appropriate" to "souvenir" and "freeze onto," would easily exceed a thousand. The same may be said of all fundamental words, the words first used by children and by "primitive" people. They seem to seek synonyms—humorous, satirical, poetical, euphemistic or merely roughly descriptive— in slang.

Catchphrases.—Catchphrases, more or less related to slang, are of many kinds, ranging from the whimsical and witty to the meaningless and inane. Popular catchphrases in the 19th century included "Who stole the donkey's bed?" addressed to men wearing straw hats; "What a tail our cat has," a bantering reference to a woman's new dress; "Does your mother know you're out?" a

supposedly funny remark of wide application. Many 19th-century examples were based on music-hall songs, as "Where did you get that hat?"; "Shoo, fly, don't bother me"; and "A hot time in the old town tonight." A similar inane specimen in the 20th century was "Yes, we have no bananas." Catchphrases still come and go, and still bear study, for they tend to reveal the processes, mental or emotional, of the people who utter them.

Artificial Types.—Since one of the functions of slang is to prevent members of other groups from understanding what is said, it is not surprising that attempts have been made to create systems of slang to function as secret languages. These include such artificial types as rhyming slang, back slang and centre slang. Rhyming slang was used by certain Cockney groups as early as 1840; it spread to the Australian underworld and after World War II became popular among the young "teddy boy" gangs in London and other English cities. In rhyming slang a wife becomes "trouble and strife," a road a "frog and toad" and a suit a "whistle and flute." This becomes even more secret when the rhyming word is omitted, whereby a road becomes a "frog," a suit a "whistle," a hat a "tit for" (tat) and mouth a "north" (and south). Back slang consists in saying a word backward with any alteration necessary to make it more pronounceable. In this way a game becomes an "emag," a market a "tekram," drunk "kennurd" and police "slop." In centre slang, the first syllable of the slang word is formed from the first vowel of the standard word and the consonant that follows it; the other syllables may be made in a number of ways, either with fixed or free suffixes. To call a man a fool in centre slang, such terms as "oolfoo," "oolerfer" and even "hoolerfer" could be used.

National Slang.—It is not surprising that world English shows traces of American influence. This influence, which tends to create variety but not disunity, extends into the field of slang. The americanizing process has been aided by media of mass communications, but it should be pointed out that communication is a two-way street and that Americans have not been slow to import "Briticisms." (*See* also AMERICAN ENGLISH.)

On the continent of Europe, slang was being recognized as early as the 12th century and studied at least as early as the 15th, when François Villon used it and probably added to it in some of his immortal ballades. A generation later, in Germany, Martin Luther wrote a preface to the 1529 edition of *Liber Vagatorum* (*Book of Vagabonds*) and gave his views on the vocabulary of vagabonds' slang featured in that work. Villon and Luther had one thing in common—they were both writers, and writers recognized, then as now the strong human appeal of certain types of slang.

Attitudes Toward Slang.—Some writers have condemned all slang especially that of hostile classes. They have called it vulgar, corrupted, uncultured, secret and false, the dialect of the rabble, of beggars, of gypsies, and of thieves. Others have praised slang as creative activity. The critic G. K. Chesterton said that "all slang is metaphor." The semanticist S. I. Hayakawa called slang "the poetry of everyday life." According to most students of language, slang in itself is neither good nor bad. It is part of the natural growth of language. A living language must continually change, and some of the changes first appear as slang. Some slang makes the language capable of new and delicate shades of meaning or adds vividness, clarity and directness to everyday expressions. Other slang expressions are vague in meaning, cover too much ground or duplicate better conventional vocabulary. Judgment on the propriety or effectiveness of an expression—and even whether it is or is not "slang"—requires knowledge of the context in which it appears.

When Elisha Coles issued his *English Dictionary* in 1676, he explained the inclusion of a number of slang words and expressions by saying, "'Tis no Disparagement to understand the Canting Terms: It may chance to save your Throat from being cut, or, (at least) your Pocket from being pick'd." Reasons less ominous impel the modern student, who knows that from slang much can be learned about the history, customs, fashions, the very thoughts of people far away or long ago. The study of unconventional speech is of immense value to authors and playwrights.

(A. McQ.; W. F. MY.)

SLANG TERMS

The following lists of slang terms are representative of mid-20th-century usage in the principal English-speaking countries. While all the terms are in current use, not all are of modern invention and not all are necessarily used exclusively in the country to which they are attributed. Much American slang has gained currency in the United Kingdom, and much, especially older, British slang has become established in the United States. As for Canada, where there is comparatively little native slang, most current slang derives from the adjacent United States; there is little general U.S. slang that is not used also among Canadians. British slang, on the other hand, is far less extensively used among native Canadians, although many older terms are well established. Australia has developed a vigorous native slang, much of which is shared by New Zealand; but here, too, much British slang has long been current and, particularly during and since World War II, American slang has made substantial inroads.

UNITED STATES

action *n.* profitable activity in gambling or crime
ants *n.pl.* anxiety; restlessness (antsy *adj.*)
applesauce *n.* (1) nonsense (2) flattery
bail out *v.* help out of a difficult situation
baloney *n.* nonsense
barf *n., v.* vomit
barrel *v.* move at a fast rate
bat *n.* (1) prostitute (2) unattractive woman (3) spree
bean-shave *n.* very close haircut
beef *n., v.* complain(t)
beetle off *v.* go away
biff *v.* strike; hit
bitch *n., v.* complain(t)
blowout *n.* party; spree
boob *n.* stupid person
boo-boo *n., v.* blunder
brainstorm *n.* sudden, brilliant idea
broad *n.* young woman
buck *n.* dollar
buff *n.* an expert on some subject; enthusiast
bug *n.* (1) germ; microbe (2) small car (3) mechanical defect (4) an unusual interest in; *v.* (1) annoy; badger (2) install a concealed microphone
bug out *v.* run away
bulldoze *v.* intimidate; bully
bum *n.* a vagrant or derelict
bum steer *n.* misleading advice or information
burp *n., v.* belch
bushwa(h) *n.* nonsense
buy *v.* accept; agree to
buzz *n.* phone call; *v.* to telephone
caboodle *n.* the entire lot
calaboose *n.* jail
can *n.* toilet; bathroom; *v.* dismiss; discharge
caper *n.* (1) prank; spree (2) robbery; illicit enterprise
cheesecake *n.* photos of scantily clad women in provocative poses
chick *n.* young woman
chicken *n.* coward
chicken out *v.* withdraw or quit from fear
chip in *v.* (1) give a share (2) intrude into a conversation
clam up *v.* refuse to talk
claptrap *n.* nonsense; exaggerated talk
clip joint *n.* business establishment charging unduly high prices
con *v.* trick; cheat
cornball *n.* one partial to out-of-date styles in dress, music, etc.
cough up *v.* pay or contribute money
crazy *adj.* delightful; first-rate
creep *n.* objectionable person
crummy *adj.* dirty and run-down; distasteful
cut no ice *v.* do no good; have no effect
dago red *n.* cheap red wine
deadbeat *n.* one who does not contribute his fair share; leech
dig *v.* understand
dilly *n.* something remarkable
dinky *adj.* small
dive *n.* low-class bar or dance hall
dog *n.* (1) unattractive girl (2) prostitute
donnybrook *n.* noisy argument or brawl
dough *n.* money
drag *n.* influence
dreamboat *n.* attractive person
eager beaver *n.* overdiligent person
egghead *n.* intellectual
end, the *n.* (1) the best (2) the last straw
eyewash *n.* flattery; nonsense
fall guy *n.* scapegoat
fancy pants *n.* sissy; fastidious person
fanny *n.* buttocks
fast buck *n.* money got easily, often illegally or unethically
fin *n.* five-dollar bill
fink *n.* informer; contemptible person
firetrap *n.* run-down building
flack *n.* publicity agent
flake out *v.* collapse from exhaustion
flap *n.* state of confusion
fleabag *n.* (1) bed (2) cheap hotel
flophouse *n.* very cheap lodging place
fluke *n.* lucky shot, stroke, or happening
freeloader *n.* one who drinks or eats at another's expense
frost *v.* irritate; exasperate
fuddy-duddy *n.* old-fashioned, fussy person
fuzz *n.* police; policeman
gander *n.* look; glance
get lost!: go away!
get the ax: be dismissed or discharged
get the bulge on: gain an advantage of
get the hang of: learn how to do; comprehend
get with it: get in step; become aware

gimmick *n.* device for selling something, winning attention, etc.
gizmo *n.* gadget
goldbrick *n.*, *v.* malinger(er); shirk(er)
goof *n.* dull person; *n.*, *v.* blunder
goof off *v.* avoid one's responsibilities
gook *n.* Oriental, especially a Chinese or Korean
grand *n.* one thousand dollars
gravy *n.* profit; easy money
grunt-and-groaner *n.* professional wrestler
guy *n.* fellow; man
gyp *n.*, *v.* cheat
hard-nosed *adj.* obstinate; stubborn
hassle *n.*, *v.* dispute; quarrel
heap *n.* decrepit car
hellcat *n.* wild, reckless girl
hen party *n.* gathering of women
hep or hip *adj.* well-informed; sophisticated
hi *interj.* hello
highbrow *n.*, *adj.* intellectual
hit the hay: go to bed
hogwash *n.* nonsense
hood *n.* hoodlum; gangster
hooey *n.* nonsense
hoosegow *n.* jail
horse around *v.* indulge in horseplay; act irresponsibly
inside dope *n.* confidential information
jack *n.* money
jack up *v.* increase (price)
jalopy *n.* decrepit car
jazzy *adj.* colourful; exciting
jeezly *adj.* miserable; distasteful
jerk *n.* dull or ineffectual person
joe-job *n.* menial task; undesirable duty
john *n.* toilet; bathroom
joint *n.* low-class restaurant, tavern, etc.
kale *n.* money
kibitz *v.* (1) proffer unwanted advice (2) play the fool
kicks *n.pl.* thrills; pleasure
kid *v.* tease; mislead
knock *v.* condemn; criticize
kook *n.* odd, unpopular person
Kraut *n.* German, especially a soldier
lemon *n.* (1) unattractive person (2) defective product
Limey *n.* Englishman
litterbug *n.* one who throws litter on streets, etc.
loot *n.* (1) money (2) winnings
lowbrow *n.*, *adj.* uncultured; nonintellectual
lower the boom: treat sternly; tighten up discipline
lush *n.* drunkard; alcoholic
make a go of: succeed at
meatball *n.* dull, obnoxious person
Mick *n.* Irishman
miffed *adj.* offended
moocher *n.* beggar; cadger
moola(h) *n.* money
mossback *n.* old-fashioned, unprogressive person
mutt *n.* nondescript dog
needle *v.* embarrass and irritate by making barbed remarks
not with it: uninformed; not in step
nutty *adj.* (1) crazy (2) highly enthusiastic about
off base: interfering; intruding; out of line
offbeat *adj.* unusual; nonconforming
off the beam: mistaken; not applicable
off the cuff: extemporaneously
on the ball: alert; accurate
on the blink: out of repair; not working
on the cuff: on credit
on the hummer: out of repair; not working
on the make: intent on the main chance for advancement or gratification
on the nose: exactly right; precisely
over a barrel: at someone's mercy; helpless
pad *n.* bed; lodging
pan out *v.* come about; result
pep *n.* energy
petting *n.* amorous fondling
phony *adj.* not genuine; insincere
pickup *n.* a chance acquaintance, esp. a woman
pinch-hit *v.* substitute for
platter *n.* phonograph record
pooch *n.* dog
pooped *adj.* exhausted
puddle jumper *n.* small car
pull *n.* influence
pull off *v.* initiate and carry through
punch *n.* vigour; force; power
pushover *n.* one easily duped, influenced, seduced, etc.
put the skids under: cause (a person or plan) to fail
put wise: acquaint with the facts
rate *v.* be highly esteemed; be accepted
raunchy *adj.* (1) slovenly (2) inferior
rhubarb *n.* dispute; fracas
ritzy *adj.* stylish; first-class
road hog *n.* selfish, inconsiderate driver
rough up *v.* abuse physically
rubberneck *n.* sightseer; *v.* sight-see
rumdum *n.* alcoholic; drunkard
sap *n.* idiot; fool
sawbuck *n.* ten-dollar bill
scads *n.pl.* large amount
screwball *n.* eccentric person
screwy *adj.* crazy; awry
shake a leg!: get a move on!
shindig *n.* party, dance, etc.
shiner *n.* black eye
showdown *n.* calling to account; settling of differences
shut-eye *n.* sleep
sidekick *n.* buddy; close friend
slob *n.* slovenly, careless person
smacker *n.* dollar
small-time *adj.* insignificant; petty
smidgen *n.* small quantity; trace
snap *n.* easy task
snazzy *adj.* stylish; first-rate
snide *adj.* condescending; sarcastically rude
snipe *n.* cigar or cigarette butt
snooty *adj.* snobbish; overcritical
soft soap *n.* flattery
souse *n.* drunkard
square *n.* old-fashioned person; conformist
squeal *v.* inform on
squirrelly *adj.* eccentric; slightly mad
stash *v.* hide; conceal
steady *n.* one's regular companion of the opposite sex
steal *n.* bargain
stoned *adj.* drunk
strong-arm *adj.* violent; physically coercive
sucker *n.* gullible person
swear off *v.* give up, esp. liquor
swinging *adj.* uninhibited; liberated
tad *n.* child, esp. a boy
tailor-made *n.* factory-made cigarette
take a powder: run away; leave precipitously
tip off *n.* warning; *v.* inform; forewarn
tomato *n.* attractive young woman
walkover *n.* easy task
wheel *n.* important, influential person
whiz *n.* one who performs exceptionally well
willies, the *n.pl.* feeling of fright; apprehension; nerves
wino *n.* drunkard who favours cheap wine
wire-puller *n.* one who uses influence to advance his ends
wisecrack *n.* witty remark
wise guy *n.* smart aleck
wise up *v.* find out what is going on; instruct regarding the facts
woozy *adj.* dizzy; faint
wow *n.* a great success; something excellent
zing *n.* vitality; zest

Great Britain

all in *adj.* exhausted
altogether, the *n.* the bare skin
attic *n.* head; mind
bag *v.* steal; appropriate
bags *n.pl.* (1) plenty; lots (2) trousers
bamboozle *v.* outwit; take advantage of by deceit
banger *n.* sausage
bang-up *adj.* excellent; first-rate
bash *n.* (1) party (2) try (3) blow; *v.* hit; strike
beak *n.* magistrate; judge
Bible puncher *n.* clergyman
bilge *n.* nonsense; empty talk
bird *n.* (1) girl friend (2) prostitute
bit of fluff *n.* young woman
blighter *n.* fellow; man
bloke *n.* fellow; man
blower *n.* telephone
blue ruin *n.* gin
bob *n.* shilling
bobby *n.* policeman
Bolshie *n.* revolutionary
booze *n.* alcoholic liquor
bounce *n.* bragging; boasting
brass *n.* (1) impudence; gall (2) money
brass-hat *n.* senior staff officer
brick *n.* loyal, reliable person
broke *adj.* penniless
browned-off *adj.* disgusted; fed up
bubble and squeak *n.* warmed-over cabbage and potatoes
buckshee *adj.* at no cost; free
buck up *v.* cheer up; hurry up
bumf *n.* paper, esp. toilet paper
bun-fight *n.* an informal social where food is served
bunk *v.* depart in haste; run away
buzz off *v.* go away
cackle *v.* tell a secret; blab
cadge *v.* beg
char *n.* tea
chips *n.* (1) carpenter (2) money
chisel *v.* cheat; swindle
chokey *n.* prison; jail
chump *n.* (1) fool; simpleton (2) head
civvies *n.pl.* (1) civilian clothes (2) civilians
claret *n.* blood
clever dick *n.* smart aleck
clink *n.* jail; prison
clock *n.* face
cock a snook: make a rude gesture (thumb to nosetip, fingers extended)
cocky *adj.* saucy; supremely confident
coffin nail *n.* cigarette
cold tea *n.* brandy or whiskey
coot(ie) *n.* louse
copper *n.* policeman
corker *n.* anything outstanding or exceptional
cosh *n.* small bludgeon
cove *n.* fellow; man
crackers *adj.* eccentric; crazy
creeps, the *n.pl.* uneasy feeling caused by unidentified fears
cushy *adj.* easy; safe; not demanding
cut a dash: make a display; cut a figure
cut the cackle: stop talking; shut up
dab *n.* expert
daddy *n.* backer; protector
darbies *n.pl.* (1) bracelets (2) handcuffs
date *n.* foolish person
dekko *n.* look; glance
dial *n.* face
dibs *n.* money
dicey *adj.* risky, dangerous
digs *n.pl.* lodgings
Dipper *n.* Baptist
dive *v.* pick a pocket
dotty *adj.* silly; insane
dress down *v.* chastise; rebuke severely
drop a brick: make an embarrassing mistake
dubber *n.* lockpicker
duds *n.pl.* clothes
duffer *n.* dull, stupid fellow
dumb *adj.* stupid
dustup *n.* disturbance; fight
fag *n.* cigarette
fence *n.* receiver of stolen goods
file *v.* rob
fin *n.* hand
fish-hook *n.* finger
fist *n.* handwriting
fizz *n.* champagne
flapdoodle *n.* nonsense; empty talk
flicks *n.pl.* motion pictures
flog *v.* steal and sell
flush *adj.* well supplied with money
fly *adj.* smart; shrewd
flyers *n.pl.* shoes
fly off the handle: lose one's temper
frost *n.* complete failure
fug *n.* stuffy atmosphere
gaga *adj.* senile
game or gammy *adj.* lame
gasper *n.* cigarette
geezer *n.* man, esp. an old man
get fresh: become impertinent
get the bird: be sent about one's business
glim *n.* light; candle
gob *n.* mouth
gods *n.pl.* occupants of the gallery of a theatre
gong *n.* medal
grinders *n.pl.* teeth
grouse *v.* complain; grumble
grub *n.* food
gumption *n.* practical sense
half-seas over: almost drunk
hanky-panky *n.* underhand activity; trickery
have cloth ears: be perversely inattentive
hearty *n.* enthusiast for outdoor life and games

heave *n.* rob
hoof it: walk; depart by foot
hook it: depart in haste
hooligan *n.* young rowdy
hugger-mugger *adv.* in a mess
jabber *n.* pointless chatter
Jerry *n.* (1) German, esp. a soldier (2) chamber pot
Joe Soap *n.* stupid person; simpleton
joker *n.* fellow; man
jump the gun: start prematurely
kick the bucket *v.* die
kip *n.* bed; *v.* sleep
knock *v.* steal; rob
lag *n.* convict or ex-convict
lamp *n.* eye
lark *n.* escapade; spree
lead-swinger *n.* lazy person; malingerer
lid *n.* hat; cap
lift *v.* steal
lip *n.* impudence; back talk
loaf *n.* head
lolly *n.* money
loony bin *n.* lunatic asylum
lubber *n.* heavy, dull fellow
lug *n.* ear
monkey *n.* £500
mouse *n.* black eye
muck *n.* filth; worthless stuff
muck in *v.* share; help
mug *n.* face
mum *adj.* silent
nab *v.* catch; take
nark *n.* informer
natter *n.* empty chatter
nest egg *n.* sum of money put by
nicker *n.* money
night starvation *n.* sexual deprivation
nipper *n.* apprentice; young helper
nix *n.* nothing
nob *n.* (1) head (2) person of consequence
nose-warmer *n.* stubby pipe
Nosey Parker *n.* inquisitive person
not one's cup of tea: something disagreeable or unsuited to one
number one *n.* oneself
nut-case *n.* foolish person
Old Harry or Nick: the Devil
pack in (or up) *v.* finish with; give up
Paddy *n.* Irishman
pass the buck: avoid responsibility by turning problems over to others
pecker *n.* courage; spirits
peepers *n.pl.* eyes
pickled *adj.* drunk
piece *n.* woman; girl
piece of cake *n.* easy task
pinch *n.* arrest; *v.* (1) arrest (2) steal
pins *n.pl.* legs
play by ear: proceed without formulated plan
pong *n.*, *v.* stink
pony *n.* £25
posh *adj.* smart; superior in quality or style
prog *n.* food
pug *n.* boxer
pukka *adj.* genuine; first-rate
queer *n.* male homosexual
quid *n.* pound (money)
quod *n.* prison
ramp *n.* swindle; racket
randy *adj.* sexually excited
rat *v.* desert losing side
rattled *adj.* frightened
ready *n.* ready cash
ripping *adj.* excellent; first-rate
ropey *adj.* worn out, inferior
round the bend: crazy; insane
rum *adj.* queer; odd; strange
sack *v.* dismiss; discharge; fire
sauce *n.* impertinence
screw *n.* (1) prison guard; turnkey (2) wages; pay; salary
scrounge *v.* obtain by means other than buying
scruffy *adj.* unkempt; badly dressed
shark *n.* trickster; sharper
shrimp *n.* undersized person
side *n.* conceit
skin *v.* cheat; fleece
skivvy *n.* domestic servant
sloshed *adj.* drunk
slyboots *n.pl.* one who is shrewd without appearing so
smarm *n.*, *v.* flatter(y)
smasher *n.* anything excellent or outstanding
snaffle *v.* steal; take
sock *v.* beat; hit
soppy *adj.* soft; foolish
spiv *n.* flashy sharper who lives by his wits
split *v.* inform against
sponge *v.* live at someone else's expense
spotted dog *n.* kind of raisin or currant pudding
swank *n.* snobbishness; pretentiousness
swing the lead: avoid work; malinger
Taffy *n.* Welshman
tanked *adj.* drunk
tanner *n.* a sixpence
tart *n.* (1) prostitute (2) promiscuous woman
telly *n.* television
tiffin *n.* lunch; tea
toff *n.* gentleman; well-dressed, cultured person
trap *n.* mouth
tripe *n.* nonsense
trouble and strife *n.* wife (rhyming slang)
tuck in *v.* eat hastily
tumble *v.* understand
turf out *v.* throw out; eject
twerp *n.* stupid, ineffectual person
upper story *n.* brain; head
upset one's applecart: upset one's plans
wangle *v.* obtain by cunning, deceit, barter, etc.
whack *n.*, *v.* share
wild *adj.* angry; furious
with knobs on: in the extreme; *par excellence*
wog *n.* nonwhite person, esp. a native of India

Canada

baby bonus *n.* mothers' allowance
beefmaker *n.* cattle-rancher
biffy *n.* toilet; bathroom
bingo *n.* cheap wine
Bluenose *n.* Nova Scotian
Brit *n.* Britisher, esp. an Englishman
Calgary red-eye *n.* beer and tomato juice
Canuck *n.* Canadian
cheechako *n.* greenhorn; tyro
chippy *adj.* argumentative; short-tempered
codhauler *n.* Newfoundlander
deke *v.* draw out of position by feinting; *n.* such a move
Dogan *n.* Roman Catholic
dog-puncher *n.* dogsled driver
doozer *n.* anything large or remarkable
Douk *n.* Doukhobor
duster *n.* Western movie
fight-water *n.* liquor
goof *n.* cheap wine
guck *n.* anything mucky or viscous
Grit *n.* Liberal
Hab *n.* French Canadian
hard-rock *n.* strong, rugged person
heller *n.* spirited, mischievous person; hellion
hellery *n.* mischief; deviltry
Herring Choker *n.* New Brunswicker
hightail it: depart in haste
Hogtown *n.* Toronto, Ont.
home brew *n.* an athlete trained and competing in Canada
homer *n.* (1) rabid hometown fan (2) referee who appears to favour home team
hootch *n.* liquor, esp. cheap, potent whiskey
horseman *n.* *See* Mountie
hose *v.* defeat soundly; take advantage of
Husky *n.* Eskimo
hydro *n.* electricity; electric power
import *n.* professional football player who is not a citizen of Canada but who plays there
Kipper *n.* Englishman
main drag *n.* principal street of a town, section of a city, etc.
mickey *n.* a 12-ounce bottle of liquor
mint *adj.* excellent; first-rate
moose milk *n.* drink made from rich milk and rum
moose-pasture *n.* worthless mining property
Mountie *n.* member of the Royal Canadian Mounted Police
Newfie *n.* (1) Newfoundland (2) Newfoundlander
Nitchie *n.* Indian
Pea-soup(er) *n.* French Canadian
pogey *n.* relief; welfare; unemployment insurance
porch-climber *n.* cheap wine
pumpsucker *n.* teetotaler
rink-rat *n.* a juvenile who works at a hockey rink, scraping ice, etc.
rubby-dub *n.* drinker of rubbing alcohol, hair tonic, etc.
screech *n.* (1) potent kind of cheap rum (2) cheap wine
shinny *n.* ice hockey
Siwash *n.* Indian
Skimo *n.* Eskimo
snake-room *n.* beer parlour, tavern
snit *n.* state of irritation; spell of bad temper
spieler *n.* one taking part in a curling bonspiel
Spud Islander *n.* Prince Edward Islander
stick-Indian *n.* backwoods Indian
suitcase farmer *n.* Western grain farmer who spends the off-season in warmer climes
tomalky *n.* tomato juice and alcohol
yellowlegs *n.* *See* Mountie
zilch *n.* absolutely nothing; zero
zombie *n.* conscript

(W. S. A.)

Australia

Abo *n.*, *adj.* aborigine; aboriginal
art union *n.* lottery with prizes in kind, not money
Aussie *n.* Australia; an Australian
backblocks *n.pl.* sparsely inhabited inland areas
bail up *v.* hold up and rob; corner
banker *n.* flooded river running banks high
barrack *v.* shout or jeer at opponents; also, barrack for: to support; barracker: a partisan
berley *n.* small pieces of food scattered on water to attract fish
big note *v.* laud, overpraise
billabong *n.* quasi-oasal river bend
billy *n.* tin can for boiling tea and cooking; also, billycan
blow-in *n.* newcomer
bludge *v.* impose on
blue *n.* (1) summons (2) error (3) fight
bluey *n.* swagman's bundle; also called a swag, drum, or matilda
board *n.* floor of a shearing shed
bodgie *n.* the modern Australian larrikin (*see* below; the female is a widgie)
bombo *n.* cheap wine; also called plonk
bone *v.* direct a ceremonial "death curse" against a person; also point (sing) a bone
bonzer *adj.* good, excellent
bowyang *n.* strap or string fixed below the knees of a worker's pants to keep the cuffs off the ground
brumby *n.* wild horse
bucker *n.* buckjumper, a refractory horse
Buckley's chance: no chance at all
bullocky *n.* bullock driver
bumper *n.* cigarette butt
bush *n.* forestlands; the inland country in general; also go bush: to decamp or hide; bushed: confused; bushranger: an outback bandit
chromo *n.* prostitute
chyack *n.* impudence; cheek
cobber *n.* friend
cockatoo *n.*, *adj.* small farmer; cocky
cockeye bob *n.* wild gale in northwest Australia
cooee *n.*, *v.* penetrating cry
cossie *n.* swimming costume
cow *n.* objectionable person or thing
crack hardy *v.* put on a brave face against misfortune
crook *adj.* ill; worthless; go crook: become angry
cut out *v.* complete a task
damper *n.* bush bread baked in ashes or a camp oven
dead ring of: exactly similar to
deaner *n.* shilling
Digger *n.* Australian soldier
dillybag *n.* native bag made of grasses or fur twisted into cord
dingo on (someone) *v.* betray
dinkum *adj.* true; honest; also, dinky die, fair (square) dinkum
do a perish: almost to die for want of a drink
do one's block: become angry or excited
dreaming *n.* any source of native legend
drink with the flies: to drink alone
drop one's bundle: become panicky
drunk as Chloe: extremely drunk

duffer *n.* cattle thief; whence, duffing
fizgig *n.* police informer
fossick *v.* search for surface gold; look for something; whence, fossicker
full as a goog: completely full; extremely drunk
furphy *n.* baseless rumour or canard
game as Ned Kelly: highly reckless or courageous
gin *n.* adult native woman
give it a burl: make an attempt at
go bung: peter out; die
go on the wallaby: wander in the outback
go walkabout: travel; wander
graft *n.* hard work
guyver *n.* tall talk; affectation
happy as Larry: completely happy
hard case *n.* reckless or amusing person
hatter *n.* man who lives and works alone
haven't a skerrick: to be penniless
haven't the bolter's: to have no chance at all of success
hooray! *interj:* good-bye
humpy *n.* hut; small shack
in smoke: in hiding
jackeroo *n.* young station hand, learning sheep or cattle farming
jacko *n.* (1) kookaburra; (2) jack, jackass
Jimmy Woodser *n.* solitary drinker
joey *n.* (1) young kangaroo (2) baby (3) lie
jumbuck *n.* sheep
keep nit *v.* mount guard while some illegal activity is afoot; also, to cockatoo
kidstakes *n.* pretence; foolery; nonsense
lair *n.* flashily dressed youth; lairy: flashily or showily dressed
larrikin *n.* street tough or hoodlum
lolly *n.* sweetmeat
lubra *n.* young native woman
lurk *n.* (1) scheme; racket (2) dart
Maoriland *n.* New Zealand; also, Enzed, Pig Islands, Shakey Isles
Merino (pure) *n.* first class in quality
no good to gundy: worthless
nohoper *n.* person who has no prospects of success
nuggety *adj.* short; thickset; sturdy
offsider *n.* helper or companion
Old Dart *n.* England
old identity *n.* old, noted inhabitant of a locality
outback *n.* inland country; bush
paddock *n.* any fenced area of land
pastoralist *n.* any farmer engaged in the primary industries
perform *v.* curse luridly; give way to temper
plonk *n.* cheap wine
poddy *n.* hand-fed calf, lamb, foal
point *v.* take unfair advantage of another; loaf; malinger
poke borak at: jeer at; make fun of
pommy *n.* English-born person
poofter *n.* male homosexual
possie *n.* place or position
push *n.* gang or clique
put the hard word on: lodge an urgent demand
Rafferty rules *n.pl.* no rules at all
ratbag *n.* eccentric person
ready up *v.* conspire; fake; *n.* conspiracy; "stew"
removalist *n.* person or firm engaged in shifting household effects
ridge *adj.* excellent; genuine
ringer *n.* fastest shearer in a woolshed; any expert
ropable *adj.* angry; savagely ill-tempered
rough as bags: unpolished; crude
rouseabout *n.* handy man on a sheep or cattle station
rouse on *v.* upbraid; reprove
run *n.* large farm for sheep or cattle; station
sandy blight *n.* form of ophthalmia
scale *v.* ride on a tram, train, or bus without paying
shanghai *n.* catapult
sheila *n.* girl
shicer *n.* shyster; swindler
shicker *n.* intoxicating liquor; shickered: drunk
shivoo *n.* spree; party
shoot through *v.* quit; decamp
shout *v.* stand treat
silvertail *n.* social figure; social climber
sly grog *n.* liquor sold illegally; place where it is sold
smoodge *v.* make love to; curry favour
snags *n.pl.* sausages; snorks, snorkers
sool on *v.* incite
southerly buster *n.* wild wind or gale from the south
squatter *n.* owner of a large sheep or cattle station
squib *n.* coward
stickybeak *n.* inquisitive person
stoush *v.* hit; thrash; *n.* violence
sundowner *n.* tramp of indolent habits
swag *n. see* bluey; swagger, swaggie, swagman: a tramp
Tassie *n.* Tasmania; Tasmanian
tinny *adj.* extremely lucky
tipslinger *n.* race tipster
top end *n.* far north Australia; topender: resident of the far north
tote *n.* totalizator or pari-mutuel
trot (a good) *n.* sequence of successes; bad trot: sequence of failures
tucker *n.* food
turf out *v.* reject; throw away
two-up *n.* gambling game played with two pennies
up a gum tree: in a quandary
up to putty *adj.* worthless
urger *n.* race tipster; trickster's confederate
waddy *n.* stick or club
wake-up *n.* alert person
waltz matilda *v.* go on the tramp
weekender *n.* weekend holiday residence
whacko!: ejaculation of pleasure or approval
whinge *v.* complain; grouse; whence, whinger, whinging
whip the cat: cry over spilled milk
white ant *v.* undermine or sabotage
willy-willy *n.* wild storm of tornadic type
wog *n.* germ; parasite; small insect; foreign-speaking immigrant
Woop Woop *n.* (hypothetically) most rustic of all rustic townships
wowser *n.* puritanical fanatic; bluestocking
yabber *v.* talk; chatter
yacker *n.* work
zack *n.* sixpence
ziff *n.* beard

New Zealand

All Blacks *n.* New Zealand representative footballers
batch *n.* small shack or cottage; batcher: one who lives alone
bot *n.* cadger or parasite
box of birds: happy; in full health
Enzed *n.* New Zealand; Enzedder: New Zealander
fire in the fern *n.* trouble; smouldering discontent
half-pie *adj.* worthless
have the wood on: have an advantage over
hoot *n.* money
kai *n.* food
kit *n.* shopping basket
over the fence: unreasonable
pakeha *n.* white man
pie at *adj.* expert at; efficient
wahine *n.* woman
whare *n.* small shack or cottage

(S. J. Br.)

Bibliography.—For practical purposes, books regarding slang can be divided into two classes, early (virtually unobtainable) and modern (1900 to date). The latter include a few popular works, such as *A Dictionary of Slang and Unconventional English*, rev. ed. (1961), by E. Partridge, or *Dictionary of American Slang* (1960), by H. Wentworth and S. B. Flexner. In *American Thesaurus of Slang* by L. V. Berrey and M. Van den Bark, 2nd ed. (1953), the words and expressions, though not individually defined, are minutely classified and indexed. H. L. Mencken's *The American Language*, 4th ed. (1936), and its two supplements (1945, 1948) contain much lively material on slang. *A Dictionary of Americanisms on Historical Principles* (1951), ed. by M. M. Mathews, includes many slang expressions, with dated quotations. Current studies, many of them of interest in British circles, appear in the periodical *American Speech* (1925–). Australasian slang is discussed in S. J. Baker's *Australia Speaks* (1953) and *New Zealand Slang* (1941). War slang is represented in *A Dictionary of Forces' Slang, 1939–1945* (1948) by E. Partridge, W. Granville, and F. Roberts. A good sampling of sailors' slang can be found in W. Granville, *A Dictionary of Sailors' Slang* (1962). Underworld slang is represented in H. E. Goldin (ed.), *Dictionary of American Underworld Lingo* (1950), and E. Partridge, *Dictionary of the Underworld, British and American* (1950). French slang is described, classified, and indexed in A. Dauzat's authoritative *Les Argots* (1929). German slang is likewise discussed and listed in *Wörterbuch der Kunden- und Gaunersprache* (1939) by A. Bertsch. *The Literature of Slang* by W. J. Burke (1939) is a useful bibliography.

(A. McQ.; W. F. My.)

SLATE, in geology, fine-grained argillaceous or clayey metamorphic rock which cleaves or splits readily into thin slabs having great tensile strength and durability. Some other rocks that occur in thin beds are improperly called slate because they can be used for roofing and similar purposes. True slates do not, as a rule split along the bedding, but along planes of cleavage, which may intersect the bedding at any angle, usually, in the case of good roofing slates, at high angles. The original material was a fine clay, sometimes with sand or volcanic dust, and the bedding of the sediment as originally laid down may be indicated by alternating bands, differing in colour or in lithological character, sometimes to be seen on the cleavage faces of the slates. Cleavage is a superinduced structure, the result of pressure acting on the rock at some time when it was deeply buried beneath the earth's surface. On this account slates are found chiefly among the rocks of the older geological systems, although some occur in regions where comparatively recent rocks have been folded and compressed as a result of mountain-building movements in the earth's crust.

In thin sections for microscopic examination, slates show much colourless mica in small, irregular scales, which in the best average about 2,000 to the inch in breadth and 6,000 to the inch in thickness. Green chlorite in flakes is also usually abundant, the principal other ingredient being quartz, in minute lens-shaped grains. In colour, slates may be black, blue, purple, red, green or gray; dark slates usually owe their colour to carbonaceous material or to finely divided sulfide of iron, reddish and purple varieties to the presence of oxide of iron in the form of hematite and green varieties to the presence of much chlorite.

Slates are made or split from quarried blocks about three inches thick. A chisel, placed in position against the edge of the block, is lightly tapped with a mallet; a crack appears in the direction of cleavage, and slight leverage with the chisel serves to split the block into two pieces with smooth and even surfaces. This is repeated until the original block is converted into 16 or 18 separate "slates," the thickness of which depends upon many circumstances, such as quality of rock, size required and purpose for

which it is to be used, the average thickness of a roofing tile of the best kinds of slate being about ⅙ in. The slates are afterward trimmed to size, either by hand, in which case they are cut between a fixed sharp edge and a movable knife acting on the principle of a paper cutter, or by means of machine-driven rotating knives.

Slate also is sold as dimension slate and crushed slate (granules and flour). Dimension slate is used mainly for electrical panels, laboratory table tops, roofing and flooring and blackboards. Crushed slate is used on composition roofing, in aggregates and as a filler. Principal production in the United States is from Pennsylvania and Vermont; lesser quantities are quarried in Maine, New York, Virginia, Georgia, Arkansas and California. North Wales provides most of the slate used in the British Isles. *See* also STONE. (F. J. N.; F. J. P.)

SLAUGHTERHOUSE (ABATTOIR), a place where food animals are butchered.

Methods of Slaughter.—In early times domestic meat animals were slaughtered by strangulation or by piercing the brain through the eye sockets with heated spears. The blood was cured in as an essential part of the meat. The Judaeo-Christian distaste for blood led to development of new practices: animals were dispatched by a head blow or by quick severance of the jugular vein, and the carcass was hung head downward to bleed (*see* also KOSHER). After bleeding, cattle and sheep were skinned, while pigs were dipped into vats of scalding water so that hair or hide or both could be easily removed.

In the modern abattoir, animals are usually stunned before being bled, a practice that indicates interest in the humane treatment of animals produced for slaughter. In both the United States and Europe the desirability of stunning was recognized before the end of the 19th century, and mechanical equipment for this was developed. Cattle may be stunned by means of a captive-bolt pistol or a pneumatic gun. Sheep and pigs may be stunned by pistol, by electric shock, or by anesthetizing in a carbon dioxide chamber. After World War II compressed-air stunners were commonly used for cattle and gas chambers for smaller animals. From 1960 U.S. federal legislation required federal agencies to buy meat produced only from animals slaughtered by humane methods.

In addition to stunning devices, large modern abattoirs use a wide variety of mechanical, electrical, and pneumatic equipment for the various slaughtering and carcass dressing and processing operations. The moving chain-conveyor production line system is widely used where the scale of operations warrants it.

History.—*United States.*—Early slaughterhouses were operated on a small scale, but the increasing demand for meat and the concentration of populations in the expanding urban areas led to the establishment of large abattoirs.

The first known meat packer in America was William Pynchon, treasurer of Massachusetts Bay Colony and founder of Springfield, Mass., who during the 1640s built an abattoir on the bank of the Connecticut River, 7 mi. (11 km.) S of Springfield. There he processed salt pork and beef for shipment to West Indies plantations. Pynchon's son John reduced the seasonal nature of meat production by wintertime stall-feeding of cattle, and in 1655 he trailed a herd of fat spring beeves to Boston. Since the cowboy was, and is, a commercial cattle drover, John Pynchon is regarded as the "father" of the American cowboy.

Commercial abattoirs developed along the Atlantic Seaboard, in the Southwest, and in California during the 17th and early 18th centuries. The cowpens of the Carolina and Georgia mountains supplied meat and leather for the coastal cities and the West Indies. Spanish and Indian *vaqueros* and *ciboleros* doubled as cowboys and butchers in the Far West. They and their West Indian counterparts, the boucan makers (*i.e.*, buccaneers), developed "jerky," the modern dried beef and originated the skills of the roundup and the first crude rodeo (*q.v.*).

Meat packers followed the livestock west after the Revolution. After 1830 Cincinnati, because of its location on the Ohio River waterway, became the largest western meat-packing centre, hogs and cattle being driven there for processing during the fall months. Rapid methods of dispatch and cure were necessary so that the salt and smoked meats might be shipped west and south before the waterways froze. Each workman performed a single cut or act as the carcasses, suspended head down from endless belts, were pulled slowly across the processing room. In Cincinnati the first major experiments were conducted in year-round refrigeration of meats.

The impoverishment of Texas after the Civil War, coinciding with the U.S. Army's campaigns against western Indians, sent millions of wild Longhorn cattle trailing north and northwest to the prairies and high plains (*see* LONGHORN). This, and the coincidental development of transcontinental railroads, enabled Chicago to become the nation's greatest meat-packing centre. Until the late 1860s live animals were shipped to eastern cities and to Europe. Then refrigerator cars were introduced (*see* REFRIGERATION: *Applications of Refrigeration*), and by 1870 Chicago packers processed, dressed, and cured more than 10,000,000 meat animals each year. Succeeding decades saw the development of huge abattoirs and processing plants in Chicago, Omaha, Kansas City, Denver, and other railway centres of the prairie and West.

The Cincinnati gravity system was generally followed. Cattle were driven via inclined ramps to the top floor of the plant, where they were stunned by mallet blows and hung from moving belts to travel slowly downward through the jugular-cut, bleeding, and dissembly stages of processing. Pigs and sheep were shackled, suspended, and then "stuck." The entire operation from shackling pen to carcass required less than one minute.

After 1920 the bulk of livestock transit shifted from railroad to highway truck, reducing the dependence of abattoirs on rail centres. Radio and, later, television enabled growers and feeders to keep abreast of market quotations and needs. Hundreds of auction markets thus developed, in small towns and at crossroads, where livestock was sold to middlemen or directly to packers through the British system of open bidding. By 1950 these changes had led to a decentralization of slaughtering and a decreased use of some of the large abattoirs, including the vast Chicago Union Stock Yard.

Europe.—Slaughterhouses in Europe during the 19th century were governed by purely local considerations and were small, serving consuming areas. The local town council often provided a public abattoir, thus maintaining control of the interests of public health, at which any farmer or butcher could slaughter his own animals on payment of a fee. Consequently European abattoirs did not adopt mechanical equipment as early as the large meat-packing centres in the U.S. did. Nevertheless, in the 1960s, in some European countries large and well-equipped abattoirs were in use.

Meat Inspection.—Local and central government authorities in many countries have long recognized that without adequate supervision and control during slaughter, meat can be a serious source of infection. In the United States, charges of insanitary conditions in slaughterhouses, originating in British medical journals and publicized by Upton Sinclair's novel *The Jungle*, led, after 1906, to federal legislation providing for the sanitary supervision of slaughterhouses and meat-packing plants, and of their products, by federal inspectors. Chicago's sprawling and odorous Union Stock Yard, founded in 1865, was the most famous of the livestock terminals brought under federal inspection by this legislation.

Public health authorities in nearly all countries in the 1960s operated a system of rigid inspection and control. Animal health was checked on the farms and at the abattoirs, prior to slaughter, and the carcasses were checked during and after slaughter. The risk of diseases being transmitted by meat has thus been reduced to infinitesimal proportions. (*See* further MEAT: *Inspection.*)

Refrigeration maintains the meat in good condition after slaughter. By dispelling the body heat from the carcass as quickly as possible, the risk of bacterial growth is significantly reduced. The development of mechanical refrigeration in the second half of the 19th century enabled the sheep-producing countries, notably Australia and New Zealand, to develop a substantial export trade in frozen lambs, and large abattoirs have been established in those countries to serve that trade.

Slaughterhouse By-Products.—Apart from the principal products of meat, the slaughter of animals produces many by-products, the most important of which are hides, for the manufacture of leather (*q.v.*), and fats that may be rendered into lard and used in the manufacture of soap.

When synthetic detergents and vegetable shortenings reduced the markets for these animal fats, large firms built plants where fat surpluses could be converted into fatty acids and complex chemicals for use in a great variety of products. Intestines are cleaned and processed for use as sausage skins, and glands are used in the pharmaceutical industry and for therapeutic purposes. Blood, which is rich in proteins (but which is drained from the animal to improve the keeping quality of the meat), is used in manufactured meat products, such as the English black pudding; in the manufacture of adhesives, animal feeding meals, and fertilizers; in the clarification of wines; and for other industrial purposes.

See also MEAT: *The Meat Industry;* FOOD SUPPLY OF THE WORLD; and individual articles such as BEEF; PORK; LAMB AND MUTTON; SAUSAGE AND READY-PREPARED MEATS, etc., for production, distribution, and preparation of meats. Further information on meat processing will be found under FOOD PREPARATION; FOOD PRESERVATION; and CANNING, COMMERCIAL.

BIBLIOGRAPHY.—Sigfried Giedion, *Mechanization Takes Command* (1948); American Meat Institute, *The Meat Reference Book* (1958); A. R. Miller, *Meat Hygiene* (1951; new ed., 1958); Walter Prescott Webb, Robert West Howard, *et al., This Is the West* (1957); Edward N. Wentworth and Charles W. Towne, *Cattle and Men* (1955); Robert B. Hinman and R. B. Harris, *The Story of Meat* (1939).

(R. W. HD.; JO. A. R.)

SLAVE COAST, a term denoting the low-lying deltas and sandbars of the West African coast that stretch for about 700 mi. (1,130 km.) between the Volta River and Mt. Cameroon used (but now rarely) to distinguish this shore line from the Gold, Ivory, and Grain coasts. The name derives from its former position as a principal source of African slaves for 350 years, especially the section of it that fronts on the Bight of Benin. (D. WH.)

SLAVEIKOV, PETKO RACHEV (1827–1895) and **PENCHO** (1866–1912), Bulgarian writers, father and son, who helped to enrich Bulgarian literature by establishing a modern literary language and introducing contemporary ideas from other European countries.

Petko Rachev Slaveikov, born in Trnovo on Nov. 17, 1827, became an itinerant schoolteacher at 17. His early poems were lyrical and patriotic (*Smesena Kitka, Pesnopoyka,* both 1852), and, by reestablishing the vernacular as a medium for literature (the language of his translation of the Bible in 1862 is also based on Bulgarian dialects), he prepared for the flowering of native poetry. As a patriot and politician, he helped to shape resurgent Bulgaria, producing political pamphlets notorious for their outspokenness against Turkish oppression and against the spiritual domination of the Greek patriarchate. In 1863 he moved to Istanbul, where he contributed to Bulgarian *émigré* reviews and edited satirical and political periodicals. After his country's liberation (1878), he became an active politician, both as president of the constituent assembly and as co-founder of the Democratic Party. After the 1881 *coup d'état* he went to Plovdiv, then still under Turkish rule, and there edited the newspaper *Nezavisimost* ("Independence"). He died in Sofia on June 9, 1895.

Pencho Slaveikov, the youngest of Petko Rachev's eight children, was born at Trevna (Tryavna) on April 27, 1866. He studied literature and philosophy in Germany (1892–1908) and traveled widely despite being an invalid after 1884. Pencho Slaveikov, who was inspired especially by Goethe, Heine, and Nietzsche, as well as by the simple eloquence and realism of Bulgarian folk songs, is best known for his unfinished epic poem *Kurvava Pessen* ("Song of Blood"; written 1911–12; published 1913), which describes the sacrifices of the Bulgarian people in their struggle for independence. He was also an outstanding essayist and translator of German literature. He died on May 28, 1912, at Brunate, near Como, Italy.

BIBLIOGRAPHY.—*Izbrani proizvedenya* (1945) gives a selection of P. R. Slaveikov's poetry. Pencho Slaveikov's collected works have been edited by B. Penev, 7 vol. (1921–25). *See* also G. Hateau, *Panorama de la littérature bulgare contemporaine* (1937); L. B. Picchio, *Storia della letteratura bulgara* (1957).

(L. BY.)

SLAVERY is the social sanctioning of involuntary servitude imposed by one person or group upon another. Until the beginning of the 20th century, chattel slavery, involving the legal right of property, or ownership, generally was distinguished from other forms of servitude. But the progressively widespread prohibition of such slavery was accompanied by increasing resort to various practices and institutions to circumvent that prohibition. Therefore the concept of slavery has been broadened to include its possible substitutes which, while they do not necessarily imply the right of ownership of one person by another, nevertheless permit his exploitation under conditions of bondage. Some of these substitutes, such as indentured, or contract, labour and serfdom, have played distinctive roles in the history of civilization, at different times and places and within different cultures, and these historic roles are dealt with in separate articles (*see* CONTRACT LABOUR; SERFDOM AND VILLEINAGE). Similarly, movements to abolish the slave trade and slavery and its substitutes have had historical and political significance broader than the subject of slavery itself, and these are discussed in ABOLITION MOVEMENT; FORCED LABOUR; CIVIL LIBERTIES; and HUMAN RIGHTS, among others (*see* also biographies of leading abolitionists).

The main divisions of this article are as follows:

I. INTRODUCTION

An example of the broadened, 20th-century concept of slavery is offered by a 1947 decision of the Supreme Court of the United States (*United States* v. *Ingalls;* appealing the finding of a lower court that a California couple had enslaved their maid), in which the court examined the classical definitions of slavery based on the right of ownership and observed that they did not apply to the case under consideration. The court decided, nevertheless, that the appellant had been properly adjudged guilty by the lower court of having imposed slavery on "a person wholly subject to the defendant, . . . one who had no freedom of action and whose person and services were wholly under the control of defendant and who

was in a state of enforced compulsory service to the defendant." As will be seen, a similar concept was adopted by the League of Nations and even more so by the United Nations in the "Supplementary Convention on the Abolition of Slavery, the Slave Trade, and Institutions and Practices similar to Slavery" adopted in 1956.

Even from a purely historical point of view a definition of slavery broader than the one based on ownership would seem warranted. Although there is no doubt that in a great number of societies slavery was more or less modeled after the right that an individual could exercise over an animal or an inanimate object, enslavement as an economic and social phenomenon preceded the appearance of the concept of ownership. Indeed this concept could not come to the fore until the individual was clearly distinguished from the group as an autonomous centre of rights and obligations and thus recognized as a "person before the law." The primary fact was that a man made another work in his own interest; the idea that he had a right to do so came only later.

1. Origins of Slavery.—Explanations of the origins of slavery are necessarily speculative, constituting *post hoc* rationalizations. However, it would seem that through the hunting and food-gathering stages of civilization slavery would have been unknown; the hunter or food-gatherer simply would have had no use for a slave, who would have been more of a burden than an asset for his master, as he would have been an additional mouth to feed. This reasoning would seem to be supported by the fact that such hunters' societies as have survived into historical times, as for example those of most of the American Indians and of the Australian aborigines, do not seem ever to have practised slavery.

The situation apparently was similar in certain communities living mainly from fishing: Eskimos, for example, have never had slaves. But in other instances fishing tribes have had a more sedentary life, and certain tasks could, with some profit, be assigned to another, as, for example, making and repairing nets, canoes, etc. It is possible that this explains why a few slaves have been found in some fishing communities.

In any event, slavery as an institution definitely made its appearance when tribes reached the pastoral as distinct from the hunter stage. Even then it was modest as far as the number of slaves was concerned because the care of flocks required only a few hands. Furthermore, life was simple and there was little difference between the condition of a slave and that of the wives, sons, or daughters of his master, who were also subject to the latter's absolute authority.

The exploitation of slave labour increased at the agricultural stage. But as long as agriculture retained the character of a subsistence economy, slavery continued to bear on the whole the "patriarchal" or "domestic" character of the pastoral period.

2. Expansion of Slavery.—The real change in the character of slavery, as well as its expansion, came as a result of the change from a subsistence to a market economy. When large tracts of land fell into the hands of a relatively few wealthy landowners, and only one or a few staple products were cultivated, it was generally profitable to employ the labour of slaves working in large gangs. Under these conditions the slaves became an anonymous multitude, and their treatment was quite different and much harsher than when they still were considered members of their master's household. Examples of this type of "praedial" or "plantation" slavery will be found in other parts of this article (see *Slavery in Antiquity: Greece and Rome*; and *Modern* [*or Plantation*] *Slavery*, below).

When, on the other hand, the crops raised on large estates were more diversified, and husbandry was practised side by side with agriculture, it was often economically sounder to partition the land into small plots. In this case the method of exploitation led to serfdom rather than to slavery (*see* SERFDOM AND VILLEINAGE).

In the cities slaves were used not only in the domestic service of the master and of his family in all kinds of capacities, and exploited for the satisfaction of the master's lust or of his artistic tastes, but also for work in mining, industry, commerce, and the like. Slave girls staffed the brothels and filled the harems of the ancient world. There were also at times slave policemen and slave warriors. Among the latter some achieved great power and sometimes took the place of their masters and even founded dynasties. Such cases, of course, were exceptional and usually slaves were treated as an inferior class or caste. However, here again there were considerable variations from one society to another, and even in the same society according to the social function entrusted to them and often also in accordance with the circumstances which brought about their servile condition.

3. Sources of Slaves.—It would seem that individuals have been reduced into slavery mainly by capture, or force; as punishment for some offense; by birth from slave parents; by sale for nonpayment of debt; or by sale by parents, guardians, or chieftains.

Undoubtedly capture has been the most ancient source of slaves. And once the need for slave labour increased, expeditions were undertaken with the specific view of capturing human beings to reduce them into slavery. Thus slave raiding and kidnapping (*q.v.*) became the major means of supply. They were practised in antiquity as well as in modern times—until the end of the 19th century and perhaps even later.

Punishment for crime is also a very ancient source of slaves. When society took the right to punish for crime into its own hands, criminals usually became publicly rather than privately owned slaves, as in the case of condemned criminals sentenced to the galleys (*see* GALLEY). Although slavery of this kind seems to have disappeared, a certain analogy to it may be found in penal servitude, and perhaps also in forced or compulsory labour employed by some colonial powers and imposed by certain governments upon political opponents (*see* FORCED LABOUR).

A third source of slaves—birth—was the consequence of the acceptance of the idea that the slave was the property of his master. Consequently, the child born of slave parents belonged to their owner. However, birth has often been a less important source of slaves than one would imagine: in particular under the conditions of mass slavery such as those that prevailed on the Roman *latifundia*, and later on American plantations, there were usually a far greater number of men than of women among the slaves. Furthermore it was often more profitable to buy grown-up slaves than to wait until children reached an age where they could be put to work.

Another major source of slaves was sale. This needs some explanation. Insofar as slaves were considered as objects of property, they were subject to sale—and the main purpose of slave raiding was to supply the slave market. In these cases, sale is not considered as the source of slavery, as the individuals concerned already were slaves (by capture or kidnapping, birth, sentence, or prior sale).

In many societies, however, it has been possible for a free individual to sell himself, or a person under his authority—a wife or concubine, a child or grandchild, or a ward—into slavery for money or other consideration; or a tribal chieftain might sell individuals, families, or groups under his jurisdiction, or sanction their forced removal by traders.

Although the sale of persons into chattel slavery has been virtually eliminated, it must be stressed that practices leading to similar results frequently have outlived legal prohibitions. One such practice has been the sham adoption of children, especially girls, of extremely poor households by wealthier people who pay a certain price to the natural parent and thereafter exploit the child. Such practices were widespread even in the 20th century in China and other countries of Southeast Asia, where they were known under the Chinese name of *Mui Tsai* (little sister), and also in certain regions of South America, where the "adopted" child was frequently of Indian origin. Certain customs, in Africa, Australasia, and elsewhere, also may sanction the sale of a girl as a sexual partner by her parents, guardians, or clan. In the same regions a husband may lend any female member of his household to another man as a form of hospitality; and a husband's heirs sometimes may "sell" his widow or widows to other men.

Closely related to slavery resulting from sale is slavery resulting from the nonpayment of debt. In a considerable number of civilizations an insolvent debtor either became his creditor's slave, or could be sold by the latter to a third person. Here again cer-

tain practices have perpetuated in disguise this kind of enslavement even after its official legal abolition. Usually it takes the form of debt bondage, a condition arising from a pledge by a debtor of his personal services, or those of a person under his control, as security for a loan. The main feature of such an arrangement is that the value of services furnished is not applied toward the liquidation of the debt. Consequently the individual pledged has practically no hope ever to escape his servile condition. And, as it is not unusual in such cases that the debt remains due after the death of the original debtor, devolving on his next of kin, the servile station may become hereditary. Despite measures taken to put an end to servitude for debt, it has survived in many regions even in the second half of the 20th century. It occurs mainly in underdeveloped agricultural countries, but even economically more advanced countries have not been entirely exempt from it. In agricultural communities it often happens that the lender of money or credit is at the same time a wealthy local landlord—while the debtors are the impoverished farmers. (*See* also PEONAGE.) In more industrialized communities, the moneylender either directly exploits the debtor's labour for his own profit, or hires him out to some third party.

4. Lack of Opposition.—A last observation would seem to be in order. Although slavery in its various forms was an almost universal institution, or rather because of that, little or no opposition was raised against it. There were, of course, denunciations of its excesses, at times, and attempts to remedy abuses, but the existence of the institution was not questioned until the beginning of the modern antislavery movement at the end of the 16th and the beginning of the 17th century.

As late as the 11th century, St. Anselm stated that it was only natural that children born to slaves should follow the status of their parents, and St. Thomas Aquinas (1225–74) said that slavery was one of the inescapable consequences of Adam's original sin. And when, later, slavery was introduced into the New World, the Roman Catholic Church expressly recognized the validity of the institution, although Catholic missionaries made great efforts to mitigate some of its extreme consequences. The attitude of other Christian churches did not differ much from that of the Roman Catholic. During the first centuries most of the heretic or schismatic churches admitted slavery. Similarly, when after protracted disputes (451–1053) the Eastern or Greek Church broke away from the Church of Rome, it accepted slavery. Nor did even the Reformation change the attitude in regard to slavery. Both the Anglican and the Lutheran churches, and also the Presbyterians, accepted it as a fact. But, as shall be seen, religious feelings which inspired certain Protestant sects, and in particular the Quakers and the Wesleyans, were at the origin of the antislavery movement. *See* also ABOLITION MOVEMENT. (E. N. Go.)

II. SLAVERY IN ANTIQUITY

A. THE ANCIENT NEAR EAST AND EGYPT

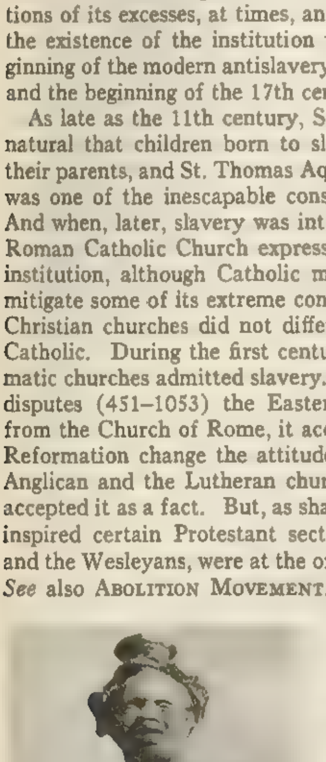
1. Sources of Slaves in Antiquity.—Warfare was the earliest source of slaves in the ancient Near East and always remained a relatively important one. Originally captives seem to have been slaughtered; later women and then men were spared to serve their captors: the Sumerian word for slave woman means literally "woman from a foreign land," and appears at an earlier date than that for a man slave. Some prisoners passed into private possession, but most of them became the property of the kings (in Egypt all captives belonged in the first instance to Pharaoh), who either kept them or dedicated them to service in temples. The numbers of royal (state) and temple slaves thus created often grew considerably. Temples constituted one of the richest financial, commercial, and agricultural units within the communities, and slaves were often "dedicated" to the temples by kings or individuals hoping to secure thereby favours from the gods. They were the only large aggregations of slaves in pre-Classical times, and there seems to have been no equivalent of the great private slave households of Roman times (*see* below).

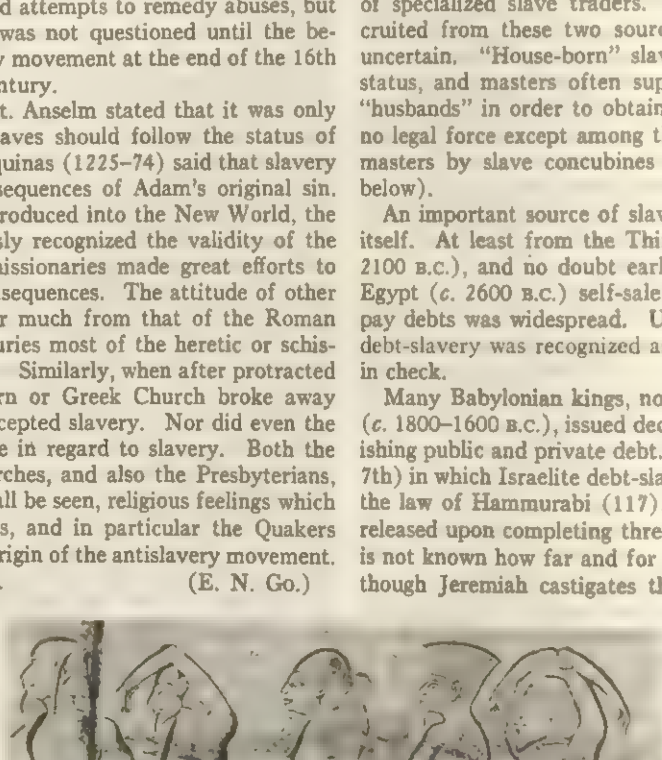
The supply of slaves by war was supplemented by purchase, usually from among neighbouring peoples. Certain tribes were famous for producing slaves of exceptional strength or beauty *e.g.*, the Gutian and Lullubu hill tribes, and merchants visiting these areas carried on a trade in slaves, but there is no evidence of specialized slave traders. How far the slave population recruited from these two sources was able to reproduce itself is uncertain. "House-born" slaves enjoyed a slightly higher social status, and masters often supplied their slaves with "wives" or "husbands" in order to obtain them. However, such unions had no legal force except among the Hittites. Children born to their masters by slave concubines fell into a different category (*see* below).

An important source of slaves lay within the native population itself. At least from the Third Dynasty of Ur in Babylonia (*c.* 2100 B.C.), and no doubt earlier, and from the Old Kingdom in Egypt (*c.* 2600 B.C.) self-sale or the sale of children in order to pay debts was widespread. Unlike the ordinary form of slavery debt-slavery was recognized as a social evil which should be held in check.

Many Babylonian kings, notably in the Old Babylonian period (*c.* 1800–1600 B.C.), issued decrees liberating debt-slaves and abolishing public and private debt. The Hebrew Sabbatical Year (the 7th) in which Israelite debt-slaves were to go free was paralleled in the law of Hammurabi (117), by which such persons were to be released upon completing three years' work for their creditor. It is not known how far and for how long these laws were observed, though Jeremiah castigates their evasion (Jer. 34:8–17). Poor Jews were forced to sell their children into slavery to pay their taxes after returning from the Babylonian Captivity (Neh. 5: 1–5). (*See* also BABYLONIA AND ASSYRIA: *Civilization: Slavery.*)

2. The Status of Slaves.—Slaves and freemen were not generally divided racially or culturally as they have been in more recent slave societies. Moreover between slave and freeman there existed a variety of semifree classes dependent on king or temple, and even free citizens were commonly liable to labour-service (*corvée*). Nevertheless, a slave was not a human being before the law, but a "head" like an animal. He seems to have enjoyed no protection from his master and if he was injured by someone else, the case was considered as one of damage to his master's





BY COURTESY OF (LEFT) THE TRUSTEES OF THE BRITISH MUSEUM, (RIGHT) THE ORIENTAL INSTITUTE OF THE UNIVERSITY OF CHICAGO

ETRUSCAN AND EGYPTIAN SLAVES

(Left) Clay effigy of an Etruscan slave from the 7th-century B.C. Castellani tombs at Praeneste, Italy. The wreath and titulus on the slave's chest indicate that he is for sale. (Right) Bound prisoners captured during the military campaigns of Ramses III and later impressed into slavery: from left to right, Libyan, Semite, Hittite, Philistine, and Semite. Carved on the wall at Medinet Habu, 12th century B.C.

property. Hebrew law provided that if a master beat his slave to death and the victim lingered on for a day or two, the master went unpunished "for the slave is his money" (Ex. 21:20–21). In the circumstances, runaways were numerous, and in Babylonia the laws of Hammurabi prescribed death for harbouring a fugitive; other societies, notably among the Hebrews, were more lenient. Probably most masters found that it paid to treat slaves reasonably well, and sometimes inducements were held out to them, such as the privilege of retaining small gratuities, or the prospect of ultimate liberation. Manumission (freedom) was often granted on condition that the freedman (who was frequently adopted) should support his ex-master during his lifetime; any failure to do so led to reenslavement. The freedman thus was in an intermediate position between slavery and complete freedom; he might be required to remain for life or a definite time with, or perform some special duty or service for, his former master or a person named by him, and his failure or refusal to do so would subject him to being again reduced to slavery. Branding of slaves appears not to have been widespread. The reference, in ancient Babylon, to the "cleansing of the forehead" is probably symbolic; some slaves appear to have been tattooed, whereas others were distinguished by small clay tablets worn around neck or wrist. In Neo-Babylonian times, the owner's name was tattooed on slaves' wrists and temple slaves were often marked with a star, the symbol of the goddess Ishtar.

Female slaves were in a most vulnerable position. They might be prostituted against their will, and seem, as a consequence, to have formed an increasing proportion of the prostitute class. Alternatively they might be passed from one slave to another in order to breed more slaves. This latter was specifically provided for under Hebrew law (Ex. 21:2–11) and contracts from Nuzi (near modern Mosul), *c.* 1500 B.C., show that parents who were obliged to sell their daughters under the guise of adoption actually wrote such a clause into the contract, as a lesser evil than prostitution. A slave girl's best hope was to become her master's concubine. If her mistress were barren, her children might become legitimate heirs, and in any case her status was improved. Clause 171 of the laws of Hammurabi provided that a slave concubine and her children were to be released on her master's death.

The capacity of slaves to carry on business on their master's behalf was recognized, under rigid safeguards, in Babylonian times. By the Neo-Babylonian period (*c.* 620 B.C.), slaves carrying on business—as merchants, bankers, artisans, tenant farmers—and living on their own were commoner; some of them even owned slaves of their own. Any property acquired by a slave, however, remained legally the possession of his master, the slave making a payment for the privilege of engaging in business.

The status of freedmen is unclear. Both in Egypt and Babylonia, slaves liberated by adoption seem to have acquired the status of their adoptive parents, and thus might become citizens. This can hardly have been so in the case of the free but illegitimate children of a master and his slave concubine, or of slaves manumitted in other ways.

3. Economic Importance.—Slaves were to be found in almost every field of work in pre-Classical antiquity, but in the absence of reliable statistics it is impossible to estimate the economic importance of slave labour. The number of slaves in Egypt before the wars of the New Empire (*c.* 1560 B.C.) was probably insignificant and they were by no means without rights. Those attached to the land, however, were sold with it, although the Egyptians do not appear to have used the word "slave" of persons attached to the soil, as much of the population was. Elsewhere numbers were larger and tended to rise. Maximum figures for slaves in private households rose from 20 or more in Old Babylonian times to over 100 a millennium later under the Assyrian Empire (745–612 B.C.). Averages were far smaller and only the well-to-do could normally afford slaves at all. Prices rose steadily despite increasing numbers, and kept pace with the rise in prices generally. Most privately owned slaves in cities were domestics; in the country they worked alongside their masters in the fields. They seem never to have presented a threat to the existence of free tenant farmers, though true slave farmers existed on large estates in the Assyrian Empire, and could be bought and sold either with the land they worked or separately. This was probably exceptional, because of the conditions of the period. Slaves seem to have been rarer in skilled trades and professions, though slave apprentices are mentioned, especially in Neo-Babylonian times, in contracts concluded between their owners and master craftsmen. Skilled workmen were also to be found among the temple slaves, but were probably exceptional there also. In general, craftsmen were freemen. It seems certain, therefore, that any threat to the economic conditions of free persons by slaves was confined to the unskilled, and it can only have been acute in times of an over-abundant supply of labour, but to judge from the retention of unpopular forced-labour service, a shortage of labour was much commoner. (D. G. E.)

B. Greece and Rome

Slaves are mentioned in the Homeric poems, which probably reflect the conditions of the early first millennium B.C., and play a prominent role in the legal code of Justinian (A.D. 527–A.D. 565). During these 15 centuries slavery was an accepted institution, never seriously challenged. In the 5th and 4th centuries B.C., it is true, philosophers questioned slavery, like all other fundamental human institutions, such as the state and the family. As Plato was not interested in the problem, we know little of the debate, but it appears from Aristotle that some philosophers declared that slavery was contrary to nature, and therefore unjust. Aristotle himself took an opposite view. According to him, many—indeed the majority—of the human race lacked those higher qualities of the soul which were necessary for freedom. Slavery was not only good for the master, who was provided with living instruments, but for the slave, who received guidance which he was incapable of providing for himself. Aristotle admitted that in practice men free by nature (*i.e.*, the Greeks) were enslaved "by accident," and natural slaves (*i.e.*, barbarians) were allowed to be free; but although the institution was not always properly used, it was fundamentally natural and good. Later philosophical schools, like the Cynics and Stoics, who preached the brotherhood of man, held that slavery was contrary to nature, but they did not therefore suggest its abolition. Freedom and slavery, like all material concerns, were, they taught, matters of indifference. To Christians, there were neither bond- nor freemen in Christ, but in this world they counseled masters to be kind to their slaves and slaves to be obedient to their masters. The doctrine of the Roman law was summed up by Justinian in the sentence, quoted from the 2nd-century lawyer Florentinus: "Slavery is an institution of the *ius gentium* (law of nations), whereby a man is, contrary to nature, subjected to the ownership of another." No other legal institution is expressly admitted to be contrary to natural law; but after this formal acknowledgment of philosophical theory, the Roman lawyers plunged into the practical rules governing slavery without demur.

1. Sources of Greek and Roman Slaves.—The sources of slaves were many and varied. In primitive Athens and Rome insolvent debtors might be enslaved by their creditors. This was forbidden by Solon at Athens in the 6th century B.C., and in the late 4th century B.C. in Rome, but the rule probably prevailed longer in many communities. With the extension of Roman law (*q.v.*) to the whole empire, slavery for debt ceased to be legal. In Greece children exposed and abandoned by their parents became the slaves of those who reared them. Roman law varied on this point until Justinian in the 6th century declared that all exposed infants were free. Poor parents also sometimes sold their children. This was prohibited by Roman law, but was nevertheless not uncommon, especially under the later empire; the sale of newborn infants was legally recognized by Constantine the Great. A free man might, under Roman law, allow himself to be sold into slavery, and could not reclaim his freedom if he was over 20 and had received a part of the price paid for him. A few offenses were punished both in Greece and Rome by sale into slavery, and under the Empire a common penalty for crime among the lower orders was slavery in the state mines and quarries.

The principal sources of slaves, however, were war, piracy and

kidnapping, breeding, and import from barbarian lands. Victors in war always possessed, and frequently exercised, the right of selling captives. In Greece combatants taken in battle were frequently ransomed, but when a city was captured the whole population was often sold into slavery. Since wars were constant many Greeks were thus enslaved. Piracy also flourished unchecked except while Athens policed the sea. But the majority of slaves in classical Greece were barbarians imported by traders, mainly from Asia Minor and Thrace and the Black Sea area. The supply from these sources was abundant and prices were low—an unskilled adult male seems to have cost a sum roughly equivalent to a year's maintenance. In these circumstances few slaves were bred, slaves being more cheaply replaceable by purchase. Female slaves in domestic service bore children to their masters and fellow slaves and the children were no doubt often reared as slaves. The majority of slaves, however, were men employed in agriculture, industry, and mining.

In the Hellenistic period (323–30 B.C.) wars were frequent and widespread, and piracy was rife. While Rome was conquering the Mediterranean world, wars increased in scale and became more ruthless and thousands of captives from Spain, Gaul, Africa, Greece, and the East were flung upon the market. With the decline of Rhodian sea power after 167 B.C., piracy flourished on a scale hitherto unknown, until in 67 B.C. Pompey cleared the seas. Slaves were imported on a vast scale to Italy and were used in ever increasing numbers in agriculture and mining.

With the establishment of the principate (27 B.C.) conditions changed. Under Augustus there were still wars of conquest and many thousands of barbarian prisoners were enslaved, but thereafter large-scale wars were rare. Peace reigned within the empire, piracy and brigandage were effectively suppressed and without these two major sources, slaves could be imported only from beyond the Rhine, the Danube, and the Euphrates. In the circumstances slaves became scarcer and their price rose; by the 2nd century A.D. they seem to have cost—in real value—eight to ten times as much as in Athens of the 4th century B.C. Breeding in these circumstances became necessary and profitable. Cato, in the 2nd century B.C., had permitted the bailiffs of his estates to keep wives, but had allowed no other women slaves on his lands, replacing the labourers by purchase. From records of the 2nd century A.D. it appears that fertile slave women were valued and commanded a higher price.

From the 3rd century onward wars with the barbarians revived and increased in intensity, bringing in their harvests of prisoners. These, however, were not always thrown upon the slave market; the government often preferred to enroll them in the army or to settle them on the land as free men, liable to military service. Many thousands of Roman citizens were, moreover, carried off as prisoners by barbarian raiders. Many were ransomed by their relatives or by the church, but a large number were reimported into the Empire as slaves, and although they were legally entitled to recover their freedom on repayment of their purchase price, most were unable to do so. In these circumstances the price of slaves dropped to about half of what it had been in the 2nd century, but slaves remained very substantially dearer than they had been in classical Greece or under the later Roman republic. It was still more economical to breed slaves than to buy them.

2. Economic Importance.—There were few activities which were not common to free men and slaves. The latter were, of course, excluded from political life and only citizens could hold magistracies (public office). Slaves were, moreover, excluded from military and naval service; only in moments of extreme crisis were slaves enrolled in the armed forces, being freed for the purpose. Domestic service, however, was reserved for slaves; it is almost unknown in classical antiquity to find a free man as a personal servant. It is often stated on the basis of statements in Plato, Aristotle, and Cicero that the Greeks and Romans regarded manual work, with the exception of agriculture, as degrading for free men and only fit for slaves. The great majority of Athenian citizens, however, were either peasants, craftsmen, or labourers, and felt no shame in the fact; they worked even in the mines. But while the hardest manual labour was tolerated, any employment which involved being subject to another man's orders was considered servile and degrading, even if the work was managerial and not manual. This attitude seems to have been universal in the Greek and Roman world and it had far-reaching effects on the economy.

It meant in the first place that slaves (or freedmen) were normally employed in secretarial and managerial posts. In 4th-century Athens, bank managers were often slaves or freedmen of the bank owners. Within the sphere of agriculture the bailiff, who was directly responsible to his owner, was normally a slave, even though the estate was worked by free tenants and free casual labour. In the early principate in Rome a curious anomaly arose. The emperor's principal secretaries and accountants, who were slaves or freedmen, inevitably became politically influential and important, and came to do the work of secretaries of state and ministers of finance. This situation caused great bitterness among the senatorial aristocracy; but it was only after a century that the problem was solved, when the offices of the imperial household came to be regarded as public rather than as domestic posts and became acceptable to members of the equestrian order, though not to senators.

Secondly, it was virtually impossible for any industrial enterprise to expand beyond the scope of a household business without the use of slave labour. A free man would accept casual jobs, usually on a contract basis, but he was generally unwilling to submit himself to regular work under a master's orders. A craftsman might employ members of his family and perhaps apprentices, but if he wished to expand his business he had to buy slaves. Except when slaves were very cheap, this meant a heavy and risky capital outlay, and as a result industry tended to be on a small scale, unless conditions were exceptionally favourable, as in mining.

There are no reliable figures for the number of slaves, either absolute or relative to free persons, but it is clear that numbers varied very greatly in different periods and areas according to the wealth and economic development of the receiving areas and the availability of alternative forms of labour. In classical Greece the Spartans had few slaves, since there was a serf population, the helots (*q.v.*), that not only cultivated the lands of the Spartans but provided them with personal service. In the more backward parts of Greece, such as Arcadia and the northwest, the majority of the population were peasants and neither needed nor could afford many slaves. Even in so rich and economically progressive a city as Athens the great majority of the free inhabitants were peasant proprietors or independent craftsmen and shopkeepers. The more prosperous of these might own one or two domestic servants, farm hands, or assistants in the workshop, but the majority probably had no slaves. Richer families had larger domestic staffs, and landowners normally cultivated their home farms and sometimes all their estates by slave labour. In industry there were factories of 20, 30, and even 120 slaves. The silver mines of Laurium were almost entirely worked by slave labour, and at times employed as many as 10,000–20,000 men. Some rich Athenians invested largely in slaves; Nicias, the 5th-century-B.C. general, is said to have owned 1,000, whom he hired out to mining contractors at a fixed daily charge. Many well-to-do Athenians owned a few slave craftsmen, whom they either leased to an entrepreneur or allowed to work independently, paying their masters a fixed revenue and keeping their other profits for themselves. Such slaves, who "lived separately," might *de facto* acquire property or incur debts, and could be sold as going concerns with their assets and liabilities. The city of Athens owned public slaves, including the famous corps of Scythian police, and others ranging from street cleaners to the public executioner and the chief accountant and keeper of the records.

Slavery reached its high-water mark in Rome in the 2nd and 1st centuries B.C. Not only were slaves very abundant and cheap, owing to the prevalence of wars and piracy, but the Roman upper classes were immensely wealthy and were building up huge estates (*latifundia*). The result was the introduction of agricultural slavery on a vast scale. The peasantry were progressively expropriated, and their lands either cultivated by slave gangs or converted into ranches manned by slave herdsmen. This is the only

period of antiquity in which large-scale slave revolts occurred, in Sicily in 135–132 B.C. and 104–100 B.C., and the rising of Spartacus (q.v.) in Italy in 73–71 B.C., and the difficulty with which these revolts were suppressed proves the large numbers involved. Slaves were also employed on a vast scale in the mines; there are said to have been as many as 40,000 in the silver mines of New Carthage in Spain.

Under the principate there was a recession as slaves grew scarcer and prices rose. On estates already manned with slaves the stock was often maintained by breeding, but landlords generally preferred to lease their lands in small holdings to free tenants, retaining slave labour only on a home farm; vineyards were commonly cultivated by slaves (the vintage being undertaken by casual hired labour or by a contractor who purchased the hanging crop), since tenants were not trusted to do the skilled work involved. Even in the mines slave labour declined. Some were worked by convicts; others were no longer leased *en bloc* to contractors but let by shafts to free working miners; in others, free indentured labour was employed. Slaves were now predominantly used in domestic service, in managerial and clerical posts, and in skilled crafts. The secretarial and financial offices of the imperial government were manned by slaves and freedmen of the emperor, who came to form a hereditary group. The *Digest* (Justinian's collection of jurists' opinions) reveals that landowners normally employed slave agents and bailiffs, and that wealthy men owned slaves engaged in commerce and industry. Rules of law were worked out to determine how far a master was responsible for the financial transactions of shippers, merchants, shopkeepers, and craftsmen whom he owned. He could, provided that he gave due notice of the fact, disclaim liability for the debts of his shopkeepers and craftsmen, and the creditors could recover only from the latters' *peculia*, the money or stock which their owner had allocated to them. Slaves were predominantly used in responsible and skilled positions, where their special usefulness or higher earnings would repay the cost of training. In the later Empire, despite the fall in the price of slaves, the position did not substantially alter. In some areas there were hereditary groups of agricultural slaves who could not legally be sold apart from the land; their status in law and practice more and more resembled that of free tenants, who on their side were being converted into serfs and were bound to a given piece of land. Convicts were used in certain state quarries, but miners were, in general, free men. The chief innovation of the age was the establishment by Diocletian, in the 3rd century, of state weaving and dyeing factories, manned by slaves; but by the 4th century these slaves had become hereditary groups whose status differed little from that of free men.

3. Social Conditions.—It is obviously impossible to generalize on the condition of slaves in classical antiquity. All slaves were liable to the arbitrary cruelty of their masters; their only legal remedy was to escape to a sanctuary and demand to be sold to another master. Antoninus Pius (A.D. 138–161) ruled that the killing of a slave by his master was homicide, but Constantine absolved the master if the slave died as a result of flogging. In practice there were great differences between various categories of slaves. In the mines slaves were often ruthlessly worked to death, and the chained gangs on the land were brutally treated. On the other hand a private secretary or a nurse might be treated as one of the family, and a craftsman in charge of a workshop or a sea captain in command of a ship enjoyed *de facto* most of the rights of a free man. Miners and agricultural slaves had little hope of freedom, while domestic servants could reasonably hope to be manumitted, and craftsmen and shopkeepers had a fair chance of buying their freedom.

The fact that slavery was universally accepted as a normal social institution made relations between slaves and masters easier. Slavery was regarded as a misfortune but not as a wrong, and slaves felt no resentment against their masters unless they were ill-treated. There was, moreover, no distinction of race or colour between slaves and freemen. In classical Greece most slaves were barbarians, but there were many Greek slaves (although no Greek was a slave in his own city) and free barbarians, and there was no marked racial or cultural cleavage between the Greeks and the neighbouring barbarian peoples. In the Roman Empire slaves were for the most part from the same stock as the provincials. Nor was there any marked distinction in the kind of work done by freemen and slaves. There were slaves in the professions and skilled trades, and free manual workers. Slaves were not distinguishable by dress or any outward mark from free persons in the same walk of life, and the intermarriage between slave and free, though legally null, was not uncommon in the lower ranks of society. A manumitted slave did not in Greek cities become a citizen, citizenship being a strictly hereditary privilege, but he ranked like any other resident alien. In Roman law he became a citizen if manumitted in proper legal form, though he did not enjoy full civic rights, being ineligible to serve in the army or to hold public office or become a city councilor. (A. H. M. J.)

III. SLAVERY IN THE OLD WORLD AFTER THE FALL OF THE ROMAN EMPIRE

1. Western and Central Europe.—Although slavery declined notably at the end of the Roman Empire, it did not disappear even in Western and Central Europe, and, although it never again reached the importance it had in Rome during the classical period, there were times when the institution was definitely on the upsurge. The first increase in the number of slaves occurred as a result of the barbaric invasions because it was customary for various Germanic tribes to reduce enemies captured in battle into slavery. The conversion of the various kings to Christianity made little difference in this respect since, as we have seen, the Church did not proscribe slavery. A second upsurge of slavery seems to have occurred between the 8th and 10th centuries when many Slavonic peoples were captured and taken as slaves to Germany. (The very origin of the word "slave" is explained by this circumstance.) At that time slaves were often given to churches and monasteries by wealthy and pious people. However, from then on slavery definitely declined in Western and Central Europe, and although no definite date can be indicated, it seems to have disappeared in these regions toward the end of the 13th century, when, under the feudal system, serfdom took its place.

2. Southern Europe and the Middle East.—But the situation was very different in the East and South. For one thing slavery had continued under the Byzantine Empire. Furthermore the appearance of Islam gave a new impetus to the institution. Mohammed found slavery well established in Arabia when he began to preach the new religion in the first years of the 7th century. His attitude toward it as revealed in the Koran was very similar to that of the Christian churches: without condemning slavery, he taught that slaves should be treated with humanity and that the liberation of a slave was a pious and meritorious act.

But the appearance of Islam was followed by wars of conquest that covered vast areas in Asia, Northern Africa, and Eastern and Southern Europe. After the Arabs came other Islam invaders, including the Ottoman Turks. All these wars led to the capture of numerous prisoners, often reduced to slavery, although usually the civilian population was left free on condition of paying taxes to the conquerors. During the Crusades, Christians also enslaved many of their Saracen and other Muslim prisoners. This brought about a considerable increase in the slave traffic, not only in Muslim lands, but in Christian Europe as well. Even the Church's strenuous efforts to prevent the sale of Christians by Christians were not always successful: Hanseatic, Venetian, and Genoese slave traders bought Syrians, Serbians, Bulgarians, Armenians, and others from the Turks and resold them elsewhere. The fall of Constantinople in 1453 was followed by a marked increase of sales. Another source of slavery in Muslim lands was the capture of seamen and passengers in the Mediterranean by buccaneers operating from North Africa. "Slave charities" were established in England and other European countries to redeem these Christian prisoners and diplomatic efforts were made to curb the Muslim slave trade. Muslim slavers also raided territories in various parts of Black Africa, or purchased slaves from African chieftains.

However, it must be noted that slavery in Muslim countries, which continued throughout centuries and survived legally, or at least in custom, in a few of them, has always been very different

from that which existed in Rome and in the Americas after their colonization by Europeans. Gang-slavery for work in the fields or in industry and mining was almost unknown in the Islamic world. Most of the slaves were employed in wealthy households for domestic service and were well treated, in accordance with the prescriptions of the Koran. The one really cruel Muslim institution was that of the eunuchs, which involved emasculation. Women slaves in harems became their masters' concubines, or even legitimate wives (*see* CONCUBINAGE). Furthermore, Muslim society was not usually in the least race- or colour-conscious. Liberated slaves of whatever origin were readily absorbed as equal members of the community and examples of slaves or former slaves reaching the highest positions were numerous. Perhaps the most famous example of all is the case of the Mameluke slave dynasty which ruled Egypt for more than two and a half centuries (1250–1517), during which period one slave ruler succeeded another in the same way as normally a son succeeds his father.

In countries of the Iberian peninsula, Spain and Portugal, slavery continued to exist, not only through the period of Arabic or Moorish dominance, but also after their reconquest by the Christians in the 15th century. Naturally, the defeated Muslims were the first to be reduced into slavery. But soon the Portuguese set out to import slaves from Africa as well. The first Negro slaves were imported in 1444 by Portuguese captains of Prince Henry the Navigator, who had obtained them as ransom in exchange for Moorish prisoners. Later the Portuguese engaged directly in the slave trade and established factories (trading posts) for that purpose on the coast of Guinea.

Negro slaves were imported in ever increasing numbers into southern Portugal and neighbouring regions of Spain, where Seville became an important slave market. These regions had greatly suffered from wars between Christians and Muslims and their populations had been largely depleted. Imported Africans were employed not only for service in wealthy households, but also for work in the fields, and for a variety of tasks in the cities, especially as stevedores in the harbours. Because the Portuguese, and to a lesser degree also Spaniards, as a result of the many conquests, had little race- or colour-consciousness, the various elements of the population mixed relatively freely and ultimately merged.

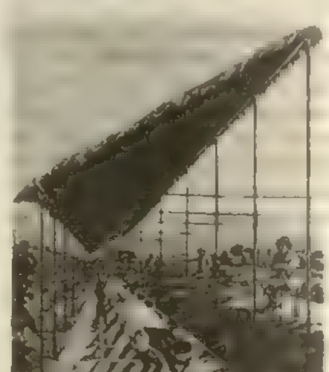
IV. MODERN (OR PLANTATION) SLAVERY

1. Early Developments in the New World.—Records show that Christopher Columbus and his companions in the course of their several voyages took a few Christianized Negroes born in Spain to Hispaniola, the first European colony in the New World. It would seem that at the time there was no conscious design to set up Negro slavery in the newly discovered lands. On the contrary, the European settlers first endeavoured to use the native Indian population as labour. In this connection it may be observed that if slavery was not a general institution in the Americas before the arrival of Europeans, it certainly existed in some regions, and particularly in parts of Central America. For example, Cortés has left a description of the great market place at Tlaltelolco (present Mexico City), and of the large number of male and female slaves brought there "for sale as the Portuguese bring Negroes from Guinea." The Spaniards themselves shipped a number of captured Indians to the home country, either as a present for the sovereigns, or for sale. In America, however, Indians were generally subjected by the Spaniards to institutions more reminiscent of serfdom, such as *repartimiento* or *encomienda*, considered more suitable for agricultural exploitation, than to chattel slavery. Only in the mines was the situation different and more akin to outright slavery.

Whatever the case, difficulties arose. Firstly, Indians proved to be less docile than had been thought and sometimes rebelled, or more often disappeared in the surrounding forests. Secondly, they often lacked physical vigour and succumbed in large numbers to diseases brought over from Europe. In addition, their employers found themselves at odds with the Roman Catholic missionaries. The missionaries' main preoccupation was the evangelization of the pagan population, and consequently they wanted it to be treated gently. The Spanish king, Ferdinand the Catholic, adopted a solution midway between the desires of the settlers for abundant and cheap labour and the much more lenient attitude of missionaries: Indians could be forced to work, but only under certain prescribed conditions, including humane treatment and the payment of a remuneration. But, of course, the last prescriptions of the edict were often disregarded. It is for this reason that Bartolomé de Las Casas, Roman Catholic bishop of Chiapas, approached the new king, Charles I (Emperor Charles V of Germany), with the proposal that each Spanish settler should be permitted to bring over a certain number of Negro slaves. In 1517 this idea was accepted, although in somewhat modified form, and a Spanish nobleman was granted, by letters patent, a licence to import each year a specified number of Africans into Hispaniola on the condition of paying duties to the royal treasury. Soon thousands of Negroes were brought, not only into Hispaniola but also into the other islands of the Caribbean and later to the mainland. Thus a measure of mercy conceived for the protection of Indian natives became the origin of one of the cruelest institutions of all time—Negro plantation slavery. Las Casas soon discovered his mistake and expressed his regrets concerning it in his *History of the Indies*.

The establishment of Negro slavery did not put an end to the exploitation of Indians and both forms of servitude existed side by side. In the Spanish possessions on the mainland, the servitude of Indians was prevalent; from it evolved peonage (*q.v.*). Certain remnants of this condition survived as late as the second half of the 20th century, thus outliving the abolition of chattel slavery. On the other hand, in the Caribbean island colonies, not only of the Spaniards, but also of the British, French, Dutch, and Danes, the aboriginal Indian population suffered total extinction in the course of time. There Negro slavery prevailed and, with the development of sugar plantations, assumed colossal proportions from the end of the 17th century.

Similar events occurred in Brazil, after the coast of that country fell under Portuguese domination. In Brazil, however, Indian aborigines were directly enslaved in large numbers with perhaps even more inhumanity than in Spanish possessions; in addition to the exploitation of the Indian population along the coast, the manhunt was pursued deep in the jungles. Jesuits and other Roman Catholic missionaries there, too, took the aborigines under their protection, an arrangement confirmed by the king of Portugal who set up a council of missions. But this did not prevent further enslavements, often under the very eyes of the missionaries. However, when sugar and later coffee plantations developed on the vast *latifundia*, the Portuguese adopted the same solution as the Spaniards and imported Negroes as slaves in ever increasing numbers.



RADIO TIMES HULTON PICTURE LIBRARY

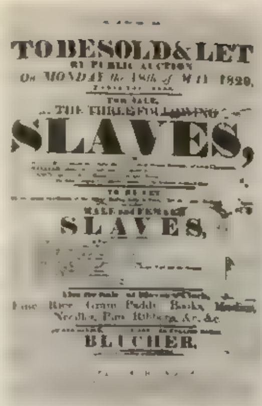
OVERSEERS SUPERVISING SLAVES WASHING FOR DIAMONDS IN BRAZIL. EARLY 19TH-CENTURY LITHOGRAPH

Thus, the development of the institution of slavery in Central and South America (and later also in parts of North America) was intimately connected with the growth of the plantation economy. Negro slaves also worked in the mines, as stevedores in the ports, and in other heavy work; served as domestic servants of all kinds; and were trained in various trades. But it was after the development of sugar plantations that the slave trade between the West Coast of Africa and the Americas reached enormous proportions, becoming the most lucrative trade of the time and constituting a considerable proportion of traffic as a whole.

2. Growth of the Triangular Slave Trade.—The English became the most important importers of slaves, although the French, the Dutch, and others also took part in the commerce, to supply their own colonies or the larger and richer Spanish posses-

sions. On the other hand the Portuguese reserved the monopoly of traffic to Brazil to their nationals.

This trade was usually triangular. Ships set out first from a home port such as Liverpool or Bristol in the case of the British for the Atlantic Coast of Africa. They carried liquor, firearms, cotton goods, and various trinkets that were exchanged for slaves brought from the African hinterland to one of the numerous factories established along what became known as the Slave Coast (Gulf of Guinea). Then came the so-called "middle voyage" from Africa toward the West Indies or one of the colonies or countries on the American continents: this leg of the trip was the slave trade proper. Although usually shortened by favourable winds, it represented a true torture for the involuntary passengers. They were closely packed in the hull to save space and often chained in order to avoid rebellion, or prevent them from jumping into the sea. Even by standards of the time, food was inadequate and water scarce. There was no proper ventilation, nor were sanitary installations provided. The only consideration which could lead to somewhat better conditions was the greed of slave traders and their desire to bring as many slaves as possible to destination in a condition which would permit their sale. But even this consideration had little effect if the ship, for one reason or another, was delayed. In such a case the mortality among the human cargo could reach appalling proportions. If a severe storm came and it was necessary to reduce the load, the sick were thrown overboard and at times also the healthy, especially, as was often the case, if the "cargo" was insured. It has been calculated that at least 20% of all the Negroes transported did not survive the voyage. Once they arrived at their destination, survivors were either directly delivered to their prospective employers or more often kept in stockades to await a purchaser. The ship then undertook the third and last leg of the voyage—homeward—loaded with the various staples produced on the American plantations. One of the chief items of the trade was molasses, from which rum was distilled and used for the purchase of more slaves on the next trip. If all went well, the profit was enormous. It was carried in bullion or in bills of exchange on banks in the home country. Later, when New England ports developed, their traders also engaged in the triangular trade, which proceeded on the lines described above, except that Boston or some other North American city was substituted as home port. The number of slaves transported under these conditions amounted to over 2,000,000 from 1680 to 1786 for the trade to British possessions and the West Indies alone. Another source indicates for the year 1790 the following figures: 38,000 in British ships; 20,000 in French; 4,000 in Dutch; 2,000 in Danish; and 10,000 in Portuguese, or a total of almost 75,000 for a single year.



(LEFT) RADIO TIMES HULTON PICTURE LIBRARY; (RIGHT, TOP AND BOTTOM) THE GRANGER COLLECTION

(LEFT) AN 1829 POSTER ANNOUNCING A SLAVE AUCTION IN THE WEST INDIES; (RIGHT) ELEVATION AND PLAN OF LOWER DECK OF AN EARLY 19TH-CENTURY SLAVE SHIP, DESIGNED TO CARRY THE MAXIMUM POSSIBLE NUMBER OF SLAVES AT MINIMAL COST TO THE SLAVE TRADER

3. Treatment of Slaves.—What was the treatment meted out to those of the slaves who reached their destination or to those who were born in captivity? The reply to this question is not simple because a considerable number of factors must be taken into consideration.

In the first place, each of the colonial powers had its own standards and policies, which varied widely. By way of generalization it may perhaps be said that slaves fared better in colonies of Catholic and Latin nations than in those of Protestant countries. One reason for this is that the link between the mother country and its dependencies was much closer for the first group than for the second. Consequently, the metropolitan authorities, through their agents, tried to exercise at least a minimum of control over planters. These countries, namely Spain, Portugal, and France, promulgated detailed laws concerning the treatment of slaves, of which the most famous is the French *Code Noir* (Black Code) put into force by King Louis XIV in 1685. This compilation contained a number of relatively mild prescriptions, however, often disregarded. Even more liberal was the Spanish Slave Code promulgated in 1789. In the British possessions, on the other hand, white settlers often were granted self-government. Consequently they were free to promulgate regulations as they saw fit. These were often extremely stringent, as in the case of the Barbados Code, imitated in Virginia and elsewhere. However, beginning with the last years of the 17th century, under the impact of the antislavery movement both at home and in certain American colonies, improvements were sometimes introduced in the British dependencies.

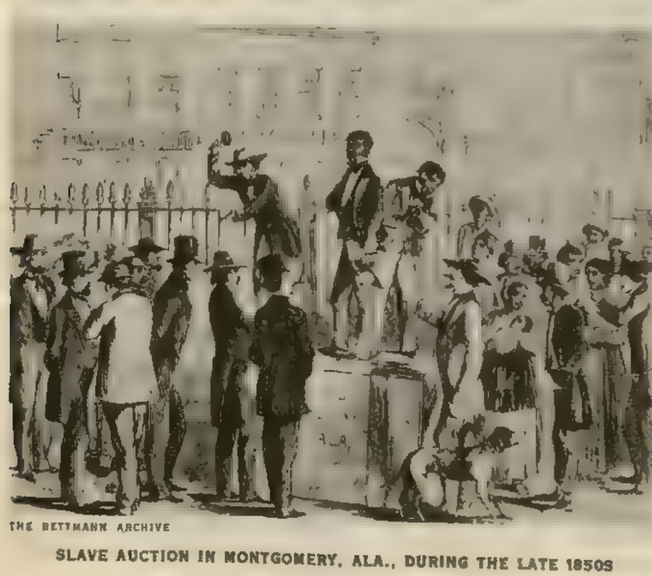
Another factor affecting the situation of Negro slaves in the colonies of Roman Catholic nations was the interest taken in their lot by the clergy. One partial expression of its influence was the enactment of already mentioned codes by metropolitan authorities. But the Roman Catholic clergy also tried to exert its authority on slaveowners on the spot. For example, it encouraged church marriages for slaves and opposed the separation of families. On the other hand, the Anglican clergy did not seem to show similar interest in the slaves, and in many British possessions and U.S. states the law expressly prohibited legally binding slave marriages. But here too the situation in some cases changed under the influence of the Quakers and of certain Protestant dissidents.

A third difference came as a result of the attitude of the peoples themselves. As already noted, the Portuguese and Spaniards had no race or colour prejudice. Miscegenation between white men and Negro women was, of course, a common phenomenon in all colonies. The rule everywhere was that the child followed its mother's condition, and consequently was born a slave. It was customary, however, in Roman Catholic lands for the master to grant manumission to his child, and sometimes also to the mother. Manumission was also frequently granted for long and devoted services of a domestic slave, in particular to nurses who had looked after the master's children. As a result, for example, in Haiti under French domination, a class of free and relatively well-to-do mulattoes gradually emerged. In Brazil, where toleration toward the half-castes and even full Negroes was much more in evidence, a number of free persons of African descent had reached positions of eminence even before the liberation which came in 1889; this, in turn, fostered still more the spirit of nondiscrimination and had its effect on the condition of slaves, particularly those in direct contact with their masters. It would be excessive to say that in Anglo-Saxon territories, including the Southern United States, similar situations never occurred. But there the colour line was drawn once and for all, and a person having even a small strain of African blood was a Negro, whether slave or free. In addition some of the colonial statutes prohibited manumission, or made it extremely difficult.

What is certain is that the condition of slaves in all countries depended greatly upon their position and the tasks they had to perform. Those in domestic service were usually a kind of privi-

leged elite, although here again there were many gradations depending upon the nature of the job, the character of the master, and personal qualities of the slave himself. The next category was that of slaves exercising various trades. The third—and most numerous—included the slaves working on plantations. Here is where one could find the worst excesses, particularly if the estate was large and the absentee landowner had to rely on an agent who hired supervisors. Nonetheless, in order to maintain a balanced picture, one must always remember that published accounts usually concerned either the best or the worst examples. Probably on the whole the situation of slaves was not much worse than that of indentured labourers brought over from Europe (*see* CONTRACT LABOUR). At times it might have been even better because the master had reason to take care of his slave, who represented capital, whereas he had no similar incentive in regard to the labourer. But, of course, the condition of the slave was permanent and even hereditary, whereas an indentured man retained important rights under colonial laws, and became free at the expiration of his contract. Furthermore, the condition of slavery itself was degrading. This was keenly felt by many Negroes, particularly those who had been brought over from Africa and sold into slavery; a person born into slavery was more likely to accept his condition with resignation. But rebellion was an ever-present threat, and uprisings occurred from time to time in various parts of the slave world—in Brazil and Haiti (which thus gained its independence from France), for example, and also in the British possessions. Nor was the danger absent in some of the Southern U.S. states, although the Civil War proved that it had been greatly exaggerated: during the entire period Southern slaves showed a surprising passivity and toward the end some of them were even called upon to serve in the forces of the Confederacy.

4. North America.—As concerns the introduction of slavery in North America, and in the United States in particular, it must be remembered that various parts of the country had originally been colonies of different European states: England, Holland, France, and Spain. Thus during the colonial period different slave codes prevailed. In the North American English possessions each colony decided for itself, but there was, of course, imitation of one by another. What is more striking is the fact that Negro slavery as a systematic institution appeared on the mainland relatively late: the first shipload of slaves arrived, almost by chance, in Virginia in 1619, and then in a Dutch and not a British ship. At first slavery developed extremely slowly and in Virginia in 1681 there were still only 2,000 Negro slaves as against 6,000 indentured labourers of European origin. It was only after the intensive cultivation of tobacco developed, and a little later also of rice, that slavery took root in the colony, so that the number of slaves reached 59,000 in 1714. From then on it increased steadily, mainly through new arrivals, reaching 263,000 in 1754. But a few years later the increase slowed down and at the time of the American



THE BETTMANN ARCHIVE

SLAVE AUCTION IN MONTGOMERY, ALA., DURING THE LATE 1850S

Revolution it even seemed that slavery might gradually disappear. The reason was that tobacco rapidly exhausted the land on which it was grown, so that it had to be used for other crops, by far less profitable. It was only with the introduction of cotton, and particularly after the invention of the cotton gin in 1793, that the trend was once again reversed and the number of slaves increased steadily till it reached the figure of almost 4,000,000 on the eve of the Civil War in 1860.

The history of the antislave movement in the United States and the sequence of events that finally brought about the emancipation are examined below. It is proper to note here that since the beginning of the 19th century the United States, from the point of view of slavery, was divided into three zones: in the North, where the economy had never been based on that institution, and which became more and more industrialized, slavery was legally abolished between 1777 and 1804 by action of the individual states; in the cotton-producing South slavery not only existed but was greatly expanded; in between was a belt of "border" states which also legally maintained the institution of slavery, but in which the number of slaves was continuously on the wane because the economy no longer had use for slave labour.

V. THE ANTISLAVERY MOVEMENT AND ABOLITION IN THE WESTERN HEMISPHERE

1. Great Britain.—The beginnings of the antislavery movement in Great Britain may be traced back to the last third of the 17th century. In 1673 Robert Baxter, a nonconformist, published his *Christian Directory* in which, without condemning slavery, he vigorously attacked the slave trade. In 1680, Morgan Goodwyn, an Anglican clergyman, described his sojourn in Barbados and the treatment inflicted upon the slaves in that island colony. Mention should also be made of the antislavery pronouncements of George Fox (1624–91), the founder of the Society of Friends, or Quakers.

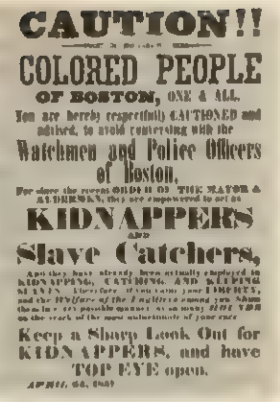
During the same period appeared Aphra Behn's novel *Oroonoko* (1688), later dramatized by the poet Thomas Southerne (1696), describing the miseries of the slaves brought over from Africa to the West Indies. The novel and the play had considerable impact on the public. Perhaps an even greater influence should be ascribed to John Locke, who attacked the institution of slavery in his *Treatise on Civil Government* published in 1690.

In the following years slavery in general and the African slave trade in particular were denounced by Sir Richard Steele, Alexander Pope, James Thomson, William Shenstone, John Dyer, Richard Savage, William Cowper, and many others. In connection with slavery in the British colonies on the American mainland, special mention should be made of the Quakers John Woolman (1720–72) and Anthony Benezet (1713–84), who was of French Huguenot origin. The latter in particular exercised his influence on both sides of the Atlantic and many of the English abolitionists were his disciples.

One of the practical problems which arose in England at the beginning of the 18th century concerned the status of slave servants brought into the mother country by planters. The controversy reached its peak in 1729 when in a legal opinion the Attorney General and the Solicitor General pronounced themselves in favour of the doctrine that a master could compel his slave to return to the colony. Chief Justice Sir John Holt (1642–1710) had held the opposite view. The matter was settled only in 1772 by the famous decision of Lord Mansfield in the case of the Negro Somersett. According to this decision a slave became a free man the moment he had set his foot on the soil of the British Islands. This decision was hailed as a victory by English abolitionists; but it had not the least effect upon either slavery practised in the overseas colonies, or the then flourishing slave trade under the British flag.

The prohibition of that trade was the first objective of British abolitionists (*see* also ABOLITION MOVEMENT). In this respect the year 1774 was marked by two important events, both brought about by the Quakers: the British Society of Friends decided in its annual convention to expel any Quaker engaged in the slave trade; the Quakers of Pennsylvania created, under the guidance of

James Pemberton and Benjamin Rush, the first antislave-trade society, and soon similar bodies sprang up in other North American colonies. Two years later, in 1776, the Quakers of Pennsylvania took the decision to set free all the slaves they held. In 1783 their English brethren created the first association in England "for the relief and liberation of the Negro slaves in the West Indies, and for the discouragement of the slave trade on the coast of Africa." Among its members were Tom Dillwyn, George Harrison, Samuel Hoare, and others. Thus both in the mother country and in the American colonies the Quakers were the first to give a systematic character to the antislavery movement.



THE BETTMANN ARCHIVE

(LEFT) "THE PECULIAR DOMESTIC INSTITUTIONS OF OUR SOUTHERN BRETHREN," ONE OF 13 WOODCUTS FROM THE "AMERICAN ANTI-SLAVERY ALMANAC" FOR 1840, A POPULAR ABOLITIONIST PUBLICATION; (RIGHT) 1851 POSTER WARNING BOSTON NEGROES TO AVOID LAW OFFICERS LEST THEY BE ARRESTED AND RETURNED TO SLAVERY UNDER THE FUGITIVE SLAVE LAW OF 1850

As concerns Great Britain, the next important event in the annals of abolition was the appearance, in 1786, of *Essay on the Slavery and Commerce of the Human Species*, by Thomas Clarkson (*q.v.*). In the process of its publication the author met a number of people sharing his views on the subject, such as Granville Sharp, the man who had brought the *Somersett* test case before the courts, James Ramsay, a clergyman who had himself published an antislavery book, and above all William Wilberforce. This group formed in 1787, together with Josiah Wedgwood, Bennet Langton, Zachary Macaulay, and others, the Society for the Abolition of the Slave Trade, which absorbed the association founded in 1783 by the English Quakers.

Under the influence of the Society and its supporters the Crown appointed in 1788 a committee of the Privy Council to inquire into the slave trade. The next year Wilberforce introduced in the House of Commons a series of resolutions intended to prohibit the slave trade under the British flag and the importation of new slaves into the British West Indies. But, despite his efforts, the first partial result was not achieved until 1806, when Parliament adopted a bill prohibiting British merchants from providing slaves to foreign colonies, and prohibiting also the importation of slaves into the newly acquired British possessions. The next year both houses finally passed a bill introduced by the government of Lord Grenville and Fox prohibiting slave trading by British vessels, as well as the importation of slaves into all British colonies. The traffic continued, however, until the Parliament adopted in 1811 another bill proclaiming slave trading to be a criminal offense. From then on the British navy pursued the now illicit traffic which was ultimately brought under control.

But this victory brought even greater suffering for those slaves who were already in the colonies. As there was no fresh supply, their masters tried to exact more work from them. In 1821, Wilberforce, assisted by Sir Thomas Buxton, undertook his second attack—this time against the institution of slavery itself. In 1823 they, with others, formed a new organization, the British and Foreign Anti-Slavery Society. Their aim was to obtain the immediate liberation of all slaves, in particular in the West Indies. However, the British government announced its plan to achieve the same objective in a more cautious manner: those of the colonies having their own legislative councils were invited to take the necessary gradual measures, although the government warned them that it would intervene directly in case of undue delay. Despite this warning there was practically no progress in the ten following years. Finally in 1833, under the pressure of public opinion, the government decided to take direct action. The Abolition Act of August 28, 1833, provided for a compensation to be paid out of the Treasury to the slaveowners. Furthermore, the slaves were not set free immediately (except when they were under 6 years of age), but had to undergo a period of apprenticeship for seven years. During this period they had to continue to work for their masters three-fourths of the day, in exchange for food and clothing. Later the transitional period was cut by two years and in 1838 all slaves in the British West Indies were set free. Later similar measures were taken in regard to other British possessions as well as in regard to India.

The example of Great Britain was gradually followed by the other European states, and some American ones had already taken action of the same kind. The immediate emancipation of the slaves in the French colonies was decreed by the Provisional Government of 1848. In 1858 it was enacted that every slave belonging to a Portuguese subject should be free in 20 years from that date, a system of tutelage being established in the meantime. This law came into operation on April 29, 1878, and the status of slavery was thenceforth illegal throughout the Portuguese possessions. The Dutch began the emancipation of their slaves in 1863. Several of the Spanish American states, on declaring their independence, had adopted measures for the discontinuance of slavery within their limits. It was abolished by a decree of the Mexican republic on Sept. 15, 1829. The government of Buenos Aires enacted that all children born to slaves after Jan. 31, 1813, should be free; and in Colombia it was provided that those born after July 16, 1821, should be liberated on attaining their eighteenth year.

Three of the most important slave systems still remained in which no steps toward emancipation had been taken—those of the Southern United States, of Cuba, and of Brazil.

2. United States.—The antislavery movement started in the North American colonies long before independence, although slavery was an accepted institution in all of them (for a discussion of slavery in the United States *see* NEGRO, AMERICAN).

In the Declaration of Independence of July 4, 1776, the 13 former colonies stated as a self-evident truth, "that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness." However, this was a purely doctrinal statement of principles having no immediate practical effect upon the institution of slavery. A denunciation of the slave trade, included in Jefferson's first draft, was deleted by the Continental Congress as being objectionable to some southern and New England delegates.

The question of slavery came up later in connection with the government of territories that did not form part of any one of the states, and the Northwest Ordinance of 1787 expressly prohibited slavery in the territory north of the Ohio River. This measure was in conformity with the then prevailing public opinion: at the Constitutional Convention that same year, Washington, Jefferson, Madison, and Hamilton argued in favour of the abolition of slavery by the Federal Constitution then under discussion. If their views and those of the majority of the delegates did not prevail, this was due to the stubborn opposition of South Carolina and Georgia, which threatened not to join the Union if slavery was to be expressly prohibited. Finally, it was decided to leave the decision whether or not slavery should be abolished to the judg-

ment of each State of the Union. The Constitution provided, however, that Congress might by federal legislation prohibit the slave trade after the expiration of 20 years (which it actually did, as from Jan. 1, 1808).

There is no doubt that in 1787 it was expected that the days of slavery were numbered and that it would come to an end within a few years even in the southern states. But an unforeseeable event—the expansion of the culture of cotton following the introduction of the cotton gin in 1793—brought a complete reversal of the trend in the South. While all northern states enacted legislation abolishing slavery between 1787 and 1804 (although sometimes only progressively), in the South public opinion opposed abolition with ever increasing determination. The main focus of the controversy, however, was not so much the states themselves as the federal territories subject to congressional legislation. The Louisiana Purchase in 1803, the annexation of Texas in 1845, and the treaty of Guadalupe Hidalgo ending the Mexican War in 1848 added vast unorganized territories for the possible expansion of slavery. The Missouri Compromise (*q.v.*) of 1820 acknowledged the southern contention that the Congress could not establish restrictions (*i.e.*, prohibit slavery) for new states that did not apply to all states; the Fugitive Slave Act of 1850 increased the severity of provisions for the return of escaped slaves; and the Kansas-Nebraska Act (*q.v.*) of 1854, repealing the Missouri Compromise, established Kansas and Nebraska as new territories with popular sovereignty, which left the decision on freedom or slavery to the settlers in the territories. An even more far-reaching effect was achieved by the famous 1857 decision of the Supreme Court of the United States in the Dred Scott case, which pronounced the Missouri Compromise unconstitutional and prohibited Congress from legislating on slavery in federal territories. By the same decision the court also rejected the plea of Scott, a slave who had been brought by his master to a "free" state, that he had acquired freedom by this fact.

Despite these controversies it would be a mistake to believe that the antislavery movement had large popular backing even in the northern states. It is true that men like Benjamin Lundy, Elijah P. Lovejoy, Wendell Phillips, Charles Sumner, John Brown, and William Lloyd Garrison (the editor of the *Liberator* and the founder of the New England Antislavery Society; 1832) were denouncing slavery with unrelenting vigour. But they often met not only apathy but outright opposition. Thus Garrison was attacked by a mob in Boston in 1835 and two years later Lovejoy was killed under similar circumstances in Illinois. However, as years passed, more and more northern intellectuals joined the ranks of those who were asking for emancipation. Among them were Emerson, Bryant, Longfellow, Whittier, and Whitman. But probably the greatest contribution in arousing antislavery public opinion was made by Harriet Beecher Stowe's *Uncle Tom's Cabin*, published in 1852 to denounce, in the form of a popular novel, not only slavery itself but also the evils of the Fugitive Slave Act of 1850.

On the other hand, Southern writers defended what they called their "peculiar institution" by all kinds of arguments, drawn not only from economic and social considerations but also from the Scripture. At least in one case arguments in favour of the plantation system and slavery were based on the writings of Karl Marx himself. Most surprising in the attitude of the South is the fact that there were relatively few slaveowners because a slave economy was not possible at all on small holdings. Thus the enormous majority of the white people of the South defended an institution in which they had no direct personal stake.

It has often been said that this defense of slavery would have gradually faded away if sufficient time had been given to southern public opinion to realize that the system had outlived its economic usefulness. Thus it has been indicated that fewer and fewer plantations were making profits, and that most of them were operating on so close a margin that they could barely show a slight profit even in the better years. However, this reasoning did not take into account a number of essential factors. From the economic point of view, whatever the profits, the 4,000,000 slaves who existed in the South in 1860 represented an enormous investment. Since the Revolution the price of slaves had gone constantly up and a slave who would have been worth $300 before the invention of the cotton gin in 1793 brought $1,400 to $2,000 in 1860. Furthermore, southerners believed that the poor state of their economy was not to be blamed on the low return of slave labour and not on the decline of the fertility of the soil, but solely on the selfish policy of the North which had imposed a "tariff of abominations" to protect its industries. But perhaps the most important aspect was of a psychological nature: slavery was for the South an essential element of its way of life, and also an element of "states' rights"—and it was therefore not prepared to give up the institution.

Whatever the case, when the election of Abraham Lincoln to the presidency in 1860 made it clear that northern policies would prevail, the South concluded that it had to leave the Union. It should be remembered that the Civil War was started not for the purpose of abolishing slavery, but with a view to bringing back the rebel states. Nonetheless, once the hostilities began, antislavery became a battle cry for the northern states, as maintaining slavery became a symbol for the Confederacy.

But even under these conditions Lincoln did not at once take measures for the emancipation of slaves. The preliminary emancipation proclamation was issued on Sept. 22, 1862, and the one that followed it on Jan. 1, 1863, applied only to those of the states that were in rebellion against the Union: Arkansas, Texas, Louisiana, Mississippi, Alabama, Florida, Georgia, South Carolina, North Carolina, and Virginia. The state of Tennessee and certain parts of Louisiana and Virginia, as well as the border states still permitting slavery, were exempted from this measure since they were not in rebellion. The proclamation was immediately applied in the regions already occupied by the Northern armies, but these did not include more than 75,000 slaves. Of course, when the Northern armies advanced, its application was extended to the newly occupied territories. But the final liberation came only after the surrender at Appomattox on April 9, 1865. Even then it was questionable whether a presidential proclamation which in its own words was a war measure justified by exceptional circumstances could have affected the legal status of slavery in states whose laws still admitted it. On Jan. 31, 1865, the House of Representatives concurred by a considerable majority in the vote of the Senate favouring the 13th amendment to the Constitution abolishing slavery throughout the Union By Dec. 18, 1865 the amendment had been ratified by three-fourths of the states and was proclaimed in force. Under this amendment "Neither slavery nor involuntary servitude, except as a punishment for crime . . . shall exist within the United States, or any place subject to their jurisdiction" (Sect. 1). Furthermore, the Congress was granted power to enforce this provision by appropriate legislation (Sect. 2). (*See* also ABOLITION MOVEMENT.)

3. France.—In France toward the middle of the 18th century philosophers and writers like Voltaire, Rousseau, Filangieri, and Raynal denounced the evils of slavery and of the slave trade either generally or with special reference to the French colonies in America. Montesquieu, who sojourned in England and was undoubtedly in contact with British abolitionists, wrote: "Slavery is just as much contrary to civil law as it is in opposition to natural law; what kind of civil statute could prevent a slave from escaping?"

The Société des Amis des Noirs (Society of Friends of the Blacks) was established in Paris in 1788. Contrary to the aims of its British counterpart, the French society demanded not only the immediate prohibition of the slave trade but also of slavery itself. The president was Condorcet and among the members were the Duc de la Rochefoucauld, Lafayette, Brissot, Clavière, and the famous Abbé Grégoire. The adoption by the Constituent Assembly, in August 1789, of the Declaration of the Rights of Man seemed to put an end to slavery in the French overseas possessions. However, it was explained, as in the case of the American Declaration of Independence, that the French Declaration was a statement of principles, and that its provisions, to become binding had to be implemented by the adoption of appropriate legislation. Furthermore, the Assembly passed a resolution in March 1790 by which it specifically excluded the colonies from the operation of

the Constitution then in preparation for metropolitan France. This was a bitter disappointment for the coloured population (freedmen of mixed ancestry) of St. Domingue (Haiti), which, although free, did not possess equality of rights with the French planters. Vincent Ogé, a mulatto, in France as head of a delegation representing his people, returned to the colony and organized the first uprising against the French. The revolt was short-lived and Ogé was sentenced to be broken on the wheel. Under the impact of this news, the Assembly adopted a resolution proposed by Abbé Grégoire recognizing the equality of "the people of colour ... born of free parents" and granting them representation in local assemblies on equal terms with the white settlers. But this was not implemented and soon a new uprising shook the French part of St. Domingue. This time many Negroes joined the mulattoes in a common cause. (*See* HAITI.)

In the meantime the Revolution deepened in France and the Convention, which had replaced the Legislative Assembly, adopted a decree on 16 Pluviôse An II (February 4, 1794) abolishing Negro slavery in all French colonies and establishing the equality of all men, without distinction as to colour. This was the first enactment in history prohibiting slavery. The decree, however, could be put into effect only in Guadeloupe and in French Guiana: the island of Martinique had fallen, in the meantime, into the hands of the British, and the colonial assembly of the island of Bourbon (Réunion) refused to apply it. St. Domingue was in a continuous state of turmoil, which finally led to the establishment of Haiti.

After the restoration of the Bourbons in 1814 measures were taken to put an end to the slave trade. But the status of existing slaves remained unaffected until the Revolution of July 1830 and the replacement of Charles X by Louis Philippe. At that time Martinique was shaken by a new Negro revolt, followed by extremely harsh repressive measures by the settlers. Liberal opinion in France was appalled and demanded reforms. Among the most vocal were the famous poet Lamartine, the Duc de Broglie, Alexis de Tocqueville, Isambert, Montalembert, and others. Under their influence a series of measures were taken by the government in 1833, 1836, 1839, 1840, and 1845: manumission was made easier, slaves were placed under the protection of the judiciary, and the separation of slave families was prohibited. But all these measures were only partly applied.

The final action came after the Revolution of February 1848 and the establishment of the Second French Republic, when a commission presided over by the under-secretary for the colonies, Victor Schoelcher, an abolitionist of old standing, prepared a text for the abolition of slavery. This text was implemented between May and August 1848 in all French colonial possessions, as well as in Algeria, where domestic slavery existed.

4. Other European Colonies and Former Colonies.—Most of the Spanish possessions in America abolished slavery when they acquired independence in the first quarter of the 19th century. Thus the government of Buenos Aires enacted a law under which all children born to slaves after January 31, 1813, were declared free. In Colombia children born after July 16, 1821, were to be liberated on attaining their eighteenth year. The Mexican Republic abolished slavery by a decree of September 15, 1829. However, slavery continued to exist in the colonies which remained under Spanish domination—Cuba and Puerto Rico in the Caribbean, and the Philippines in the Pacific. Although certain measures were taken in the course of the 19th century for the gradual abolition of slavery, the most important of which was the so-called Moret Law of 1870, the final emancipation came only after the Spanish-American War of 1898 and the loss of these colonies by Spain.

Portugal had lost its main colony, Brazil, in 1822. However, the Portuguese dynasty continued to rule until 1889 when the last emperor was overthrown and went into exile. Although measures were taken to prevent the slave trade, and as early as 1826 Brazil signed a convention with Britain to that effect, the traffic continued illegally at least until the middle of the 19th century. There, as in the British colonies, the end of the trade brought greater hardship for the existing slaves who were compelled to work longer hours in the mines and on the plantations. However, as already noted, slavery was less cruel in Brazil than in most other countries of the Americas. In 1871 a partial abolition act was passed. Under it the state-owned slaves were liberated at once. The others remained in their servile condition, but the children born to slave mothers after the law was passed were free, except that they had to work for the masters of their mothers for a term of 21 years. A clause provided that a certain sum should be annually set aside from fines to aid each province in emancipating slaves by purchase. It should also be noted that a great number of private slaveowners granted manumission to their slaves and that when the end of slavery was proclaimed by law in 1888, there were not more than 700,000 slaves in the country.

In other Portuguese possessions slavery was abolished in 1858, but the existing slaves remained with their masters for 20 years, a system of supervision and tutelage being instituted for the interim period.

As to the Dutch possessions, the slaves were also liberated, beginning in 1863.

VI. INTERNATIONAL EFFORTS TO SUPPRESS SLAVERY THROUGHOUT THE WORLD

The fight against slavery on an international scale started immediately after the end of the Napoleonic Wars, in 1814, and was still continuing in the second half of the 20th century. It passed through three successive stages: during the first period, from 1814 to 1885, the action was mainly directed against the slave trade; during the second, from 1885 to the end of World War I in 1919, it was, in addition, also directed against slave raiding in Africa; during the third, which covers the periods of the League of Nations and of the United Nations, it was further expanded to cover the Middle East and the Far East, including not only chattel slavery, but also other similar forms of servitude throughout the world.

A. Action Against the Slave Trade

As has been noted, legislative action against slavery in Great Britain started with an attack not against the institution itself, but against the main source feeding it—the slave trade.

Soon after the adoption of the British antislave-trade laws of 1807 and 1811, it became clear that they could be fully effective only if other maritime and colonial countries followed the British example, and international cooperation was established in this field.

1. Search and Seizure on the High Seas.—As the slave trade toward the Western Hemisphere had to use the high seas, the British thought of using, for the purpose of combatting it, two accepted doctrines of international law: the one according to which a pirate vessel is not entitled to the protection of any state; the other according to which warships can arrest, search, and seize any merchantman under any flag guilty of running a blockade or of carrying contraband of war. Clearly if slave trading could be assimilated either to piracy or to blockade running, the British Navy would be entitled to arrest and seize vessels transporting slaves. Such an extension of either of the two doctrines, however, could be achieved only by way of international agreement.

The first treaty of peace signed at Paris on May 30, 1814, included a clause according to which the defeated French undertook to bring slave trading under the French flag to an end. In addition, the French promised to assist the British in their endeavours at the forthcoming Congress of Vienna "to induce all Powers of Christendom to decree the abolition of the slave trade." Talleyrand, however, refused to support Wellington's proposal that international action should take the form of reciprocal recognition of the right to arrest and search vessels suspected of being engaged in the slave trade in the Atlantic north of the equator (the usual route of slave traders).

When the Congress of Vienna convened in November 1814, for a time it still seemed possible to come to an agreement under which the slave trade would have been assimilated to piracy. But here again opposition developed, and finally the Congress adopted a declaration merely enunciating a number of principles to be followed by each of the parties (Austria, France, Great Britain, Prussia, and Russia) for the abolition of the slave trade. The

declaration recognized, however, the need for regard "to the interests, the habits, and even the prejudices" of the subjects of the signatory states, and that "therefore, it was impossible to prejudge the period when this end could be achieved." The declaration envisaged only individual action by the signatories and did not establish a reciprocal right of arrest and search for their warships. The essence of the Declaration of Vienna was again repeated, in somewhat stronger terms, in the declaration of the Congress held at Verona in October 1822.

The British did not give up their efforts. They made first an attempt to claim the right to search foreign slave-trading vessels on the basis of general provisions prohibiting this trade. But the High Court of Admiralty declared that such a right did not exist. Beginning with 1817 they therefore negotiated and signed bilateral treaties with various colonial and maritime powers conceding to the navies of the parties a mutual and reciprocal right of search and seizure over merchantmen engaged in the slave trade in specified waters. Such treaties were concluded with Portugal, Spain, Brazil, and many other countries. In 1841 a treaty of the same type was signed at London between Great Britain, Austria, Prussia, Russia, and France. But France refused to ratify it and another treaty was signed with the French in 1845, establishing only a measure of cooperation between the British and the French navies. Similarly, the United States refused to concede a right of search to the British until the outbreak of the Civil War; a treaty was signed at Washington in 1862.

On the strength of the treaties recognizing a reciprocal right of search and seizure, the Royal Navy undertook energetic action to put an end to the slave trade. British naval squadrons were dispatched for this purpose to the West and East coasts of Africa, where the trade originated, and to waters adjacent to Cuba and Brazil, the main countries of destination of the traffic. The suppression of the trade was not easy because it brought enormous profits to the slavers even if only one voyage out of several was successful. Furthermore, the slave traders often benefited by the complicity of men in power in their respective countries, even when they had declared the traffic illegal and signed treaties with the British for its suppression. Thus the traffic to Brazil was not brought to an end before 1850, and the traffic to Cuba continued until the Treaty of Washington made it possible to search ships under the U.S. flag. The treaty also brought to an end the smuggling of Negro slaves into the southern part of the United States.

There were accusations at times that British naval officers themselves favoured the traffic in order to increase the bounty money they were paid by the Crown for captured and liberated slaves. But even if such abuses occurred, they were undoubtedly contrary not only to the proclaimed but also to the real policy of the British government. There is no better proof for this than its constant insistence to search and seize ships equipped for the transportation of slaves, even when they did not actually have slaves on board. Another accusation leveled against the British in the second half of the 19th century was that, while pursuing the suppression of the African slave trade, they were encouraging a similar traffic in so-called "indentured" labourers. Such labourers were recruited under contract in India, China, certain parts of Australasia and of South America, and transported to British and other European colonial possessions in other parts of the world. It was pointed out that indentured labour was often recruited by unscrupulous means and even outright deceit. Furthermore the employers used all kinds of devices to retain the workers beyond terms fixed in the original contracts, and consequently they either died in the country of employment, or were returned to their native land at an age or in a state of health which could make them only a burden for their communities. (*See* also CONTRACT LABOUR.)

2. Overland Trade in Africa.—Parallel to the action against the maritime slave trade, the British also undertook efforts against the overland traffic in slaves in Africa. Since the exploration of the heart of the Dark Continent started only in the second half of the 19th century, their first efforts were concentrated on the terminal points at which slave caravans reached the shores of the Mediterranean, the Red Sea, the Persian Gulf, and the Indian Ocean. Some of the slaves brought there were sold for local employment, but most were sent farther, usually in small native craft to Turkey, Arabia, Iran, and other eastern countries. Whereas the sea traffic from the West Coast and from Mozambique was mainly confined to adult male slaves to be employed on American plantations, land traffic affected women and young girls as well, whose destinations were the Oriental harems; young boys were also often sold into slavery as eunuchs.

Since the most important slave market was then located in Zanzibar, the ruler of this island was singled out by the British for special attention. The first treaty with him was signed in 1822 by Captain Moresby with a view to limiting the reexportation of slaves. A series of further treaties followed between 1845 and 1876, mainly negotiated by Sir John Kirk, British Consul at Zanzibar. But these agreements were only partially effective and the clandestine traffic in slaves toward exterior markets was brought under control only when the British took the island under their protectorate. More or less similar agreements were also concluded by the British with sheikhdoms on the Persian Gulf, as well as with a number of Somali chieftains. But here too the clandestine traffic continued despite the efforts of the British naval patrols.

B. Action Against Slave Raiding in the Heart of Africa

1. Slave Raiding in Africa.—The exploration of Africa revealed all the horrors of slave raiding and of the traffic in human beings which was going on completely unchecked in a vast territory covering the basins of the Nile and the Congo, and the region of the great African lakes. Dr. Livingstone, Heinrich Barth, Sir Samuel Baker, and many other explorers and missionaries who followed them, described the scenes they had witnessed. This literature had an enormous impact not only on the reading public of England and of the other English-speaking countries, but also in continental Europe. The world learned that hundreds of thousands of men, women, and children were submitted to a systematic manhunt. Only some of them were captured; others were slain or left behind maimed, sick, and deprived of all means of livelihood. Further thousands of the captives perished during the long trek across the African wilderness, or during the sea voyage. In this respect, the surveillance exercised by the British and other navies made matters worse rather than better: because of the risk of capture, the traders packed their crafts to capacity, disregarding not only the comfort of the involuntary passengers but also any security precautions. A further toll was taken by the castration to which boys and young males were subjected and which was performed under unsanitary conditions.

While revealing all these facts, the exploration of Africa, and its aftermath—the occupation, at least nominally, of the newly explored lands by European powers—offered an opportunity to strike at the root of the evil. The new authorities, particularly the British, worked to halt the overland traffic and slave raiding.

An international diplomatic conference at Berlin in 1885, called for the purpose of delimiting the spheres of influence in the Congo basin, presented a first opportunity for the coordination of the action already undertaken by individual European administrations. The General Act of Berlin, which applied only to the so-called "Conventional Basin of the Congo," required all of the powers concerned "to watch over the conservation of the indigenous populations and the amelioration of their moral and material conditions of existence" and "to strive for the suppression of slavery and especially of the Negro slave trade" (art. 6). Under another provision, the contracting parties agreed that the territories "may not serve as a market or way of transit for the trade in slaves of any race whatever" (art. 9).

The next step came in 1889–90 when two further conferences met at Brussels. The Conference of 1889 had only a limited purpose: to recognize Leopold II of Belgium as sovereign of the newly created so-called Congo Free State. On this occasion, Leopold reaffirmed the obligations that he had already assumed as president of the International Association of the Congo, created under his auspices by Stanley, who had signed treaties and agreements with various African chiefs.

2. The General Act of Brussels.—The Second Brussels Conference of 1890 was a far more important event. Its outcome was the General Act of Brussels, which for a quarter of a century constituted the fundamental charter of international efforts to suppress slavery. One of the important facts about the General Act of Brussels was the number of its signatories; in addition to all the major European powers, the United States of America and Turkey, Persia (Iran), and Zanzibar were parties thereto. The participation of the last three countries was particularly significant because at the time they still recognized the institution of slavery. Another characteristic of the General Act was the extent and importance of the territories covered by it: it applied to all possessions of the signatories, although special attention was paid to African regions where slave raiding and slave trading were still flourishing; to the countries of transit; and to those that constituted potential markets for the trade. But perhaps the major feature of the General Act was the broadness of the subjects covered by its provisions. These included recommendations for taking various measures of vigilance in countries of origin and transit, such as the establishment of permanent military posts and the use of patrols for stopping slave raiding and the conveyance of slaves by land. These acts, and some others connected with them, were to be treated as criminal offenses. Other provisions dealt with measures to be taken for the welfare of liberated or fugitive slaves, and their repatriation. The usual provisions concerning the repression of maritime slave trade were inserted; but the mutual right of arrest and search was limited to small native craft (under 500 tons) and in a specified zone along the coast of the Indian Ocean, in the Persian Gulf, and in the Red Sea (by 1890 the traffic toward America had been practically stopped). Under a special provision any slave taking refuge on board a warship of a contracting state was to be set free. The most novel part of the General Act concerned the creation of permanent international institutions for coordinating the fight against the traffic in slaves—the International Maritime Office set up at Zanzibar, and the International Bureau located at Brussels. Laws and regulations enacted for combatting slavery and the slave trade, as well as statistical data concerning this trade, the number of slaves detained and liberated, and similar facts, were to be communicated by the contracting parties, or otherwise collected. The General Act recommended also the creation by the states of various organs and missions for the purpose of protecting liberated or runaway slaves and rehabilitating them through educational and other measures.

Although the Conference of 1890 had been formally convened by the king of the Belgians, it, as well as the General Act it produced, came about largely as a result of the relentless efforts of the international antislavery movement. In particular, one of the members of the Committee of the Anti-Slavery Society of London introduced a motion before the British House of Commons in March 1889, suggesting the convening of such a conference. There is no doubt that the General Act of Brussels played a major role in the years 1890–1914 in the fight against the worst aspects of slavery: the capture of Africans, and their conveyance through the Dark Continent to ports of embarkation and hence by sea to countries where slavery was still legally practised.

3. Extension of Abolition.—Another factor contributing to the reduction of the slave trade during this period was the abolition of the institution of slavery itself in a number of Islamic and Oriental countries, such as Egypt, the Ottoman Empire, Siam (Thailand), and China.

4. New Forms of Forced Labour.—But as against these gains, mention should also be made of factors which led to an increase, if not of slavery proper, of certain other varieties of involuntary servitude. As concerns Africa in particular, in a number of cases European occupation led to the establishment of new forms of forced or compulsory labour, at times even more cruel than those existing under local custom. To quote only one example, in 1897, a missionary leveled grievous accusations against the state of affairs on the rubber plantations in the Congo Free State. In 1905, King Leopold II appointed an international commission which confirmed, at least in part, some of the allegations (*see* BELGIAN CONGO: *The Congo Free State*). As a result, under various pressures, the king transferred the territory to Belgium as a colony. This made possible the control of the colonial administration by the Brussels Parliament, and conditions improved notably in the ensuing years. The latter arrangement remained in force until June 30, 1960, when the (Democratic) Republic of the Congo came into being as an independent state.

C. Direct Attack Upon Slavery and Other Forms of Servitude

The outbreak of World War I in 1914 brought to a standstill the work of the international organs set up under the General Act of Brussels, the International Maritime Office at Zanzibar and the International Bureau at Brussels. At the close of the war the principal Allied powers signed a new convention at Saint-Germain-en-Laye on September 10, 1919, intended to release the signatories (Belgium, the British Empire, France, Italy, Japan, Portugal, and the United States) from their obligations under the Brussels Act, and substituting a declaration that they would "endeavour to secure the complete suppression of slavery in all its forms and of the slave-trade by land and sea." It remained at least doubtful whether or not these or the other signatories of the Brussels Act of 1890 were still bound by its provisions.

1. The League of Nations.—The Saint-Germain Convention coincided with the creation of the League of Nations. The international work carried on by the League for the suppression of slavery may be considered under the following headings: (a) action concerning mandated territories; (b) action in respect to certain other countries; and (c) general activities aimed at combatting the slave trade, and slavery and similar institutions and practices, including enforced prostitution.

Former Ottoman territories and German colonies were placed under the mandate (*q.v.*) system. When the instruments for their administration were drawn up, provisions were inserted for the suppression of the slave trade and for the prohibition of forced labour, except for essential public works and in return for adequate remuneration. Class B mandates, which applied to most of the former German territories in Africa, also included undertakings on the part of the mandatory powers to provide for the eventual emancipation of all slaves and for as speedy an elimination of domestic and other slavery as social conditions would allow. The mandatory powers had to present annual reports on their administration, including information on such problems as slavery, the slave trade, forced labour and other forms of servitude, and on measures taken for their suppression. These reports were examined first by the Permanent Mandates Commission, composed of experts, and thereafter by the Council of the League of Nations, which made various recommendations from time to time.

The work of the League in regard to other countries concerned, in the most notorious instances, Ethiopia and Liberia. When Ethiopia applied for admission to the League, some delegations opposed this application on the ground that slavery existed in that country. The Ethiopian government gave a pledge, accepted by the League's Council, to abolish slavery within a definite time and to at once take preliminary steps to that end. Despite the Italian-Ethiopian War of 1934–36, and the temporary loss of independence, Ethiopia lived up to this undertaking and slavery was abolished in 1942, once the country recovered its independence. In the case of Liberia, following certain allegations, an international commission was sent to that country in 1930. The commission stated in its report that classic slavery no longer existed in Liberia, although intertribal slavery and some other forms of servitude, including forced labour, were found to be serious problems. The commission made recommendations including action by the government of Liberia and financial and technical assistance by the League. The Liberian government accepted these recommendations in principle, and availed itself of the help offered. The League also made recommendations to countries having large Muslim populations in order to check more carefully the movement of pilgrims to the Hejaz. It had received information that pilgrims sometimes sold servants and even their own children into slavery in the country of their destination. Incidents of this kind have been reported as late as in the 1950s, in particular in a report

presented in 1955 by M Emmanuel La Gravière to the Assembly of the French Union, recorded in Document No. 75 of the Assembly of the French Union, Session of 1955–56.

The League of Nations took up the general problem of slavery in 1920. In 1924 the League set up a Temporary Slavery Commission of experts to collect information on slavery proper (including so-called domestic slavery), slave raiding and the slave trade, and also on such other institutions and practices as serfdom, purchase of girls as brides, sham adoption of children to exploit their work, various forms of indenture, or reducing to servitude, of persons for debt or for other reasons, and systems of compulsory labour, both public and private. The commission recommended the conclusion of an international convention covering all these matters.

However, when the League of Nations Assembly adopted the International Slavery Convention in 1926, it created ambiguity as to its scope. It decided not to include the subject of forced or compulsory labour, which it referred for further study and action to the International Labour Organization (*q.v.*). The other institutions and practices described by the Temporary Slavery Commission were all, with the exception of serfdom, to be brought under the operation of the Convention. At least this was stated in the report of the Assembly's committee which prepared the final text. But the Convention itself included a provision defining slavery as "the status or condition of a person over whom any or all the powers attaching to the right of ownership are exercised." When the Convention came into force, several of the contracting states considered, on the basis of this text, that it concerned only "chattel" slavery, and that they were not bound by it in regard to other institutions which did not imply the right of ownership over a person.

The substantive provisions of the Slavery Convention of 1926 did constitute real progress as compared with those of the General Act of Brussels—and fell short of them in many cases. For example, the parties undertook to suppress the slave trade in all territories under their control, but promised only "to bring about progressively and as soon as possible the abolition of slavery in all its forms." Even the traditional provisions of former treaties concerning a mutual and reciprocal right of arrest and search of vessels suspected of slave trading were not inserted in the new convention, although it was envisaged that such provisions could find their place in another international instrument. Nor did the Convention establish permanent international organs on the lines of those that had been set up under the Brussels Act for the implementation of its provisions by concerted action; it provided only for exchange of information, mainly of a legal character.

The absence of a permanent international organ dealing with slavery was felt as the major shortcoming of the 1926 Convention and the British government took the initiative in filling this gap. On its proposal, the Assembly in 1932 decided on the establishment of the Advisory Committee of Experts on Slavery as a permanent organ of the League. Experience had shown that states were often reluctant to supply information on objectionable practices within their own borders; the Advisory Committee therefore was authorized to rely also on the special knowledge of the various aspects of slavery and of the slave trade of its own members, who were independent from governments. Furthermore, the Committee was to proceed in a confidential manner, and even its reports could be published only to the extent decided by the Council of the League. This assured considerable freedom in its debates, as well as the possibility of taking discreet diplomatic action, often more conducive to results than direct and open pressures on governments. The Advisory Committee worked from 1934 until 1938; only the events that led ultimately to World War II prevented it from holding further sessions. In its last report (April 1938) the Advisory Committee pointed out that there had been considerable improvement in the main problems related to slavery. In this connection, it should be indicated that, in addition to what has already been said about the League's action in Ethiopia and Liberia, during the period between the two World Wars the legal status of slavery was abolished in Afghanistan (1923), Iraq (1924), Nepal (1926), Transjordan, under British Mandate, and Iran (1929), and Bahrain (1937).

2. The United Nations and Human Rights.—When after the conclusion of the Second World War the United Nations organization replaced the League the question of combatting slavery in its different forms came once again to the fore. The General Assembly, at its third session (1948), adopted the Universal Declaration of Human Rights which in article 4 states that "1. No one shall be held in slavery; slavery and the slave trade shall be prohibited in all its forms; 2. No one shall be held in servitude."

Following the fourth session of the General Assembly, in 1949, an Ad Hoc Committee of four members was set up "to survey the field of slavery and other institutions or customs resembling slavery." This body, in 1951, presented a report to the Economic and Social Council outlining a program for action. The basic ideas were that the League of Nations' 1926 Slavery Convention, binding forty-five states, should be maintained in operation and brought into the framework of the UN, which, however, should also prepare a new international instrument to make sure that not only chattel slavery but also other institutions and practices already studied by the League of Nations, and servitude of women resulting from marriage practices, should be brought to an end. In addition, mindful of another experience of the former League, the Ad Hoc Committee recommended the creation of a Standing Committee of Experts to supervise the application of both the old and the new conventions.

3. Continued Existence of Slavery.—The UN implemented the first recommendation in 1953 by executing a protocol transferring to its Secretary-General the powers and duties of the League's Secretariat in regard to the 1926 Slavery Convention. The conclusion of a supplementary convention was implemented in 1956, after a number of further studies, summarized in a report presented in 1955 by Hans Engen of Norway. In his introduction he pointed out that he endeavoured "to extract . . . only those statements which, in his view, would assist the [Economic and Social] Council in surveying the extent to which slavery, and practices resembling slavery, exist in the world today." He stated, however, that the documentation was in some respects incomplete and unverified. It must also be added that a number of governments concerned either denied the information in the report, or stated that it was inaccurate or out of date, or indicated measures had been taken to combat the objectionable practices reported. The material in the report on the continued existence of slavery (including "domestic slavery") and the slave trade concerned a number of countries in the Arabian Peninsula. Mention was also made of a few countries in Southeastern Asia, Africa, and South America. The material on serfdom concerned mainly countries in Asia, although certain regions in Africa and one country in Central America were also mentioned. The material related to "traditional forms of unpaid or underpaid personal services exacted by landowners and other employers of labour, or their agents" concerned a number of South and Central American countries, certain regions in Africa, in East and Southeast Asia, and in the north of the Arabian Peninsula. The material related to debt bondage, including pledging and pawning of third persons as security for debt, concerned a number of countries in East and Southeast Asia, in the north of the Arabian Peninsula, in various parts of Africa, and Latin America. The material related to the exploitation of children, notably under the guise of adoption, concerned various countries of East and Southeast Asia, South America, and one country in Africa. The material related to the "purchase" of wives and the "inheritance" of widows concerned mainly territories in Africa and in Northern Australasia, as well as one country in South America and a few in Asia.

4. Attack on Other Forms of Servitude.—The Supplementary Convention on the Abolition of Slavery, the Slave Trade, and Institutions and Practices similar to Slavery, based on a draft submitted in 1954 by the United Kingdom, was approved by a diplomatic conference at Geneva in 1956. The major feature of the Supplementary Convention lay in the fact that, although it maintained the 1926 definition of slavery, this definition had been supplemented by a description of the various institutions and practices similar to slavery, referred to above, in respect to which the signatory states agreed to "take all practicable and necessary legis-

lative and other measures to bring about progressively and as soon as possible [their] complete abolition or abandonment" (art. 1). Further, with a view to bringing to an end marriage institutions and practices leading to involuntary servitude for women, the states undertook to prescribe minimum ages of marriage and to encourage the use of facilities assuring the free consent of the parties to marry. (In 1962 the General Assembly adopted a convention drafted by the Commission on the Status of Women, calling on the states to assure that all marriages meet the requirements of minimum age, consent of both parties, and official registration.) The Convention of 1956 prohibited the mutilation, branding, or other marking of slaves and other persons of servile status. It also made punishable acts which might lead to the enslavement of a person, or to placing him into one of the servile conditions mentioned above. In some other respects the Supplementary Convention of 1956 was less satisfactory. In particular, once again governments were unable to agree on a clause authorizing their warships to exercise a reciprocal and mutual right of search on vessels suspected of slave trading. But under one provision, any slave taking refuge on board of any vessel of a state party to the Convention becomes *ipso facto* free. The Convention also established cooperation between party states and prescribed communication by them of certain information to each other and to the Secretary-General for submission to the Economic and Social Council "as part of documentation [in] view of making further recommendations for the abolition of slavery, the slave trade or the institutions and practices which are the subject of this Convention."

See also references under "Slavery" in the Index.

BIBLIOGRAPHY.—*General:* A. Tourmagne, *Histoire de l'esclavage ancien et moderne* (1880); H. J. Nieboer, *Slavery as an Industrial System*, 2nd ed. (1910); J. A. Saco, *Historia de la esclavitud desde los tiempos más remotos hasta nuestros días*, 2nd ed., 6 vol. (1936–45); Lady K. Simon, *Slavery* (1929); Sir G. MacMunn, *Slavery Through the Ages* (1938); P. Keti, "Slavery, Ancient and Modern," *Modern Review* (June 1948); E. Westermarck, *The Origin and Development of the Moral Ideas*, vol. 1, 2nd ed. (1912–17); C. Letourneau, *L'évolution de l'esclavage* (1897); C. W. W. Greenidge, *Slavery* (1958).

Slavery in Antiquity: A. Wallon, *Histoire de l'esclavage dans l'antiquité*, 3 vol., 2nd ed. (1879); I. Mendelsohn, *Slavery in the Ancient Near East* (1949), *Legal Aspects of Slavery in Babylonia, Assyria and Palestine* (1932); R. P. Dougherty, *The Shirkûtu of Babylonian Deities* (1923); A. M. Bakir, *Slavery in Pharaonic Egypt* (1952); A. Barnes, *An Inquiry into the Scriptural Views of Slavery* (1857); M. J. Raphall, *Bible View of Slavery* (1861); Z. Kahn, *L'esclavage selon la Bible et le Talmud* (1867); Ch. Polsky, *Slavery in the Bible and Later* (in Hebrew) (1959); N. D. Fustel de Coulanges, *La cité antique, étude sur le culte, le droit, les institutions de la Grèce et de Rome* (1864); W. L. Westermann, *The Slave Systems of Greek and Roman Antiquity* (1955); W. Richter, *Die Sklaverei im Griechischen Altertume* (1866); T. Frank, *An Economic History of Rome*, 2nd ed. (1927); P. Louis, *Ancient Rome at Work*, Eng. trans. by E. B. F. Wareing (1927); R. H. Barrow, *Slavery in the Roman Empire* (1928); W. W. Buckland, *The Roman Law of Slavery* (1908); M. I. Rostovtzev, *The Social and Economic History of the Roman Empire* (1926); M. I. Farley (ed.), *Slavery in Classical Antiquity* (1960); A. H. M. Jones, *Athenian Democracy* (1958).

Medieval Slavery: F. Pijper, "The Christian Church and Slavery in the Middle Ages," *American Historical Review*, vol. xiv (1908–09); J. W. Thompson, *An Economic and Social History of the Middle Ages, 300–1300* (1928); R. R. Miller, *Slavery and Catholicism* (1957); Ch. Verlinden, *L'esclavage dans l'Europe médiévale* (1955).

Muslim and African Slavery: M. Gaudefroy-Demombynes, *Les Institutions musulmanes* (1921); R. Roberts, *Das Familien-, Sklaven- und Erbrecht im Qorân* (1908); M. Ali, "Islam and Slavery," *Islamic Review* (Dec. 1957); C. W. W. Greenidge, "Slavery in the Middle East," *Middle Eastern Affairs* (Dec. 1956); S. Baker, *Ismailia* (1874); W. H. Ingrams, *Arabia and the Isles* (1942); R. O'Shea, *The Sand Kings of Oman* (1947); S. Bailes, *Slavery in Arabia* (1952); Baron Lugard, *The Dual Mandate in British Tropical Africa* (1922); *Africa Emergent* (rev. and expanded edition, 1949); E. Noel-Buxton, "Slavery in Abyssinia," *International Affairs*, vol. xi (1932); R. Coupland, *The Exploitation of East Africa, 1856–1890* (1939); E. de Jonghe, *Les formes d'asservissement dans les sociétés indigènes du Congo belge* (1949); Ch. Tisserant, *Ce que j'ai connu de l'esclavage en Oubangui-Chari* (1955).

Slavery in the Far East: R. Lingat, *L'esclavage privé dans le vieux droit siamois* (1931); Sir R. Reid, "Hill Tribes of Assam," *Journal of the Royal Emp. Society* (1943); B. Lasker, *Human Bondage in Southeast Asia* (1950); A. Wennington, *The Slaves of the Cool Mountains* (1959); D. R. Banaji, *Slavery in British India* (1933); G. S. Ghurye, *Caste and Class in India* (1957); R. Sh. Sharma, *Sudras in Ancient India* (1958).

The Antislavery Movement and Abolition of Slavery in the Western Hemisphere: Great Britain: B. Edwards, *History, Civil and Commercial, of the British Colonies in the West Indies*, 5th ed., 6 vol. (1819); T. Clarkson, *History of the Rise, Progress and Accomplishment of the Abolition of the African Slave Trade by the British Parliament* (1839); R. Coupland, *Wilberforce* (1923), *The British Anti-Slavery Movement* (1933); Sir T. Fowell Buxton, *African Slave Trade* (1839), *The Remedy—a Sequel* (1840); Charles Buxton, *Memoirs of Sir T. F. Buxton* (1848); F. J. Klingberg, *The Anti-Slavery Movement in England* (1926); W. L. Mathieson, *Great Britain and the Slave Trade, 1839–1865* (1929), *British Slavery and its Abolition, 1823–38* (1926); W. L. Burn, *Emancipation and Apprenticeship in the British West Indies* (1937); L. J. Ragatz, *Fall of the Planter Class in the British Caribbean, 1763–1833* (1928); E. E. Williams, *Capitalism and Slavery* (1944); H. A. Wyndham, *The Atlantic and Emancipation* (1937); H. Russell, *Human Cargoes; a Short History of the African Slave Trade* (1949); Ch. Lloyd, *The Navy and the Slave Trade* (1949); O. A. Sherrard, *Freedom from Fear; the Slave and his Emancipation* (1959); Great Britain, Foreign Office, *Correspondence with the British Commissioners at Sierra Leone, Havana, the Cape of Good Hope, Loanda, and New York; and reports from British Vice-Admiralty Courts and from British Naval Officers, relating to the Slave Trade.*

United States: J. E. Cairnes, *The Slave Power* (1862); H. Greeley, *The American Conflict* (1865); W. E. B. du Bois, *The Suppression of the African Slave-Trade to the United States of America, 1638–1870* (1896); G. S. Merriam, *The Negro and the Nation* (1906); H. H. Johnston, *The Negro in the New World* (1910); Mrs. N. W. Eppes, *The Negro of the Old South* (1925); E. Donnan (ed.), *Documents Illustrative of the History of the Slave Trade to America*, 4 vol. (1930–33, reissued 1965); U. B. Phillips, *American Negro Slavery* (1918); A. Y. Lloyd, *The Slavery Controversy, 1831–1860* (1939); G. H. Barnes, *The Antislavery Impulse, 1830–1844* (1933); H. Henkle, *Let My People Go; the Story of the Underground Railroad and the Growth of the Abolition Movement* (1941); H. Aptheker, *American Negro Slave Revolts* (1943); G. Myrdal *et al.*, *An American Dilemma* (1944); J. H. Franklin, *From Slavery to Freedom*, 2nd ed. (1956); F. Tannenbaum, *Slave and Citizen, the Negro in the Americas* (1947); R. B. Nye, *Fettered Freedom* (1949); T. E. Drake, *Quakers and Slavery in America* (1950); H. Aptheker (ed.), *A Documentary History of the Negro People in the United States* (1951); B. Mandel, *Labor: Free and Slave* (1955); K. M. Stampp, *The Peculiar Institution; Slavery in the Ante-Bellum South* (1956); W. D. Weatherford, *American Churches and the Negro* (1957); S. Alexander (ed.), *Quaker Testimony Against Slavery and Racial Discrimination* (1958); S. M. Elkins, *Slavery; a Problem in American Institutional and Intellectual Life* (1959); F. L. Olmsted, *The Slave States Before the Civil War* (1959); F. Bancroft, *Slave Trading in the Old South* (1959); J. R. Spears, *The American Slave Trade* (abridged edition) (1960); D. L. Dumond, *Antislavery: The Crusade for Freedom in America* (1961), *A Bibliography of Antislavery in America* (1961); D. P. Mannix, *Black Cargoes* (1962); Harvey Wish (ed.), *Slavery in the South*, a collection of contemporary accounts (1964); J. M. McPherson, *The Negro's Civil War* (1965); the *Journal of Negro History* (1916 *et seq.*) and numerous articles in other periodicals.

France: G.-Martin, *Histoire de l'esclavage dans les colonies françaises* (1948); M. D. Kennedy, *Lafayette and Slavery, from his Letters to Thomas Clarkson and Granville Sharp* (1950); L. Peytaud, *L'esclavage aux Antilles françaises avant 1789* (1897); P. Bande, *L'affranchissement des esclaves aux Antilles françaises* (1948); P. F. J. Grunebaum-Ballin, *Henri Grégoire, l'ami des hommes de toutes les couleurs* (1948); M. Satineau, *Schoelcher, héros de l'abolition de l'esclavage dans les possessions françaises* (1948).

Latin America: G. Scelle, *La traite négrière aux Indes de Castille*, 2 vol. (1906); A. Alcalá y Henke, *La esclavitud de los negros en la América española* (1919); H. H. S. Aimes, *A History of Slavery in Cuba, 1511 to 1868* (1907); R. Cepero Bonilla, *Azúcar y abolición*, 2nd ed. (1960); J. A. Saco, *Historia de la esclavitud de los Indios en el nuevo mundo* (1932); M. W. Williams, "The Treatment of Negro Slaves in the Brazilian Empire, a Comparison with the United States of America," *Journal of Negro History*, vol. xv (1930); G. Freyre, *The Masters and the Slaves*, Eng. trans. by S. Putnam, 2nd ed. (1956); International Labour Office, *Indigenous Peoples, Living and Working Conditions of Aboriginal Populations in Independent Countries* (1953).

International Action for the Suppression of Slavery: Many of the works mentioned in other sections of this bibliography deal also with this question. In addition the following United Nations documents may be consulted: *The Suppression of Slavery—Memorandum Submitted by the Secretary-General*, No. 1951. XIV. 2. (1951): In addition to a summary of international action since the beginning of the 19th century, contains the texts of the most important international treaties and conventions, including the League of Nations Slavery Convention of 1926 as well as the list of documents of the League of Nations relating to slavery; *Report of the First Session of the Ad Hoc Committee on Slavery to the Economic and Social Council*, doc. E/1660 of 27 March 1950 (1950); *Report of the [Second Session] of the Ad Hoc Committee on Slavery*, doc. E/1988 of 4 May 1951 (1951); *Replies of Governments to the Questionnaire on Slavery and Servitude, circulated by the Ad Hoc Committee on Slavery under Resolution 276 (X) of the Economic and Social Council*, documents E/AC.33/10 and Adds. 1 to 67 (1950–51); *Protocol Amending the Slavery Convention*

signed at Geneva on 25 September 1926, Resolutions adopted by the General Assembly at its Eighth Session during the Period from 15 September to 9 December 1953—General Assembly Official Records: Eighth Session, Supplement No. 17, doc. A/2630 (1953); *Report of the Rapporteur* [on the question of Slavery] *appointed under Resolution 525 A (XVII) of the Economic and Social Council,* Economic and Social Council, Official Records, Annexes, Nineteenth Session—Agenda item 8 (1955); *The Final Act of the United Nations Conference of Plenipotentiaries on a Supplementary Convention on the Abolition of Slavery, the Slave Trade, and Institutions and Practices Similar to Slavery,* Geneva, 13 August–4 September 1956, doc. E/CONF.24/23 (1956); *United Nations Conference on Slavery, 1956,* 57. XIV.2(1957). *See* also *United Nations Yearbook* (annual); *Yearbook on Human Rights* (annual). *See* also bibliographies under FORCED LABOUR; HUMAN RIGHTS. (E. N. Go.)

SLAVIC (SLAVONIC) **LANGUAGES.** The Slavic languages represent a western branch of the eastern, or satem, division of Indo-European (*q.v.*) and are closely connected with the Baltic languages, which embrace Lithuanian, Latvian, and the now extinct Old Prussian.

Classification.—Slavic languages are spoken in parts of East Germany, in Poland, Czechoslovakia, Yugoslavia, Bulgaria, and in most of the European part and large areas in the Asiatic part of the U.S.S.R. There are 13 living Slavic languages, classified in three groups. (1) Russian (more specifically, Great Russian), Ukrainian (formerly called Little Russian or Ruthenian), and Belorussian (or White Russian) comprise the East Slavic group (2) Polish, Slovak, Czech, and the High and Low Sorbian (also called Lusatian or Wendish) of East Germany form the Western group. (3) Slovene, Serbo-Croatian, Macedonian, Bulgarian, and Church Slavonic make up the Southern group. Church Slavonic is the liturgical language for all Orthodox and some Catholic Slavs, and its use is by no means confined to the South Slavic area. The two Sorbian languages were guaranteed a place in schools and public life by East German laws promulgated in 1948 and 1950. The other ten are official state languages in their respective areas.

Two additional languages may be noted, Polabian and Pomeranian. The former, once widely spread in the Elbe basin, became extinct in the 18th century. The latter is the conventional name for dialects once spoken along the Baltic, known from two remnants, Slovincian and Kashubian, which survived into the 20th century.

In spoken Slavic dialects, there are scarcely any true linguistic frontiers, for transitional dialects connect all types, excepting, of course, in the area where the South Slavs are separated from the others by non-Slavic Austrians, Hungarians, and Rumanians. Almost any two intelligent Slavs can manage fairly quickly to establish elementary communication on a simple conversational level, for the languages have preserved a remarkable degree of uniformity in over-all grammatical pattern and in the vocabulary of everyday life. However, the myriad differences in the details of phonetics, grammar, and above all of vocabulary cause misunderstandings even in the simplest of conversations, and the difficulties increase in the language of journalism, of belles-lettres, and especially in technological treatises, which tend to use newly created terminology. For a Slav to master effectively a second Slavic language demands time and work.

History.—For at least a millennium before the 3rd century A.D. the Slavs lived north of the Carpathians. (*See* further SLAVS.) Scholars have not been able to reconcile archaeological and linguistic evidence to arrive at a wholly satisfactory delimitation of the area, but they generally agree that its core extended from the Vistula to the Dnieper (Dnepr). Linguistic affinities indicate that the western neighbours of the Slavs were very likely first Illyrians, later perhaps Celts, and then Teutons. In the southeast there was apparently contact with Iranian and various Altaic peoples. In the northeast were Finno-Ugric tribes, and in the north the Balts. In the first centuries A.D. the Slavs began to expand rapidly, first to the west and east, then southward through and around the Carpathians, appearing in the 6th century as a serious threat to the Roman Empire all along the Danube. By the 9th century they occupied Europe east of a line running roughly Kiel-Brunswick-Magdeburg-Bamberg-Passau-Trieste, including all of the Balkan mainland except for parts of the littoral in Dalmatia and Greece. During the last millennium they have either retreated before or been absorbed by Germans in the west, Greeks and Albanians in the south, and Hungarians and Rumanians in the Danube basin. Only in the east was Slavic expansion a continuous process, which carried the Russians to the Pacific in the 18th century.

A long period, fully three millennia, separates recorded Slavic from any conceivable type of Indo-European from which Indic, Greek, and Slavic could all have sprung. The phonetic and morphological affinities now observable between Baltic and Slavic demonstrate at least a long period of contiguity between Slavs and Balts, and perhaps, as some hold, an ancient Balto-Slavic group. In any case, the Slavic of the 9th century appears still to have been a single language. It was of course divided, like all languages, into regional dialects, but the points of divergence were relatively few in number and quite insignificant in quality.

The 8th–11th centuries brought enormous changes to the Slavs. Slavic states arose and clashed with one another and with their non-Slavic neighbours. Social and economic conditions were radically transformed. In language, too, change was relatively rapid; and by the 10th century the western and southern dialects had split into something like the present linguistic divisions, although in the east a greater uniformity prevailed.

No credible evidence attests to Slavic writing before the 860s when a mission was sent from Byzantium in answer to the request of Prince Rostislav of Moravia for teachers "in our own language." The mission was

Phonological Divergences Among the Slavic Languages

Language	*candle*	*border*	*day*	*sleep*	*five*	*husband*	*fort*	*bank*
C.S.	*světja	*medja	*dĭnĭ	*sŭnŭ	*pęti	*mǫži	*gordŭ	*bergŭ
O.Ch.S.	svěšta	mežda	dĭnĭ	sŭnŭ	pętĭ	mǫžĭ	gradŭ	brěgŭ
Bulg.	sveštà	meždà	den	săn	pet	măž	grad	brjag
Mac.	svèḱa	mèǵa	den	son	pet	maž	grad	breg
Ser.-Crt.	svéća svijèća	mèđa	dân	sân	pêt	mûž	grâd	brêg brȉjeg
Sln.	svę̂ča	méja	dân	sən	pě̂t	mǫ̂ž	grâd	brę̂g
Cz.	svíce	meze	den	sen	pět	muž	hrad	břeh
Slk.	svieca	medza	den	sen	pät	muž	hrad	breh
H.Sorb	swěca	mjeza	dźeń	són	pjeć	muž	hród	brjóh
L.Sorb	swěca	mjaza	źeń	son	pěś	muž	grod	brjog
Pol.	świeca	miedza	dzień	sen	pięć	mąż	gród	brzeg
B.R.	svjačà	mjažà	dzen'	son	pjac'	muž	hòrad	bèrah
Ukr.	svičà	mežà	den'	son	p'jat'	muž	hòrod	bèreh
Russ.	svečà	mežà	den'	son	pjat'	muž	gòrod	bèreg

Language	*plow*	*wolf*	*heart*	*sister*	*brother*	*worker*	*button*
C.S.	*ōrdlo	*vĭlkŭ	*sĭrdĭce	*sestra	*brat(r)ŭ		
O.Ch.S.	ralo	vlĭkŭ	srĭdĭce	sestra	brat(r)ŭ	?	?
Bulg.	ràlo	vălk	sărcè	sestrà	brat	rabòtnik	kòpče
Mac.	ràlo	volk	srce	sèstra	brat	ràbotnik	dùgme
S.C.	rȃlo	vȗk	sȓce	sèstra	brat	rȃdnîk	dùgme
Sln.	rálo	volk	srcê	séstra	brát	dẹ́lavəc	gȗmb
Cz.	rádlo	vlk	srdce	sestra	bratr	dělník	knoflík
Slk.	radlo	vlk	srce	sestra	brat	robotník	gombík
H Sorb	radło	wjelk	serce	sotra	brat	dźěłaćeř	bubla
L.Sorb	radło	wjelk	serce	sotša	brat	źěłaśař	bublin
Pol.	radło	wilk	serce	siostra	brat	robotnik	guzik
B.R.	ràla	vowk	sèrca	sjastrà	brat	rabòčy	gùzik
Ukr.	ràlo	vovk	sèrce	sestrà	brat	robitnyk	hùdzyk
Russ.	ràlo	volk	sèrdce	sestrà	brat	rabòčij	pùgovica

led by a brilliant scholar and diplomat, Constantine (later known as St. Cyril, d. 869), and his brother Methodius (d. 885) of Thessalonica, a Greek city entirely surrounded by Slavs. (*See* CYRIL AND METHODIUS, SAINTS.) They created a new alphabet, the Glagolitic, and began to translate church books. Within two decades the Bible and many homiletic and liturgical texts had been translated, and some original works, including poetry, had been written in the new literary language. This language is called Old Church Slavonic (O.Ch.S.) because its local Macedono-Bulgarian origin is of less importance than its use as a literary language in the west (Moravia, Bohemia, Pannonia), the south (Bulgaria, Macedonia), and the east (Russia) during the next two centuries. In time it was modified locally, and the Church Slavonic in use today shows clearly the hand of Russian editors.

LETTER* CAPITAL	LETTER* LOWER CASE	EQUIVALENT	LETTER* CAPITAL	LETTER* LOWER CASE	EQUIVALENT	LETTER* CAPITAL	LETTER* LOWER CASE	EQUIVALENT
А	а	a	Ј	ј	j	Ч	ч	č or ch
Б	б	b	К	к	k	Џ	џ	dž or dzh ¶ ‖
В	в	v	Ќ	ќ	ǩ ‖	Ш	ш	š or sh
Ґ	ґ	g ♀ ‡	Л	л	l	Щ	щ	št or sht † ¶ šč or shch (all others)
Г	г	h ♀ ‡ g (all others)	Љ	љ	lj ¶ ‖	Ъ	ъ	ə or ă † ″ ("hard" mark) ♂ §
Ѓ	ѓ	ǵ	М	м	m	Ы	ы	y
Д	д	d	Н	н	n	Ь	ь	′ ("soft" mark)
Ђ	ђ	dj ¶	Њ	њ	nj ¶ ‖	Ѣ	ѣ	ě
Е	е	e	О	о	o	Э	э	è
Ё	ё	e	П	п	p	Ю	ю	ju or yu
Є	є	je ♀	Р	р	r	Я	я	ja or ya
Ѥ	ѥ	je §	С	с	s	Ѳ	ѳ	f
Ж	ж	ž or zh	Т	т	t	Ѵ	ѵ	i
З	з	z	Ћ	ћ	ć ¶	Ѧ	ѧ	ę §
Ѕ	ѕ	dz ‖ §	У	у	u	Ѩ	ѩ	ję §
И	и	y ♀ i (all others)	Ѹ	ѹ (ȣ)	u §	Ѫ	ѫ	ǫ § ə †
І	і	i	Ў	ў	w ‡	Ѡ	ѡ	o
Ї	ї	ji ♀	Ф	ф	f	Ѭ	ѭ	jǫ §
Й	й	j	Х	х	kh			
			Ц	ц	c			

* LETTERS ARE RUSSIAN AND BULGARIAN EXCEPT AS INDICATED BY THE FOLLOWING SIGNS: † BULGARIAN; ‡ BELORUSSIAN; § CHURCH SLAVONIC, ‖ MACEDONIAN; ¶ SERBO-CROATIAN; ♀ UKRAINIAN, ♂ RUSSIAN.

Note: Where alternative transliteration is given, the first is the preferred form.

CYRILLIC ALPHABET

In the West Slavic area O.Ch.S. was suppressed during the 11th century, but its influence is to be seen in the vernacular literature which sprang up in Bohemia in the 13th century and flourished in the 14th–15th centuries. Czech culture was exported to Poland at this time, and the Poles quickly developed their own standard language and produced superior literature in the 16th and 17th centuries. The Slovaks, if writing Slavic at all, used Czech until the national revival at the beginning of the 19th century. Sorbian texts began with Protestant translations in the 16th century.

In the south, Church Slavonic, in its progressively modified local forms, remained the exclusive written language for the Orthodox Serbs, Macedonians, and Bulgars, and for those Croatian Catholics who clung to the Slavonic rite. It was the Renaissance which gave the impulse to vernacular literature in Croatia in the late 15th century, and the Reformation which produced the first works in Slovene in the 16th century (except for fragmentary texts from the 11th). The Serbs discarded Church Slavonic in favour of the vernacular early in the 19th century, and by 1850 a common Serbo-Croatian literary language had been elaborated. The development of modern standard Bulgarian had its roots in the 18th but progressed little before the mid-19th century. Macedonian is the youngest Slavonic literary language, and can hardly claim to any history before 1944, when it was proclaimed the official language of the new Republic of Macedonia.

In the east, Church Slavonic continued in use well into the 18th century, although strongly vernacularized modifications were used for business documents and occasionally other writing. In Russia, the modern literary language grew out of a compromise style which utilized Church Slavonic elements to enrich the native Great Russian vernacular. The Ukrainians, on the contrary, rejected Church Slavonic in principle and based their writings chiefly on the spoken dialects. The first modern works appeared at the very end of the 18th century, but the real development of the contemporary Ukrainian literary language dates from the 1840s. The Belorussians had used their own distinctively modified type of Church Slavonic in the 16th century, but literature in the vernacular came only three centuries later. It is difficult to speak of a Belorussian literary language before the 20th century. Its most significant growth came about after it achieved official status in 1918.

Alphabet.—Slavic languages in the east and southeast are written in slightly variant forms of the so-called Cyrillic alphabet (*see* table), which seems to have been devised in Bulgaria at the end of the 9th century, and which soon displaced Constantine's Glagolitic. In the west, Roman Catholic Slavs use the Roman alphabet, mostly with the diacritics introduced by John Huss. Serbo-Croatian is written in Cyrillic by Serbs and Montenegrins and in Roman by Croats.

Vocabulary.—Much of the Slavic lexicon is Indo-European in origin ("mother, house, be"), a significant portion may be called Balto-Slavic ("head, hand"), and many words seem to be purely Slavic ("human, dog"). Ancient borrowings from Iranian include important words of spiritual culture ("god, sacrifice"), and a somewhat later stratum from Germanic pertains to material culture ("coin, buy"). The vocabulary of the individual languages reflects more recent local influences, the most important being German in the west and Turkish in the Balkans. Loan translations from Greek in the east and Latin in the west have served as patterns for creating new terminology with Slavic elements in many languages. Inter-Slavic borrowings are manifold and often difficult to trace.

Morphology.—Common Slavic (C.S.) retained the Indo-European case system almost intact, the only innovation being that the ablative merged with the genitive. Still discernible in late C.S. was the older system of paradigms based on purely

formal distinctions of stem and suffix-type: *o*- and *jo*-stems, *a*- and *ja*-stems, *u*-stems, *i*-stems, and, in restricted numbers, *ū*-stems and several kinds of consonant-stems. As early as O.Ch.S., however, a strong tendency prevailed to reorganize the system on the basis of gender. Thus *o*-, *jo*-, and *u*-stems were united in a masculine-neuter paradigm which has slightly variant suffixes according to whether the stem ends in a palatal ("soft") or nonpalatal ("hard") consonant, the usual feminine declension includes *a*- and *ja*-stems ("hard" and "soft"), and the *i*-stems are predominantly feminine. This principle of classification triumphed almost completely in the modern languages, except for Bulgarian and Macedonian, where the only remnant of declension is in vestiges of a general objective case. The West Slavic languages kept hard and soft variants of the masculine-neuter and the feminine *a*-declension, but in the Eastern and Southern groups there are only three paradigms for all but irregular nouns. The vocative disappeared completely in Russian and is preserved only with a limited number of nouns in Slovak, Slovene, and Low Sorbian. Broadly speaking, the force of the cases is weaker in the West, while in Russian some scholars see a strengthening and extension of the case system in the relatively new forms which have specifically partitive or locative functions, over against the forms which are generally genitive or prepositional.

The C.S. verb lost the voices and moods of Indo-European, retaining special forms for two past tenses (the imperfect and aorist), the present, imperatives (derived from the old optative), and five participles, two passive and three active. The active resultative *l*-participle was used with auxiliaries to form a series of periphrastic perfects. Constructions with the reflexive pronoun, and to a lesser degree with passive participles, render the passive and middle voices of classical languages.

The most significant innovation was the development of *aspect*, which remains the pivotal feature of all Slavic verb systems. Every verb is either of perfective aspect, denoting an action viewed in its entirety, or of imperfective aspect, denoting an action not necessarily so viewed. In principle, every verbal idea is represented by a pair of verbs, one for each aspect. Most modern languages have lost aorist and imperfect, using the old perfect instead. The two are, however, completely living forms in eastern Serbo-Croatian, Macedonian, and Bulgarian, where the perfect has come to signify a past action not witnessed by the speaker.

Phonology.—As a satem language, Slavic has *s* and *z* for the Indo-European palatal stops $\hat{k}$ and $\hat{g}$. In a development paralleled by, but not identical with, Baltic and Iranian, Slavic changed *s* to *x* (*kh*) after *i*, *u*, *r*, and *k*. As in Baltic and Germanic, *o* and *a* coincided in Slavic, the short vowel eventually resulting in *o* and the long in *a*. With Baltic, Slavic completely lost aspiration and maintained a sharp distinction between voiced and voiceless consonants. Specifically Slavic features are: an insistence on the distinction between front and back vowels which led to mutual adaptations between vowels and adjacent consonants and resulted eventually in a distinction between soft and hard syllables; the change of short *i* and *u* to reduced, ultrashort vowels called *jers;* the gradual elimination of closed syllables by dropping final consonants or simplifying clusters, by reducing diphthongs ending in *i*, *u*, or nasal, and by various dialectal changes in diphthongs ending in *l* or *r;* the replacement of quantitative vocalic distinctions by qualitative ones, and the subsequent development of a new quantitative system with rising or falling intonation possible on short as well as long vowels.

The table gives examples of some common words to show typical phonological divergences among the Slavic languages, while the relatively modern equivalents for "industrial worker" and "button" give a hint as to differences in roots and in word-formation. Some details are obscured by the different types of orthography current. Particularly significant for the classification of the languages is the development of *tj*, *dj*, *ě*, *ǫ*, *ę*, *ĭ*, and *ŭ*.

See also separate articles on the various Slavic languages and literatures; and references under "Slavic Languages" in the Index.

Bibliography.—R. Jakobson, *Slavic Languages* (1955), includes a concise account of historical phonology and grammar and carefully selected bibliography. R. G. de Bray, *Guide to the Slavonic Languages* (1951), furnishes practical information. K. Horálek, *Úvod do studia slovanských jazyků*, 2nd ed. (1962), in Czech, is the only detailed introduction. A. Meillet, *Le Slave commun* (1934), is the best description of Common Slavic.

(H. G. L.)

SLAVOPHILS, 19th-century Russian intellectuals who extolled the superiority of the typically Slav, or Russian, way of life. A group in the sense of intellectual affinity rather than organization, they opposed the modernization of Russia in the Western image as advocated by the so-called Westernizers. The Slavophils argued that Russia was basically different from the West in national character, that its foundations were unique and superior, and that some day Russia would emerge as the saviour of the West.

Slavophil ideas may be traced back to medieval times and to the doctrine that Moscow was the Third Rome, the eternal capital of the only true Christian ruler. They were rooted in the criticism of the West that had mounted concurrently with Peter the Great's efforts at westernization; in the new nationalism exemplified by A. S. Shishkov (*q.v.*) and fanned by Napoleon's invasion of Russia in 1812; in the writings in praise of autocracy by such historians as N. M. Karamzin (*q.v.*); and in the deliberations of the Lovers of Wisdom, or Lyubomudry, a circle of Russian devotees of the German idealist philosopher F. W. von Schelling (*q.v.*). The Slavophil movement took shape in the 1830s and early '40s, was at its height between 1845 and 1860, and faded toward the end of the 19th century.

The Slavophils were members of the gentry, a small circle of friends and relatives. Their antagonism toward the West was not the product of ignorance. They were well educated, both in Russia and abroad, were widely read in Western thought, and were influenced by many foreign authors such as Schelling and Hegel (*q.v.*). Philosophers by avocation, they tended to be less systematic and scholarly than haphazard and polemic in their writings.

The dominant figure of the Slavophil movement was A. S. Khomyakov (*q.v.*), a man of remarkable versatility, whose many writings ranged from Oriental philosophy to technology. His lyrical dramas *Ermak* (published 1832) and *The False Dimitri* (1833) were expressions of his belief in Russia's great mission and role as leader of all Slavs—a doctrine which he propagated repeatedly in his other writings and in debate with fellow intellectuals. Close to Khomyakov in importance and ideology, though more romantic, lonely, and depressed in personal makeup was I. V. Kireevski (*q.v.*), "the philosopher of Slavophilism," who formulated a new Slavophil philosophy in his *Answer to Khomyakov* and *On the Necessity and the Possibility of New Principles in Philosophy*. His brother Petr enhanced the movement by contributing to the study of Russian folklore through a large collection of folk songs. Two other brothers, the Aksakovs, were in the forefront of the Slavophils (*see* Aksakov): Konstantin, the historian of Slavophilism, was the most extreme member of the movement, dogmatically opposed to anything foreign, be it language, thought, or dress; and Ivan used his post as secretary and then as chairman of the Moscow Slavonic Benevolent Committee to make himself the foremost advocate of Pan-Slavism. Ivan Aksakov was also a leading figure in the Society of the Friends of Russian Letters and in the Orthodox Missionary Society; but he is most noteworthy as the popularizer of Slavophilism, since his writings reached more readers than the works of the other Slavophils. The most important Slavophil in terms of public position was Yuri Fedorovich Samarin (1819–76), a prominent theoretician. More liberal than the other Slavophils and a leading figure in the movement for the emancipation of the serfs, Samarin in his booklet *Revolutionary Conservatism* opposed constitutionalism as giving power to the landed gentry at the expense of the people. He was a readable, prolific, and versatile writer, propagating the Slavophil ideology in every field.

The Slavophil ideology embraced different, at times contradictory, ideas. To speak generally, the Slavophils subscribed to the following views:

1. Man is by nature religious. The religious element in him is more significant than his national or racial composition. Faith is more important than fatherland, and the church should be free from state interference.

7. The Orthodox Church is the essence of Slavdom. It is the only true church. The Russians are the only true Christians—simple, frugal, humble, and human (indeed, Humanity itself), with a high mission.
3. Russian life finds its fullest expression in the Christian brotherly love of the peasant communes, which encompass the majority of the Russian population. Russia is the mass of the people on the land, a classless society.
4. Land is the property of the communes, not (as in Roman law) of individuals. Private ownership of land as well as serfdom is wrong.
5. The commune is the natural level of administration in economic, religious, and other areas. Central government is a necessary evil. The tsar performs a useful function, but the ruling class is antinational.
6. The Russian people is by nature nonpolitical. Participation in government would corrupt it. Hence autocracy, with its lack of popular participation, is desirable. The role of the tsar is not one of glory, however, but one of sacrifice and martyrdom. He must not infringe the spiritual freedom of the people.
7. The history of Russia is one of organic growth, free of internal strife and divisions. Russia is an organic state, built on the glorious past of the Slavs and held together by Orthodoxy, in contrast with the West, where one-sided rationalism has led to political and social upheaval.
8. Russia is unique, different from and in contradiction to the West.
9. The Russian spirit and soul are expressed in the Russian language, which is closer than other tongues to the original language of mankind.

The relationship of Russia to the West has been a perennial problem in Russian history. The Westernizers, like the Slavophils, recognized the differences between Russia and the West. Unlike the Slavophils, however, they advocated the transformation of Russian society in the Western image. The Slavophils admitted the advantages of Western technology, but inveighed against the materialism of Western life. Influenced by Hegel, they viewed the relationship of Russia and the West in terms of dialectical opposites, though they expected the struggle to result in the dominance of Russia (the thesis) over the West (the antithesis) rather than in the evolution of a new culture (the synthesis). They had a "we-and-they" approach, praising the Russian and condemning the foreign. Deeply steeped in Western learning, notably in German idealism and in the Romantic idea of the nation and of the national mission, the Slavophils yet hated the West and were convinced that the feeling was mutual.

The Slavophils exaggerated the historical importance of the Slavs. Khomyakov, for example, regarded the Slavs as the original settlers of the European continent. He labeled the Angles, the Trojans, the Huns, and the Bulgars as Slavs and adopted Siegfried, Parsifal, Thor, Apollo, and Venus as Slavic figures. He asserted that the Slavs, with their Orthodox Christianity and their agricultural communes, had lived in an idyllic state of harmony until the reforms of Peter the Great had opened the floodgates for alien ideas—legalism, materialism, rationalism, and compulsion—fostering class division and conflict. Thus the Slavophils were able to discern enemies in St. Petersburg as well as abroad. The conservative bureaucrats at the Russian court were to them as alien and dangerous as the Westernizers, liberals, and radicals.

Convinced of the ultimate victory of truth, the Slavophils were opposed to censorship, whether of themselves or their rivals (the social scientists and unreligious philosophers). Their advocacy of freedom of conscience and speech and their attacks on the "anti-Russian" government invited repression. But the seeds of failure lay within Slavophilism itself. Slavic nationalism and Christian universalism were irreconcilable, the former ultimately emerging victorious. The Slavophil view of Russia as above social struggle proved false, the Slavophils having mistaken passivity for harmony. The elevation of the commune above the individual was as vain a prospect for freedom as the belief that the state would eventually wither away. Nevertheless the significance of the Slavophils was far from ephemeral. Slavophil themes pervade the writings of F. M. Dostoevski (*q.v.*); and elements of Slavophil thought passed into the mental makeup of Russians of the 20th century. History has proved the Slavophils wrong in many of their interpretations and predictions, but their emphasis on the fundamental problems of human faith, motivation, and life, and their attempt to find the basis for a true Christian society remain of value.

See also PAN-SLAVISM.

BIBLIOGRAPHY.—N. V. Riasanovsky, *Russia and the West in the Teaching of the Slavophiles* (1952); R. Hare, *Pioneers of Russian Social Thought* (1951); V. V. Zenkovskii, *Russian Thinkers and Europe*, Eng. trans. (1953); H. Kohn, *Pan-Slavism* (1953); M. B. Petrovich, *The Emergence of Russian Panslavism, 1856–1870* (1956).

(G. A. LN.)

SLAVS, the most numerous ethnic and linguistic body in Europe, with a considerable extension in northern Asia. Linguistically, they belong to the group known as Indo-European. (For their languages, *see* SLAVIC LANGUAGES and separate articles on the individual tongues.) The main block territory inhabited by the Slavic peoples includes almost all eastern Europe between the White, the Baltic, the Black, and the Caspian seas; westward it extends to the Oder (Odra) and the Lusatian Neisse (Nysa Łużycka) rivers, to the Erzgebirge (Krušne Hory), and to the Böhmer Wald (Český Les); and eastward it stretches across Siberia as far as the Pacific Ocean. In the Balkans there is the further Slavic area represented by Yugoslavia and Bulgaria, but this is separated from the main block by non-Slavic Hungary and Rumania.

The number of Slavs in eastern Europe or in northern Asia was estimated in the early 1960s at 226,600,000. They may be subdivided into East Slavs (Russians, Ukrainians, and Belorussians), 161,000,000; West Slavs (Poles, Czechs, Slovaks, and Lusatians), 43,700,000; and South Slavs (Serbs, Croats, Slovenes, Macedonians, Montenegrins, and Bulgars), 21,900,000. In addition there were about 10,000,000 people of Slavonic origin in the Americas, in western Europe, and in Australia.

The Slavs (Russ. *Slavyane*, Pol. *Słowianie*) were known in ancient times under two names: one which they themselves used and from which the English form Slav is derived; and another which non-Slav peoples used and from which the term Wends comes. The former, traceable to the 4th–5th century A.D., is of controversial etymology: some Slavists consider that it describes people of the same ethnic origin, others that it was applied to peoples dwelling in the proximity of rivers and lakes. A form of the name Wends was used by the Celts of the 1st century B.C. mainly to designate the West Slavs living east of the Elbe River; Pliny the Elder (1st century A.D.) recorded it as Venedi; later the Goths called the Slavs Wenden or Winden; and even today the Finns and the Estonians call Russia Venäjä. (GE. L.)

PHYSICAL ANTHROPOLOGY

Territorial expansion (*see* below, *Slav History and Culture: The Great Migrations*) has produced a considerable differentiation among the Slavs from the anthropological point of view. The following groups may be distinguished:

1. Nordic, with fair, almost pink, complexion, generally blue eyes and blond hair, long head, narrow and straight nose, tall stature, and slender body.
2. Laponoid (or Alpine), with dark eyes, dark or black hair, sallow complexion, round head, large face, broad nose, and short and thick-set body.
3. Mediterranean, with darkish eyes and hair, sallow complexion, rather long head and oval face, narrow and straight or aquiline nose, and small but well-proportioned physique.
4. Armenoid, with dark eyes and black hair, pale complexion, short head with a flattened back to the cranium and a face narrowing toward the chin, strong hooked nose, and tall stature.
5. Paleo-European, with fair eyes, hair, and skin, elongated head, large face and nose, heavy physique, and tall stature.

The Nordic type prevails numerically over the Laponoid, and Mediterranean admixtures are commoner than Armenoid. This quantification is valid in general for the northern part of the West Slavonic territory and for most of the East. Among the Ukrainians, however, Armenoid admixtures are commoner than Mediterranean. Among the southeastern Russians the Nordic majority is mixed more with the Mediterranean than with the Laponoid type, an obvious consequence of the absorption of the Iranian steppe peoples, descending from the Scythians, Sarmatians, and Alans.

In the Sudetic-Carpathian area, the anthropological structure is different. In Upper Lusatia, in southern Poland, and among the

Czechs and Slovaks, the Laponoid element is stronger than the Nordic and the Armenoid admixture greater than the Mediterranean. This is to be explained by the relatively late slavicization of this area. North and northeast of the Carpathians, save in Polesie, the presence of the Laponoid element is rather exceptional.

The predominance of the Mediterranean type is the main anthropological characteristic of the Bulgars as well as of the populations (only partly Slavic) living northwest of the Black Sea. The Macedonians, however, although linguistically akin to the Bulgars, are anthropologically different from them. The reason is that the Bulgars arrived in a territory with an archaic Mediterranean population, while in the western part of the Balkan Peninsula this stock had been overlaid by an Armenoid one, resulting from an expansion of Asianic (Anatolian) tribes. The occupation by the Serbo-Croats of a section of the latter area resulted in a very considerable absorption of the Armenoid element. The Slovenes, geographically neighbours and linguistically akin to the Serbo-Croats, are very different from them anthropologically, the Nordic element being predominant among them.

The Paleo-European type, a considerable admixture among the European population in the Mesolithic Age, is only slightly represented among the Slavs. It is most noticeable where archaic Finno-Ugric indigenous populations have been russified.

This account of the present anthropological character of the Slavs reflects their ancient contacts with the Germanic tribes as well as the consequences of their expansion eastward and southward. Very little is known about their anthropological structure before their territorial expansion. How the process of differentiation worked itself out in detail is obscure, but the main trends of the process are clear enough: they were the quantitative reduction of the Mediterranean and Paleo-European admixtures and the increase of the Nordic element. The territorial expansion of the Slavs effected a considerable change in the population structure of the area previously occupied by the Balts (*q.v.*); and west of the Oder the Slavs arrived in such great numbers that the subsequent German colonization caused no perceptible changes in the anthropological type. In the Balkans, though linguistic slavicization made great progress, the Slavic layer failed to submerge ancient anthropological differentiations, as the immigration was relatively small. In the Ukraine, which underwent many Asiatic invasions, the present anthropological situation is the result of a new wave of Slavic colonization that began in the 17th century, more or less simultaneously with the Russian penetration of Siberia, where anthropological slavicization is heavy because of the sparseness of the indigenous populations. (J. Cz.)

SLAV HISTORY AND CULTURE

Prehistory.—The original habitat of the Slavs, as of the ancestors of all other Indo-European speakers, was Asia. They migrated in the 3rd or 2nd millennium B.C. and, together with Balts, Scythians, and Sarmatians (*qq.v.*), populated eastern and central Europe. Archaeologically, it seems possible to trace the route of their migrations by finds of a painted pottery decorated with spirals and meanders, belonging to a culture that flourished between the Neolithic and Bronze ages. Balts and Slavs were originally members of the same linguistic family, but a differentiation started *c.* 1000 B.C., leading to the formation of two different linguistic groups.

In the basins of the Vistula and, chiefly, of the Oder River the so-called Lusatian culture was born (18th–8th century B.C.). Many archaeologists, anthropologists, and linguists identify this cultural area with the habitat of the early Slavs, but some archaeologists are inclined to associate it with other Indo-European peoples, above all with the Illyrians, or Veneti (*see* ILLYRIA), and to maintain that the original habitat of the Slavs was on the middle Dnieper River. There seems, however, to be no radical contradiction between the two theories. For in the whole area between the Dnieper and the Elbe and between the Baltic Sea and the Carpathian Mountains, the names of rivers and lakes are obviously of a common origin, and there is likewise much similarity among the various cultures found there. If the people of the Lusatian culture originally constituted a linguistic group distinct from the Balto-Slavs (though akin to them), it would seem eventually to have become totally subject to that influence, with the result that the Balto-Slavs of the Lusatian cultural area probably developed into the new ethnic and linguistic group known as Slavs or Wends.

Herodotus (5th century B.C.) seems to refer to Slavs in his description of Scythia when he mentions the Neuri living west of the middle Dnieper. Pomponius Mela, Pliny, Strabo, and Tacitus, writing in the 1st century A.D., give some information on Wends inhabiting an area along the Vistula River and on the Baltic shores. Ptolemy (2nd century A.D.) mentions the "considerable people" of the Venedai and the Venedic Gulf on the Baltic, by which he probably meant the Gulf of Danzig. By that time the Slavs were certainly inhabiting the area between the Dnieper and the Oder, the Vistula basin being their main habitat.

The Great Migrations.—The lands of the Slavs were crossed by many peoples forced by economic conditions to migrate. In the middle of the 1st millennium B.C. they were traversed by Celts and Germans on their way toward the Mediterranean and by Scythians and Sarmatians going to the Danubian basin. In the following centuries the Slavs were in cultural relations with these peoples.

Celtic tribes settled along the upper Oder (in Silesia) and Germanic tribes (Goths, Gepidae, Rugii, Vandals, and Burgundians) on the lower Vistula and lower Oder, usually without displacing the Slavs there. From the Goths the Slavs took many loanwords of cultural and social significance; *e.g.*, the Gothic *kuniggs*, "king," which became *knez*, "prince," in Slavonic. At the beginning of the 3rd century A.D. the Goths and the Gepidae left the lands of the lower Vistula and moved toward the Black Sea, by the shores of which in the 4th century they formed a powerful state in contact with the Slavic tribe of the Antae. Around 370 their state was invaded by the Huns (*q.v.*), who subdued Slavs and Germans alike.

The collapse of the Hunnish empire (455) and the departure of the Germans started the great migration of the Slavs. In the 6th century the historians Jordanes and Procopius divide the Slavs into three tribes, Antae, Sklaveni, and Venedi. The Antae were the East Slavs, the Sklaveni the South Slavs, and the Venedi the West Slavs. This division, however, was somewhat artificial, since at that time the Slavs still formed a very largely homogeneous community. Differentiation was the final result of the migration of groups of Slavs in three main directions: (1) westward, into the country between the Oder and the Elbe-Saale line; (2) southward, into Bohemia, Moravia, Pannonia (the later Hungary), Dalmatia, the lower Danube basin (Rumania), Bulgaria, and northeastern Greece; and (3) northward, along the upper Dnieper (*See* also EUROPE: *History: Greeks, Romans and Barbarians: The Migrations.*)

The First Organized States.—When the migratory movements had ended, there appeared among the Slavs the first rudiments of a state organization, headed by a prince with a treasury and a defense force, and a beginning of class differentiation. The Slavs formed such states somewhat later than the Germans did, as the latter had moved into territory where they inherited the forms of political organization from the Romans.

The oldest Slavic state was that of Samon or Samo, who at the beginning of the 7th century succeeded in grouping the Moravian and Czech tribes into a political unit as a result of a national rising against the Avars (*q.v.*). After Samon's death (658) the Moravians and the Czechs were again subjugated by the Avars. When the Avar state was finally destroyed by the Franks under Charlemagne (805), another Moravian, or Great Moravian, state was founded in the same area by Mojmir. Under Prince Svatopluk (870–894) this Great Moravia included the present Bohemia, Moravia, southwestern Poland, western Slovakia, and western Hungary. This state was destroyed in 907 by the Magyars. (*See* MORAVIA.)

The Catholic Franks and the Orthodox Byzantines had competed for influence in Great Moravia. The missionary work of Saints Cyril and Methodius (*q.v.*), though it failed along the middle Danube, led to the foundation of an Orthodox Church which

embraced the East Slavs and the more easterly South Slavs (Serbs, Macedonians, and Bulgars), while the West Slavs and the more westerly South Slavs (Slovenes and Croats) turned to Roman Catholicism.

The westernmost boundary of Slavic land was the Elbe-Saale line. The Polabian Slavs (*see* POLABS) living east of that boundary established their own principalities, the most important being those of the Lusatians, of the Obodrites, and of the Wilcy or Veleti. They all fell victims to the German ***Drang nach Osten*** or "drive to the East" in the 12th–13th centuries.

The *Drang nach Osten* may be regarded as a process continuing from the subjection of the Polabian lands by Germans in the Middle Ages to the participation of Prussia and of Austria in the destruction of Poland in the 18th century (*see* below). It affected all the Slavic lands, not excluding Russia. Military conquest was not the only form that it took: other aspects were settlement by German colonists, economic domination, and cultural influence. In the earlier period, contact with the German world was economically and culturally beneficial to the Slavs; but later subjection caused a strong national resentment.

During the 9th and 10th centuries, the Balkan Slavs, especially the Bulgars and the Croats, formed their own independent states. The first Bulgarian state (*see* BULGARIA: *History*) was founded in the 7th century by non-Indo-European invaders of Hunnish descent. They became slavicized, and under their tsar, Simeon the Great (893–927), who assumed the title of "emperor and autocrat of all of the Bulgars and Greeks," the Bulgarian empire became a civilized power. Destroyed in 1014 by the Byzantines, it was restored in 1185 to be destroyed again in 1392–96 by the Ottoman Turks.

On the territory of modern Yugoslavia the first Slavic kingdom was Croatia (*q.v.*), whose prince, Tomislav, received a royal crown from Rome (925). But conflicting influences—Orthodoxy and Catholicism on the one hand, the Holy Roman Empire, Hungary, and Bulgaria on the other—long prevented the Slavs of Yugoslavia from forming a unified independent state. In 1102 Croatia was united with Hungary, and this union eventually brought Croatia with Hungary to the Austrian Habsburgs. Slovenia (*q.v.*) had lost its independence much earlier, becoming a march or mark (frontier zone) of the Holy Roman Empire in the 10th century.

The first Serbian state was formed in the 9th century on the territory of the present Bosnia and Serbia (*see* BOSNIA-HERCEGOVINA; SERBIA). It became an empire under the Nemanya dynasty in the 12th century but was destroyed by the Ottoman Turks after the Battle of Kosovo (1389).

Magyars, Pechenegs, and Kumans (*qq.v.*) destroyed the Slavonic settlements in the present Hungary and Rumania. At the beginning of the 11th century Slovakia (*q.v.*) was also incorporated into Hungary.

The Czechs and the Poles succeeded in consolidating their independent states. Bohemia (*q.v.*) had to become part of the Holy Roman Empire in the 10th century, but in 1085, 1158, and 1198 the Přemyslide dynasty secured royal crowns, and under Přemysl Otakar I, in 1198, the royal title became hereditary. From 1310 the Bohemian throne was occupied by Germans of the Luxembourg dynasty. One of them, who became Holy Roman emperor as Charles IV in 1355, made Prague one of the foremost cultural centres of Europe. The Luxembourgs, however, lost Bohemia in 1419 during their wars against the Hussites (*q.v.*) who stood for Bohemian nationalism. In 1458 George of Poděbrady (*q.v.*) was elected king. After his death (1471) the Bohemian throne was coveted by Austria, Hungary, and Poland. Poland obtained it for Prince Wlandyslaw (Vladislav), son of King Casimir IV; but in 1526 it passed to the Habsburgs.

The cradle of the Polish state was Great or Old Poland (Wielkopolska, with Gniezno). Other parts were Small or New Poland (Malopolska, with Cracow), Silesia, Mazovia, and Pomorze. Unification was the work of the Piast dynasty, particularly of Prince Mieszko I (*c.* 962–992) and his son King Boleslaw I the Brave (992–1025); but Prince Boleslaw III the Wry-Mouthed (d. 1138) divided Poland among his four sons. The kingdom was restored in 1295, when Prince Przemysl II was crowned king.

Przemysl II's kingdom did not embrace Silesia or Mazovia (*qq.v.*), which remained autonomous under Piast princes. The Silesian princes subsequently recognized the suzerainty of the Bohemian crown, and Silesia passed with Bohemia to the Habsburgs in 1526 and was surrendered by them to Prussia in 1740. Mazovia, on the contrary, was reincorporated into the Polish realm by stages.

On the Baltic coast, Pomerania, or Pomorze (*q.v.*), was divided from 1181: both the western and the central part, under the Slavic dynasty of Gryf, had to accept the protection of the Holy Roman Empire, but the eastern or Vistulan part, with Gdańsk as its capital, remained linked with Poland—to become integrated with it in 1294. It was conquered by the Teutonic Order in 1308 but was returned to Poland in 1466. On the southern frontier, meanwhile, the principality of Halich-Volodymir (*see* GALICIA) was incorporated into the Kingdom of Poland in 1340.

In 1386, faced by a common German danger, Poland and Lithuania concluded a personal union under the Jagiello dynasty; and in 1569, to oppose Muscovite expansion, this union became a real one. It survived till the partitions of Poland among Russia, Prussia, and Austria in the last quarter of the 18th century (*see* POLAND: *History*).

The oldest state of the East Slavs was that of Kiev, in the Ukraine, formed at the end of the 9th century (*see* RUSSIAN HISTORY). Historians who argue that the first Russian state was a creation of Varangian (Norman or Scandinavian) rulers cannot deny that from the beginning this state was ethnically and culturally Slavic. Under its grand prince Vladimir the Saint (reigned 978–1015) the Rus—as the Kievan state was called—accepted Christianity from Byzantium in 988–89. Under Yaroslav the Wise (reigned 1019–54) the Rus attained considerable power, and Kiev became the great metropolis between Constantinople and the Baltic. In 1240 the Tatars destroyed the Kievan state, and the Russian princes became vassals of the Golden Horde. (*See* MONGOL EMPIRES.)

While the Tatar yoke paralyzed the development of Russian culture for two centuries, the principality of Moscow rose to political importance (*see* MOSCOW, GRAND PRINCES OF). In 1472, less than 20 years after the fall of Constantinople to the Turks, the grand prince of Moscow, Ivan III the Great, married Zoë (Sophia) Palaeologus, niece of the last Byzantine emperor; and in 1480 Ivan proclaimed himself "autocrat of all Russia" and declared Moscow to be the "Third Rome." Ivan IV the Terrible was crowned as first "tsar of all Russia" in 1547. The 17th century saw the settlement of Russians and Ukrainians in Siberia (*q.v.*) and in the steppes north of the Black and the Caspian seas. Peter (*q.v.*) the Great, who proclaimed himself emperor in 1721, and the empress Catherine II (*q.v.*) the Great, who reigned from 1762 to 1796, consolidated Russia's position as a great power. At the beginning of the 19th century Russia was the only independent Slavic state in the world.

Political Restoration in the 19th and 20th Centuries.—The weakness of the Ottoman Empire at the end of the 18th century, together with the conflicting interests of the Great Powers in the area controlled by it, led to the restoration of Slavonic states in the Balkans (*see* EASTERN QUESTION). Serbia, proclaimed autonomous in 1830, became independent and territorially aggrandized in 1878 and was declared a kingdom in 1882. It was further aggrandized through the Balkan Wars (*q.v.*) of 1912–13, which also gave an extension of territory to Montenegro (*q.v.*), a small state independent from 1878 and a kingdom from 1910. After World War I, Serbia, with an expanded area that included Montenegro, merged itself into the Kingdom of the Serbs, Croats, and Slovenes, which in 1929 changed its name to Yugoslavia. Bulgaria, meanwhile, having become autonomous under Ottoman suzerainty in 1878 and having extended its control over Eastern Rumelia in 1885, had become an independent kingdom in 1908 and had also expanded as a result of the Balkan Wars.

Poland's struggle to recover independence found a powerful echo among other Slavic nations, especially the Czechs and the Slo-

vaks. Toward the end of the 19th century, in Russian Poland, it assumed partly a social revolutionary character, whereas in Prussian and in Austrian Poland the movement was essentially national. Nationalism was also the main force in the resistance of other Slavs within the Austro-Hungarian monarchy—Czechs, Slovaks, Croats, and Slovenes. The defeat of the Central Powers in World War I led to the restoration of Poland, as well as to the creation of Czechoslovakia and of the Kingdom of the Serbs, Croats, and Slovenes.

Culture.—The oldest forms of Slavic culture, material, social, and spiritual, are known somewhat superficially from archaeological discoveries. The higher forms of that culture are closely linked with the spiritual life (religious cults, the cult of ancestors, and sacred architecture), a link that persisted for many centuries and was especially strong in the Middle Ages. Moreover, the most ancient writings are concerned with religious life. Only when the first organized states emerged did any historiography appear, as well as collections of laws. The Slavs of the Orthodox Church early began to write in their national languages, particularly the Russians and the Bulgars; but the Roman Catholic Slavs wrote in Latin till the Reformation.

The lay trend in Slavic culture is above all represented by popular arts. Its beginnings can be traced in prehistoric remains, and its development, not only in the plastic arts but also in music and in poetry (songs, ballads, and epics), ran parallel to that of the work produced for the church or for the upper classes. Popular poetry flourished everywhere among the Slavs, most richly in Russia and in the Balkans.

For the literature of the East Slavs *see* Russian Literature and Ukrainian Literature; for that of the West Slavs, Czechoslovak Literature and Polish Literature; for that of the South Slavs, Bulgarian Literature and Yugoslav Literature. The Slovak poet Jan Kollár was a pioneer exponent of Pan-Slavism (*q.v.*), while the reactionary Slavophils (*q.v.*) tried to persuade non-Russian Slavs that "Holy Russia" was their motherland.

Slavic music owes very much to popular inspiration: Frédéric Chopin, Aleksandr Borodin, Mikhail Glinka, Modest Mussorgsky, Nikolai Rimski-Korsakov, and Bedrich Smetana provide examples. In painting, Jan Mateiko in Poland and Ilya Repin in Russia are famous for their scenes of national history. Among sculptors, the Croat Ivan Mestrović is universally known. In the physical sciences were the astronomer Copernicus and the radiologist Marie Curie (née Sklodowska). Among Russian scientists the polymath Mikhail Lomonosov, the mathematician Nikolai Lobachevski, the chemist Dmitri Mendeleyev, and the physiologist-psychologist Ivan Pavlov must be mentioned here. Among the Czechs, Jan Purkyně (Purkinje) and Jiří Procházka were both important physiologists.

See also articles on the various Slavic countries and references under "Slavs" in the Index.

Bibliography.—L. Niederle, *Manuel de l'antiquité slave*, 2 vol. (1923–26); S. H. Cross, *Slavic Civilization Through the Ages* (1948); F. Dvornik, *The Making of Central and Eastern Europe* (1949) and *The Slavs: their Early History and Civilization* (1956); A. Mousset, *The World of the Slavs* (1950); J. Czekanowski, *Wstęp do historii Słowian*, 2nd ed. (1957); S. A. Nikitin, *Istoriya yuzhnykh i zapadnykh Slavyan* (1957). (Ge. L.)

SLAVYANSK, a town of the Donetsk (formerly Stalino) Oblast' of the Ukrainian Soviet Socialist Republic, U.S.S.R., stands at the confluence of the Kazënnyy Torets and Sukhoy Torets, tributaries of the Severnyy (Northern) Donets. Pop. (1959) 82,784. The town is the main centre of the northwestern part of the Donets Basin (Donbass) industrial area. The presence of saline and mud springs and of rock salt and coal makes Slavyansk an unusual combination of health resort and industrial town. There is an important chemical industry, producing chiefly caustic soda. Armatures, other machinery, and pencils are also made. Slavyansk has a peat-burning power station. The town was founded in 1676. (R. A. F.)

SLEEP is a recurring state of inactivity, decrease of consciousness and decrease in responsiveness to events in the environment. The ease with which sleep can be terminated differentiates it from coma, general anesthesia, alcoholic stupor and the seasonal lethargy of certain animals (*see* Hibernation and Estivation). On the other hand, there are conditions of semiawakeness ("brown study," postepileptic automatism, hypnotic trance, fugue, etc.) that are intermediate in character, possessing some features of both sleep and wakefulness. Moreover, in the routine of living there are regular diurnal gradations in the depth of slumber and the degree of alertness, which alternate with each other like the crest and trough of a 24-hour wave.

From observation and experiment, theories of sleep have been elaborated and a hygiene of sleep developed.

Observational Data.—A conspicuous characteristic of sleep, for man and many animals, is the horizontal posture of repose which permits a relaxation of the body musculature. When a person is awake, muscular contractions not only produce overt movements but also maintain station, or posture, through tonic activity —a state of partial contraction which manifests itself in resistance to passive or externally imposed movement. In raising gently an arm of a sleeping person, however, little opposition is met, and when the grip on it is released, the arm drops back limply. A rather annoying (to others) aspect of this general relaxation of sleep is that it also involves the muscles of the jaw, causing the mouth to open; and then the soft palate, with its appendage, the uvula, flutters with passage of the air in and out of the lungs, producing the well-known sounds of snoring.

Muscles are supplied with two types of nerve fibres: (1) efferent, motor, centrifugally conducting fibres, which carry impulses from the nervous system, resulting in overt as well as tonic contractions; and (2) afferent, sensory, proprioceptive ("self-feeling"), centripetally conducting fibres, which furnish conscious and unconscious information concerning the state of the different parts of the body and their relation to each other. This muscle sense, literally a sixth sense, contributes greatly to the continuous stream of afferent impulses from the sense organs to the cerebral cortex. In order to sleep, a person shuts off the streams of visual and auditory impulses by retiring to a dark, quiet room, and decreases cutaneous impulses by lying down on a soft, smooth surface, but the proprioceptive impulses, coming as they do from the body itself, are still there, gradually decreasing only when and as the body musculature is relaxed.

Muscular relaxation and immobility, however, are not absolute. From time to time, unless mechanically restrained, the sleeper moves one part of the body or another, or turns over completely changing his position. A person's positional repertory may be small or large, and the total amount of motility varies with the degree of fatigue, environmental temperature, amount of bed coverings, width of the bed and the quantity and composition of food and beverage taken before going to bed. As a rule, the longer the duration of sleep, the greater the frequency of movement. Occurring altogether about 20 to 40 times in the course of one night, this motility entails at least a semiawakening and a partial awareness of external relations. The sleeper does not usually recall these multiple interruptions of rest, and the total time spent in motility is small, amounting to about 30 seconds per hour, or a few minutes for a whole night's sleep. The periodic change in the position of the body is probably brought on by, and relieves, the pressure on the skin in contact with the bed surface. Alcoholic intoxication, deadening sensitivity to discomfort, leads to decreased motility during sleep, often resulting in soreness and stiffness in a muscle or even a partial paralysis from pressure exerted on a nerve for several hours.

It is possible to sleep in a semireclining position (as on a reclining chair), sitting upright (at a lecture, theatre or in an easy chair at home) or, when very tired, while standing up or riding horseback. Some animals regularly sleep in a standing position.

Of greatest interest are the changes in the activity of the nervous system during sleep. In older children and adults this manifests itself in the abolition, or at least depression, of critical reactivity to external events. In the waking state the impulses coming from the different sense organs to certain areas of the cerebral cortex are analyzed in the light of the person's previous experience, and appropriate responses (which include movement

as well as refraining from overt muscular activity) are elaborated or integrated in other cortical areas. Identical afferent impulses from sense organs will not elicit the same response in different persons. This individuality of reaction is lost during sleep and is replaced by stereotyped predictable reflexes from the lower centres of the nervous system. Furthermore, compared with wakefulness, the reflex excitability of the lower centres themselves is decreased during sleep. A more powerful stimulus is required to obtain a response of a certain magnitude.

During sleep the amplitude and frequency of the electrical brain waves undergo characteristic changes, and it has been possible to associate particular electroencephalogram (EEG) patterns (*see* ELECTROENCEPHALOGRAPHY) with certain stages of sleep. Epileptic persons show typically peculiar brain waves while awake, but even more revealing are their EEGs during sleep. Furthermore, whereas an epileptic fit in the waking state is almost always ushered in by a definite EEG complex, the same complex in sleep does not lead to overt convulsions unless the sleeping epileptic is awakened at that time. This suggests that in sleep there may be a functional de-efferentation of the lower centres of the nervous system.

In contrast to the general inactivity, muscular relaxation, lack of critical reactivity, and depression of reflex excitability—all involved in the external or animalistic relations of the organism—the internal, household, visceral or vegetative functions are either unaffected or somewhat enhanced during sleep. Circulation of the blood goes on, although the heart rate is decreased and the arterial blood pressure lowered. These changes are in part related to the assumption of the horizontal position, and often occur upon simply lying down and remaining awake. In sleep there are also modifications in the rate and depth of respiration, indicative of a decreased excitability of its regulating centre, along with that of other reflex centres, in the nervous system. As a result there are certain changes in the composition of the blood, which is less thoroughly purified than in the waking state. Digestion—the movements of the different parts of the gastrointestinal tract and the secretory activity of its glands—goes on unabated. The same seems to be true of the liver and kidneys, although some decrease has been detected in the functioning of these vital organs, as well as in the metabolic rate (oxygen consumption) of the organism. There is a marked drop in body temperature, often amounting to 1° F., which is undoubtedly related to muscular relaxation and a lowering of metabolic processes. Sweating, in general, as a means of cooling off the body by evaporation, may actually be accentuated during sleep, especially in warm weather, but is diminished in skin areas of the palms of the hands and soles of the feet.

Variation in each of the above-enumerated concomitants of sleep in the course of the night has been employed as an index of its depth. Unfortunately, there is no agreement among the several depth-of-sleep curves constructed in this manner. The most commonly used method is to determine the loudness of a sound required to awaken the sleeper. The figures obtained suggest that during the night sleep gets to be very deep in the first two or three hours and then becomes progressively lighter. Similar curves could be derived from data on quiescence (the reverse of motility), the increase in the carbon dioxide of the air in the lungs and the predominance of slow (3 to 0.5 per second) delta EEG waves. Using arterial blood pressure or body temperature as a criterion, sleep is deepest in the middle of its course. Lastly, the lowest heart rate is not reached until the sixth or seventh hour of sleep. Thus, by proper selection of an index, it can be shown that sleep is deepest in the first, second or third quarter of the night. More important, study of the figures on which the different curves are based reveals that in any one night there may be repeated fluctuations in the depth of sleep. Instead of one curve there are several—each representing a basic rest-activity cycle. This cycle, first detected in the sleep of infants, in whom it repeats itself every 50–60 minutes, gradually lengthens to 80–90 minutes in adults. In young kittens the short-term sleep cycle is 20–30 minutes; in mature cats, 30–45 minutes.

One phase of the repetitive depth-of-sleep wave has been called paradoxical, because it is characterized by an EEG fast-activity pattern usually seen in wakefulness, yet the sleep is very deep, as judged by the intensity of the stimulus required to awaken the animal. In man the corresponding phase of the short-term sleep cycle shows a modified alpha EEG pattern, somewhat irregular and a trifle slower than the waking EEG. This phase may be termed quasi-paradoxical: using a neutral sound to arouse the sleeper, the intensity needed is very high, suggesting deep sleep of the delta EEG pattern; but with a meaningful sound stimulus, *e.g.*, the subject's name or a modulation indicating punishment for failure to respond, the sleeper can be awakened as easily as in the early light stage of sleep. During the quasi-paradoxical phase there are, in addition to the modified alpha EEG, cardiac and respiratory accelerations as well as bilaterally synchronous rapid eye movements (REMs), distinctly visible even under closed eyelids but also recordable electronically as changes in the corneoretinal potential.

If the sleeper is awakened during this phase of the short-term cycle, he nearly always reports dreaming, but rarely does so in other phases of the cycle. The quickening of the pulse and breathing may be related to emotional elements of the dream episode, and the REMs probably represent scanning the scene of dream action, as their direction conforms to the movements of objects and persons seen in the dream.

The length of the dream story is proportional to the lapse of time between the appearance of the modified alpha EEG and the arousing and questioning of the sleeper. Contrary to anecdotal reports, the course of time during dreaming is about the same as in real life. Sound or light stimuli too weak to awaken the sleeper, if applied while dreaming is in progress, are occasionally incorporated into the dream story. After the cessation of the objective signs of dreaming, the ability to recall details of the dream content rapidly diminishes, and in a few minutes the sleeper may forget that he dreamed at all. That explains why many persons maintain that they seldom, or never, dream. However, even the latter invariably recall dreaming if aroused while showing a modified alpha EEG sleep pattern. The objective criteria of dreaming can be used to plot the incidence and duration of dreaming episodes during a night of undisturbed sleep. There are usually five or six periods of dreaming, the first, and shortest, occurring about one hour after the onset of sleep, and the succeeding ones, corresponding to 80- to 90-minute cycles of EEG pattern variation, progressively longer. The average duration of dreaming periods is from 5 to 50 minutes, with a total of about two hours for a whole night's sleep. (*See* also DREAM AND DREAMING.)

Experimental Data.—Sleep deprivation (keeping a person awake by constant watching and prodding) has been a favourite method used in study of the sleep problem. After about 60–90 hours of enforced wakefulness (four to six times the usual span of 16–17 hours), the most prominent effect is extreme muscular weariness. The subjects of such experiments want most to be allowed to close their eyes and lie down, but it is precisely this enforced muscular activity that enables them to remain awake. Lying down means failure of the experiment. Nothing is more illustrative of the importance of proprioceptive impulses in the maintenance of wakefulness and of muscular relaxation in the precipitation of sleep.

Among other features of behaviour in sleep deprivation are irritability to the point of irascibility in normally even-tempered subjects, and a mental disorganization, leading to dreaming while awake, hallucinations and automatic behaviour, occasionally bordering on temporary insanity. It is easy to understand why the third-degree method of continuous interrogation for many hours will make a person sign a "confession," even if innocent of the crime he is accused of having committed. He wants to be permitted to sleep, and he fails to realize the seriousness of his self-incrimination.

Experiments on animals include inducing lesions in, and stimulation of, different parts of the central nervous system. Dogs surgically deprived of the cerebral cortex still show alternation of sleep and wakefulness. Such decorticated animals lose the greater

part of their learning ability and no longer profit from their previous experience. They would starve to death, in the presence of an abundance of food and water, unless artificially fed. They do not recognize their keeper, and their behaviour in many ways resembles that of very young puppies. Like the latter they go to sleep after they are fed, become restless only when hungry, and their sleep-wakefulness cycle loses its usual relation to night and day.

Nature performs similar experiments on human beings, in the form of anencephalous infants (born without a cerebral cortex) who often remain alive for months or even years. Unlike normal infants, who learn in a few weeks to consolidate several short naps into one long night's sleep, anencephalous ones, as long as they live, sleep at irregular intervals. Their sleep-wakefulness cycle, like that of decorticated dogs, seems to be geared largely to the satisfaction of hunger. Evidently, primitive sleep-wakefulness alternation does not require cerebral cortical activity in either animals or man.

By modern electrocoagulation procedures, it is possible to burn out discrete regions of the brain stem without removing or injuring the cerebral cortex. Extensive damage to the hypothalamus, in the region of the mammillary bodies, in cats and monkeys induces profound somnolence or frank sleep. It would appear that the waking state is maintained by impulses sent from the hypothalamus to the lower centres of the nervous system. Conversely, afferent impulses from the periphery are capable of stirring the wakefulness centre of the hypothalamus into activity. Again, nature's experimentation, in the form of lethargic encephalitis (sleeping sickness), seems to confirm the laboratory findings. In the latter disease the cerebral cortex is undamaged, but there are extensive lesions in the brain stem in the region of the wakefulness centre. Significantly, sleeping sickness patients can at times be aroused briefly, but they are unable to maintain wakefulness. No naturally occurring destruction in the brain stem or elsewhere in the nervous system has ever produced permanent insomnia, militating against the existence of a sleep centre.

A sleep abnormality that points to the importance of afferent impulses from the sense organs in wakefulness is narcolepsy, a disease in which bouts of almost uncontrollable drowsiness may occur several times daily for years. The incidence of these attacks increases when the narcoleptic is alone and is engaged in uninteresting activity, conditions conducive to drowsiness in nearly everyone. Of added interest is that narcolepsy is usually associated with cataplexis, a sudden complete loss of muscle tonus, causing the person to sink to the ground, with consciousness usually preserved. This sudden muscular relaxation, seldom lasting more than a few minutes, is brought on by some emotional stress, particularly laughter. The causes of narcolepsy and cataplexis are unknown, but their occurrence in combination suggests that a defect in the maintenance of tonus in the skeletal musculature and failure, at times, of wakefulness may have a common cause.

Theories of Sleep.—Partial theories pertain to the mechanism of going to sleep and attempt to explain how sleep is brought about; awakening is usually ascribed to a reversal of the changes that produce sleep. General theories deal with the reason for sleep, often without attention to the mode of its onset or termination. Complete theories try to explain both the mechanisms of and reasons for the phenomenon, sometimes by proposing the same scheme for both.

Circulatory theories, of the partial kind, ascribe sleep to a shifting of the blood out of or into an important organ, usually the brain. These theories are obsolete, as the absence of such a shift of blood has been definitely established. Also discarded are the neural theories, postulating a physical break in the chains of nerve cells that serve as conducting pathways in the nervous system. Inhibitory theories implicate widespread inhibition which decreases or abolishes the activity of the cerebral cortex, but they are clearly inadequate when it is considered that decorticated animals and newborn infants (whose cerebral cortex is not yet functioning) can sleep. Deafferentation theories are based on the fact that shutting off incoming sensory impulses to the cerebral cortex is incompatible with wakefulness, but they likewise fail to explain sleep after decortication. Biological theories are general in that they consider sleep an instinct and, as such, a positive act rather than a mere cessation of wakefulness. The vagueness of these theories is at once a strength and weakness; it is impossible either to refute them or to prove their correctness. Humoral or chemical theories stem from established changes in the composition of the blood and urine, modification of metabolic processes in the tissues and a possible variation in the secretory activity of the glands of internal secretion, occurring during sleep as compared with wakefulness. Either some substance required to maintain the waking state is exhausted, and sleep serves to replenish it; or, on the contrary, some toxic substances accumulated in wakefulness are excreted or destroyed in sleep. However, alertness and efficiency of performance are not at their best at the time of getting up in the morning, nor at their worst at bedtime at night.

An evolutionary theory distinguishes between the roles of the hypothalamic centres and the more recently evolved cortical centres in the maintenance of the waking state. According to this theory, sleep as a state of inactivity requires no explanation. Wakefulness of the primitive type, seen in the newborn infant or in decorticated animals, designed to meet the minimum needs for keeping alive and for protoplasmic growth, is a subcortical, probably hypothalamic, function. The higher type of wakefulness, with its attendant alertness, critical reactivity and wider interests, as well as the adaptation of the sleep-wakefulness cycle to the 24-hour alternation of night and day, is a function of the cerebral cortex. Wakefulness maintained by the cortex depends upon its receiving impulses from both the hypothalamic centres and the sense organs, particularly the skeletal muscles.

Hygiene of Sleep.—Although no proposed theory of sleep has met with universal acceptance, enough factual information is available for a rational hygiene of sleep. Acculturation to the family and community pattern of living begins at a very early age. Even in the first few days of life the human infant sleeps more at night than in the daytime. A continuous long period of sleep at night is achieved, with some training, in two or three months. Supplementary daytime naps are reduced to two at six months (by which time a distinct diurnal body temperature curve is in evidence) and to one by the end of the first year; they are given up entirely at the age of three to six years. The belief that very young babies sleep nearly all the time—about 20 or 22 hours out of 24—has been shown to be incorrect. The average total duration of sleep of a newborn infant is only about 14–15 hours and this is reduced to 13–14 hours at the age of six months. Thereafter there is a continuous decline to 7–9 hours of sleep per 24 hours in the adult.

The nightly sleep "ration" is made up of two components: an obligatory need for sleep, which varies with the physical, mental and temperamental characteristics of the individual, and an accessory indulgence in sleep, influenced by age, sex and day-to-day fluctuating environmental factors. A person's irreducible sleep fraction is best expressed by its complement, the maximum sustainable duration of wakefulness. At one extreme of the population distribution are persons capable of remaining alert for more than 18 hours daily, at the other those who succumb to uncontrollable drowsiness in less than 14 hours. Everyone, however, tends to overindulge in sleep in the absence of stimulating events in his immediate environment, simply as an escape from boredom. Thus, in routine uninteresting existence most people sleep longer than is necessary for their well-being. By the purely empirical test of freedom from sleepiness during the customary waking hours, each person can determine his minimum sleep requirement.

Although sleep is in no sense a habit, the adjustment to hours of retiring and getting up, to certain types of bed and bedding and to a host of environmental conditions depends on individual experience and training. The diurnal sleep-wakefulness rhythm, it will be recalled, is cortically conditioned, and by following a regular schedule of hours for various tasks and recreation, meals and sleep, this rhythm is fortified. The reward is twofold: (1) a

more prompt occurrence of sleep at night and spontaneous awakening in the morning, and (2) greater alertness during wakefulness, especially at customary hours of activity.

Insomnia usually involves a difficulty in falling asleep. When it is caused by pain or physical discomfort, a hypnotic drug, prescribed by a physician, may have to be resorted to. Occasionally, grief over past events or worry over possible future misfortunes produces a muscular tension that prevents the onset of sleep. Under these conditions, anything that takes a person's mind off the disturbing thoughts promotes muscular relaxation and is thus conducive to sleep. A series of routine chores, such as grooming and bathing, for an hour or two before going to bed helps induce drowsiness. Reading in bed (if the subject matter is not emotionally charged) or listening to soothing music is soporific for some persons. However, what puts one person to sleep may keep another awake.

A particular difficulty faces those who are engaged in around-the-clock operations of some industries, public utilities, transportation and communication companies, police and fire departments, hospitals and military services. These persons have to work (and therefore also sleep) at unconventional hours and are subject to many inconveniences. The common practice of rotating work shifts at frequent, often weekly, intervals all but prevents the workers from developing a suitable individual sleep-wakefulness rhythm.

The temperamental make-up of the individual also affects his ability to conform to the community pattern of living. Early to bed and early to rise is easy for the "morning" type, whose body temperature and efficiency reach their maxima around noon. A marked "evening" type dislikes getting up at the conventional hour in the morning, depends upon an artificial mode of awakening, does not reach his height of efficiency until late in the afternoon (when the work day is practically over) and hates to go to bed at night. A satisfactory adjustment can be made by nearly everyone, though not with equal ease.

BIBLIOGRAPHY.—H. Piéron, *Le Problème physiologique du sommeil* (1913); N. Kleitman, *Sleep and Wakefulness* (1939; rev. ed., 1963); G. E. W. Wolstenholme and M. O'Connor (eds.), *Ciba Foundation Symposium on the Nature of Sleep* (1961); I. Oswald, *Sleeping and Waking* (1962); E. Bünning, *The Physiological Clock* (1964).
(N. K.)

SLEEPING SICKNESS: *see* TRYPANOSOMIASIS.

SLEEPWALKING: *see* HYPNOSIS; HYSTERIA.

SLEET, in the United States and Canada the popular name for precipitation of transparent, roundish, hard ice particles one to four millimetres in diameter, which in other countries are properly called grains of ice, ice pellets or frozen raindrops. These are often confused with graupel, small hail, soft hail, granular snow and snow pellets, which resemble ice pellets somewhat in size and general appearance but are whitish, softer and less dense. Ice pellets mixed with rain and falling on surfaces having temperatures below freezing result in a pebbly surfaced crust of glaze (clear ice) forming over the surfaces.

Ice pellets are produced by raindrops formed in clouds in an upper warm-air layer becoming frozen while falling through a lower (or surface) air layer having a temperature well below freezing.

In British countries (except Canada) sleet refers to a mixture of rain and snow. *See* HYDROMETEOR. (R. G. SE.)

SLEIDANUS, JOHANNES (1506–1556), German historian, the Protestant annalist of the Reformation, whose account remained a basic work until the archives were opened in the 19th century, was born near Aachen at Schleiden, from which his name was later taken (his father's name was Philippi). From 1533 to 1542 he studied classics and law in France, adopting Protestant opinions about 1541. He was an intermediary in Francis I's futile negotiations for an anti-Habsburg alliance with the princes of the Schmalkaldic League and in 1545 ambassador for those princes in England. From 1551 to 1552 he represented Strasbourg and other imperial cities at the Council of Trent. He died in Strasbourg on Oct. 31, 1556.

Martin Bucer had proposed to Philip of Hesse that Sleidanus be appointed historian of the Reformation, with a salary and access to documents, and the work—*De statu religionis et reipublicae Carolo V Caesare commentarii*, the first account of the Reformation to be based on documents (primarily those of Hesse and Strasbourg), and including many excerpts—appeared in 1555; a 26th book was added posthumously in 1558, bringing events up to 1556 (English translation 1560, 1689). The work pleased nobody: Catholics, then as later, complained of bias, while few of the Protestant princes could welcome so frank an account. Even Melanchthon sighed that Sleidanus had told much that had been better buried in eternal silence.

A convinced believer in the divine nature of the Reformation, Sleidanus nevertheless made a real attempt at objectivity, in spite of a perhaps historically inevitable failure to understand the motives of the papacy and the emperor. His book is the work of a layman versed in imperial law. In the line, though hardly of the stature, of Machiavelli, he views the state as an organism independent of the church and brings out the political and European aspects of the religious conflict. Sleidanus wrote also *De quattuor summis imperiis* (1556; English translation 1563), a textbook of universal history widely used up to the 18th century.

BIBLIOGRAPHY.—*Briefwechsel*, ed. by H. Baumgarten (1881); *see* also H. Baumgarten, *Ueber Sleidans Leben und Briefwechsel* (1878); W. Friedensburg, *Johannes Sleidanus* (1935). (M. KL.)

SLESSOR, SIR JOHN COTESWORTH (1897–), British air marshal, was born at Rhanikhet, India, on June 3, 1897, and educated at Haileybury College, Hertfordshire. In 1915 he was commissioned in the Royal Flying Corps and served as a pilot in France, Egypt, and the Sudan. Appointed to a permanent commission in the Royal Air Force in 1920, he served two years in India and afterward in various home appointments. In 1935–37 he commanded the wing supporting the Indian army in the Waziristan operations, and for this service was awarded the distinguished service order. On his return to England he served as director of the plans branch of the Air Ministry. In 1940, during World War II, Slessor visited the United States on a special mission. In 1942 he took the new and important post of assistant chief of the air staff (policy), and attended the chief Allied conferences. The following year he was appointed air officer commander in chief, Coastal Command, and was able, in close cooperation with the Royal Navy and the U.S. forces, to contribute to the defeat of the German submarines in the Battle of the Atlantic. In 1944 he became deputy air commander in chief of the Allied forces in the Mediterranean and a year later was made air member for personnel on the Air Council. In 1948 he became commandant of the Imperial Defence College before holding the highest of all posts in the RAF—that of chief of the air staff. He retired in 1952.

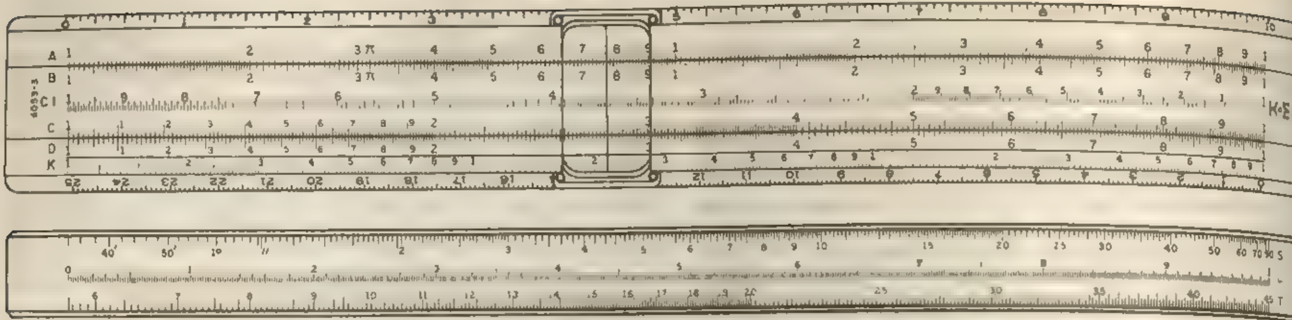
Air Marshal Slessor became well known also for books and articles on air strategy. He was created knight grand cross of the Order of the Bath in 1948. His book *The Central Blue* is largely autobiographical. (E. B. BN.)

SLEZAK, LEO (1873–1946), Czech tenor known for his performance of Wagnerian operatic roles, was born at Sumperk, Moravia, on Aug. 18, 1873. He studied under Adolf Robinson and made his debut at Brno in *Lohengrin* in 1896. In 1908 he studied under Jean de Reske in Paris and in the following year established his reputation as a heroic tenor in the part of Othello in London and New York, his impersonation being helped by his imposing height. He was later esteemed for his interpretations of Wagnerian roles throughout Europe and the U.S. In later years Slezak abandoned singing and became known as a film comedian in Austria. He died at Egern, Bavaria, on June 1, 1946. His son, WALTER (1902–), became a well-known U.S. actor.

See L. Slezak, *Song of Motley* (1938); W. Slezak, *What Time's the Next Swan?* (1962).

SLICK, SAM: *see* HALIBURTON, THOMAS CHANDLER.

SLIDE RULE, a rule consisting of graduated scales capable of relative movement, by means of which simple calculations may be carried out mechanically. In ordinary slide rules these operations include multiplication, division and extraction of square roots, as well as, in some cases, calculation of trigonometrical functions and logarithms. The slide rule has become an essential



BY COURTESY OF KEUFFEL & ESSER CO.

FIG. 1.—MANNHEIM SLIDE RULE

tool in the mathematics of science and engineering and is widely used in business and industry as well.

The logarithmic slide rule is a compact device for rapidly performing calculations with limited accuracy. The invention of logarithms in 1614 by John Napier of Merchiston, Scot., and the computation and publication of tables of logarithms made it possible to effect multiplication and division by the simpler operations of addition and subtraction. (*See* LOGARITHMS.) Napier's early conception of the importance of simplifying mathematical calculations resulted in his invention of logarithms, and this invention made possible the slide rule as we know it. In 1620 Edmund Gunter plotted logarithms on a two-foot straight line. With such scales, multiplication and division were performed by addition and subtraction of lengths by a pair of dividers.

William Oughtred, according to his own statement (1633), constructed and used as early as 1621 two of these Gunter's lines sliding by each other so as to do away with the need for dividers. The lines were used in both the straight and circular forms. In the former the scales were held against one another by the hands; in the latter, dividers were replaced by an "opening index"—really a pair of dividers fixed centrally on the circular scale. Oughtred's two scales were the forerunners of the C and D scales, the basic scales on nearly all slide rules from then on.

The first known slide rule in which the slide worked between parts of a fixed stock was made by Robert Bissaker in 1654. Others were due to the enterprise of Seth Partridge (1657), Henry Coggeshall (1677)—a slide in a two-foot folding rule adapted to timber measure—and Thomas Everard (1683) for gauging purposes. The usefulness of the slide rule for rapid calculation became increasingly recognized, especially in England, during the 18th century, and the instrument was made in considerable numbers, with slight modifications.

Improvements in the direction of increased accuracy in graduation, etc., were initiated by Matthew Boulton and James Watt from about 1779 in connection with calculations in the design of steam engines at their works at Soho, Birmingham.

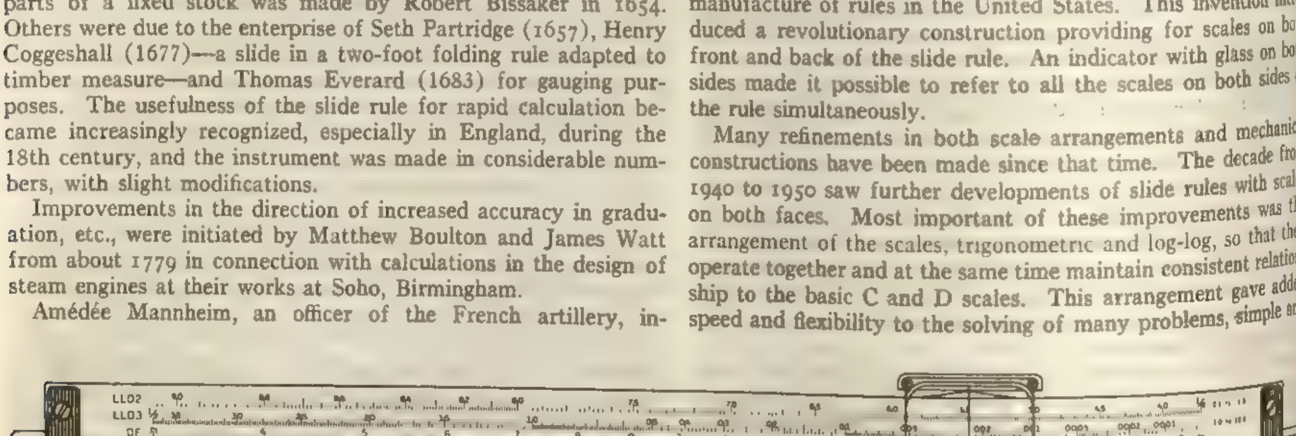
Amédée Mannheim, an officer of the French artillery, invented in 1859 what may be considered the first of the modern slide rules. This rule had scales on one face only and although it was quite simple is basically of a type still made and designated by his name. The disposition of the scales in the Mannheim rule is the arrangement still adopted in the great majority of rules made in the 20th century (*see* fig. 1). This rule, which also brought into general use a cursor, or indicator, was much used in France, and after about 1880 was imported in large numbers into other countries.

Up to this period the rule had been constructed usually of boxwood and occasionally of brass or ivory, but a great improvement was introduced in 1886 by Dennert & Pape in Germany by dividing the scales on white Celluloid, which gave a much greater distinctness in reading. This material was later almost universally adopted, and the slide rule attained a high degree of perfection.

In 1815 Peter M. Roget invented his "log-log" slide rule for performing the involution (*q.v.*) and evolution (*q.v.*) of numbers (*see* fig. 2). The fixed scale, instead of being divided logarithmically, is divided into lengths which are proportional to the logarithm of the logarithm of the numbers indicated on the scale; the sliding scale is divided logarithmically.

Before 1890 slide rules were made only in England, France and Germany, but at that time an invention by William Cox led to the manufacture of rules in the United States. This invention introduced a revolutionary construction providing for scales on both front and back of the slide rule. An indicator with glass on both sides made it possible to refer to all the scales on both sides of the rule simultaneously.

Many refinements in both scale arrangements and mechanical constructions have been made since that time. The decade from 1940 to 1950 saw further developments of slide rules with scales on both faces. Most important of these improvements was the arrangement of the scales, trigonometric and log-log, so that they operate together and at the same time maintain consistent relationship to the basic C and D scales. This arrangement gave added speed and flexibility to the solving of many problems, simple and



COURTESY OF KEUFFEL & ESSER CO.

FIG. 2.—THE LOG-LOG SLIDE RULE: TOP, FRONT FACE; BOTTOM, REVERSE FACE

complex alike, since it produced solutions by continuous operation, without the need of intermediate readings.

See J. N. Arnold, *The Slide Rule; Principles and Applications* (1954).
(A. W. K.)

SLIEMA, a town of Malta and a suburb of Valletta, the capital, lies on a headland that forms the northwesterly approach to Marsamxett (Marsamuscetto) harbour. Pop. (1961 est.) 24,014. The town has two aspects, one facing Valletta southward across Sliema creek with Fort Tigné dominating the entrance and the other looking northward over the Mediterranean. Sliema is perhaps the most English part of Malta, but there is a reminder of other days in the watch tower against corsairs on the northern sea front. A large number, at times a majority, of its inhabitants are British servicemen (both naval and military) and officials.
(W. B. Fr.)

SLIGO (Contae Shligigh), a county of the Republic of Ireland in Connaught Province, covers 693 sq.mi. (1,795 sq.km.). It is varied in scenery and one-third is rough pasture, chiefly in the mountains and hills and in the lowland peat bogs. On the east the boundary with County Leitrim runs south through the coastal lowland, about 5 mi. wide, to the fine upland of the Dartry Mountains (Ben Bulbin, 1,722 ft. [525 m.]), which is made up of limestone and has spectacular scarped edges. The boundary then crosses Glencar and another area of plateau to Lough Gill, a lake surrounded by hills and having rich woods which include the arbutus, or strawberry tree, with oak and birch. The little island of Innisfree, made famous by the lyric of W. B. Yeats, is in Lough Gill. To the south of Lough Gill there is a long ridge broken by gaps called *alts* of which one, at Collooney, carries the main road and railway north to Sligo. West of this gap the ridge, formed of ancient sedimentary and metamorphic rocks, becomes the Ox Mountains, a peat moorland 7 to 10 mi. wide and rising above 1,700 ft. (518 m.) (Knockalongy, 1,778 ft. [542 m.]). These mountains have the northeast–southwest, or Caledonian, trend so prevalent in Ireland.

In the area north of the Ox Mountains there is a continuously farmed lowland, 2 to 5 mi. wide, from the mouth of the River Moy to the Leitrim border. Sligo Bay has three long estuaries leading to Drumcliff, Sligo, and Ballysodare, which receive the waters of the Drumcliff, Garavogue, and Owenmore rivers. There is a long tradition of fishing, but only salmon is profitable now: a limited number of coasting steamers come to the port of Sligo. The lowland is divided into small farms, mainly of 25–30 ac. (10–12 ha.) but formerly much smaller, which depend on cattle, sheep, and poultry. Peat is cut in the neighbouring plateau areas and, especially, the Ox Mountains. Noted tourist resorts are Rosses Point, with a famed golf course, and Strandhill beside the isolated hill of Knocknarea (1,078 ft. [329 m.]). Some of the valleys have farms, but the Ox Mountain ridge makes an almost complete break between the inland and the coastal areas of the county.

Inland from the Ox Mountain ridge, the eastern boundary with Leitrim strikes southeast across the plateaus and then bends southwest through varied country to Lough Arrow, crosses the Curlew Hills (828 ft. [252 m.]) to Lough Gara, and finally runs west through farmed lowlands to the Ox Mountains. The farmed lowlands are broken by the attractive lakes and by limestone hills; they are mostly limestone-floored, covered with drumlins and drift which include some soils of considerable fertility, with good natural drainage. The farms are small, mainly about 30 ac. Both dairy and dry cattle are kept, but many beasts are sold young in the local fairs. In some areas there is daily milk sale to creameries.

Sligo was anciently MacDermott country. After the Anglo-Norman invasion it was dominated by the De Burgh (or De Burgo) power. With the murder in 1333 of the "Brown Earl," William de Burgh, earl of Ulster, the O'Connor Sligo became lord of the town and county under O'Donnell patronage and suzerainty, and the O'Dowds were established in the west. Donal O'Connor Sligo surrendered in 1567 the "captaincy" he had acquired by Irish custom over the territory and received from the queen a grant for life of the whole area. It was created a county in 1579. Forming part of Connaught, which was reserved by Cromwell for Irish proprietors, the district retained its Irish social pattern longer than most, though passing through the big changes brought about by the penal laws and the land acts.

In 1841 the county's population was 180,886. By 1956 it had fallen to 56,850, and in 1961 there was a further 6% decrease to 53,561. A county council meets at Sligo, but the county is joined with Leitrim under one county manager. Sligo, the county town and only urban district (pop. [1961] 13,145), has a mayor and corporation. It also has harbour commissioners. The county is joined with Leitrim to form one constituency which elects five members to the *Dail*. Sligo is the only town of any size in the county, the others being small market centres with fewer than 1,000 people. Ballymote (965) is the biggest market centre and in the Franciscan Friary (now in ruins) the Book of Ballymote was compiled in the 14th century; the book is now in the Royal Irish Academy at Dublin. There are mills at Ballysodare and Collooney and a shoe factory at Tobercurry.

See Terence O'Rorke, *History of Sligo: town and county*, 2 vol. (1889); W. G. Wood-Martin, *History of Sligo . . .*, 3 vol. (1882).
(T. W. Fr.; Hu. S.)

SLIGO (Sligeach) is a seaport and the county town of County Sligo, Republic of Ireland. Pop. (1961) 13,145. It lies on Sligo Bay and the River Garavogue between Lough Gill and the sea, 136 mi. (219 km.) NW of Dublin by road. The Roman Catholic Cathedral (1869–74) serves the diocese of Elphin, and there is also a Church of Ireland Cathedral. Summerhill College stands on a hill overlooking the town.

A castle, built in 1242 by Maurice Fitzgerald, lord of Offaly and justiciar of Ireland, was destroyed in 1270; it was rebuilt in 1310 but was again partly destroyed in 1369 and 1394. Maurice Fitzgerald also founded a Dominican friary (known as Sligo Abbey) in 1252. It was accidentally burned down in 1414 and later rebuilt, but was again destroyed in 1641 when the town was sacked. The ruins of the church consist of nave, choir, south transept, and central tower; and three sides of the cloisters have been preserved. Early in the reign of James I the town received a market and two annual fairs; in 1613 it was incorporated and received the privileges of a borough; and in 1621 it received a charter of the staple. In 1641 it was besieged by the parliamentary forces, but was afterward evacuated and occupied by the royalists until the termination of the war. In 1688 it declared in favour of James II and, after being captured by the Enniskilleners, was retaken (1689) by Gen. Patrick Sarsfield, but ultimately surrendered to the earl of Granard. The borough was disfranchised in 1870.

Sligo takes rank with Galway and Limerick as one of the three principal ports of the west coast of Ireland. Some cattle are exported, and coal, iron, timber, and provisions are imported. The harbour has commodious quays. Sligo is a centre of salmon fishing.

At Carrowmore, 3 mi. SW of the town, is a remarkably large collection of megalithic monuments—cairns, stone circles, and dolmens. On Knocknarea (1,078 ft. [329 m.]), west of Sligo, is a cairn which tradition sets down as the burial place of Queen Mab (Maeve of Connaught).

SLIM, SIR WILLIAM JOSEPH (1891–), British field marshal, best known for his work in the Burma campaign in World War II, was born in Bristol on Aug. 6, 1891, and educated at King Edward's School, Birmingham. In World War I he volunteered as a private, but within a few days became a commissioned officer. He saw active service in Belgium, France, Iraq, and the Dardanelles. In 1920 he was granted a regular commission and transferred to the Indian Army.

In 1940 Slim commanded the 10th infantry brigade of the famous 5th Indian division, which landed in northeast Africa and advanced into Eritrea. He was wounded, but recovered in time to lead the 10th Indian division in Iraq and in Iran, where he decisively defeated the enemy and made the first contact with the Russian Army at Teheran. For this service he was awarded the Distinguished Service Order. In 1942 he commanded the 1st Burma Corps and assisted in conducting an orderly retreat in Burma. Later, as commander in chief of the 14th Army, he inflicted a crushing defeat on the Japanese; after repulsing the enemy at Imphal and Kohima and capturing Mandalay, he drove the Japanese down the Irrawaddy River to Rangoon. In 1945 he became

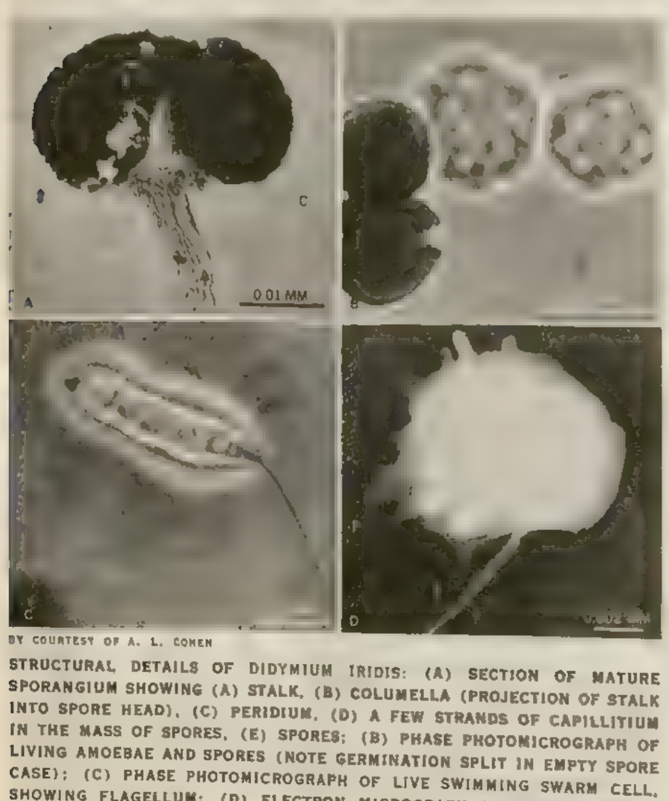
commander in chief of Allied land forces in southeast Asia. In 1946 he was appointed commandant of the Imperial Defence College. He became chief of the Imperial General Staff in 1948 and was promoted to field marshal. In 1953 Slim was appointed governor general of Australia. In 1956 he published *Defeat Into Victory*, which gave a remarkable picture of the Burma campaign.

(E. B. Bn.)

SLIME MOLDS (Slime Fungi). Science fiction did not invent the slime molds, but it has borrowed from them in using the idea of sheets of liquid, flowing protoplasm, giant voracious amoebae, engulfing and dissolving every living thing they touch. What fiction could only imagine, nature has evolved, and only their sharp dependence on coolness, moisture and darkness has kept the slime molds from ordinary observation, for they are common enough.

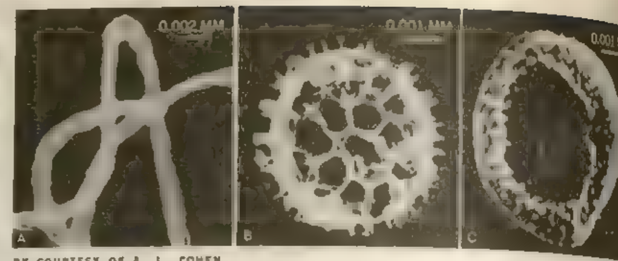
Coolness, moisture and darkness are found in the soil and in logs, twigs and leaves decaying on the forest floor; there also are found the slime molds. An old log turned over—preferably a few days after a warm summer rain—may reveal on its underside a white or yellowish fan of almost liquid consistency, a sheet of slowly flowing protoplasm. The drier upper side of the log may be covered with what look like pinhead-sized toadstools that when touched give off clouds of dust-fine spores. The creeping, feeding fan, the plasmodium, or vegetative phase of a slime mold, has the characteristics of a primitive animal, a large terrestrial amoeba. The reproductive stage, or sporangium, with its dusty spores, has the characteristics of a lower plant, a mold. This double life is reflected in the equivocal scientific synonyms for slime molds, derived from the Greek: Myxomycetes, meaning "slime mold" or "slime fungi"; Myxomycophyta, the "-phyta" merely stressing the plantlike aspects; Mycetozoa, meaning "fungus animals." The last synonym was coined in 1858 by the great German botanist H. A. De Bary, when their dual role as plants and animals was elucidated.

The term slime mold embraces a heterogenous assemblage of organisms whose juxtaposition reflects a historic confusion between superficial resemblances and actual relationships. In this article the term Myxomycetes covers those forms with a plasmodial stage and definite fruiting bodies (true slime molds or



BY COURTESY OF A. L. COHEN

STRUCTURAL DETAILS OF DIDYMIUM IRIDIS: (A) SECTION OF MATURE SPORANGIUM SHOWING (A) STALK, (B) COLUMELLA (PROJECTION OF STALK INTO SPORE HEAD), (C) PERIDIUM, (D) A FEW STRANDS OF CAPILLITIUM IN THE MASS OF SPORES, (E) SPORES; (B) PHASE PHOTOMICROGRAPH OF LIVING AMOEBAE AND SPORES (NOTE GERMINATION SPLIT IN EMPTY SPORE CASE); (C) PHASE PHOTOMICROGRAPH OF LIVE SWIMMING SWARM CELL, SHOWING FLAGELLUM; (D) ELECTRON MICROGRAPH OF SWARM CELL IN TRANSITION BETWEEN AMOEBOID AND FLAGELLATE PHASE SHOWING (F) FLAGELLUM, (S) SECOND FLAGELLUM AND (P) PSEUDOPODIA



BY COURTESY OF A. L. COHEN

SCULPTURING OF CAPILLITIUM AND SPORES: (A) CAPILLITIUM OF HEMITRICHIA CLAVATA SHOWING SPIRAL MARKINGS; (B) RETICULATED SURFACE OF SPORE OF RETICULARIA LYCOPERDON; (C) EMPTY SPORE CASE OF STEMONITIS SPLENDENS SHOWING GERMINATION PORE AND SHORT SPINES

Mycetozoa in the strict sense); the term Acrasieae includes those forms in which individual amoebae aggregate without fusing to produce a fruiting structure (simpler slime molds, cellular slime molds); and the term Plasmodiophoreae covers a group of root parasites of higher plants having a plasmodial stage but lacking definite fruiting bodies. Other organisms, in most cases imperfectly known, have been included among the slime molds, but their treatment and placement are still uncertain.

TRUE SLIME MOLDS (MYXOMYCETES)

Sporangium.—Although some Myxomycetes form sporangia from plasmodia that creep over the surface of wood and leaves (*e.g.*, *Didymium*), others such as the common and beautiful *Stemonitis* and *Arcyria* form the fruits from plasmodia buried within the decaying wood. There is no indication that a log or tree stump is inhabited until some drops of creamy liquid appear on the surface—as if the dead wood were "sweating" drops of living protoplasm.

This is the first sign that the plasmodium, whose strands are thrown throughout the porous wood, is finishing its days as an active, growing, feeding network. These droplets coalesce and grow, expanding to a thick cushion that eventually becomes a column, perhaps half an inch wide and an inch high. A purplish tinge, gradually darkening to black, spreads over the column as it begins to take shape as a bundle of cylinders, each supported by the stalk that it has formed. The bundle, now almost black, dries out, and each sporangium, densely filled with dark spores, separates from its neighbours. The tenuous wall enclosing the sporangium, the peridium, disintegrates on drying. A touch of the hand or a puff of wind disperses the spores like smoke. Finally, only the delicate skeletons of the sporangia are left, each consisting of its stalk with an intricate network of branches, the capillitium, which held the spores. This whole feathery tuft of sporangia, resting on a silvery sheet of dried slime, the hypothallus, has formed and fulfilled its mission of reproduction within the span of a single day.

The startling transformation of a shapeless, ever-changing plasmodium into an intricately organized fruiting structure is the remarkable feature of the Myxomycetes. Not all follow the same pattern as exemplified by *Stemonitis*.

The capillitium of different species is formed variously, but a composite picture of its formation may be given as follows. Within the mass of protoplasm enclosed by the peridium, lines appear that later become chains of beads of clear droplets, or vacuoles.

The vacuoles coalesce to form hollow tubes whose walls become coated with dense material. Some capillitia are what the name implies, fine hairlike processes. Some have the walls beautifully sculptured with spirals and cogs, as in the genera *Trichia*, *Hemitrichia* and *Arcyria*, in which, on drying, the capillitium expands into a fluffy mass scattering the spores as it does so. In others the capillitium is permeated with lime.

Not all sporangia are stalked. Some are sessile, the plasmodium breaking up into droplets that invest themselves with a sporangial wall and complete the development without ever rising from the substrate. In the plasmodiocarps the plasmodium hardly breaks up at all, but the profound interior change of liquid protoplasm to dusty spores belies the exterior shape. The largest sporangia are

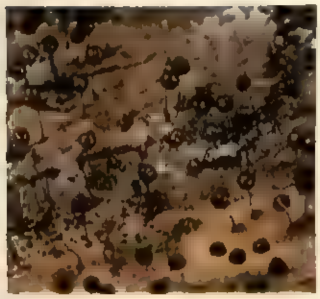
aethalia, compound sporan[illegible]gia composed of individual sporangia incompletely separated from each other. In the ubiquitous *Fuligo* and in *Brefeldia* they for[illegible] great cushiony masses, those of *Brefeldia maxima* sometimes [illegible]aching the remarkable length of one foot. The aethalia may [illegible]nge from the easily delineated individual sporangia of *Tubifera* thr[illegible]ugh the tortuous interwoven plasmodiocarps of *Ceratiomyxa* t[illegible] spherical fruit of *Lycogala*, in which individual sporangia are [illegible] reduced that the sphere is practically a single entity.

Although many invest[illegible]rs have recognized the potential research value of organism[illegible]t have the two aspects of growth—increase in mass and d[illegible]pment—separated in time, few have undertaken a study of [illegible]hysiology underlying morphogenesis. In 1938 W. D. Gray [illegible]ed that fruit formation of pigmented plasmodia occurred onl[illegible]en stimulated by light. Subsequent papers by a number of w[illegible]s have confirmed and extended these findings; for example [illegible] Fergus and R. D. Schein showed in 1963 that although no [illegible]res of *Physarum gyrosum* fruited in complete darkness, as li[illegible] one footcandle of light for 180 hours allowed fruiting, with [illegible]ime required for higher intensities. Since 1955 evidence ha[illegible]n accumulating that there is a more or less profound shift i[illegible]mount and in the type of metabolism as the plasmodium goes [illegible]ruiting.

Spores and Swarm C[illegible]s.—The spores are formed in the sporangium by successive [illegible]ing, or cleavage, of the protoplasm; these furrows redivide [illegible]s until each little block contains but one nucleus. Each blo[illegible]nds up, surrounds itself with a dense spore wall and shrinks [illegible]y from its fellows as it dries. The spores of Myxomycet[illegible]y vary slightly in shape from oval to round, and in size fr[illegible] $\frac{4}{1,000}$ to $\frac{15}{1,000}$ of a millimetre in diameter. The colour m[illegible]e black, gray, brown, brilliant red or yellow, rarely white [illegible] surface is often covered with ridges or spines.

If the spores are pla[illegible]n water, they swell, and in a half hour to a week, depending [illegible]es, and on factors such as age of the specimen, the spore w[illegible]her splits open or a pore is dissolved in it, and a minute, nak[illegible]ss of protoplasm slips out. This little globule, the swarm cell [illegible]ts out a whiplike process, a flagellum, becomes pear-shaped an[illegible]th the flagellum lashing in front of it, swims away. The swar[illegible]ll then may suddenly come to rest on any surface, put forth p[illegible]dopods and creep along. In this stage it can feed as any amoe[illegible]oes, by engulfing bacteria. The flagellum may be retracted and the myxamoeba is then practically indistinguishable from any o[illegible]e many species of small amoebae that teem in the soil. Or it m[illegible]y put forth its flagellum and suddenly take off, the myxamoeb[illegible] transforming again into the swarming stage.

This discussion has assumed thus far that the swarm cell is monoflagellate, and indeed one flagellum is all that is usually seen. However, in many species careful examination shows a shorter, second flagellum, pressed against the cell body and thus normally hidden from view. While such minutiae may appear unimportant, they provide clues that often indicate profound evolutionary relationships. The swarm cell is a sex cell, a gamete, which fuses with another swarm cell to initiate the plasmodium. In only relatively few species has this fusion been observed in the laboratory, even when germination is abundant, and in even fewer have plasmodia been grown successfully from spores.

ARTHUR L. COHEN

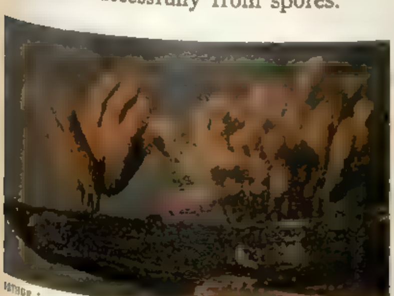
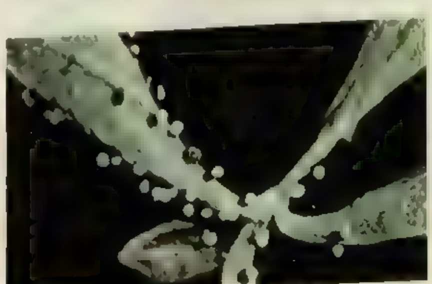
(LEFT) DELICATE FRUITING BODIES OF LAMPRODERMA IN THE ACT OF FORMING; (RIGHT) THE DARK MATURE SPORANGIA

Are the gametes alike (isogametes), so that any gamete may fuse with any other gamete, or are they of different types (heterogametes), so that what may be designated as a + gamete can fuse only with a − gamete? The discussion, particularly sharp in the 1930s, has been opened again and it appears that *Didymium iridis* is heterothallic (producing heterogametes) whereas *Fuligo cinerea* is homothallic (producing isogametes). It is possible that there are strain differences within the species, which would explain the flatly contradictory results of earlier workers, who, while working with the same species, may have been unknowingly comparing hetero- and homothallic strains.

Plasmodium.—In the primitive sexual act wherein two identical appearing swarmers fuse, the nuclei also fuse. The flagella are permanently retracted, and the amoeboid zygote (fertilized cell) begins to grow. Ordinarily when a cell has grown to a certain size it divides into two. The myxomycete zygote departs from this behaviour in that only the nucleus divides repeatedly: the cytoplasm, which remains together, becomes by definition a plasmodium, a multinucleate protoplasmic mass. As the plasmodium flows over the damp undersurface of logs or within the rotting wood itself, over leaves and, in some species, through the soil, it feeds on bacteria, molds and sometimes large fungi. What it cannot digest by engulfing, the plasmodium digests by enveloping with its slimy film.

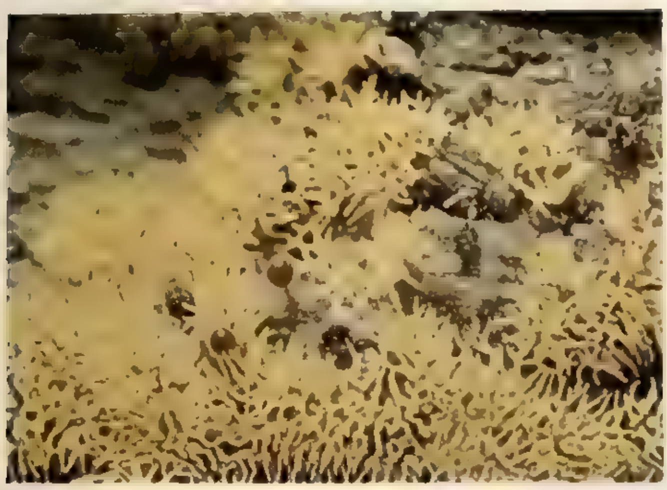
On a smooth surface a vigorous plasmodium generally takes the form of a fan with trailing veins, creeping in the direction of the broadly curved front edge. Under the microscope the fan itself is like a broad delta, with a network of channels in which the protoplasm streams now slowly, now so rapidly that the particles are but blurred streaks. These channels gradually separate into distinct veins from front to rear. The streaming is not all in one direction. In any vein or channel, the protoplasm slows, pauses and then often reverses direction. Generally a little more flows toward the front than flows back, advancing like an incoming tide creeping up a beach.

While the description above is that of the "classical" plasmodium, other types have been observed. As classified by C. J. Alexopoulos in 1962 they are: the protoplasmodium—minute, sluggishly amoeboid, without distinct veins, each plasmodium giving rise to a single sporangium (example, *Echinostelium minutum*); the aphanoplasmodium—colourless, consisting of fine veins forming a network, or sometimes lying parallel to each other (the aphanoplasmodium, typical of some of the Stemonitales, lacks the fan structure described before; it is almost invisible and may be


ARTHUR L. COHEN

EXAMPLES OF MYXOMYCETE FRUITS: (left) *Arcyria incarnata*, after an active life in dead wood, emerges to form fruits; (centre) *Badhamia macrocarpa* erupting on plant debris; (right) *Tubifera ferruginea*, with individual sporangia crowded together, illustrates the transition from solitary fruits to the aethalium



ARTHUR L. COHEN

(ABOVE) TWO FORMS OF CERATIOMYXA GROWING TOGETHER: SPONGELIKE PORIOIDES AND BRANCHED CORALLOIDES; (BELOW) YOUNG FRUITING COLONY OF CERATIOMYXA FRUTICULOSA. NOTE YOUNG BROWN FRUITING BODIES OF ANOTHER MYXOMYCETE AMONG THE YELLOW CERATIOMYXA BRANCHES

of fairly frequent occurrence.); the phaneroplasmodium—with a fan, usually more or less opaque, often coloured; in other words the typical plasmodium. (To these the author would tentatively add a fourth type, a sluglike or limacine plasmodium, as in *Cribraria,* in which the plasmodia creep about like minute black slugs, apparently never forming a fan or network.)

If the plasmodium is jarred or cut, flow ceases and the protoplasm sets to a jelly; then hesitantly, but with increasing vigour, flow resumes. If a strand of a vigorous plasmodium is cut, a drop of the liquid protoplasm oozes out and usually immediately sets to a gel. This droplet, if gently detached and placed on a clean surface, may, in the course of a few hours, become a miniature plasmodium.

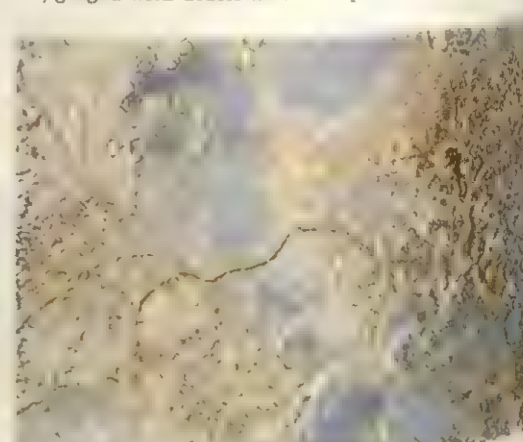
While many cells show this mysterious streaming of the liquid living substance, only the myxomycete plasmodium shows it on such a large scale. The theory has long been held that the streaming is due to contraction of the walls, which squeeze the inner liquid plasm from one part to another.

In an apparent confirmation of this muscular action of the more solid parts there has been isolated from the protoplasm of *Physarum polycephalum* a protein called myxomyosin, which resembles in its properties myosin, the contractile protein of vertebrate muscle. By 1962 choline esterase, an enzyme intimately concerned in the propagation of nerve impulses and in muscular contraction, was isolated from plasmodia of the same species. This finding is one more example of the essential biochemical unity of life—the molecular machinery that commands and drives the artist's hand is similar to that which sends the sheets and veins of myxomycete protoplasm pulsing through the disintegrating litter of the forest floor. In 1939 it was shown that a number of slime molds could thrive on laboratory media whose chief constituents were killed bacteria or yeasts, or extracts of microorganisms, although in nature slime molds presumably feed on living microorganisms as well as dead vegetable matter. While such a medium is [illegible] it is not completely known chemically. The next step [illegible] completed until the early 1960s, when J. W. Daniel, H. [illegible] h and their colleagues reported the continuous culture of [illegible] n of *Physarum polycephalum* plasmodia on completely [illegible] media. The medium for optimal growth contains a total of nearly forty constituents including minerals, amino acids and [illegible]. The majority of Myxomycetes have not been successfully cultivated in the laboratory at all; most of the rest have been [illegible] ted on complex vegetable media with other organisms pr[illegible] few have been kept in pure culture on similar media of unk[illegible] composition.

Oriented responses (tropisms) are [illegible] shown by plasmodia. Plasmodia will, as stated previously [illegible] creep off a stale substrate to a fresh one (chemotropism) [illegible] will move away from light (negative phototropism), except [illegible] they are preparing to fruit, in which case they seek the ne[illegible] posed region, probably attracted by light (positive phototrop[illegible]. The classical illustration of their behaviour is positive rhe[illegible] m—a plasmodium will creep upstream against gently flowin[illegible]

The pigments of most plasmodia [illegible] ntained in granules in an otherwise colourless protoplasm. [illegible] dia of different species may be rarely blue, black or green; [illegible] quently white, colourless, cream, maroon or brown; an[illegible] frequently yellow or orange. Colour often deepens with [illegible] d may be varied by a change in diet. Thus the plasmod[illegible] *Fuligo septica,* usually lemon-yellow in nature, may becom[illegible] on certain laboratory media and remain yellow on other[illegible] hese yellow and white plasmodia of the same parent source [illegible] nger fuse when brought into contact, but each can merge wit[illegible] plasmodium of its own colour. The same interesting phenom[illegible] n has been observed in other plasmodia, but if there is littl[illegible] our difference the independence of each is difficult to see a[illegible] ir veins cross in a confusing tangle.

The plasmodium is not an individual in the full sense. It behaves as a unit, but it can be cut into small pieces that may either fuse together again or grow into new plasmodia. The fluid and protean nature of the plasmodium was arrestingly shown by Lister's classical experiment in the late nineteenth century. A plasmodium, gorged with debris and food particles, was placed on



ARTHUR L. COHEN

(LEFT) PLASMODIUM OF FULIGO, WHICH HAS CREPT OUT OF A ROTTED WOOD CHIP; (RIGHT) ENLARGEMENT SHOWING THE DELICATE VEINLIKE CHARACTER OF THE FLOWING PLASMODIUM

moist cotton. It strained itself through the fibres, leaving the debris behind, and reconstituted itself, thoroughly scoured, on the other side of the barrier.

Sclerotium.—If a plasmodium is rapidly dried, it becomes a dull, lifeless, horny mass. But if it is given some time to dry, perhaps twenty-four hours, then it contracts, becomes somewhat waxy and usually retains its colour, having thereby transformed into a sclerotium. Under the microscope it can be seen that this sclerotium is composed of many cells, each containing a number of nuclei, with each cell separated from its neighbours by only the thinnest of walls. If the sclerotium is moistened, within a few hours the walls dissolve, and what earlier looked like a splash of dry, orange paint crawls off and resumes activity. Those plasmodia that form sclerotia easily may be filed away in envelopes for a year or more, to come alive again when dampened.

Habitat.—*Stemonitis, Lycogala, Tubifera, Ceratiomyxa* and a host of other genera are practically always found on wood, while many of the lime-bearing types, such as species of *Physarum, Mucilago* and *Badhamia,* are found on leaves and straw. *Physarum cinereum* typically lives in the soil, and sometimes its gray fruits and dark spores cover lawns overnight with a cindery, sooty film that just as quickly disappears with the next wind or rain. Only *Fuligo septica* can throw up its formless yellow or brown aethalia anywhere—on a tree trunk six feet above ground, in the crack of a concrete pavement, on desert sand and, as once recorded, on a whale's skull set to bleach on the grounds of the British Museum.

Classification and Relationships.—The Myxomycetes include approximately 60 genera and 500 species. The various orders, families, genera and species are distinguished from each other in rough order of importance by: spore colour, presence of a capillitium, presence of lime in the peridium or capillitium and size and detail of fruit and spore structure.

Confusion in the nomenclature reflects an earlier misunderstanding of the nature of the Myxomycetes. While it is true that the sporangia of such forms as *Lycogala* resemble puffballs, the differences in structures and the life cycles show that the primitive Myxomycetes and the advanced puffballs are entirely different. The myxomycete sporangium is formed from an undifferentiated, naked, fluid mass of protoplasm; the sporangia of the higher fungi are formed from the interwoven filaments, the characteristic hyphae. The myxomycete sporangium wall is a secretion around the spores; the fungal sporangium grows as a cell or mass of cells. Although some of the lower true fungi have motile swarm cells or gametes, these sooner or later transform into the typical hyphae, forming a feltlike vegetative structure, the mycelium. The myxomycete vegetative structure is the naked, amoeboid plasmodium.

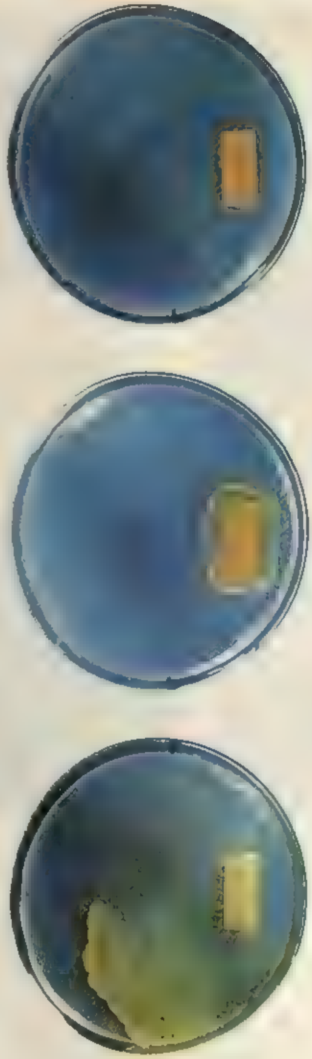
Classification of the Myxomycetes as plants or animals depends on whether the categorizer places more emphasis on the animal-like creeping, feeding plasmodia and myxamoebae, or whether he stresses more the funguslike fruiting body and the dry spores.

CELLULAR SLIME MOLDS (ACRASIEAE)

In 1869 Brefeld described a remarkable organism, *Dictyostelium mucoroides,* which to the naked eye looks very much like a common mold, a mass of spores set on a slender stalk. The basic structure and the mode of formation, however, are completely different from that of any mold, in fact, different from that of any other group of organisms known.

On a suitable substrate the spore germinates, and from the split spore case there emerges a small amoeba. For a time each amoeba feeds by engulfing bacteria, and then divides into two. These amoebae increase prodigiously in numbers under proper conditions and then begin to converge. Until this time of aggregation they are virtually indistinguishable from the many other amoebae that swarm in the soil.

If a plate of agar very dilute in nutrients is streaked with suitable bacteria such as the ordinary *Escherichia coli,* and a few of the spores of *D. mucoroides* are placed on the plate, the spores germinate, each yielding a very small amoeba. These amoebae feed on the bacteria, and divide again until the plate is practically divested of bacteria, the amoebae forming almost a complete covering. The amoebae then begin to congregate, climbing up on each other as they form a little mound. The centre pile becomes a hemisphere with a nipple, like half a lemon set on end, and in its interior a sheath forms around a core of amoebae each of which fills with a fluid, swells, secretes stiff cell walls, dies and becomes in structure and function a central pith. This cylinder of sacrificed amoebae forms the stalk on which the remaining amoebae climb. The stalk becomes narrower until finally it is a chain of single cells; the survivors of the ascension transform into oval spores, held in a drop of mucus.



ARTHUR L. COHEN

GERMINATING SCLEROTIUM OF BADHAMIA UTRICULARIS, PHOTOGRAPHED OVER A 9-HOUR PERIOD

Dormant orange sclerotium, daubed on a piece of absorbent paper and placed in a moist petri dish, within 4 hours (top) has absorbed water and begun to show activity; 1 hour later (centre) the fully active plasmodium has begun to spread out from the paper; (bottom) after 4 more hours the bulk of the plasmodium has left the paper and formed the typical veined plasmodium fan

In 1935 another species of *Dictyostelium* (*D. discoideum*) was described. This remarkable organism has one further stage of the welding of separate living entities into one composite organism. When the myxamoebae have aggregated to the conical stage with its little nipple at the top, the whole pyramid falls over on its side, and with the nipple (*papilla*) raised and leading the way, the "slug" crawls over the substrate toward the light before continuing further development.

It was found that if the myxamoebae were fed the brilliant red bacterium *Serratia marcescens,* they retained the indigestible pigment and formed pink aggregating and migration stages. When pieces of pink migrants were transferred to the normal white forms, they retained their identity. This knowledge was utilized in studying the simpler slime molds. If the papilla is cut off, the little slug ceases migrating at once; when a new papilla is grafted on, migration resumes. The discoid foot that gives the name to this species forms from the papilla; the mid-region forms the stalk and spores.

These little Acrasieae, repeating their life cycle in the laboratory, quickly, easily, dependably, without the vexatious errant ways of the Myxomycetes, have become popular experimental organisms for the study of morphogenesis.

Maurice Sussman and his co-workers have produced evidence that aggregation is started by "initiator" cells around which "responder" cells rally. The initiator amoebae, as seen in time-lapse cinemicrographs, are larger and more active than the more numerous responder cells. They are presumably the primary producers of an attractant called acrasin. In a series of experiments Sussman has demonstrated that mutant amoebae, whose lack of initiators allows them a chronic freedom, may be forced to aggregate by providing them with initiators from another source. The more initiators, the more fruits there are from a given number of amoebae.

Whether or not sexual reproduction exists in the Acrasieae has not been answered. Particularly in the late 1950s and the early 1960s there has been sharp written and oral discussion. The findings and arguments were judiciously reviewed in 1963 by L. S. Olive who finds no very convincing evidence in either direction.

ARTHUR L. COHEN

(LEFT) EMERGING RED PLASMODIUM OF TUBIFERA FERRUGINOSA: WITHIN HOURS THE RED WILL TURN TO UMBER AS THE PLASMODIUM MATURES; (ABOVE) COMPOUND FRUITING BODY OF FULIGO SEPTICA ON DEAD LOG; (BELOW) PLASMODIUM OF MUCILAGO SPONGIOSA ON BLADES OF GRASS

There are several different genera and species of Acrasieae, which differ strikingly in colour and shape. If the amoebae of these different forms are mixed, each sort separates itself out and forms its own fruiting structures. But what happens when they have started to aggregate? If the early aggregations, say, of the purple, branched *Polysphondylium violaceum* and of the white, single-stalked *Dictyostelium discoideum* are mixed, again the amoebae sort themselves out and build their own pure sporangia. If the aggregations have continued almost to the point of differentiation this sorting out no longer occurs; a hybrid fruiting structure is produced. But when the spores are sown, each produces amoebae after its own kind; the mixed decks unshuffle themselves.

The Acrasieae, first thought to be inhabitants of dung, are now known, with possible exceptions noted below, to be inhabitants of the soil. Isolation methods developed in the 1930–40s indicate that they are common, being numerous in species and numbers in forest soil, but also occurring in garden, lawn and field soils.

Despite the fact that the Acrasieae as a group can be carried through their life cycle much more easily than the Myxomycetes, knowledge of their nutritional requirements still lags behind that of the Myxomycetes. A number of species have been cultured on media containing killed microorganisms, and in 1963, H. Hohl and K. B. Raper reported the cultivation of two strains of *Polysphondylium pallidum* on an exceedingly complex medium similar to the one used for the culture of the myxomycete *Physarum polycephalum.*

Unlike the Myxomycetes, the Acrasieae consist of relatively few forms. The vast majority isolated from soil are species of either *Dictyostelium* or the closely related *Polysphondylium.* In 1884 P. E. L. van Tieghem described a remarkable organism consisting of stalks that supported a single row of spores rather than a spherical head, and which he called *Acrasis granulate,* after which the Acrasieae are named. The organism has apparently not been reisolated since that time, and the Acrasieae have gone without a type genus until 1960, when L. S. Olive and C. Stoianovitch reported the discovery of another species of *Acrasis, A. rosea.*

There are a few genera, such as *Guttulina* and *Guttulinopsis,* that fruit on dung of various animals, in which separation into stalk and spore head is highly imperfect, the spore cells and stalk cells shading into each other. It is tempting to construct an evolutionary scale from the probably accidental heaps of free-living soil amoebae through *Guttulina* and *Acrasis* and finally to *Dictyostelium* and *Polysphondylium,* each genus showing increased differentiation of cell types. But until more information is gleaned about these fascinating organisms, the lines of descent can only be suggested.

PLASMODIOPHOREAE

This group, all of which are root parasites of higher plants have only two economically important members, *Plasmodiophora brassicae,* the cause of club root of cabbage and related plants; and *Spongospora subterranea,* the agent of sweet potato powdery scab. Their relation to the Myxomycetes is indicated by the presence of flagellated swarm-spores that transform into myxamoebae and one, possibly two, plasmodial stages. They differ from the Myxomycetes in the absence of a well-developed sporangium with peridium, capillitium, etc., and in their strict dependence upon parasitic life.

No member of the Plasmodiophoreae has been carried indisputably through its life cycle in the laboratory, but a composite cycle may be made by piecing together the processes that have been observed. Upon the decay of the host tissue, spores are released into the soil and produce flagellate swarm cells, which invade root hairs and there transform into small plasmodia. By a complex process these plasmodia produce the gametes that fuse. The resulting amoeboid zygote penetrates the root itself, there to grow into a plasmodium that stimulates the characteristic tumourlike growth of the host cells. The plasmodia divide up into spores that are released on the death of the host tissues.

BIBLIOGRAPHY.—*General:* C. J. Alexopoulos, "The Myxomycetes, II," *Bot. Rev.,* 29:1–78 (1963), *Introductory Mycology,* 2nd ed. (1962); J. T. Bonner, *The Cellular Slime Molds* (1959); W. Crowder, "Marvels of Mycetozoa," *Nat. Geogr. Mag.,* 49:421–443 (1926); R. Hagelstein, *Mycetozoa of North America* (1944); N. Kamiya, *Protoplasmic Streaming* (1959); J. S. Karling, *The Plasmodiophorales* (1942); T. H. Macbride and G. W. Martin, *The Myxomycetes* (1934); G. W. Martin, "Myxomycetes," *N. Amer. Flora,* vol. 1, part 1 (1949), "The Myxomycetes," *Bot. Rev.,* 6:356–388 (1940); K. B. Raper, "The Environment and Morphogenesis in Cellular Slime Molds," *The Harvey Lectures,* Series 57 (1963).

Special: C. J. Alexopoulos, "Gross Morphology of the Plasmodium and Its Possible Significance in the Relationships Among the Myxomycetes," *Mycologia,* 52:1–20 (1959); O. R. Collins, "Heterothallism and Homothallism in Two Myxomycetes," *Amer. J. Bot.,* 48:674–683 (1961); J. W. Daniel *et al.,* "Organic Requirements and Synthetic Media for Growth of the Myxomycete *Physarum polycephalum,*" *J. Bact.,* 86:324–331 (1963); C. L. Fergus and R. D. Schein, "Light Effects on Fruiting of *Physarum gyrosum,*" *Mycologia,* 55:540–548 (1963); H. Hohl and K. Raper, "Nutrition of Cellular Slime Molds. III. Specific Growth Requirements of *Polysphondylium pallidum,*" *J. Bact.,* 86:1314–20 (1963); G. W. Martin, "The Systematic Position of the Myxomycetes," *Mycologia,* 52:119–129 (1960); H. Nakajima and S. Hatano, "Acetylcholinesterase in the Plasmodium of the Myxomycete *Physarum polycephalum,*" *J. Cell Comp. Physiol.,* 59:259–263 (1962); L. S. Olive, "The Question of Sexuality in Cellular Slime Molds," *Bull. Torrey Bot. Cl.,* 90:144–153 (1963); J. C. Sobels and A. L. Cohen, "The Isolation and Culture of Opsimorphic Organism II. Notes on Isolation, Purification, and Maintenance of Myxomycete Plasmodia," *Ann. N. Y. Acad. Sci.,* vol. 56, art. 5, pp. 944–948 (1953); M. Sussman and R. R. Sussman, "Aggregative Performance," *Experimental Cell Research,* Suppl. 8, pp. 91–106 (1961).

(A. L. Cn.)

SLING, an implement for propelling missiles, probably the earliest device by which force and range were given to the arm of a thrower. It consisted of a small strap or socket of leather

to which two cords were attached; the slinger held the ends of the cords in one hand, whirled the socket and missile rapidly round his head, and, loosing one cord sharply, dispatched the missile. In another type the sling was attached to a short staff that was held in both hands. This type was used for heavier missiles, especially in siege operations during the Middle Ages.

There are many references to slings and to slingers in the Bible; the left-handed slingers of Benjamin were famous (Judg. 20:16). Assyrian monuments show the sling of the ordinary type; slings were also used in the ancient Egyptian army, but not before the 8th century B.C. The sli[illegible] not mentioned in Homer; Herodotus (vii, 158) speaks of the s[illegible]gers in the army offered by Gelon to serve against the Persians. The sling seems to have been a weapon chiefly used by barbarian troops; the Acarnanians, however, were expert slingers (Thucydid[illegible] i, 81), and so also were the Achaeans, who later invented the s[illegible] that discharged a shaft with an iron bolt head (Livy xlii, 65, [illegible]m Polybius). In the Roman army by the time of the Punic W[illegible] the slingers (funditores) were auxiliaries from Greece, Syria [illegible] Africa. The Balearic islanders, who were in Hannibal's arm[illegible] were always famous as slingers. In medieval times the sling [illegible] much used in the Frankish army, especially in defending trenc[illegible] while the staff sling was used against fortifications in the 14t[illegible] century. Until the 17th century, slings were used to throw gre[illegible]

In modern times the w[illegible] has taken on other meanings. It denotes a hanging loop to support a wounded arm, a chain with hooks for raising or lowering he[illegible]y goods, or a shoulder strap for carrying a rifle or carbine. A variation of the ancient sling as a hand weapon is the slingshot, a [illegible]ked stick with an elastic band attached for hurling small pellets.

SLIPHER, VESTO MELVIN (1875–   ), U.S. astronomer whose discovery of the rotation and extraordinary space velocities of the spiral nebulae formed the first observational evidence for the expanding universe theory. He was born on a farm in Clinton County, Ind., Nov. 11, 1875, graduated from Indiana University, Bloomington [illegible] 1901, and received his Ph.D. degree there in 1909. For about [illegible]0 years he directed the Lowell Observatory at Flagstaff, Ariz., where he organized and guided the search which resulted in the di[illegible]very of the outermost planet, Pluto, predicted by the founder of the observatory, Percival Lowell.

Slipher made extensive investigations by methods of spectroscopy which resulted in the determination of the rotation periods of the planets; his discovery of molecular bands in the planetary spectrum ultimately led to the identification of their atmospheric composition. He demonstrated that many diffuse nebulae shine by the reflected light of the nearby stars. He also discovered the permanent sky and auroralike radiations of the night sky, their nocturnal intensity changes and the existence of interstellar sodium and calcium scattered throughout the depths of space.

(H. L. G.)

SLIPWAY, the waterfront space in a shipyard allotted to the building of a ship. It is fitted with foundations for keelblocks, shoring, and inclined launching ways, and with cranes, which may be of many types, for handling materials. When a gate or side walls are provided to exclude the water from the lower or outboard end of the slipway, it is sometimes called a semisubmerged slipway. The term slipway is also used for an inclined railway extending into the water and fitted with a cradle, on wheels or rollers, on which a vessel is hauled out for cleaning, painting, and repairs. Such a railway is usually called a marine railway in the United States and sometimes a patent slip in Great Britain.

A slipway should be distinguished from a slip, which is a space of water alongside a pier or wharf at which a ship may lie, as for loading or unloading.

A wet slip or wet basin in a shipyard is a berth alongside a pier, usually fitted with crane facilities, in which a vessel may lie afloat during completion after launching. (J. P. Ck.)

SLIVEN, a town and capital of the *okrug* of that name in eastern Bulgaria, lies in the southern foothills of the Balkan Mountains, 60 mi. (97 km.) W of Burgas. Pop. (1961 est.) 60,129. It was an old settlement which arose at a strategic point at the entrance of the Balkan passes. Later it developed as a manufacturing centre; the first woolen textile factory of Bulgaria was founded there in 1834. The town is now one of the chief centres of the textile industry, mainly woolen; other industries include woodworking and the manufacture of electric bulbs and silk and knitwear articles. Sliven is on the Sofia-Levskigrad-Burgas Railway. The town is associated with the Bulgarian revival of the 19th century, the movement for liberation from Ottoman rule.

Sliven *okrug* had a population (1963 est.) of 218,700, and an area of 1,388 sq.mi. (3,595 sq.km.). (An. Be.)

SLOAN, JOHN FRENCH (1871–1951), U.S. painter, the most brilliant of the artists associated with Robert Henri (*q.v.*), and a figure of some influence in U.S. art. Sloan was born in Lock Haven, Pa., on Aug. 2, 1871, and became a commercial newspaper artist in Philadelphia, where he studied with Henri. Eventually Henri and four of his associates went to New York (Sloan in 1904), and in 1908 these five, with three others ("The Eight"), held a group exhibition that was a significant forerunner of the modern art movement in the United States. Sloan's realistic paintings of urban scenes gave rise to the epithet "ashcan school," at first applied derisively to "The Eight" and subsequently in a more general and less pejorative sense to realistic modern artists generally. The Eight had little in common except their rejection of traditional and academic standards of the beautiful; they did not constitute a school of painting and had no further group exhibitions, but several of the artists, including Sloan, helped organize the Armory Show of 1913 in New York (*see* Painting: *United States*). For most of his life Sloan supported himself by illustrating, etching, and intermittent teaching; until the United States entered World War I he did illustrations for the socialist periodical *The Masses*. In 1939 he published *The Gist of Art*. He died in Hanover, N.H., on Sept. 7, 1951. His important paintings date from the period 1900–1920; representative examples are "Sunday, Women Drying Their Hair" (1912; Addison Gallery, Andover, Mass.); "McSorley's Bar" (1912; Detroit Institute of Arts); and "Backyards, Greenwich Village" (1914; Whitney Museum, New York).

See Lloyd Goodrich, *John Sloan* (1952); Van Wyck Brooks, *John Sloan* (1954); *John Sloan's New York Scene, from the Diaries, Notes and Correspondence 1906–1913*, ed. by Bruce St. John (1965).

SLOANE, SIR HANS (1660–1753), British physician whose collection of books, manuscripts, and curiosities formed the basis of the British Museum, was born on April 16, 1660, at Killyleagh in County Down, Ire. After studying medicine in London he traveled in France, taking his M.D. degree at the University of Orange in 1683. During a visit to Jamaica (1687–88) he collected about 800 new species of plants, of which he published an elaborate catalogue in Latin in 1696. Sloane was created a baronet in 1716, being the first medical practitioner to receive a hereditary title. He became first physician to George II and president of the Royal Society in 1727. His great stroke as a collector was to acquire the cabinet of William Courten (1642–1702), who had made collecting the business of his life. When Sloane retired from active work in 1741 his library and cabinet of curiosities had grown to be of unique value and on his death on Jan. 11, 1753, he bequeathed his collection to the nation, on condition that Parliament pay his executors £20,000. The bequest was accepted, and went to form the collection that was opened to the public as the British Museum in 1759.

SLOOP: *see* Boat.

SLOTH, a tree-dwelling mammal of proverbial sluggishness, belonging to the order Edentata (*q.v.*). The few extant species of sloths, forming the family Bradypodidae, of the infraorder Pilosa, or hairy edentates, are confined to the forests of tropical America; their extinct relatives, the ground sloths, once ranged into North America.

Sloths have rounded heads (which can be turned through an arc of 270°), inconspicuous ears, and grotesque, flattened faces. The body is about $2\frac{1}{2}$ ft. long; the forelimbs are longer than the hindlimbs, with long, curved claws sharp enough to inflict deep wounds. The three-toed sloth, or ai (*Bradypus*), has three toes on each limb; the two-toed sloth, or unau (*Choloepus*), has two on the forelimbs but three on the hindlimbs. Although most mam-

mals have seven neck vertebrae, the ai has nine and the unau has six or seven.

These most arboreal of mammals seldom voluntarily descend to the ground, but hang upside down from the branches and move with the greatest deliberation, hand over hand, through the trees, feeding on the vegetation. They are more active by day than by night. Although they can swim well, on the ground sloths can only drag themselves awkwardly along by their arms—making them easy prey for jaguars and other predators. They are almost invisible among the foliage, especially during the rainy season, when a green alga grows among their stiff hairs. Generally silent, sloths can utter a shrill cry.

They occur alone or in pairs. A single young, born at the beginning of the dry season, is carried clinging to the belly fur of the mother for about five weeks. In captivity the unau has lived more than 20 years; the ai, however, rarely survives more than a few months when taken from its habitat, partly because of its apparent dietary preference for *Cecropia* leaves. (L. H. M.)

SLOUGH, a municipal borough (1938) in the Eton and Slough parliamentary constituency of Buckinghamshire, Eng., lies 20 mi. (32 km.) W of London on the Great West Road. Pop. (1961) 80,781. It developed into an industrial centre after World War I, when an industrial park (the first in Britain and a model for many others) was established on the site of a wartime Army mechanical transport depot. It consequently became also a residential centre and is now the largest town in Buckinghamshire. It has over 300 factories and many housing developments. Among the notable buildings and many open spaces in the borough, the chief are the Norman church of St. Laurence, St. Mary's Church, Baylis House (*c.* 1695) and grounds, Salt Hill Pleasure Ground, Slough Lido, Herschel Park, and Upton Court Park. Slough is on the main Western Region railway and the London airport is 8 mi. SE.

(N. T. Be.)

SLOVAKIA (Slovensko), a region and former country of east-central Europe, from 1945 a component of the restored Czechoslovakia (*q.v.*). With an area (from 1947) of 18,922 sq.mi. (49,009 sq.km.), it is bounded north by Poland, east by the Ukrainian S.S.R., south by Hungary, west by Austria and by Moravia. The capital is Bratislava (*q.v.*).

The country was inhabited in the first centuries A.D. by Illyrian, by Celtic, and then by Germanic tribes. The Slovaks—Slavs closely akin to, but possibly distinct from, the Czechs—probably entered it from Silesia in the 6th or 7th century. For a time they were subject to the Avars, but in the 9th century the area between the Morava River and the central highlands formed part of Great Moravia, when the population accepted Christianity from Cyril and Methodius (*q.v.*). In the 890s, however, the German king Arnulf called in the Magyars to help him against Moravia. The Moravian state was destroyed in the first decade of the 10th century, and after a period of disorder Slovakia became one of the lands of the Hungarian crown in the 11th.

The main ethnic frontier between Magyars and Slovaks ran along the line where the foothills merge into the plain, though there were also Magyars settled in the larger valleys; later, the landlord class and much of the urban population in the whole area was Magyar. On the other hand, as the country suffered from chronic overpopulation, a constant stream of Slovak peasants moved down into the plains, where they usually were Magyarized in two or three generations. In the east, part of the mountain population was Ruthene (Ukrainian), and the mining areas of Spis had been colonized in the 12th century with German settlers, who founded flourishing cities.

In the 15th century Czech Hussites overran the country and for a time ruled parts of it, leaving behind them chiefly a cultural legacy. During the Reformation the German and Slovak converts adopted the Lutheran tenets, while the Magyars followed those of Calvin. The liturgical language used by the Slovak Protestants was Czech. The Counter-Reformation, however, was strongly pressed home in Slovakia by the Habsburgs, who succeeded to the Hungarian crown in 1526. When most of the Hungarian lands were occupied by the Turks, the Habsburgs retained possession of Slovakia.

Late in the 18th century a national renaissance began in Slovakia; it affected, of course, only a tiny fraction of this essentially peasant population. In 1803 an Institute of Slavonic Language and Literature was established at the Lutheran *lycée* of Bratislava: Juraj Palkovič (1769–1850) filled the first chair of Slavonic studies there. Two leading figures in the Slovak literary movement were Jan Kollař (1793–1852) and Pavel Josef Šafarik (1795–1861). Kollař, one of the earliest Pan-Slav enthusiasts, regarded the Slovak as indistinguishable from the Czech language, while Šafarik revered Slovak only as a [illegible] literary form. Other leading Slovaks of this period were eager defenders of a Slovak language and nationality separate from those of the Czechs. Ljudevít Štúr (1815–56), the successor of Palkovič in Bratislava, was expelled by the Magyars in 1844. It was after this that he adopted the dialect of central Slovakia as the best literary form of Slovak if any form of the Slovak language was to be defended against Magyar pressure. Štúr became the foremost leader of the Slovaks in 1848 when, after his literary "separatism," he entered after all into the closest alliance with the Czechs, especially with Karel Havliček. He and Šafarik were active members of the Pan-Slav Congress in Prague. (*See* Bohemia.)

During the period of Austrian absolutism after the suppression of the Hungarian revolution in 1849 Slovakia was for several years *de facto* separated from the rest of Hungary under a regime which favoured the non-Magyars. After the Austro-Hungarian *Ausgleich* or Compromise of 1867 the Hungarian government reversed this trend and pressed Magyarization forward. Secondary education in Slovak was entirely suppressed; between 1869 and 1911 the Slovak primary schools were reduced from 1,921 to 440 in number; and the schools law of 1907, associated with the name of Count Albert Apponyi, established Magyar control over the confessional schools, which were supported solely by Slovak voluntary contributions. Though the electoral system was heavily weighted against non-Magyars and the Hungarian police made difficulties for their supporters, a few Slovak deputies sat in the Hungarian Parliament; on the eve of World War I one of these was Milan Hodža. Meanwhile the difficulties which they encountered in Hungary, the poverty and suppression by comparison with which the Czechs under Austrian rule were thoroughly well-off, drove large numbers of Slovaks to emigrate, particularly to America. (*See* Czechoslovakia: *History*.)

Only on May 24, 1918, was it possible for the Slovaks in Hungary to declare openly for participation in a future state embracing Slovakia and the Czech lands. On May 30, the American Czechs and Slovaks concluded on agreement at Pittsburgh, to which T. G. Masaryk (*q.v.*) subscribed. On Oct. 30, at Turčiansky Svaty Martin, the Slovak National Council repeated the declaration of May 24; and on the same evening Hodža arrived from Budapest with the news of the collapse of the Hungarian regime with which he had been negotiating. He then persuaded the Slovaks to drop a demand for separate representation at the Peace Conference. Three Slovaks, namely Hodža, Vavro Šrobar and Milan Ivanka, became members of the first Czechoslovak government, headed by Karel Kramař: Šrobar was appointed minister for Slovakia. The peace settlement drew the frontiers of Czechoslovakia so generously as to leave a large Magyar minority in southern Slovakia.

Release from Magyar oppression gave the Slovaks leisure to develop irritation against the Czechs, whose regime proved more centralist than the Pittsburgh agreement had led the Slovaks to expect. However, the Prague government at any rate supplied the Slovaks with educational facilities: the first Slovak university was opened at Bratislava in 1919.

On Oct. 6, 1938, after the Conference of Munich (*q.v.*), the Slovaks declared themselves autonomous within a federal Czecho-Slovak state. The Slovak autonomist leader, Andrej Hlinka (*q.v.*), had died in the previous August; his heir was Josef Tiso, who now became prime minister. On Nov. 2, by the first Vienna award Germany and Italy compelled Tiso to surrender the southern rim of Slovakia, with the major part of the Magyar minority, to Hungary. After the German occupation of Prague, Slovakia on March 18, 1939, became a nominally independent state under

German protection; and in October, when World War II had broken out, Tiso was chosen to be president of Slovakia, with Bela Tuka as premier. For the next five years the Slovak regime was characterized by the balance preserved between the cautious clerical Fascism of Tiso and the more Nazi policy of Tuka and his friends, but the balance tilted increasingly in Tiso's favour. At first German protection brought Slovakia unprecedented prosperity. But the Slovaks sympathized with the Poles, against whom Hitler forced them to declare war; and later they resented having to send a contingent to fight against the Russians in the U.S.S.R. After the Battle of Stalingrad, Slovak sentiments became confused, and by August 1944 there was enough opposition to bring about a serious rising against the Germans, which lasted for two months.

With the arrival of the Soviet and Czechoslovak armies early in 1945, the Slovak National Council declared for the government of Edvard Beneš (*q.v.*), while claiming the complete equality of the Slovaks with the Czechs. The frontier of 1919 with Hungary was restored, with a small extension south of Bratislava. Some of the Magyars of the recovered territory were expelled.

After the Communist seizure of power in Czechoslovakia in February 1948, Slovakia was more severely subjected to Prague. On July 31, 1956, however, a new constitutional law made the Slovak National Council (counterpart of the Czechoslovak National Assembly) the organ of state power in Slovakia, while its former legislative powers were increased. The National Council received the right to nominate and recall the Board of Commissioners (counterpart of the Czechoslovak Council of Ministers), which was to have full executive and administrative powers in Slovakia, except for national defense, foreign affairs, foreign trade, and some branches of heavy industry and of rail transport.

The Communist regime made special efforts to speed up the industrialization of Slovakia. By 1956 the industrial output was 5.3 times greater than in 1937 (this increase was begun under German protection during World War II). Periodically, nevertheless, there were indications of the persistence of the old separatism.

In July 1960 Slovakia was subdivided into three administrative regions: West Slovak (chief town Bratislava), Central Slovak (chief town Banská Bystrica) and East Slovak (chief town Košice).

BIBLIOGRAPHY.—E. Denis, *Les Slovaques* (1917); R. W. Seton-Watson, *The New Slovakia* (1924); (ed.) *Slovakia: Then and Now* (1931); C. A. Macartney, *Hungary and Her Successors* (1937). See also *Slovensko 1945–1957: hospodarsky a kulturny rozvoj* (1958).
(E. Wi.)

SLOVAK LANGUAGE. Slovak became in 1918 the official language of the eastern part of Czechoslovakia. In 1955 the estimated number of speakers of Slovak in Europe was about 3,500,000, and nearly 500,000 Slovaks in the U.S. continued to use their native tongue to some extent.

The Slovaks were under Hungarian rule from the 10th century until 1918, and for centuries there was no serious attempt to write Slovak. The flourishing Czech culture of the 14th and 15th centuries produced a written language which differed from most Slovak dialects little more than these dialects differed among themselves. By the 16th century a number of Slovak towns were keeping their records in Czech.

An attempt in the 1780s to raise the western Slovak dialects to the position of a literary language aroused little interest. In the 1840s a group led by L'udevít Štúr (1815–56) began to write in the central Slovak dialects. Štúr's language, as modified and codified by M. Hattala in 1852, rapidly gained approval and was accepted as standard. (*See* further CZECHOSLOVAK LITERATURE: *Slovak Literature*.)

Classification.—The contemporary literary Slovak and Czech languages differ in so many points of detail that even the shortest utterance can be identified as either Czech or Slovak (*see* also CZECH LANGUAGE). During the 1920s the official designation of both languages was Czechoslovak, but because of the striking and pervasive differences the term was abandoned. On the other hand, the over-all agreement in vocabulary, in word formation, and in syntax reveal such a close relationship between the two that the linguist is justified in speaking of a Czechoslovak subgroup of the West Slavic languages (*see* SLAVIC LANGUAGES).

The spoken Slovak dialects, as opposed to the standardized language of culture and state, fall into three major groups. The central one is the most distinctive type, and possesses certain features in common with South Slavic. Eastern and Western Slovak are closer to each other than to the central dialects. Nowhere, however, is there a sharp linguistic frontier, for every local type of speech may be viewed as a transition dialect between the neighbours on every side. Eastern Slovak has certain affinities with Polish (loss of long vowels, some consonantal features), but it passes into the typical Slovak of central Slovakia, which in turn is transitional to the Western Slovak dialects that shade into the Czech Moravian dialects and on to the typical Czech of Bohemia. The dialects are all mutually comprehensible, except perhaps the extreme Eastern Slovak and Bohemian Czech.

Characteristics.—Like Czech, Slovak has a system of long and short vowels and an automatic, nondistinctive stress which always falls on the first syllable of a word. The two vowel systems were doubtless almost identical in the early Middle Ages, but since the 14th century Czech has undergone far-reaching changes, while Slovak has been much more conservative. Slovak has retained long syllabic *l* and *r*, and a series of diphthongs (*ie, iu, ia, uo*). It possesses the open front vowel *ä* (*päť* "five," *mäso* "meat"; Cz. *pět, maso*). In the consonantal system, Slovak did not develop the distinctive sibilant vibrant *ř* of Czech, but did retain the soft *l'* (*strieľať* "to shoot"; Cz. *střileti*). It developed *dz* and *dž* as independent phonemes. In morphology, Slovak has suppressed certain consonantal alternations: *e.g.*, *veľký* "big," pl. *veľki* (Cz. *velci*); *ruka* "hand," dative *ruke* (Cz. *ruce*); *platim* "I pay," *platený* "paid" (Cz. *placený*). The first person singular of all verbs ends in *m*. The vocative has been lost except in a few words.

BIBLIOGRAPHY.—*Dictionaries:* J. J. Konuš, *Slovak-English and English-Slovak Dictionary* (1941); Š. Peciar (ed.), *Slovník slovenského jazyka*, 4 vol. (1959–65). *Grammar:* E. Pauliny, J. Ružička, and J. Štolc, *Slovenská gramatika* (1955). *History:* E. Pauliny, *Dejiny spisovnej slovenčiny* (1948), for the standard language; J. Stanislav, *Dejiny slovenského jazyka*, 3 vol. (1956–57), covers all dialects, with bibliography; E. Pauliny, *Fonologický vývin slovenčiny* (1963), is a structural history.
(H. G. L.)

SLOVENE LANGUAGE, a member of the South Slavic group of the Slavic languages (*q.v.*), is one of the state languages of Yugoslavia. It was spoken in 1955 by about 1,500,000 people in the northwestern corner of Yugoslavia and neighbouring areas in Austria and Italy, and by some of the 180,000 Slovenes in the U.S.

History.—The earliest record of Slovene is generally agreed to be found in the Freising manuscript, from about A.D. 1000, which contains a confessional form, a short homily, and a confessional prayer. After that the Slovenes, subject to the Austrians, did not write their own language until the Reformation, when a group of Protestants, led by Primož Trubar, translated some of the Scriptures and wrote tracts. Their religious works were destroyed early in the 17th century, during the Counter-Reformation, but their grammatical and orthographical usage served as a model for the few Catholics who wrote Slovene during the next 200 years. The Enlightenment brought a new interest in Slovene, and at the end of the 18th century a Slovene translation of the Bible appeared, followed quickly by grammars (by Jernej Kopitar in 1808 and by Valentin Vodnik in 1811), dictionaries, and literature of all sorts. By the middle of the 19th century a standard written language was in use. (*See* YUGOSLAV LITERATURE.)

The agreement on matters of orthography and on the words and grammatical forms suitable for literature was not accompanied by an agreement on standard pronunciation. More than a century of discussion culminated in an authoritative handbook of orthography and pronunciation (1950) that was reissued in drastically revised form (1962). Yet great diversity of pronunciation persisted, apparently little affected by the unifying influences of the schools, radio and television, and modern transportation.

The reason for the lack of agreement lies in the extraordinary diversity of Slovene dialects, which developed during a millen-

nium when the Slovenes had no major administrative or economic centres, no unifying political organization, and lived in scattered communities with little communication between one another. The Alpine villages in the north and west were particularly isolated, and the multiplicity of dialects is correspondingly greater there. The leading Slovene authority, Fran Ramovš, distinguished seven dialect groups, with no fewer than 46 individual dialects. While there are significant differences in consonantism, morphology, and vocabulary, it is the vowels and the accent which vary most widely. Slovenes tend to pronounce the literary language with the vowels of their own local dialect when speaking formally, and in informal conversation they often add dialect grammatical forms as well.

Slovene is closely related to its eastern neighbour, the Serbo-Croatian language (*q.v.*), and the transition from eastern Slovenian dialects to the *kaj* Serbo-Croatian of Croatia is gradual. Practically speaking, however, the two standard languages are more remote. Slovenes have some Serbo-Croatian in school and understand it at least passively, but the average Serb or Croat, without special preparation, can read Slovene only with great difficulty and understands little or nothing of the spoken language.

Phonology.—The recommended pronunciation is a compromise between that of the Lower Carniola dialect group (*dolenjski*), which covers the largest area, and the Upper Carniola group (*gorenjski*), which includes the dialect of Ljubljana, the capital of Slovenia. It distinguishes seven long vowels (*i, ẹ, e, a, o, ọ, u*) plus syllabic *r*, occurring only under stress, with either rising or falling intonation, and six short vowels (*i, e, a, o, u, ə*) plus syllabic *r* which may be stressed or unstressed. If there is a long vowel in a word, it must be stressed; otherwise the stress falls on the last syllable. It is a very strong stress; unstressed vowels tend to be reduced, and some of them disappear altogether in the pronunciation of many speakers. Normal spelling, utilizing only the five letters *i, e, a, o, u,* unfortunately gives little help in pronunciation of vowels.

Slovene is characterized by the development of Common Slavic $ǫ > o$, of both *ŭ* and *ĭ* to *a* in long and *ə* in short syllables, $tj > č$, $dj > j$, syllabic $l > ou̯$ (spelled *ol*), soft $r > rj$, and a general tendency for long vowels to be closed, short vowels open. None of these features is unique among the Slavic languages, but they occur together only in Slovene.

Other Features.—In grammar, Slovene has preserved the dual in nouns and verbs, and the supine is usually distinct from the infinitive. The vocative has been lost. The future for both aspects is periphrastic: *bom* (< **bǫdǫ*) + *l*-participle; *e.g., bom kupil/kupoval,* "I shall buy/be buying." The vocabulary includes many borrowings from German, while Czech and Serbo-Croatian models have been important in creating modern technical terminology.

BIBLIOGRAPHY.—*General:* R. G. de Bray, *Guide to the Slavonic Languages* (1951). *Dictionaries:* M. Pleteršnik, *Slovensko-nemški slovar* (1894–95), is the most complete source, although it lacks modern terminology; J. Kotnik, *Slovenian-English Dictionary* (1956). *Grammar:* A. Breznik, *Slovenska slovnica* (1934). *Spelling and Pronunciation: Slovenski pravopis* (1962). *History:* F. Ramovš, *Historična gramatika slovenskega jezika,* of which only two volumes appeared (vol. ii, *Consonantism;* vol. vii, *Dialects*), but supplementary material is found in Ramovš's *Kratka zgodovina slovenskega jezika* (1936) and *Morfologija* (1952). *Dialectology:* F. Ramovš, *Dialektološka karta slovenskega jezika* (1931). (H. G. L.)

SLOVENIA (SLOVENIJA), a country of southeastern Europe, politically organized in 1946 as one of the six people's republics of Yugoslavia (*q.v.*). Bounded west by Italy, north by Austria, northeast by Hungary, and southeast by Croatia, it has an area of 7,819 sq.mi. (20,251 sq.km.). Of the population, 95% are Slovenes, a South Slav race with its own language.

Geography.—Slovenia is largely mountainous and wooded, with deep and fertile valleys and numerous rivers. Its northwestern tip reaches into the Julian Alps, where Mount Triglav towers over a region of great natural beauty. From this region the Sava River flows southeastward and on its banks are the steel town of Jesenice and the textile centre of Kranj. It bypasses Ljubljana (*q.v.*), the capital of the republic, and then cuts a gorge through the hilly country of the coalmining towns of Trbovlje, Zagorje, and Hrastnik.

The Karavanke (Karawanken) Mountains stand on the Austrian border. Farther to the south and southeast the Kamnik mountain group emerges. Along the Drava River the country descends into the hilly region of Pohorje and Kozjak. Maribor (*q.v.*), on the Drava, is the second largest town of Slovenia. The ancient town of Ptuj, further down the Drava, has aluminum works in its neighbourhood. Eastward the hills yield to the fertile Pannonian plain, where oil is extracted.

Slovenia between the Mura, Drava, Savinja, and Sava rivers is a good wine-growing region and is also noted for mineral springs. Celje is a major town in this region.

West and southwest from Ljubljana, along the precipitous course of the Soča River (It. Isonzo), the climate and the general character of the country become less continental and more Mediterranean. Southwest from Idrija, where mercury is mined, a limestone plateau eventually leads to a strip of seacoast south of Trieste.

History.—The Slovenes arrived in their present territory and farther north in the 6th century A.D. Subjected to Bavarian domination *c.* 743, they were subsequently incorporated into the Frankish Empire of the Carolingians. With the partitioning of that empire in the 9th century, the country was assigned to the German kingdom. It was eventually divided between the marks or marches of Carantania (Carinthia), Carniola, and Styria. The Germans reduced the Slovenes to serfdom, and most of the Slovene settlements north of the Drava were germanized. That the Slovenes preserved their identity through centuries of German rule was largely due to intensive educational work by the native intelligentsia, most of whom were Roman Catholic priests. The suzerainty of the Habsburgs (*q.v.*) over the Slovenian lands was gradually established from the last quarter of the 13th century onward. The counts of Cilli (Celje), powerful in the Middle Ages, died out in 1456.

Sporadic risings of peasants in the 15th and 16th centuries were exceptions in an uneventful history. One of these risings (1573), made in alliance with Croatian peasants led by Matija Gubec, had political and nationalist undertones, but most were motivated entirely by social and economic grievances. The reforms decreed by the empress Maria Theresa and by her son Joseph II in the 18th century improved the lot of the peasantry.

From 1809 to 1814 a large part of the Slovene territory was included in the Illyrian provinces of Napoleon I's French Empire, and the French encouraged local initiative and favoured the use of Slovene as an official language. After Napoleon's defeat, Habsburg rule was restored. In 1848 the first Slovene national program was formulated; it demanded a unified Slovene province within the Austrian Empire. Hopes for a political union of South Slavs (Slovenes, Serbs, and Croats) were expressed in the 1870s. Political parties—the Slovene People's Party (Catholic), the Progressive (Liberal) Party, and the Socialist Party—were formed in the 1890s. Members of the Catholic clergy also established large-scale cooperative movements among peasants and artisans.

On May 30, 1917, during World War I, the Slovene and other South Slav deputies in the Austrian *Reichsrat* put forward their declaration in favour of "the unification of all territories of the monarchy inhabited by South Slavs in one independent political body, under the sceptre of the Habsburg dynasty"; but in 1918 the Slovene political leaders, on a wave of popular enthusiasm collaborated in the formation of the Kingdom of Serbs, Croats, and Slovenes (*see* YUGOSLAVIA). At the Paris peace conference, however, the Allies awarded Gorizia (Gorica) and adjacent territory, with a Slovene population of about 300,000, to Italy. Moreover, the Treaty of Saint-Germain (*q.v.*), between the Allies and Austria, assigned only a small part of southern Carinthia outright to Yugoslavia. For the rest of southern Carinthia it was decided to hold a plebiscite, in two zones; but since on Oct. 10, 1920, the more southern zone chose Austria, no plebiscite was held in the northern zone (Klagenfurt). Both zones were thus left to Austria.

There was some resentment in Slovenia at Serbian hegemony in Yugoslavia before World War II, though gratitude was felt for progress made in the economic and educational fields (the first

Slovene university was established in Ljubljana).

In World War II Slovenia was partitioned: Italy took the southwest, with Ljubljana; Germany took the northeast, with Maribor; and Hungary recovered the Prekomurje (a small area north of the Mura attached to Slovenia in 1920). Slovene resistance movements sprang up, by far the most significant being the Communist-led Liberation Front. Yet Communist partisans combined activities against the invaders with a ruthless struggle against other potential opponents, particularly members of the former Slovene People's Party. Anti-Communist military units were subsequently organized under the auspices of the occupation authorities.

After the Allied victory of 1945 the old Slovenia returned to Yugoslavia. Under the Paris peace treaty of 1947, Italy ceded additional territory in the west (but not Gorizia); and in 1954 some of the former Free Territory of Trieste (*q.v.*) was annexed. The federal constitution of Yugoslavia (1946) made Slovenia an autonomous republic with its own People's Assembly as the supreme organ of power. The assembly consisted of two houses: the Republican Council (elected on a general ballot) and the Council of Producers (elected by workers and salaried staff of Slovenia's industries). The assembly appointed an Executive Council. Though these constitutional arrangements were circumscribed in practice by the dictatorship of the Yugoslav League of Communists, Slovenia enjoyed a fair degree of independence in the economic and cultural field. For administrative purposes Slovenia was divided into 8 districts and 83 municipalities.

BIBLIOGRAPHY.—A. Melik, *Slovenija*, 6 vol. (1935–60); B. Grafenauer, *Le Développement national des Slovènes de Carinthie* (1946) and *Zgodovina slovenskega naroda*, 3 vol. (1954–56); E. Kardelj, *Razvoj slovenskega narodnega vprašanja*, 2nd ed. (1957); J. A. Arnez, *Slovenia in European Affairs* (1958). (K. LA.)

SLOWACKI, JULIUSZ (1809–1849), ranks with Adam Mickiewicz and Zygmunt Krasinski in Poland's triad of Romantic messianic poets (*see* POLISH LITERATURE: *Romanticism*). Born at Krzemienec on Aug. 23 (old style; Sept. 4, new style), 1809, he was educated at Wilno (now Vilnius, Lith.), at the *Gymnasium*, and (1825–28) at the university. In 1829 he entered the civil service in Warsaw but, after the failure of the November revolution, resigned and went to Dresden, and thence (July 1831) to London and Paris with dispatches from the insurrectionary government. During 1833–35, he was in Switzerland and, in 1836, in Italy, where he wrote the love idyll *W Szwajcarii* (1839; Eng. trans. *In Switzerland*, 1953). His travels in the Middle East (1837–38) are described in *Podróż na Wschód (Podróż do Ziemi Świętej*; "Voyage to the Holy Land"; publ. 1866), a narrative poem in the manner of Byron's *Childe Harold*.

Supremely egocentric, Slowacki has been called "a poetic peacock," "the most romantic of all the Romantics." Rivaling Mickiewicz's *Dziady III* and *Books of the Polish Nation and Its Pilgrimage*, in *Kordian* (1834), a romantic drama, and in *Anhelli* (1838; Eng. trans. 1930), a poem in prose, he scanned the recent past and the present, finding in a projection of his own self the promise of Poland's delivery. His palingenetic view of history, formed c. 1844, was to have found expression in the uncompleted visionary poem *Król Duch* ("King Spirit"; part publ. 1847; full text, 1925).

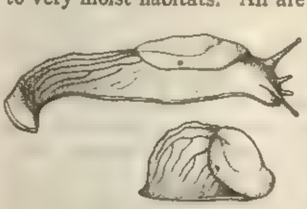
As a dramatist he provided the Polish theatre with a variety of plays (some published posthumously in 1866), largely inspired by Shakespeare (*Balladyna*, written 1834), classical drama (*Lilla Weneda*, 1840), Calderón (*Ksiądz Marek*, written 1843; *Sen srebrny Salomei*, 1844), and by contemporary productions seen in Paris. *Fantazy* (written 1843) is an anti-Romantic comedy. These constituted a repertoire of serious drama equal in merit to that of western Europe and still preeminent in Poland. His dramatic work was the chief native influence on Stanislaw Wyspianski (*q.v.*).

Slowacki was master of a wealth of verse forms traditional as well as adopted, but excelled in conveying his own feelings in lyrics and in lyrical or polemical digressions, such as those in *Beniowski* (written and part publ. 1841; full text, 1866), a verse narrative indebted for its framework to Ariosto and to Byron's *Don Juan*. His technical virtuosity, the richness and inventiveness of his vocabulary, and the evocative nature of his imagery made him a precursor of the Symbolists. The poets of the "Young Poland" movement and after acknowledged him as their master.

BIBLIOGRAPHY.—*Dzieła*, new critical ed. of complete works, by J. Krzyżanowski, 14 vol. (1952); *Korespondencja*, ed. by E. Sawrymowicz, 2 vol. (1962–63); *Kalendarz życia i twórczości J. Słowackiego* (a detailed chronological survey of Slowacki's life and work), ed. by E. Sawrymowicz (1960). *See* also M. Kridl, *The Lyric Poems of J. Slowacki*, a critical essay (1958); C. Backvis, "Slowacki's Place in Polish Drama," in *The Slavonic Review*, vol. XXVII, no. 71 (1949); J. Kleiner, *J. Słowacki. Dzieje twórczości*, 4 vol. (1919–27); J. Krzyżanowski, chapter on Słowacki, in *Polish Romantic Literature* (1930); J. J. M. Bourrilly, *La Jeunesse de J. Słowacki, 1809–33: La vie et les oeuvres* (1960); S. Treugutt, *J. Słowacki, Romantic Poet* (Eng. trans. 1954). (L. R. LR.)

SLUG, any snail with the shell reduced to an internal plate, a series of granules or completely absent. Generally the term refers to a land mollusk, but members of the marine order Nudibranchia are called sea slugs.

Ranging in size from ½ in. to more than 6 in., slugs are restricted to very moist habitats. All are usually active only at night or on overcast days. Eggs are buried in the soil, and the animals live one to four years. All species are both male and female, some with one, others with two genital pores. Most slugs have the respiratory and excretory ducts combined into a single external opening. Usually slugs are mottled with brown, gray, green or buff, but some species have brilliant markings of orange, red, purple or yellow.



COMMON SLUG: (ABOVE) IN CRAWLING POSITION, (BELOW) INACTIVE

The approximately 300 species of land slugs are not closely related, since loss of an external shell happened several times during snail evolution. The common slugs (families Arionidae, Limacidae and Philomycidae) of temperate-zone fields, forests, gardens and greenhouses feed on fungi, decaying plants and occasionally the leaves of living plants. Often a nuisance in gardens and greenhouses, they can be controlled by spreading ashes or a bait composed of bran and metaldehyde (30 to 1). The tropics have a few European species introduced by man, and the very primitive plant-eating Veronicellidae. Other vegetarian slugs include the relict Athoracophoridae of Melanesia, Australia and New Zealand and the intertidal pulmonates of the family Oncidiidae. Slugs which feed on other snails or earthworms include the European Testacellidae, South African Aperidae and Malayan-Indonesian Rathouisiidae.

The sea slugs (nudibranchs) live in shallow reef waters, sand flats or rocky pools where they browse on minute algae. Delicately coloured and often translucent, they are among the most beautiful sea creatures. Retiring in habit, they are seldom seen.

See SNAIL; GASTROPODA. (G. A. S.)

SLUIS (SLUYS), **BATTLE OF,** a naval engagement of the Hundred Years' War fought on June 24, 1340, which ended in an overwhelming defeat of the French fleet by English forces under the command of King Edward III. Sluis, now in the Zeeland province of the Netherlands, was in the 14th century a flourishing port on the estuary of the Zwyn River (now silted up) in northern Flanders, and the encounter took place in the mouth of the estuary below Sluis. After four months in England, Edward was returning to Flanders to resume his campaign against France. He had intended to cross with a modest force, but, warned that the French naval commanders, Hugues Quiéret and Nicolas Béhuchet, were waiting to prevent his landing, he hastily impressed more ships and soldiers. Edward sailed from Orwell, Suffolk, on June 22, and was joined at sea by Robert, Lord Morley, admiral of the northern fleet. The English force may have comprised about 147 ships (*Chronicle of Lanercost*), but the number of soldiers on board is unknown. It is possible that Edward also received some Flemish support during the battle, but this is uncertain. The French, with a fleet of about 200 ships manned by over 20,000 soldiers and sailors, had a numerical advantage, but this was wasted through the unwise decision to fight from a defensive position within the estuary.

Shortly after midday on June 24, with the wind in their favour and the sun behind them, the English sailed in on the French fleet, which was massed in three tight lines. The fire of the English archers was fierce and accurate, and when the ships met, fighting with sword, lance, and axe began. A bitter struggle followed, which lasted into the night. By the following morning all except 24 enemy ships had been captured or sunk, and French loss of life had been great. The scanty evidence about English losses tends to confirm Edward III's claim that they were comparatively small. This victory seriously weakened the naval resources of France, and added greatly to Edward III's military prestige. The battle secured for England a temporary ascendancy at sea, but not, as has often been claimed, a command of the Channel which lasted for many years. *See* also HUNDRED YEARS' WAR.

See Sir N. H. Nicolas, *History of the Royal Navy*, vol. ii (1847); C. Bourel de la Roncière, *Histoire de la Marine Française*, vol. i (1898). (J. W. SE.)

SLUM: *see* HOUSING.

SLUPSK (German STOLP), a town of northern Poland in Koszalin *wojewodztwo* (province) and a district capital. Pop. (1960) 53,000. It lies in the coastal plain on the Slupia River, 11 mi. (18 km.) from the Baltic. About 35% of the town was destroyed during fighting in World War II, the old buildings suffering most. There are many factories; one of the largest in all Poland produces furniture for export. The town is on the Gdynia-Szczecin Railway, with a branch northward to Ustka on the coast.

A fortress was erected at Slupsk in the 8th and 9th centuries, which came under Poland at the end of the 10th century with the rest of Pomorze (Pomerania). After a short period of independence under its own duke, it shared the fate in the 11th–13th centuries first of Danzig Pomorze, then of Western Pomorze. In 1648 Slupsk and the rest of the dukedom was seized by Brandenburg. It was returned to Poland in 1945. (T. K. W.)

SLUTER, CLAUS (*c.* 1350–1406), the greatest sculptor of his time, was probably born at Haarlem in Holland. He entered the Brussels stonemasons' guild about 1380, and in 1385 was employed in the workshop of Jean de Marville, chief sculptor to Philip the Bold, duke of Burgundy. On Marville's death in 1389, Sluter succeeded to his position and to two major works then in progress, the portal of the ducal chapel at the Chartreuse de Champmol, Dijon (*q.v.*), and the duke's tomb, which was to be installed in the same building. In 1392 he visited Paris, and in 1393 Mehun-sur-Yèvre, where he inspected the works executed for Jean, duc de Berry, by André Beauneveu. In 1395 the duke of Burgundy commissioned a monumental Crucifixion group to crown a well (the *Puits de Moïse* or "Well of Moses") in the cloister of the Chartreuse. On Sluter's death, the direction of the workshop passed to his nephew and pupil, Claus de Werve.

GIRAUDON

DETAIL OF AN ANGEL FROM THE "WELL OF MOSES" BY CLAUS SLUTER. IN THE DIJON MUSEUM, FRANCE

The three works mentioned above still exist, albeit in damaged condition. The tomb (Dijon Museum) has suffered severely and the effigy of the duke is a 19th-century replacement. The substructure, an open arcaded gallery along which passes a numerous procession of diminutive mourners, is an ingenious and original conception which is probably due to Marville. Two of these mourners were in existence in 1404 and others may have been completed by Sluter before his death; the best of them, despite their small scale, exhibit a masterly breadth and realism. The tomb was not finished, however, until 1410 and many of the figures were doubtless made under Werve's direction.

The life-size statues of the Chartreuse portal (complete by 1397) comprise the Virgin and Child (sometimes attributed to Marville) on the central pillar of the doorway and, on either side figures of the duke and duchess (the latter finished by 1393) protected respectively by St. John the Baptist and St. Catherine (both finished by 1391). The imposing physical presence and intense psychological interaction of these figures seem almost to annihilate their architectural frame. The celestial personages hover in a swirl of near-Baroque drapery; the ducal pair are living portraiture, reverend and grave. The Calvary group has been reduced to a battered torso of Christ and a base surrounded by six life-size prophets (executed 1400–05), which rank among the greatest masterpieces of European sculpture. Transcendental beings swathed in endless convolutions of Gothic drapery, they are at the same time living persons whose presence asserts itself like a physical shock. The sculptor's chisel, minutely exploring the wrinkles of ancient flesh, has drawn a map of the soul unparalleled in art, except in the late portraits of Rembrandt.

Sluter's work represents the culmination of a tradition—that of the Franco-Flemish realist sculpture whose older masters were Jean Pepin of Huy and Jean de Liège—but his genius was intensely personal. Both the painting and sculpture of northern Europe would have worn a different aspect in the 15th century but for the prophets of the *Puits de Moïse*.

See Henri David, *Claus Sluter* (1951). (D. Ko.)

SMALL ARMS, MILITARY. In military language, the term small arms includes a great variety of weapons designed to be carried on the person and held in the hands when used, or fired from light supports. Small arms are light, portable weapons distinguished from heavy artillery-type weapons that must be fired from mobile carriages or fixed mounts. Not all of them are small. During the first half of the 20th century the U.S. army fixed the dividing line between small arms and artillery at calibre .60, but this line is no longer universally observed because of the development of portable rocket launchers and recoilless rifles. Small arms are often grouped in three broad categories: handguns (pistols and revolvers), shoulder weapons (muskets, rifles, carbines and others) and edged weapons (swords, daggers, pikes, lances bayonets, etc.). After the invention of machine guns late in the 19th century, these weapons were assigned to the small-arms category although they were far heavier than rifles and normally had to be fired from a bipod, tripod or other support. A narrow interpretation of the term has been adopted for this discussion. There is a separate article on MACHINE GUN, another on PISTOL AND REVOLVER and still others on edged weapons (*see* BAYONET; DAGGER; DIRK; LANCE; and SWORD). This article deals primarily with shoulder firearms, chiefly muskets and rifles. It treats the subject in broad historical terms to show the development of the basic principles of these weapons and their ammunition from their first appearance in the 14th century to the present day.

The following outline indicates the main topics covered in the article:

I. HISTORICAL DEVELOPMENT TO 1900

A. Early History

1. Hand Cannon.—The first portable military firearms probably developed in the second quarter of the 14th century. The earliest written records date from the 1360s, but it would seem that these refer to weapons already in use. Variously called hand bombards, hand culverins, *gonnes*, *Handbüchsen* (handguns) or, in modern parlance, hand cannon, these primitive small arms usually consisted of a tube 8–12 in. long, closed at the breech end, with a touchhole bored through the top about two inches from the breech. Usually they were made of brass or bronze, but there are a few references to iron barrels.

To protect the shooter's hand from the heat of the discharge, and also to provide a lever for controlling the recoil, these barrels were attached to a straight wooden haft from five to eight or more feet in length. There were several means of attachment. Sometimes the tube was simply lashed to the haft with iron bands; sometimes it was set into the haft as well as being lashed by the bands; sometimes a socket was fastened to the breech end into which the haft could be inserted; and sometimes the breech was extended in a spike or tang which could be driven into the haft. All these variations are found in specimens made before 1400.

A German manuscript of about 1390 survives to indicate some of the details of loading and firing. The barrel was filled three-fifths full of powder well rammed down. Then a wooden sabot or wad was added, and finally the ball. The charge was ignited by inserting a red-hot wire through the touchhole in the top of the barrel or by using a glowing coal held in tongs. An improvement that was probably in use by 1400 was the use of a "match," a length of loose hemp rope soaked in a solution of saltpetre or spirits of wine so that it would burn slowly and steadily, much in the manner of the modern punk used to light firecrackers.

There were several methods of handling these guns, depending somewhat on their size. Small ones could be managed by one man, who clamped the haft between his left arm and his body and supported the piece just behind the breech with his left hand while he applied the wire or match with his right hand. In other instances the butt was rested on the ground and the fore end supported by a forked rest; the shooter had only to steady the weapon while he applied the fire. Still other guns needed the services of two men to hold and shoot them.

As time passed, improvements came rapidly. The touchhole was moved from the top of the barrel to the side, where it was better protected from the weather. A pan to hold a small amount of powder was affixed outside the touchhole to provide for surer ignition; the pan was soon fitted with a movable cover. The barrel was lengthened. The haft was shortened and broadened so that it could be held against the shoulder, and gradually a curve was introduced so that the recoil was directed upward instead of directly back against the user. The curved stock is thought to have developed first in France just before 1500.

Some years earlier, certainly before 1425, a most significant advance occurred with the addition of a pivoted S-shaped bar of metal with a pair of jaws on the top end for holding the match. Now all that the soldier needed to do was to raise the lower end of the lever, or serpentine, and the lighted end of the match would be automatically depressed and brought into contact with the powder in the pan. Only the pressure of a finger was necessary. One hand could be used for supporting the gun while the other controlled the serpentine. The shooter could look where he was aiming rather than fix his attention on the pan to make sure he touched the coal to it.

2. Matchlock.—The final major advance in the development of small arms fired by a lighted match came about 1450 with the development of the matchlock. This improvement consisted of adding a trigger and connecting the serpentine to it by a simple link arrangement so that it acted against a spring. Pressure on the trigger depressed the serpentine and fired the gun. When the trigger was released the spring raised the serpentine and held it away from the pan. All working parts were now placed inside the lock plate and protected. It was at last a true lock. Further developments consisted only in lightening and improving the design of the weapon for easier handling and in replacing the early lever trigger with one that acted against a sear much in the manner of modern firearms. This last advance took place shortly before 1600.

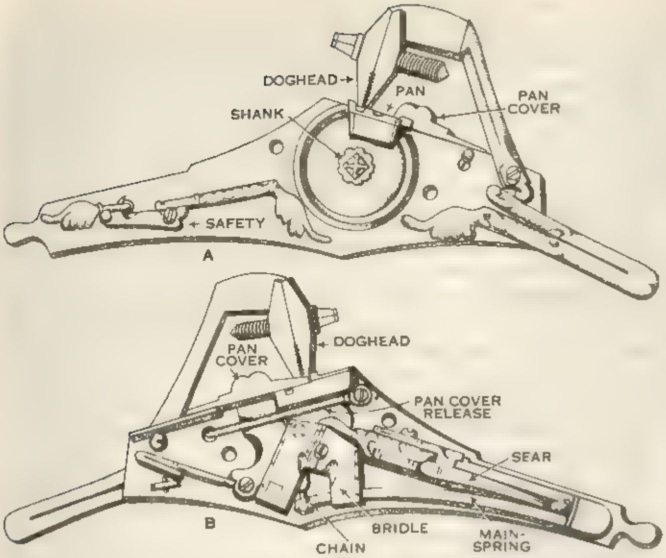
Many names were applied to these early matchlocks, including arquebus (*q.v.*), *Hak-büchse*, hacquebut, hagbutt, hachbuss and other similar terms, as well as caliver and musket (*q.v.*). Because of the similarity of the terms and the indefinite and careless way in which they were used by contemporaries, much has been written in an attempt to obtain accurate definitions. Actually, they seem to have meant different things in different countries and in different periods. This is particularly true of the word arquebus and its cognates. Some authors applied it indiscriminately to any firearm. Other and more careful writers used it to designate a gun that was lighter than a musket and could be fired without a rest, thus making it for a short period synonymous with caliver. Still later, at the beginning of the 17th century, it was used to designate a wheel lock (*see* below) as opposed to a matchlock.

The term musket has always referred to a heavy military firearm. It is generally believed that the term was first used by the Spanish (*mosquete*) to denote a military firearm developed shortly before 1550 and introduced into the Spanish service by Fernando Álvarez de Toledo, duque de Alba. These first muskets weighed about 20 lb. and had a bore of 8 or 10 gauge. (In this sense, gauge indicated the number of spherical lead balls of the same diameter as the bore that were required to reach a weight of one pound.) As years passed, the size was reduced until by the middle of the 17th century the standard English musket weighed about 16 lb. and fired a 12-gauge ball from a 10-gauge bore.

Although the matchlock was a tremendous improvement over the primitive hand cannon, it still had many very serious drawbacks. It was slow and clumsy to use. The lighted match was a decided liability! It was a constant hazard in the presence of powder, and could not be used in the wind or rain; its glow and smell prevented its use in an ambush; it could not be lighted and used at a moment's notice but had to be kept burning whenever it seemed there was even a remote possibility of action.

Despite these many disadvantages, the matchlock remained the principal military firearm of Europe for many years after better and more efficient arms were developed. It was not completely supplanted in many armies on the continent until about 1700, although in America the change had taken place as much as 75 years earlier.

3. Wheel Lock.—Chronologically, the first of the improved ignition systems that eventually supplanted the matchlock was the wheel lock. Possibly invented by Johann Kiefuss of Nürnberg, Ger., about 1515, this lock worked on the same principle as the modern cigarette lighter: the spark was produced by holding a piece of iron pyrite against a revolving rough-edged wheel. The mechanism devised to produce this action was considerably



FROM H. L. PETERSON, "ARMS AND ARMOR IN COLONIAL AMERICA"; REPRODUCED BY PERMISSION OF THE STACKPOLE COMPANY, HARRISBURG, PENNSYLVANIA

FIG. 1.—GERMAN WHEEL LOCK MECHANISM (C. 1565), PHANTOM VIEW: (A) RIGHT SIDE, (B) LEFT SIDE

more complex than the matchlock, yet it was sturdy and not likely to get out of order easily. The power for revolving the wheel was supplied by a heavy V-shaped mainspring. This spring was attached to the spindle of the wheel by a short chain of two or three links. When the wheel was wound by hooking a key or spanner over the outside shank of the wheel, this chain was wrapped around the spindle, and the mainspring was compressed and locked in position by a laterally acting sear that engaged a hole in the inner surface of the wheel.

The flashpan of a wheel lock was attached to the lock plate just as it was in the later matchlock. It differed from the pan of a matchlock in that it was pierced in the bottom to allow a portion of the wheel to intrude; and instead of a pivoted cover it possessed a sliding cover that was connected with the internal lock mechanism. The pyrite was held in the doghead, a miniature vise mounted on an arm that pivoted just in front of the pan. A spring fastened on the outside of the lock plate just below the arm of the doghead acted upon the arm in such a way that when the head was placed against the pan cover in firing position a constant downward pressure was maintained.

The loading and firing of a wheel lock was a comparatively simple and sure process. First, the wheel was wound. Then the charge was placed in the barrel in the normal manner. Next, the pan was filled with priming powder, the cover was slid shut and the doghead pushed down against it. The trigger was then pulled. The pressure on the trigger released the sear from its lodging in the hole in the wheel. The wheel thus freed began to turn, and as it did so a cam attached to it struck against a bar connected to the pan cover and automatically opened the pan. The doghead with its piece of pyrite, which had been held against the pan cover by its spring, was thus forced against the part of the revolving wheel that intruded through the slot in the pan. A strong series of sparks was produced, and the priming powder ignited.

The wheel lock was a tremendous improvement over the matchlock. The lighted match was no longer needed. The weapon could be carried ready to fire at a moment's notice. It could be fired with one hand, thus making the pistol practical; and it could be fired from horseback. As cavalry could now be armed with firearms, radical changes in tactics resulted, especially in Germany and northern Europe.

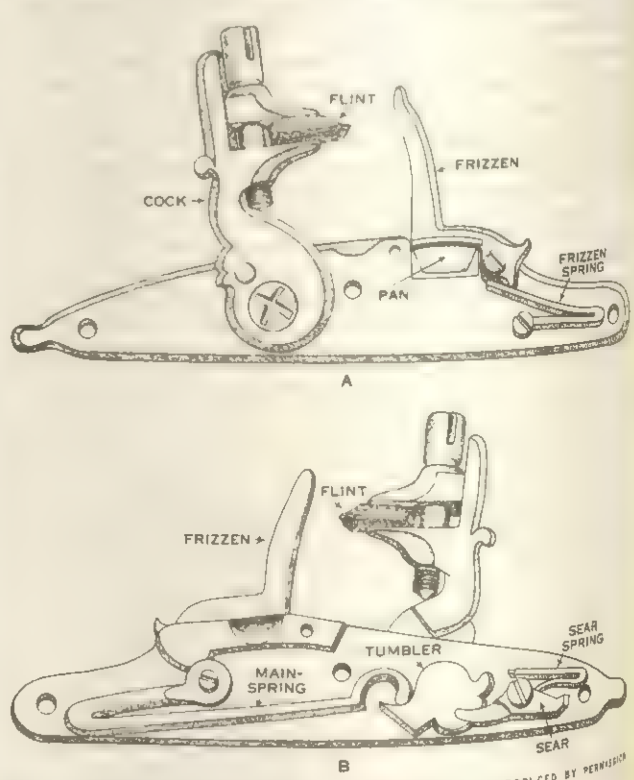
But the wheel lock was an expensive weapon; it never succeeded in supplanting the matchlock as the standard European military firearm. In America it was widely used because conditions there demanded a faster and more reliable gun than Europeans found necessary. In Europe generally it remained the weapon of horsemen and special troops.

4. Flint Arms.—The ignition system that finally supplanted the matchlock was a simple mechanism that produced its spark by striking flint against steel in a procedure similar to that for lighting household fires. In firearms the flint was held in a vise on one end of an arm, known as the cock. The other end of the arm was pivoted on the lock plate so that the flint-bearing end could be swung in an arc in the direction of the steel. The steel, which was also called the battery (now sometimes called the frizzen), was mounted on another pivoted arm and placed in a position opposing the cock. The flashpan was placed directly below the battery. When the trigger was pulled, the cock, impelled by a strong spring, moved forward in a short arc. The flint, held in the jaws of the cock, struck the steel a glancing blow producing a shower of sparks that dropped into the priming powder in the pan. If all went well, the flash of the priming powder penetrated to the powder in the bore and fired the weapon. But sometimes this failed to happen and there was only a "flash in the pan."

Modern students recognize at least six distinct types of flintlocks: the snaphance, English lock, dog lock, Scandinavian snap lock, miquelet lock and the "true" flintlock. These terms denote both different evolutionary stages in the development of the flintlock and regional variations. The first of these forms to appear was the snaphance, believed to have been developed in Scandinavia and the Low Countries about the middle of the 16th century. It was characterized by a separate battery and pan cover. Its name was derived from the Dutch words describing the action of the cock, *snaphaan*, or "snapping cock." This term was also adopted by English writers and used by them for over 150 years to indicate any form of firearm with a snapping cock, thus causing much confusion among arms historians in later years.

The miquelet lock developed almost as early as the snaphance, but it was confined to southern Europe, principally Spain and Italy. An exceptionally strong and simple lock, it remained virtually unchanged from the time of its invention until the opening years of the 19th century. Its main characteristics were an external mainspring and a battery and pan cover made in one piece.

The Scandinavian snap lock was also a product of the 16th century. It, too, had an external mainspring, and sometimes the battery and pan cover were combined as in the miquelet. There were, in fact, many points of similarity between the Scandinavian



FROM H. L. PETERSON, "ARMS AND ARMOR IN COLONIAL AMERICA"; REPRODUCED BY PERMISSION OF THE STACKPOLE COMPANY, HARRISBURG, PENNSYLVANIA

FIG. 2.—FLINTLOCK MECHANISM (C. 1660–70), BASED ON LOCK PROBABLY OF DUTCH ORIGIN; PHANTOM VIEW: (A) RIGHT SIDE, (B) LEFT SIDE

and southern European types, unusual for such widely separated areas. The principal differences were in the shapes of the parts and in minor mechanical details.

The English lock and dog lock were variations of the Dutch snaphance in which the battery and pan cover were made in one piece. The internal mechanism was the same as that of the snaphance. There was no half-cock or safety position, and thus on the dog lock a dog catch was pivoted on the outside of the lock to engage the tail of the cock and provide such a safety position. Both these forms were in concurrent use in England from the first quarter of the 17th century until about 1650, when the French flintlock began to supersede them. The dog catch, however, persisted on many guns until after 1700, though it no longer served a necessary function.

The true flintlock was invented in France probably by Marin Le Bourgeoys about 1610–15. It combined features of both the snaphance and the miquelet and produced a simple, safe and reliable weapon. The battery and pan cover were combined in one piece so that the striking of the flint on the steel automatically opened the pan, and the internal mechanism was altered so that there was a half-cock or safety position. From France this new lock quickly spread throughout most of western Europe, although it never succeeded in displacing the miquelet or, to a lesser degree, the Scandinavian locks. It soon became the standard system for all the better guns and after 1700 for almost all guns.

The chief advantages of the flintlock over the wheel lock lay in its simplicity and cheapness of construction and in the fact that the user no longer had to carry a separate spanner or wrench to operate it. Special troops were armed with the flintlock in Europe by 1650, and after 1700 it had completely replaced the matchlock as a military weapon. In America the change was somewhat more rapid; every colony had banned the matchlock by 1677.

5. Standardization of Arms.—The beginning of the 18th century also brought the standardization of military small arms. Some attempts at standardization had been made earlier, it is true, but real success was not achieved until well into the first quarter of the century. By that time royal manufactories and systems of inspection had been set up in several countries and official models adopted. Muskets were the first arms to be standardized in most countries. Pistols and carbines or musketoons followed.

The average musket of the period was a flintlock smoothbore weighing about 12 lb. Calibres ranged from .69 in France and Spain to .75 in England and even slightly larger in some of the German states. With such a musket a well-trained recruit was expected to be able to load and fire in 15 seconds.

These muskets were not noted for accuracy. The best that could be expected was to hit the figure of a man at 80 to 100 yd. The military men of the period were more interested in rapidity of fire than in accuracy; they operated on the principle of laying down fields of fire much in the manner of modern machine gunners. Carbines and musketoons were shorter and lighter versions of the musket and were carried by cavalry, artillery and other special troops. Pistols were still shorter arms, also smoothbore, with calibres almost as large and even less accurate. Most cavalrymen scorned them and looked on the sword as their only true and effective weapon. (*See* also PISTOL AND REVOLVER.)

6. Percussion System.—A Scottish clergyman, Alexander John Forsyth, revolutionized all firearms theory with the invention of a percussion compound that would explode when struck a sharp blow. Working at his manse, Forsyth made his first percussion lock in 1805 and patented it in 1807 after some months of experimenting at the Tower of London. His compound was basically potassium chlorate. When placed in a tube communicating with the bore of a gun and struck by the hammer it produced a flash strong enough to ignite the charge. Separate priming powder and free sparks were no longer necessary. Forsyth thus laid the basis for later self-contained metallic cartridges and for contact fuses in shells and opened a whole new field of possibilities.

The next task was to perfect methods of utilizing this new material efficiently. Pellets and tubes filled with the compound were tried in the so-called pill and tube locks. Then came the percussion cap claimed by several as an invention. Joshua Shaw, an Anglo-American living in Philadelphia, developed the device in the United States but was refused a patent because he was not a U.S. citizen at the time. The first caps, shaped somewhat like the top hat of the period, were made first of iron (1814), then of pewter (1815) and finally (1816) of copper. The bottom of the cavity contained a minute amount of the percussion mixture covered by a disk of tin foil; the whole was sealed and made waterproof by a drop of shellac. In operation, this cap was placed over a tube, known variously as a cone or nipple, that projected from the barrel. The hammer, acting exactly in the manner of the cock on flint arms, struck the cap a sharp blow and produced a flash that was directed by the tube into the bore where it set off the propellent charge.

Despite the obvious advantages of the percussion system, a number of years passed before it was adopted as standard for military small arms. Military men could appreciate the fact that percussion guns were simpler to make, were surer to go off, and were not affected by wind or even light rain. They saw, too, that it was no longer necessary to carry extra flints and to chip and adjust them as they wore down. Nevertheless, they were concerned about the supply of percussion caps, the methods of carrying them and the delicate process of applying them to the gun. Thus it was not until 1842 that either the United States or Great Britain began the manufacture of standard military muskets of the percussion type, though the United States had adopted the system for special weapons as early as 1833 and Great Britain in 1836.

B. Development of Rifling

Rifling, or the cutting of spiral grooves in the barrel of a firearm to impart a spin to the projectile and thus improve its accuracy, was invented between 1450 and 1500. Some authorities claim the invention for Leipzig, Ger., in 1498, while others assert that the honour belongs to Vienna. Be that as it may, a specimen exists bearing the arms of the emperor Maximilian I which can be dated definitely between 1493 and 1508. The grooves in many of these very early rifles were straight instead of spiral, and some students maintain that the purpose of the grooves was to receive the fouling from the gunpowder and not to impart stability to the projectile. Powder fouling was a serious problem, especially with the earliest types of black powder, and such grooves would collect the powder somewhat so that more shots could be fired between cleanings. At the same time, however, it should be noted that the theory of stabilizing a projectile by causing it to revolve on its axis was well understood at that period and was utilized on crossbow bolts and some arrows. This, plus the fact that many of the early grooves are spiral, would seem to indicate that perhaps both purposes were intended.

If a ball were to take the spin imparted by the rifling, it had to fit the bore tightly. This meant that the ball had to be forced down the bore with repeated blows of the ramrod, or it had to be wrapped in a greased patch of cloth or leather that would ease its passage while still permitting a tight enough fit for it to take the rifling. Both methods were used from an early date, and descriptions of them can be found in manuals of the late 16th and early 17th centuries.

Rifling a gun barrel was a comparatively expensive process, and there were other drawbacks about these guns that prevented their rapid adoption as military weapons. The tight fit required of the ball made loading a slow process, even when the greased patch was used. This was a distinct handicap from the military standpoint. Ramrods were apt to break because of the need to exert considerable pressure on the ball, and a gun without its ramrod was worthless in battle. The thicker barrels of the rifles tended to make them muzzle-heavy; they were even harder to balance with a bayonet. Use of the early plug bayonet that fitted into the bore was, of course, out of the question with rifles. Finally, it took more skill to use a rifle to best advantage, a decided drawback in the eyes of the men who had to drill raw re-

cruits and levies.

Despite these handicaps, some rifles saw military service at an early date. Christian IV of Denmark is said to have been the first to arm troops with this weapon early in the 17th century; soon thereafter the French royal horse guard could boast that they were armed with eight rifled carbines per troop. By the middle of the 18th century the use of the rifle as a military arm had spread widely on the continent, and most Scandinavian and Germanic countries had organized special troops provided with rifles. The Swedish model of 1761 was even equipped with a bayonet.

Great Britain lagged far behind the other nations of Europe in the adoption of rifles for military service. Some of the continental troops employed by the British in the various European wars of the 18th century had had rifles, it is true, but it remained for the American rifle, encountered by British troops during the French and Indian War and the American Revolution, to wield the deciding influence in introducing the weapon to the regular British service. Copies of the American rifles were made in small quantities and issued to special troops; the Highlanders serving in America had them at least as early as 1760, and thereafter the practice spread. The breech-loading Ferguson rifle, described below, was tried briefly in 1776–77 and again late in the century, followed by the famous Baker rifles for the first official rifle brigade in 1800. Interestingly, these Baker rifles followed German patterns more closely than they did American and were equipped with bayonets.

1. American Rifle.—The American rifles that so influenced the British had been developed by German and Swiss immigrants on the eastern seaboard, especially in Pennsylvania. For this reason they are frequently called Pennsylvania rifles and sometimes Kentucky rifles because of their popularity with the settlers of the territory that later became the states of Kentucky and Tennessee. The earliest of the rifles made by these immigrants in the 18th century were indistinguishable from the short, large-calibre rifles of their homelands. As the years passed, however, the colonists lengthened the barrels, decreased the size of the bore and added a brass patch box to hold the greased pieces of cloth with which they wrapped their bullets. In so doing, they achieved a distinctly American gun. Stocked in handsomely grained curly maple, it reached its fullest development between 1770 and 1810.

The American rifle was a supremely accurate weapon. In the hands of practised marksmen it could find its mark at unheard-of distances, striking a target at 300 yd., almost three times the effective range of the musket. For this reason it was an exceptionally valuable weapon for special troops such as light infantry and snipers who were supported by regular infantry with muskets. But because the rifle took considerable time to load and had no bayonet it could not replace the musket as the regular infantry arm. The tactics of the day called for rapidity rather than accuracy in laying down a field of fire before the enemy closed in with the bayonet and decided the issue in hand-to-hand fighting. This point has been overlooked frequently in the past by popular writers who were not trained soldiers and did not understand the tactical situation. A whole body of literature has developed criticizing the military authorities for not recognizing the superiority of the rifle and adopting it immediately as the standard infantry arm. George Washington, Anthony Wayne, Peter Muhlenberg, Daniel Morgan and other American generals, some of them trained riflemen themselves, recognized fully the values of the rifle but realized also that it needed improvement before it could replace the musket.

2. Ammunition for Rifles.—Attempts to make improvements in rifles occupied the time of leading gun designers throughout the western world for half a century following the American Revolution. The principal problem was to increase the speed of loading. Two notable attempts were made with the round balls that had always been standard for rifles. In Great Britain, the Brunswick rifle was designed to fire a ball with a raised band around its equator to fit the gun's two-groove rifling. The British government adopted the Brunswick rifle in 1836, but it fouled badly and was often called by its contemporaries the worst military rifle in Europe. In France Henri Gustave Delvigne designed a rifle barrel with a narrow chamber at the breech just large enough to contain the powder charge. When the ball was dropped loosely down the barrel so that it rested against the opening of this chamber, several sharp blows with the iron ramrod would expand its diameter and make it fit the bore tightly. This distortion of the ball affected its accuracy, however, and the loading process, though more rapid than that with the patched ball, was still slower than the loading of a musket.

The answer lay not in the round ball but in the elongated projectile. As early as 1823 a Captain Norton of the British 34th regiment found the right approach in a bullet with a base hollowed in such a manner that the gases produced by the explosion of the charge would expand it and cause it to press tightly against the sides of the bore. In 1836 the British gun designer William Greener developed an elongated projectile with a conical plug fitted into the base cavity. On firing, the plug was driven far enough forward into this cavity to expand the walls of the missile into the rifle grooves. Both these bullets were tested by the British government and rejected.

In France, Delvigne, who had given up his experiments on round balls and switched to cylindroconoidal projectiles, teamed up with Colonel Thouvenin of the French artillery to produce a new system, usually known today as the Thouvenin system or the *carabine à tige*. In this rifle a stout pin or *tige* just long enough to pass through the powder charge was fixed to the centre of the breech plug. When the cylindroconoidal projectile was dropped down the barrel, its base struck against this pin, which acted as an anvil for flattening the bullet under blows from the ramrod. Since the ramrod had a shaped cavity in its tip, the bullet retained its form and its accuracy. The pin had a tendency to bend after repeated use, however, and fouling presented a particularly difficult problem. The hammered bullet did not provide the answer.

It was another Frenchman who finally produced the solution. In 1849 Capt. Claude Étienne Minié perfected the projectile that made his name a household word throughout the world. In developing his bullet, Minié followed the pattern set by Norton and Greener—a cylindroconoidal slug with a hollow cavity in the base. At first Minié inserted in this cavity an iron cup that was driven forward by the force of the explosion to expand the base of the bullet until it fitted the bore tightly. Later cups of other materials were tried, and finally it was discovered that if the cavity itself were properly designed, no cup of any kind was necessary.

Here at last was the device that made the rifle a practical military weapon for all branches of the service. The Minié ball, as it came to be called, could be dropped down the barrel and rammed home as easily as a musket ball. The rifle could thus be fired as rapidly as the musket. As the rifle had the advantage of greater accuracy, only a few years passed before the smoothbore disappeared from the military scene. Rifled arms of the Minié system were adopted as standard by Great Britain in 1851 and by the United States in 1855.

In addition to the main course of rifle projectile evolution, there were also one or two side developments using specially shaped projectiles and bores that achieved some success. These were the oval-bored and hexagonal-bored rifles. In these systems, instead of the usual grooves, the whole bore twisted to impart the proper spin to the bullet. Among these types, the oval bore has the longer history. Of the many guns that attempted to utilize it, one of the best known was the Greene bolt-action rifle used occasionally during the American Civil War. The hexagonal bore and projectile are associated primarily with the English gun designer Sir Joseph Whitworth, who produced both cannon and small arms utilizing this system. His guns were supremely accurate, and many saw service in the American Civil War in the hands of Confederate troops.

C. Breechloaders and Repeaters

Attempts to produce breech-loading and repeating firearms date back almost to the first appearance of small arms. Breech-

loading cannon, in fact, predate the first handguns. Multishot firearms were known at least as early as the first part of the 16th century. Most of the primitive breechloaders used separate chambers that were wedged or screwed into place. A few used metal tubes surprisingly similar to modern metallic cartridges. All suffered from an excessive leakage of gas and flame and from the fact that fouling soon made them difficult or impossible to use. Multishot arms most often employed multiple barrels and locks, revolving cylinders or a Roman candle effect in which the priming charge set off the foremost explosive charge in a series of charges loaded one on top of the other in the barrel and a powder train carried the flame onward, setting off one shot after another until the barrel was empty.

1. **Lorenzoni System.**—During the first half of the 17th century an especially good breech-loading and repeating flintlock firearm was developed, possibly by the Florentine gunmaker Michele Lorenzoni, and its use soon spread throughout Europe and, by the early 18th century, to America. In Europe the mechanism is known today as the *système Lorenzoni;* in the United States it is called the Cookson type, since the first such gun described in an American publication bore the name of the British gunsmith John Cookson. In this firearm, the balls and powder were placed in separate tubular magazines in the stock. A single backward-and-forward motion of a lever on the reverse side of the gun caused a special revolving breechblock to select a ball and a charge of powder from their respective magazines, place them in their proper position in the barrel, prime the flashpan and set the gun at half cock.

Some guns based on this system were designed to fire five, six or seven shots. It was an excellent and efficient system, but its construction required the skill of a master gunsmith. It also suffered from the hazard of the possible ignition of the whole powder magazine if the breech were not closed properly at the time of discharge. Despite these drawbacks, guns of this type were still being manufactured as late as 1800.

2. **Ferguson Rifle.**—Among the breechloaders, the 18th century witnessed the development of several systems which, as far as is known, were the first such arms used by soldiers. France tried one in 1723, and Austria armed its dragoons and light cavalry with them in 1770. Both experiments were of brief duration. In England an excellent breech-loading rifle was developed slowly during the first half of the century, reaching its peak in time for the American Revolution. The man who brought this rifle to its fullest development was Patrick Ferguson of the British army, and 100 or 200 men under his command were armed with this rifle for a short time during the American Revolution.

The Ferguson system was simple and efficient. A plug operating on a screw passed vertically through the breech of the barrel. The lower end of this plug was attached to the trigger guard, which acted as a handle. One revolution of the trigger guard in a clockwise direction lowered the plug until its top was flush with the bottom of the chamber in the breech. The plug being thus depressed left a hole in the top of the barrel which communicated directly with the bore. In order to load the gun the barrel was pointed slightly downward, the trigger guard revolved and the plug accordingly depressed. A ball was dropped into the hole in the top of the barrel whence it rolled forward until stopped by the lands of the rifling at the end of the chamber. The ball was followed by a charge of powder, measured simply by filling the cavity of the chamber behind the ball. The trigger guard was then revolved in a counterclockwise direction, closing the opening in the top of the barrel and forcing out any excess powder that might have been poured in. The pan was then primed separately, and the piece was ready for firing. The Ferguson rifle was probably the finest military firearm used during the American Revolution. It could be fired six times in a minute. It was accurate and was equipped with a bayonet. Nevertheless, it did not meet with official favour, and its active career was brief.

3. **Hall Rifle.**—It remained for the United States in 1817 to adopt the first breech-loading rifle as a standard military arm that saw long service. This was the Hall rifle, a flintlock weapon invented by John H. Hall of Yarmouth, Me., in 1811. In Hall's system a pivoted breechblock containing the lock mechanism and the chamber could be tipped up, loaded and then closed and locked in position by a spring catch. Since the junction of the chamber and the barrel was a simple butt joint, there was considerable gas leakage. Nevertheless, the Hall system, in both flint and percussion, was used for rifles and carbines for almost 25 years, the last-known carbine contract being awarded in 1850. The Hall rifle also enjoyed the distinction of being the first military firearm made in America with completely interchangeable parts.

4. **Dreyse Rifle.**—In Europe the breechloader was also making its appearance on the scene. Between 1827 and 1829 Johann Nikolaus von Dreyse invented his famous needle gun, which was adopted by the Prussian army in 1848. Dreyse's gun used a long, sharp firing pin to pierce the charge of propelling powder and strike the fulminate primer located in the base of the bullet. This needlelike firing pin gave the gun its name and also proved to be the arm's major weakness, for it tended to break or warp during heavy use. Dreyse's breech-closing mechanism included the bolt action; it was the first appearance of that very important device and was probably Dreyse's greatest contribution to firearms design. Despite the weakness of the firing pin, the needle gun remained standard in Prussia until it was replaced by the Mauser in 1871. In France, Antoine Alphonse Chassepot developed a rifle similar to Dreyse's but with a shorter pin and an improved cartridge. Elsewhere, Norway adopted a breechloader with a tip-up breech designed by F. W. Scheel in 1842, and in other countries experimentation continued.

5. **Sharps Rifle.**—In the United States, Christian Sharps of Philadelphia, who had worked under Hall making the latter's breech-loading flintlock rifles, invented the first really satisfactory breech-loading system in 1848. In Sharps's guns the breechblock was raised and lowered in a vertical mortise by the action of a lever that usually served also as a trigger guard. When the breechblock was lowered, the chamber was exposed and a paper or linen cartridge could be inserted. As the breechblock was raised, it sheared off the rear of the cartridge, exposing the powder to sparks from the priming pellet, which was not contained in the cartridge but was fed automatically over the nipple by the action of the hammer. The Sharps action was very strong and tight. It allowed no gas leakage, and it would not bend or crack, even with excessively large charges. The basic Sharps principle, in fact, is still in use today on some heavy arms and small cannon in which strength is needed.

6. **Other Types.**—During the 1850s and 1860s a large number of breech-loading arms were developed in the United States, and most of them saw service in the Civil War. It was only the war, in fact, that permitted most of them to get into production, since they were not good enough to have survived in a civilian market. In 1863 the United States government decided to look toward the adoption of a breechloader for its principal infantry and cavalry arms. In 1866 it began the conversion of muzzle-loaders then in use, and in 1873 the first new breechloaders firing .45–70 (calibre .45 with 70 gr. of powder) metallic cartridges were manufactured. On these rifles and carbines, the top of the breech tipped up in a trap-door action allowing the insertion of the cartridge. The new metallic cartridge eliminated all worries about gas leakage, and the action was strong. The .45–70 Springfield remained the principal U.S. military rifle until it was replaced by the Krag-Jörgensen in 1892.

Meanwhile experimentation with repeating weapons had not ceased with the *système Lorenzoni* discussed above. The United States tried the Roman-candle principle briefly during the Revolution and again during the War of 1812. Sliding locks that could fire superimposed loads one after another also were tried. The British experimented with multibarreled carbines, and there were other attempts to obtain greater firepower.

7. **Pistols and Revolvers.**—The pistol makers were the first to achieve real success, first with the revolving barrels of the pepperbox and then with the revolving cylinders of the Colt revolver. The pepperbox pistols consisted of a series of barrels grouped around a central axis; they could be fired one after another by a

single hammer. The system had developed slowly over a century and a half, reaching its height during the 1830s and 1840s. By this time, all were percussion arms and all were muzzle-loading. On some the barrels revolved automatically; on others it was necessary to turn them by hand. Normally these pistols had from 3 to 6 barrels, but occasionally there were more—8, 10, 12 or even 18. When first developed, they were the fastest-firing guns of their time.

Like the pepperbox, the revolver also developed slowly over a period of centuries. It was made a really practical arm by Samuel Colt in 1830 when as a boy of 16 he whittled out a model of a pistol in which the cocking of the hammer automatically revolved the cylinder. Colt's revolvers were all percussion until after the close of the American Civil War. Each chamber in the cylinder was loaded separately with a combustible cartridge, and the individual nipple for each chamber was capped by hand. The Colt revolver was an excellent weapon and one of America's greatest contributions to small-arms technology. As such it was soon copied by a host of other designers, in both the United States and Europe, who sought in various ways to avoid the patents that Colt held in the United States, Great Britain and France and that protected his system until they expired in 1856. By that time the revolver was firmly established. (*See* also PISTOL AND REVOLVER.)

8. Later Development of Repeaters.—Further advances in the field of repeating arms waited upon developments in ammunition. The evolution of the cartridge is discussed below, but its tie-in with the invention of practical repeating arms is particularly important here, through the work of two U.S. gunsmiths, Horace Smith and Daniel B. Wesson. These two young artisans became friends in 1852 and together developed a new repeating pistol and rifle using a new cartridge. The repeating mechanism was an improvement on one devised some time before by Walter Hunt and Lewis Jennings. As strengthened and simplified by Smith and Wesson, it formed the basis for the Volcanic pistols and rifles of 1855 and, later, with further improvements by B. Tyler Henry, for the famous Henry and Winchester rifles.

Before the Henry rifle reached its full development, it acquired a serious competitor in the form of the Spencer rifle and carbine. Christopher M. Spencer patented his gun March 6, 1860. It was an excellent firearm with a tubular magazine for seven metallic cartridges in the butt. It was sturdy, did not get out of order easily and was a great favourite with troops during the Civil War, when more than 100,000 Spencer arms saw active service. After the war the Spencer company was purchased by Winchester, and manufacture of Spencer arms was discontinued.

During the second half of the 19th century almost all nations experimented with various repeating rifles and carbines. By 1900 almost all had adopted one form or another as their principal small arm. The Norwegians and Danes adopted the Krag-Jörgensen, as did the United States, after much experimentation with other systems, in 1892. The British adopted first the Lee-Metford in 1888, followed by the Lee-Enfield in 1895. The Germans chose the Mauser, and so on throughout Europe. All were bolt-action rifles with magazines for five or six shots. A few nations preferred the Winchester system, however, either manufacturing their own variations or buying arms directly from Winchester. Only the most backward clung to the old single-shot guns.

D. Ammunition

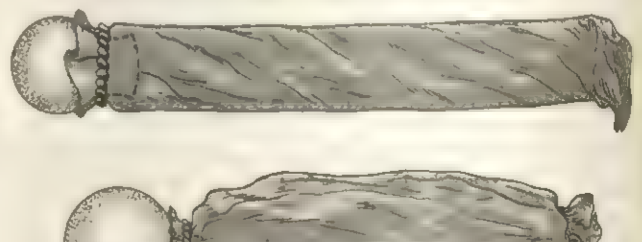
Many of the steps in the evolution of firearms design have been directly related to corresponding improvements in the types of ammunition available. It is impossible to understand one without the other.

1. Gunpowder.—First, there were improvements in gunpowder (*q.v.*) itself. From the time of its discovery in Europe, sometime in the 13th century, until the appearance of smokeless powder late in the 19th century, all gunpowder was composed of three essential ingredients—saltpetre, charcoal and sulfur. The earliest method of making powder was to grind these three ingredients and mix them together. This resulted in a black substance of about the texture and consistency of fine coal dust. This form of powder was known as meal or serpentine powder. It was explosive, but it had many distinct disadvantages. It was first of all unstable. It was also hygroscopic, and when stored in a damp place it formed lumps that would burn slowly but would not explode. If the powder were rammed into the barrel with too great pressure it would form a lump and lose much, if not all, of its explosive force. Even when the complete charge remained in its mealy form, not all of it burned, and a gummy residue quickly fouled the gun barrel.

There were other disadvantages that related primarily to shipping and storing serpentine powder. Because of the dusty nature of its contents, a quantity of highly explosive dust was given off every time a container of the powder was shifted or jarred, which created a distinct safety hazard. Also, since the mixture of the three ingredients was purely a mechanical alignment, there was a tendency for these materials to settle out according to their specific densities if the powder were allowed to stand undisturbed for any period of time. Consequently, the periodic inverting of powder kegs was an important part of the routine of every well-run magazine.

Despite these obvious drawbacks, serpentine powder continued in use as the main propelling charge in firearms until about the middle of the 16th century, when "corned" powder began to supersede it. Corned or granulated powder was made by moistening the mixture and squeezing the resultant mass through sieves under considerable pressure. Corned powder overcame most of the disadvantages of the serpentine powder, and, in addition, it burned much more rapidly and thus added considerable force to the explosion. For these reasons, corned powder quickly supplanted the earlier variety as the main propellant for firearms. Serpentine powder, however, remained as the principal material for priming for at least another century.

At first, gunpowder was carried in flasks or horns. During the first half of the 16th century a device known as the bandolier appeared and became popular in many parts of Europe and in America. It consisted of a number of cylindrical containers, each holding enough powder for one charge, suspended from a belt that could be worn over the shoulder. Usually there was also a little flask for priming powder, and a bag for the bullets, which were carried separately.



FROM H. L. PETERSON, "ARMS AND ARMOR IN COLONIAL AMERICA": REPRODUCED BY PERMISSION OF THE STACKPOLE COMPANY, HARRISBURG, PENNSYLVANIA

FIG. 3.—EARLY TYPES OF POWDER AND BALL AMMUNITION: (A) LATE 16TH-CENTURY PAPER CARTRIDGE WITH BALL ATTACHED BY FLANGE; (B) 17TH-CENTURY PAPER CARTRIDGE WITH BALL ATTACHED BY SPRUE

2. Cartridges.—None of these methods was entirely satisfactory. There was still the need for greater speed and safety. The answer lay in the individual cartridge. The first cartridges appeared in the second half of the 16th century. At first they were simply charges of powder wrapped in paper. The ball was still loaded separately. Before 1600, however, several methods had been devised for attaching the ball, and during the next century these were refined until by 1700 the ball was wrapped inside the paper with the powder in one self-contained unit. These paper cartridges were in widespread use in America by 1650, and they were adopted as standard by most European armies during the opening years of the 18th century. They remained standard as long as muzzle-loading arms were used. To use one of these

cartridges, the soldier simply bit off the end of the paper tube, poured a little powder into the pan (if the gun was a flintlock), dumped the rest down the barrel and then rammed the ball and the paper down on top.

The early breech-loading and multishot weapons brought forth a great variety of cartridges. Some were wrapped or covered with paper, some with linen, animal tissue, collodion, metal, rubber and even other materials. All, however, lacked an internal primer that would supply the spark to ignite the charge.

In 1846 a Paris gunsmith named B. Houiller patented two cartridges that represented the first completely self-contained metallic-cased ammunition for small arms. One, known today as the pin-fire cartridge, held a small percussion cap mounted sideways in its base with a pin poised above it and protruding through the case itself. A blow from the hammer drove the pin into the cap, exploding it and setting off the charge. The other was the first of the rim-fire cartridges with a compound of fulminate of mercury in a cavity that encircled the rim of the cartridge base. A blow from a sharp-nosed hammer set this cartridge off in the same manner as the percussion cap that utilized the same explosive compound. In these first rim-fire cartridges the fulminate of mercury also supplied the propelling force, and, as a consequence, the charges were weak. In 1856 Smith and Wesson produced a rim-fire cartridge with a suitable charge of black powder as well as the detonating compound, and thus developed a truly useful cartridge.

The rim-fire cartridge gained widespread acceptance for every sort of firearm, but it had certain defects which still had to be overcome. It was difficult to spread the detonating compound evenly all around the rim cavity, and if the rim cavity was not filled all the way around there was a chance that the hammer would strike a blank space and produce a misfire. Also it was difficult to produce a brass case strong enough to withstand the pressures developed by the heavy charges in large-calibre guns and still weak enough in the rim to indent easily under the blow of the hammer.

The final answer to these problems was found in the centre-fire cartridge that was developed at about the same time as the rim-fire. Many men in England, France and the United States contributed to its evolution. In 1852 Charles Lancaster of Great Britain invented a cartridge with an internal perforated disk forming a small cavity at the base. This cavity was filled with fulminate and exploded when a blow from the hammer crushed the compound between the base of the case and the inner disk. This system was improved by A. M. Pottet of France in 1857 with a design that allowed a percussion cap to be inserted in a domed cavity in the base. Colonel Boxer of the British Royal laboratory improved Pottet's design in 1867 and produced the basic form of one of the two standard types of centre-fire cartridges that are the universal standard—the type in which the anvil or inner surface against which the priming compound is crushed is made separately from the cartridge case. The other standard type was developed by Col. (later Gen.) Hiram Berdan of the United States in 1866. In it the anvil was formed as an integral part of the case. After 1867 there were minor improvements in design and manufacturing techniques, and smokeless nitrocellulose powder replaced black powder as the propelling charge, but there were no basic changes in these two systems. (Hd. L. P.)

II. 20th-CENTURY SMALL ARMS

By the year 1900, foot soldiers in the armies of all leading military powers had at their disposal a wide variety of high-powered repeating rifles, lightweight carbines, semiautomatic pistols and revolvers and fully automatic machine guns. All these weapons fired ammunition that was far superior to that in use a generation earlier, much of it being filled with the propellant known as smokeless powder. Lead bullets covered with a thin jacket of cupronickel or other metal could be fired with great accuracy and lethal effect at man-size targets more than 500 yd. distant; extreme ranges ran up to 3,000 yd. and more. Military small arms had not only become lighter, more dependable, more accurate and longer ranging than those of earlier years, but they had become faster firing. Rates of small-arms fire rose tremendously in the period 1890–1910 as magazine-fed repeating rifles and automatic machine guns became standard military weapons.

These developments in small arms changed the whole face of war during the first two decades of the 20th century. They helped bring to an end the long reign of the cavalry arm. They turned frontal attacks by infantry on well-defended positions into mass slaughter. The new small arms, along with improved artillery, forced armies to build elaborate fortifications or seek safety in underground trenches. For a time they so strengthened the defense that no army could hope to take the offensive against a determined enemy and gain more than a few miles without suffering heavy casualties. The new weapons led to a long period of tactical stalemate on the western front during World War I, a condition that stimulated development of a new weapon, the armoured tank, that was destined to play a major role in World War II. (*See* Tank.)

During the period after 1880 ingenious inventors brought forth such a great variety of small arms that a mere listing of them all, with but brief description of each, would fill a large book. As machine guns, pistols and revolvers are treated in separate articles, they are mentioned here only incidentally; this article is focused mainly on rifles, carbines and submachine guns, but does not attempt to describe or name all the countless varieties that have seen military service. It is limited to a discussion of broad trends of development, with specific mention of some of the most notable types. As the development of ammunition for small arms has been so closely related to development of the weapons themselves, some discussion of small-arms ammunition has been included.

A. Repeating Rifles

Among the many different types and models of military rifles that appeared during the 1890s and early 1900s, certain common characteristics stand out. They all weighed between 8 and 10 lb., had barrels ranging from 24 to 30 in. in length, and were between 43 and 52 in. in over-all length. The new rifles were shorter and lighter than their predecessors of the mid-19th century, and they fired ammunition of about calibre .30 (7.62 mm.) rather than the heavier bullets of earlier years. They were magazine-fed and were usually loaded by five-round clips or chargers. They were designed to be rugged enough to withstand hard usage and to keep firing even under the worst weather conditions, though further improvements along these lines were constantly being sought.

The military rifle most widely used and copied during the first quarter of the 20th century was the German Mauser. It had first appeared in Germany in the 1880s and over the years had been adopted by a score of countries other than Germany. No other military rifle ever enjoyed such wide use. A bolt-action piece loaded with a five-shot charger, the Mauser was produced in many models, including both a standard rifle (*Gewehr*) and a shorter carbine (*Karabiner*). Improved models of 1898 served as the basic weapons of the German infantry through World War II.

1. Bolt Action.—The term bolt action calls for some explanation because it is fundamental to an understanding of all modern rifles. The bolt or breech mechanism of a rifle operates on the same principle as the sliding bolt used to lock a door, though it is far more complicated and its manufacture calls for much more precise workmanship. The typical rifle bolt is a steel cylinder containing a firing pin; it slides back and forth in the frame or receiver. The bolt has a solid head and, like the door lock, is fitted with a projecting handle with a round knob at the end. One or more lugs near the front or rear of the bolt fit into slots in the receiver and hold the bolt firmly in place against the base of the cartridge in the chamber when the weapon is to be fired. When the trigger is pressed, a long slender firing pin within the bolt is pushed forward by a spring to strike the primer cap in the base of the cartridge. After firing, the shooter grasps the round knob, raises it slightly to turn the bolt and disengage the lugs, and then slides the bolt back to extract the fired case. Forward movement of the bolt strips a fresh cartridge from the magazine and forces it into the chamber; a downward pull on the knob locks the bolt in place. Straight-pull rifles such as the Austrian Mannlicher and

Canadian Ross require no turning to permit the bolt to slide back.

2. U.S. and British Rifles.—In 1892 the U.S. army adopted the Norwegian Krag-Jörgensen rifle in preference to the Mauser and many other weapons tested. Produced as both a rifle and a carbine, it underwent several modifications during its short period of service. In 1903 it was replaced by the Springfield rifle, a modified copy of the Mauser manufactured at the U.S. army's Springfield (Mass.) armoury and at Rock Island (Ill.) arsenal. An excellent rifle, the Springfield served as the basic infantry weapon of the U.S. army for the next 35 years and helped win for the U.S. soldier a world-wide reputation for rifle marksmanship. As late as World War II, modified Springfields with telescopic sights continued to serve as sniper rifles. The 1903 Springfield was a calibre-.30 weapon with a 24-in. barrel, about 6 in. shorter than the Krag-Jörgensen. For this reason, no carbine type of Springfield was considered necessary for mounted troops.

In the British army, too, during the South African War, there was a demand for a shorter rifle with a magazine feed that could serve for both infantry and cavalry, thus eliminating the carbine. A committee was appointed in 1900 to study the matter, and in Dec. 1902 a new rifle described as the "short, magazine, Lee-Enfield, Mark I" (S.M.L.E.) was adopted. It replaced the Lee-Metford (1888) and the long Lee-Enfield (1895). Its over-all length of 44½ in. was approximately 5 in. shorter than the "long rifle." The S.M.L.E. combined the bolt action developed by an American, James P. Lee, with the Enfield type of rifling. It fired rimmed calibre-.303 ammunition, and its bolt locked on the rear as did that of the Krag-Jörgensen. An improved model known as Mark III was adopted in 1907, followed by Mark III* in 1918, a model with no cutoff (a device to permit reloading with single cartridges) and no long-range sights. The S.M.L.E. was fed from a ten-shot box magazine loaded by five-shot clips or single rounds. Though not quite so accurate as the Mauser or Springfield at long range, the S.M.L.E. held twice the number of cartridges and was faster firing.

At the outbreak of World War I in 1914 the British army was about to adopt a new Enfield rifle chambered for calibre-.276 rimless ammunition. But, in the face of war, the British government abandoned its plans for mass production of a new rifle and new type of ammunition. Instead, the .276 rifle was modified to fire the standard calibre-.303 rim-type cartridges, and orders for its production were placed with factories in the United States. Production was slow, and by 1917, when the United States entered the war, British need for these rifles had declined while the U.S. forces were desperately short of Springfields. To meet this emergency the British rifles in production were converted to fire U.S. calibre-.30 ammunition and were standardized as the M1917, or American Enfield. Over 2,000,000 of these rifles were produced during the war at three U.S. plants. The Enfield thus became, in terms of quantity production, the leading U.S. rifle of World War I. The rifling of the Enfield, unlike that of most rifles of its era, consisted of five grooves with a left-hand twist.

Other notable military rifle types of the early 20th century were the Austrian Mannlicher, Russian Mosin-Nagant, Japanese Arisaka, French Lebel and Canadian Ross. As noted above, the Ross and the Mannlicher were of the straight-pull type. They suffered from frequent jamming in mud or sand, and the Ross was dropped during World War I because of frequent malfunctioning.

B. Semiautomatic Rifles and Carbines

All the rifles described above were repeaters, not automatics or semiautomatics. They fired one shot each time the trigger was pulled (or squeezed, to use the preferred military term), but after each shot the user had to pull back the bolt by hand to eject the spent case and then slide it forward to bring another cartridge into the chamber. By 1900, and even much earlier, firearms designers had begun to think seriously of developing semiautomatic or self-loading rifles that would automatically bring a fresh cartridge into firing position after each shot, using the force of recoil or gas pressure for the purpose. France and Germany experimented with semiautomatics on a small scale during World War I, France issuing a few of the St. Étienne model and Germany some Mondragons designed by a Mexican officer of that name. But neither type functioned well in combat. The United States was the first nation to achieve real success in this field.

1. Garand Rifle.—After a generation of experiment and test (mostly with calibre-.276 rifles during the 1920s) the U.S. army in Jan. 1936 adopted a calibre-.30 semiautomatic rifle developed by John C. Garand, a civilian engineer at the Springfield armoury. Replacing the 1903 Springfield, the Garand rifle, or M1 as it was officially designated, became the basic weapon of the U.S. infantry in World War II and the Korean war. It was a calibre-.30 gas-operated weapon fed by an eight-round clip and weighing about 9½ lb. Its 24-in. barrel was rifled with four grooves.

As the gas-operating mechanism of the M1 differed from the manually operated bolt described above, it merits brief explanation at this point. A small hole or gas port on the underside of the barrel near the muzzle permitted part of the propellent gases to escape into a small cylinder holding a piston that was connected to the bolt. As the gas pressure forced back the piston and the bolt, the empty cartridge case was ejected and the hammer was cocked. A spring then forced the bolt forward. As it moved forward the bolt stripped the top cartridge from the magazine and seated it in the chamber ready to fire. Gas pressure thus performed automatically the reloading task formerly done by hand. For this reason, weapons of this type are often called self-loading or autoloading.

With the Garand rifle the U.S. army adopted a new cartridge of the same dimensions as its predecessor but with a lighter bullet —152 gr. as compared with 172 gr. Some critics considered this a backward step but military authorities defended it on two main counts: first, the lighter bullet caused less recoil and thus helped the training of recruits in marksmanship; second, the extreme range of the heavier bullet—over 3,000 yd.—was of little military value and proved dangerous in target practice on 1,000-yd. ranges. Rifles were seldom used in combat beyond about 600 yd.

2. Johnson Rifle.—Soon after the Garand rifle was adopted in 1936, and while Springfield armoury was tooling up for its manufacture, a serious competitor appeared on the scene and precipitated a major rifle controversy in the United States. This was the semiautomatic rifle designed as a private venture by Melvin M. Johnson. It had a 22-in. barrel, weighed about 9½ lb., operated on the short-recoil principle rather than the gas system of the Garand and was fed with a ten-shot rotary magazine. Largely because of defects that appeared in the first production models of the Garand, there arose a demand that the Johnson be adopted in its stead. But, after much public discussion and extensive tests in 1939–40, the U.S. army decided to retain the Garand after improving its gas take-off. The Johnson rifle meanwhile was adopted by the Netherlands Indies forces and saw some service with U.S. marine corps units in World War II.

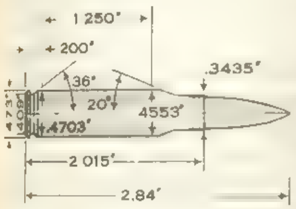
3. German and Soviet Rifles.—The German and Soviet armies also employed semiautomatic rifles to some extent in World War II, but their performance fell short of expectations. The German semiautomatics that saw most service were the models *Gewehr* 41, 41-W and 41-M, *Gewehr* 43 and *Karabiner* 43. All were gas-operated and fired 7.92 mm. cartridges. The 41 and 41-W did not have a gas port in the barrel but employed a cone attached to the muzzle to trap some of the expanding gas. As early as 1936 the Soviet army adopted the 7.62-mm. Simonov gas-operated rifle, which was capable of both semiautomatic and fully automatic fire. Its gas port and piston were on top of the barrel, thus resembling the German *Gewehr* 43. Another Soviet gas-operated semiautomatic rifle, the 7.62-mm. Tokarev, appeared in 1938 and 1940 models. But throughout World War II most German and Soviet troops carried manually operated repeating rifles or carbines.

4. Carbines.—Early in World War II the U.S. army again adopted a carbine. This time the carbine did not serve as a cavalry weapon, as had carbines of the 19th century, but as a defensive weapon for service troops and operators of crew-served guns whose positions might be overrun in mobile warfare. For such troops the pistol had too short a range and too little accuracy, while the standard rifle was unnecessarily heavy and

powerful. A gas-operated weapon developed by the Winchester Repeating Arms company, the U.S. carbine M1, was similar in design to the Garand rifle but was 8 in. shorter. It weighed only about 5½ lb. and fired a different type of ammunition from a 15-round magazine. The M2 carbine that appeared late in the war had a selector switch that gave the user a choice of fully automatic or semiautomatic fire; it was fed from a 30-round magazine. Some airborne troops carried a carbine (M1A1) with a folding skeleton metal stock.

The British army also employed light rifles or carbines for certain types of troops in World War II. Rifle No. 4, Mark I* weighed only 6¾ lb. and was sometimes described as a carbine. Rifle No. 5, Mark I* was a British carbine especially designed for jungle fighting.

Most rifles and some carbines were equipped with detachable bayonets or bayonet knives for hand-to-hand combat, though bayonet fighting was not common in World War II or the Korean war and many observers considered the bayonet obsolete. Another attachment, the grenade launcher, enabled troops armed with rifles and carbines to hurl antitank and antipersonnel grenades for short distances. (*See* BAYONET; GRENADES.)



FROM J. S. HATCHER, "HATCHER'S NOTEBOOK"; REPRODUCED BY PERMISSION OF THE STACKPOLE COMPANY, HARRISBURG, PA.

FIG. 4.—7.62-MM. NATO CARTRIDGE

C. Rifle Developments After World War II

At the end of World War II small arms were overshadowed by the atomic bomb and long-range guided missiles, but they were by no means forgotten. All leading military powers sought to improve their small arms in the light of World War II experience. The main objectives were to make rifles lighter and faster firing. The United States and Great Britain, joined in 1949 by other members of the North Atlantic Treaty organization (NATO), made determined efforts to improve their military rifles and also to standardize their equipment, particularly ammunition. Agreement in this field was not reached overnight. The Korean war (1950–53) was fought in the main with small arms and ammunition of the World War II type.

1. New Cartridges.—Acceptance of a common small-arms cartridge came first, though not without some difficulty. The British government proposed that a rifle cartridge of calibre .280 (7 mm.) with lead-core bullet be adopted. The U.S. government favoured a more powerful cartridge of calibre .30 with steel core. After extensive tests the U.S. cartridge, known as T65 during its development at Frankford arsenal, gained NATO acceptance in Dec. 1953. Formal action to standardize the new ammunition for all NATO countries came in 1954. In line with the policy of using the metric system, the new round was officially designated Cartridge, NATO, Calibre 7.62-mm. Its length, 2.84 in., was ½ in. shorter than the existing U.S. M2 cartridge, and its weight was about 10% less. The saving in length and weight resulted from the use of improved propellants, such as ball powder, that required less space. The 165-gr. bullet of the new cartridge was capable of inflicting a fatal wound on personnel protected by standard steel helmets and body armour at a distance of approximately 1,000 yd. It was standardized for both rifles and machine guns. Many older weapons were converted to take the new round.

After years of studying ways and means of increasing the infantryman's chances of hitting his target when firing a small arm, the U.S. army in 1964 adopted the 7.62-mm. duplex ball cartridge (M198) that carried two bullets instead of one. The two bullets (each weighing 83 gr.) were generally similar in size and shape except that the base of the first contained a cavity into which the point of the second fitted. When the cartridge was fired the first bullet followed a normal trajectory but the second, which had a slightly slanted base, followed a spiral course around the flight path of the first. At 100 yd. the second bullet would strike somewhere within 11 in. of the point of impact of the first; at 300 yd. it would hit within about 31 in. of the first. The duplex cartridge could be fired in any 7.62-mm. rifle or machine gun.

2. FN Rifle.—The British government meanwhile brought forth a rifle, designated EM2, of unusual design and appearance, with pistol grip placed ahead of the magazine. It had a straight stock, weighed about eight pounds and was chambered to fire the .280 cartridge. The Belgian firm Fabrique Nationale d'Armes de Guerre entered the competition with another rifle, popularly known as the FN and officially designated the T48 in U.S. tests. It also was designed to fire the .280 cartridge. The ordnance department of the U.S. army carried on extensive development and test work at Springfield armoury with designs based generally on modifications of the Garand rifle to provide both automatic and semiautomatic fire. After field tests and prolonged discussion, reaching up to the level of the British prime minister and U.S. president in 1952, the British government in 1954 adopted the FN rifle, which meanwhile had been altered to fire the 7.62-mm. cartridge and designed in light-barrel and heavy-barrel models. Belgium, Canada and Australia also adopted the FN rifle in 1954.

The FN was a gas-operated weapon with selector switch for either semiautomatic or fully automatic fire. It weighed about 8¾ lb., had a 20-round magazine and extended to 44.4 in. over-all length, including flash suppressor. In appearance the FN was markedly different from weapons of the Garand type, for it had a separate pistol grip behind the trigger guard and a carrying handle that could be swung up above the barrel. Because its butt could be swung downward on a hinge, thus permitting access to the receiver mechanism, the FN was easy to field-strip and clean.

3. M14 and M15 Rifles.—In 1957 the U.S. army announced

Principal Dimensions and Weights of Representative Military Rifles

Country	Rifle	Calibre	Approximate weight (pounds)	Length over-all (inches)	Type
Switzerland	Schmidt-Rubin (1889)	7 5 mm.	9.8	51 5	straight pull repeater
U.S.	Krag Jorgensen (1892)	.30 in.	9 2	49	turning-bolt repeater
Austria	Mannlicher (1895)	8 mm.	8.4	50	straight pull repeater
Germany	Mauser (1898)	7 92 mm	9	49 4	turning bolt repeater
Russia	Mosin Nagant (*c.* 1900)	7.62 mm.	9	51 8	turning bolt repeater
U.S.	Springfield (1903)	.30 in.	8.7	43.2	turning bolt repeater
Japan	Arisaka (1905)	6.5 mm.	9	51	turning-bolt repeater
Great Britain	S M L E Mark III (1907; originally adopted 1902)	.303 in.	8.7	44.5	turning-bolt repeater
France	Lebel (1907; originally adopted 1886)	8 mm.	9.2	51.5	turning-bolt repeater
U.S.	Enfield (1917)	.30 in.	9.2	46 3	turning-bolt repeater
U.S.	Garand M1 (1936)	.30 in.	9.6	43.6	gas operated semiautomatic
Japan	Model 99 (1939)	7.7 mm.	7.7	45	turning-bolt repeater
U.S.S.R.	Tokarev (1938–40)	7.62 mm.	8.8	47.8	gas-operated semiautomatic
Germany	*Gewehr* 41-W (1941)	7.92 mm.	10.8	45	gas operated semiautomatic
U.S.	Carbine M1 (1941)	.30 in.	5.5	35.6	gas operated semiautomatic
Germany	*Gewehr* 43 (1943)	7.92 mm.	9	43	gas-operated semiautomatic
Great Britain (also Canada, Australia and Belgium)	Fabrique Nationale (FN), light-barrel version (1954)	7.62 mm.	8.7	44.4	gas-operated semiautomatic and automatic
U.S.	M14, light barrel (1957)	7.62 mm.	8.7	44.1	gas-operated semiautomatic and automatic
U.S.	M15, heavy barrel (1957)	7.62 mm.	13.7	45.5	gas-operated semiautomatic and automatic
U.S.	M16 (1960)	5.56 mm.	6.4	39	gas-operated semiautomatic and automatic
U.S.S.R.	AK (Avtomat Kalashnikov)	7.62 mm.	10.5	34.2	gas-operated semiautomatic and automatic
U.S.S.R.	SKS (Simonov carbine)	7.62 mm.	8	40.1	gas-operated semiautomatic

that it had adopted a new semiautomatic rifle designated the M14, known during its test period as the T44. A gas-operated weapon, the M14 was similar in design to the Garand, or M1, which it was eventually to replace. Its weight was 8.7 lb., approximately 1 lb. less than the Garand. Like the Garand, it was developed at Springfield armoury. It fired the standard 7.62-mm. NATO cartridge, had a magazine capacity of 20 rounds and could be fitted with a selector switch to permit semiautomatic or fully automatic fire. A notable feature of both the U.S. rifle and the FN was a slotted flash suppressor at the muzzle. The new U.S. rifle, like the FN, was manufactured also in a heavy-barrel version (M15) that weighed approximately 13¾ lb. and was better adapted for fully automatic fire. This heavy-barrel version was equipped with a bipod support for the muzzle.

Combined with the M60 ground machine gun, the new rifle formed a complete family of small arms for the U.S. army, all firing the same ammunition. The M14 rifle was designed to replace the Garand rifle, the carbine and the M3 submachine gun, except as the latter was used on vehicles. The heavy-barrel M15 was to replace the Browning automatic rifle (BAR) as the automatic weapon of the infantry squad. The 23-lb. air-cooled light machine gun was to replace three existing models of calibre-.30 ground machine guns. The new rifle and machine gun would thus reduce the number of weapons in the U.S. army's small-arms system from seven to three, greatly simplifying logistical support and shortening the time needed for troop training.

4. M16 Rifle.—A calibre-.22 (5.56-mm.) rifle came into the U.S. military picture in 1960 when the air force adopted it as a replacement for the calibre-.30 carbine carried by guards at air bases. Designated the M16, it had been developed commercially in the late 1950s and had come to be known as the AR-15. It was a gas-operated, magazine-fed rifle capable of both semiautomatic and fully automatic fire, and weighed only 6.4 lb. The army procured comparatively small quantities of the M16 for use by airborne troops and special forces.

5. Soviet Rifles.—The Soviet Union also adopted a new family of infantry weapons during the 1950s, including a rifle, submachine gun (*see* below) and light machine gun. All fired a standard short 7.62-mm. cartridge. The Soviet rifle, described as a semiautomatic carbine, was a gas-operated weapon with ten-round magazine and an attached folding bayonet. It weighed about 8½ lb. and had a stock of laminated wood. The Soviet light machine gun was designed to provide fully automatic fire for the rifle squad. It weighed about 14½ lb. and was fed by a drum-type magazine.

D. Automatic Rifles

Automatic rifles, sometimes classed as light machine guns, represent attempts to combine the burst fire of the machine gun with the mobile flexibility of the rifle. Their combat employment reached its peak in World War II, though several models saw service in World War I. The typical automatic rifle was a shoulder-fired weapon weighing between 20 and 30 lb. and equipped with a bipod to support the muzzle when firing from a prone position. One of the earliest weapons of this type was the 28-lb. gun developed by the U.S. army officer Isaac N. Lewis. The British army adopted the Lewis gun as a platoon weapon just before the outbreak of World War I, and it proved to be a major factor in British infantry firepower during 1914–18. It was later adopted by the U.S. navy but not by the U.S. army.

The most famous U.S. army weapon of this type, the calibre-.30 Browning automatic rifle (BAR), appeared too late to see service in World War I, but it served well for the next 40 years as the basic automatic weapon of the rifle squad. It was a shoulder-fired, gas-operated rifle weighing about 20 lb. complete with bipod support. A comparable British weapon that saw extensive service in World War II was the 22–26-lb. Bren machine gun, a weapon that by U.S. standards would be classed as an automatic rifle. The Bren took its name from Brno, Czech., where it was originally manufactured, and Enfield, Eng., where it was further developed. It was basically a calibre-.303 weapon but was produced also in 7.92-mm. size. A selector switch permitted either semiautomatic or fully automatic fire.

The German army used the MG34, developed secretly in 1934 in violation of the Versailles treaty, and the MG42, which had the extremely high cyclic rate of 1,500 shots per minute. For paratroops the Germans in 1944 adopted a special gas-operated 7.92-mm. rifle named *Fallschirmjägergewehr* 42 (FG42). It had a selector switch that permitted either semiautomatic or fully automatic fire. With sling, bayonet and bipod it weighed about 10½ lb. A 13-lb. light machine gun developed in 1940–41 by Melvin M. Johnson saw some service with U.S. troops, the Dutch and other Allied nations in World War II.

E. Submachine Guns (Machine Carbines)

Since their introduction by the Germans and Italians in World War I, these weapons have been known variously as machine carbines in Britain, machine pistols in Germany and submachine guns in the United States. They are small, light weapons standing between the pistol and the rifle. They are far simpler in design than machine guns but they offer rapid-firing action, either semiautomatic or fully automatic. In spite of their limited range and poor accuracy they have proved to be useful for paratroops, commandos, tank crews and others who require a light gun able to deliver a high volume of fire at close quarters. Submachine guns normally fire pistol-type ammunition—calibre .45 in the United States and 9 mm. in western Europe—at rates ranging from 450 to 900 rounds per minute. Nearly all submachine guns have short barrels and heavy bolts and operate on the plain blowback principle. (*See* Machine Gun.)

Among the earliest weapons of this type to see military service were the Bergmann Muskete, the Villar-Perosa gun and the Beretta machine carbine. In the United States the most famous submachine gun was that patented in 1920 by a retired army officer, John T. Thompson. Popularly known as the Tommy gun, it gained notoriety as the weapon of U.S. gangsters during the prohibition era. The U.S. army adopted it in 1928, and the British army also used it. The Thompson submachine gun weighed about ten pounds and fired calibre-.45 ammunition from a circular drum holding 50 rounds or a box magazine holding 20 rounds.

During World War II and the Korean war, all major armies employed submachine guns in considerable numbers. They all developed weapons of simple design intended for cheap mass production and weighing between seven and ten pounds. Best known among them were the British Sten, Australian Austen and Owen, Finnish Suomi and German Schmeisser (MP38 and MP40). The latter with folding stock, weighed about nine pounds, fired 9-mm. ammunition and was carried by paratroops and motorcycle units of the German army in its 1940 blitzkrieg. Later types firing 7.92-mm. ammunition were known as MP43, MP44 and *Sturmgewehr* 44. Following the German example, the U.S. army first simplified the Thompson gun and then turned in 1942 to a new all-metal submachine gun weighing about nine pounds and firing calibre-.45 pistol ammunition from a vertical magazine under the bolt. Officially designated the M3, it soon took the nickname "grease gun" because of its resemblance to the tool used to grease automobiles. The Soviet army also replaced its 1940 model (calibre 7.62 mm early in the war to speed production. First came the M41 PPSh. fed by a 71-round drum, and then the PPS-43 with folding metal stock and curved 35-round box magazine. The M41 gun was extensively used by the North Koreans during the Korean war. A new calibre-7.62 mm. Soviet submachine gun appeared in the 1950s. It weighed about 9½ lb. and was fed by a 30-round curved magazine. It was produced in two models, one with a conventional wooden stock and the other with a folding metal stock.

F. Antitank Weapons

To meet the need for a powerful infantry weapon to attack armoured vehicles, the British army adopted the Boys antitank rifle. It was much like an ordinary bolt-action magazine rifle but fired a heavy, calibre-.55 armour-piercing bullet at high velocity. A little over five feet in length, the rifle weighed 36 lb. and could penetrate 24-mm. armour at 100 yd.

The most revolutionary light antitank weapon that appeared in World War II was the shoulder-fired U.S. rocket launcher famil-

iarly known as the bazooka. It was one of the simplest weapons employed during the war, and one of the most effective. The bazooka was an outgrowth of efforts to fire a powerful antitank grenade from a rifle and thus give the infantryman a means of defense against tanks. As the open-end rocket launcher had no recoil, it could fire a heavier round than could a conventional rifle. The bazooka gained its remarkable armour-piercing capability from the use of ammunition designed on the shaped-charge principle. Its chief limitations were its short range, lack of accuracy and backblast. Though its bore diameter of 2.36 in. far exceeded the normal upper limit of calibre .60, the bazooka was usually classed as a small arm, or, more properly, an infantry weapon. The same was true of the 3.5-in. bazooka that saw service in the Korean war (*see* BAZOOKA).

During the 1960s the U.S. Army adopted a 66-mm. high-explosive, antitank rocket that could be fired by one man from a bazookalike carrying case, which was then discarded. Known as a light antitank weapon (LAW), the rocket and case weighed only 4½ lb. but were capable of knocking out a heavy tank.

G. Recoilless Rifles

Another new type of infantry weapon, the recoilless rifle, made its appearance in combat toward the end of World War II and saw much service in the Korean war. It had some of the characteristics of an artillery piece but, because of its light weight, was often classed as a small arm. Ordnance designers had long dreamed of creating a powerful weapon that would have no recoil and would therefore need no heavy recoil-absorbing mechanism. The dream was realized by designing a gun that permitted some of the propellent gases to escape through exhaust ports in the breech, thus giving the weapon a forward thrust equal to the normal recoil. The 57-mm. recoilless rifle developed for the U.S. army weighed less than 50 lb. and could be fired by one man from the shoulder (though not in the usual rifle firing position) or from a bipod or tripod mount. A larger 75-mm. recoilless rifle also saw some service with the U.S. army in World War II; after the war, 105-mm. and 106-mm. models were developed to be mounted on a jeep or fired from a ground support. In 1959 the U.S. army announced a new 90-mm. recoilless rifle said to be capable of knocking out the heaviest tanks. It was a 35-lb., four-foot-long weapon that fired a shaped-charge projectile to an effective range of 500 yd.

Recoilless weapons are characterized by their simplicity, lightness and vented breechblocks. They fire high-explosive and antitank ammunition of unique design. Their cartridge cases are perforated to permit a portion of the propellent gases to escape to the rear; their shells are pre-engraved (*i.e.*, with rotating bands notched to fit the rifling). Recoilless rifles have lower muzzle velocity and much shorter range than do guns or howitzers of the same calibre. One of the greatest drawbacks to their tactical use is the rearward blast that betrays their location and makes them unsuitable for use in confined spaces.

H. Ammunition Developments

Remarkable progress was made in the improvement of small-arms ammunition during the period 1880–1910, with still further development during the next half century. Smokeless powder, improved brass cartridge cases, better bullet design, jacketed bullets, and clips and chargers for loading magazines—these were among the most noteworthy developments of the pre-World War I era. The earlier shift from bullets of spherical to cylindroconoidal shape had stimulated a trend toward smallbore rifles, since an elongated bullet had less diameter than did a spherical bullet of the same weight.

1. Bullets.—After P. M. E. Vieille's discovery of so-called smokeless powder in the 1880s, this propellant came into general use, ending the long reign of black powder. It gave off less smoke than black powder but was not completely smokeless. The high muzzle velocities (over 2,000 ft. per second) attained with smokeless powder were too much for conventional soft lead bullets to withstand. Bullets tore loose from the rifling, failed to spin properly and emerged from the muzzle unbalanced and hopelessly inaccurate. The answer to this problem lay in the jacketed or metal-patched bullet—a lead core covered with a thin shell of steel or cupronickel. Credit for making great improvements during the 1880s in the work of earlier pioneers in this field is usually given to Eduard Rubin, a Swiss army officer. For a time the French army adopted a different solution, using a solid bronze bullet in its 8-mm. round for the Lebel rifle, but virtually all other nations adopted jacketed bullets, or compound bullets, as they were sometimes called.

Soon after the turn of the century another change in bullet design occurred when the pointed bullet, known as the spitzer type, gradually supplanted the round-nosed bullet that had been standard in all armies for many years. Another apparently trivial but actually important development came in the early 1900s with the discovery that further streamlining of bullets, accomplished by giving them a slight taper or boat tail at the base, increased their range. An early example of this type appeared in World War I, the French 8-mm. boat-tailed round known as *Balle D.* Some nations followed the French example, but others kept flat-based bullets in service for many years. Meanwhile the substitution of gilding metal (copper and zinc) and allied compounds for steel and cupronickel in bullet jackets was found to lessen bore erosion and metal fouling. Later, during World War II, when copper was scarce, the United States turned to the use of steel jackets coated with a thin deposit of copper or gilding metal.

Most early-20th-century rifle cartridges produced chamber pressures between 35,000 and 50,000 lb. per square inch and muzzle velocities between 2,000 and 2,800 ft. per second. They fired bullets that were a little more than one inch long and weighed between 135 and 244 gr. Their powder charges normally ranged between 30 and 50 gr. Cartridges varied in calibre from 6 mm. (.236 in.) in the U.S. navy's Lee rifle to 8 mm. (.315 in.) in the French Lebel, and in over-all length from about 2.90 in. to 3.35 in. Fired from a firm support at a target 200 yd. distant, the best ammunition would group ten shots within a circle about five inches in diameter. Using U.S. Springfields, expert riflemen could consistently hit a 20-in. bull's-eye at 600 yd.

2. Chargers and Clips.—The introduction of magazine rifles in all armies put an end to the older method of feeding from a tube in the butt or under the barrel, and the use of chargers or clips for loading became standard practice. These terms are often used loosely and interchangeably, but, for the sake of historical accuracy, a distinction is best observed. A clip was originally a light metal openwork box in which a small number of cartridges, usually between five and ten, were placed, one on top of the other, against a spring at the bottom. They all faced the same way and their bullet ends were exposed. The entire clip filled with cartridges was slipped into the magazine, the so-called *en bloc* system of loading. It remained there until all its cartridges were fired, then dropped out at the bottom or, on some models, popped out at the top. In contrast, a charger was merely a flat strip of metal with its edges curled slightly to hook over the rims or extractor grooves of a row of cartridges, usually five. To load a rifle, the bolt was drawn back, the charger holding a row of cartridges was slipped into position above the receiver, and the cartridges were pushed down into the magazine against a spring-supported follower. The empty charger was then discarded. The Garand rifle was loaded with a double-row, eight-round clip; the Springfield and other Mauser-type rifles were loaded from five-round chargers.

3. Types of Cartridges.—Three types of central-fire cartridges —rimmed, semirimmed and rimless—deserve some explanation. The first was the type developed on the pattern of the earlier rimfire case but operated on a different principle. It was a central-fire case and its rim was a flange at the base that could be gripped by the extractor hook to remove the spent case from the chamber after firing. The rimless type had no protruding flange but had instead an annular groove just forward of the base into which the extractor could fit; the groove thus served the same purpose as the rim. The semirimmed case was essentially rimless, but its base section had a somewhat greater diameter than the main body of the case, thus giving it some of the characteristics of both the rimmed and rimless types. The widely used Mauser family of

rifles fired rimless cartridges, as did the U.S. Springfield. British, French and Russian rifles fired rimmed ammunition.

Another feature of most modern cartridge cases is their bottleneck shape. As bullets of smaller calibre came into use at the end of the 19th century, ordnance designers had to narrow the mouths of cartridge cases and fit the bullets into them like corks in bottles. Retaining the standard size for the main body of the cartridge case provided room for a large powder charge without resorting to use of a much longer and narrow case. Most 20th-century military rifle cases assumed the bottleneck shape, but that for the U.S. carbine did not, nor did ammunition for pistols and submachine guns.

Three years after adoption of the Springfield rifle the U.S. army standardized the calibre-.30 M1906 cartridge, familiarly known as the .30-'06. It had a rimless case and fired a 150-gr. flat-based bullet. After World War I this cartridge was replaced by a more powerful type, with a 172-gr. boat-tailed bullet, known as M1. It had an extreme range of 3,500 yd., or nearly two miles. Its lead core was hardened with antimony and its jacket was of gilding metal rather than cupronickel. As noted above, the M1 cartridge was supplanted by a lighter (152-gr.) flat-based round, known as M2, in the late 1930s when the Garand semiautomatic rifle came into use. It remained the basic rifle round of the U.S. army until the 1950s when its replacement by the 7.62-mm. NATO cartridge began (*see* above).

During the second quarter of the 20th century another major advance in small-arms ammunition came with the introduction of noncorrosive primers. In the early 1920s manufacturers of sporting ammunition had brought forth cartridges containing no potassium chlorate, a substance that left a moisture-attracting residue in the bore and caused rusting if the weapon were not cleaned promptly after firing. U.S. military authorities were reluctant to accept new primer compositions without exhaustive testing, particularly in view of their problems with defective primers during World War I. The first large-scale production of noncorrosive military primers in the United States came during World War II when commercial firms manufacturing carbine ammunition employed their own primers with good success. Germany made extensive use of noncorrosive primers during World War II for 9-mm. and 7.92-mm. ammunition. In 1949 the U.S. army standardized a noncorrosive styphnate mixture for small-arms primers.

4. Special Ammunition.—The most noteworthy trend in small-arms ammunition development in the 1910–60 period was the appearance of many different types of ammunition to serve special purposes. The impetus for this development came largely from the employment of machine guns for aircraft and antiaircraft fire. In addition to the familiar type known as ball ammunition there were added others known as armour-piercing, incendiary, tracer, armour-piercing-incendiary and special types for launching rifle grenades. Whatever the type, the term cartridge was used to designate the complete round, including cartridge case, propellant, primer and bullet.

Ball ammunition is most closely akin to the lead spheres of an earlier age and is chiefly an antipersonnel round. Its core is of lead alloy or common steel. Despite its name, the bullet fired by a round of ball rifle ammunition is not actually ball-shaped; it more nearly resembles a short cigar, with the front end pointed (but not sharply pointed) and the other end flat or slightly tapered. The bullet of the calibre-.45 ball cartridge for pistols, revolvers and submachine guns is more nearly ball-shaped, with rounded nose and flat base.

Armour-piercing (AP) cartridges assumed great importance during and after World War I as a means of attacking tanks, planes and other armoured vehicles. In the U.S. army there was a strong tendency during and after World War II to use AP ammunition for all purposes and do away with ball ammunition. The bullets of AP cartridges normally consist of a hardened steel core, gilding metal jacket and a small amount of lead to fill the space between the point of the core and the point of the jacket.

Tracer cartridges contain a chemical agent that emits light or smoke to enable the shooter to see where his shots are going. Such ammunition proved to be of special value in adjusting the aim of machine guns in aerial combat and, to a lesser extent, in setting fire to gasoline tanks of enemy planes.

Incendiary cartridges contain a chemical agent that ignites upon striking the target. They are highly effective against aircraft with light armour protection carrying heavy loads of fuel.

Grenade cartridges are designed to propel grenades (*q.v.*) from rifles and, in some cases, from carbines. They have no bullet but contain a powder charge that propels the grenade from a launcher fitted to the muzzle.

Efforts to improve rifle ammunition during the decades following World War I were directed toward reducing barrel erosion, eliminating muzzle flash and attaining improved ballistic performance. One of the most notable developments was a propellant known as ball powder that was widely used in British ammunition and in U.S. carbine ammunition during the 1940s and in many other types in later years. Brass remained the preferred metal for cartridge cases in spite of extensive experiments with steel, aluminum and plastics. Brass was strong, nonrusting and could easily be worked into the desired shape; it was also sufficiently elastic to expand slightly under the pressure of firing and then to contract for easy extraction.

Bibliography.—*Early Development:* As the literature dealing with firearms is huge, readers with special interests would do well to consult Ray Riling, *Guns and Shooting: a Selected Chronological Bibliography* (1951), which lists almost 3,000 titles. Some of the most generally useful books are the following: Josef Alm, *Eldhandvapen* (1933); Torsten Lenk, *Flintlåset* (1939); J. N. George, *English Pistols and Revolvers* (1938) and *English Guns and Rifles* (1947); James E. Hicks, *Notes on French Ordnance, 1717 to 1936* (1938); J. Margerand, *Armement et équipement de l'infanterie française* (1945); W. Keith Neal, *Spanish Guns and Pistols* (1955); Harold L. Peterson, *Arms and Armor in Colonial America, 1526–1783* (1956); John G. W. Dillin, *The Kentucky Rifle* (1924 and subsequent editions); Ned H. Roberts, *The Muzzle Loading Cap Lock Rifle*, 3rd ed. (1948); Arcadi Gluckman, *United States Martial Pistols and Revolvers* (1939) and *United States Muskets, Rifles and Carbines* (1948); Berkeley R. Lewis, *Small Arms and Ammunition in the United States Service* (1956); William A. Albaugh and E. N. Simmons, *Confederate Arms* (1957); James E. Serven, *Colt Firearms, 1836–1954* (1954); Alden Hatch, *Remington Arms in American History* (1956); Lewis Winant, *Pepperbox Firearms* (1952), *Firearms Curiosa* (1955), *Early Percussion Firearms* (1959); Charles E. Chapel, *The Gun Collector's Handbook of Values* (constantly being revised); and Herschel Logan, *Cartridges* (1948). (Hd. L. P.)

20th Century: British War Office, *Text Book of Small Arms* (1929); Edward C. Crossman, *The Book of the Springfield* (1932); Melvin M. Johnson, *Rifles and Machine Guns* (1944); Philip B. Sharpe, *The Rifle in America*, 2nd ed. (1947); Gladeon M. Barnes, *Weapons of World War II* (1947); Julian S. Hatcher, *The Book of the Garand* (1948); Arcadi Gluckman, *United States Muskets, Rifles, and Carbines* (1948); W. H. B. W. G. B. Allen, *Pistols, Rifles and Machine Guns* (1953); W. H. B. Smith, *Small Arms of the World*, 5th ed. (1955); Julian S. Hatcher, *Hatcher's Notebook*, 2nd ed. (1957) and *The Book of the Garand* (1948); Frank C. Barnes and John T. Amber (ed.), *Cartridges of the World* (1965). *See* also the magazines *Ordnance* and the *American Rifleman*, both published in Washington, D.C., and the annual *Gun Digest*, published in Chicago, Ill. (H. C. T.)

SMALL BUSINESS, a term used to denote units of nonagricultural business enterprise below an arbitrarily determined size limit. Since it is arbitrary, the size limit varies from country to country and from time to time. The designation is established to serve as a guideline in carrying out government policies that aid smaller business firms in one way or another, or that exempt such firms from certain regulations that apply to large firms. The basic concept underlying most efforts to aid small business is that the existence of such enterprises is desirable and that they need special assistance if they are to meet the competition of big business.

In the United States, during and after the war, firms with fewer than 500 employees were classified as small businesses. In Western Europe a much smaller figure was generally used. In France, for example, the General Confederation of Small and Medium Enterprises considered manufacturing firms as small if they had from 6 to 50 employees; those with fewer than 6 employees were legally classified as artisan enterprises. In Great Britain the Board of Trade classified as small those manufacturing enterprises with 10 or fewer employees and those retail firms with fewer than 5 branches and 25 employees.

Most students of the subject argue that quantitative criteria are unsatisfactory and that some qualitative characteristic such

as function or form of ownership, should be used in defining small business. Thus it has been held in the United States that all independently owned and operated enterprises that are not dominant in their field should be considered small businesses. In Europe some specialists have contended that all firms engaged in "particular" production as distinct from "mass" production should be included in the category of small business. Such qualitative criteria extend the coverage of the term to a larger segment of the business population than do most quantitative definitions.

In the United States, the antitrust legislation of the 1890s was advocated in part as a means of protecting small firms; other laws adopted in the early 1900s were intended to aid small businesses, but it was not until the depression of the 1930s that small business legislation gained much recognition. The economic difficulties experienced by the great majority of small firms during the depression led to demands for governmental action in their behalf. Many state governments adopted laws that discriminated against chain stores and outlawed so-called "unfair trade" practices of big firms. The federal government passed the Robinson-Patman Act that restricted the discounts that manufacturers and wholesalers could grant to large retailers for mass purchases. Various measures to make credit more readily available to small firms were provided. Special small business committees were established in both houses of Congress; these committees conducted extensive hearings on small business problems, sponsored numerous studies, and proposed a variety of legislative solutions. During World War II the Smaller War Plants Corporation was created to assist small firms to obtain war contracts or subcontracts. The Servicemen's Readjustment Act made small loans available to returning servicemen who wanted to establish their own businesses. This legislative interest in small business continued after World War II, culminating in the establishment of the Small Business Administration in 1953. This agency extended loans to small firms that were unable to obtain credit from commercial banks. It also assisted qualified small firms to obtain government contracts, promoted research on small business problems, and gave advice to small firms.

Governmental measures in behalf of small business were usually defended in terms of the need to preserve competition or to encourage new investment. Often the several million small businesses in the United States were cited as the bulwark of the middle class and their preservation was called for on that account. Most of the legislation adopted in the name of small business, however, affected only a small percentage of the firms classified as small businesses, principally the larger ones. In part this was a result of the tendency to expand the size limits within which firms qualify as small, and in part it was a result of the greater ability of firms with several hundred employees to present their cases and to utilize the services provided. A few authorities frankly said that only such firms were worth helping.

Nearly all the governments of the nonsocialist countries of Europe adopted policies to assist or protect small producers and tradesmen. In many of these countries there were governmental or private agencies whose function it was to study and promote the interests of artisans and other so-called independent elements in the society. The depression of the 1930s stimulated demand for measures of protection. An important element in the demagogic program of German National Socialism was directed to these groups, but their interests were soon sacrificed to the requirements of war. In the years following World War II, as before, much of the interest in the small operator in production, trade, and services was explicitly related in most European countries to his putative role in the middle classes. Thus Belgium had its Ministry of the Middle Classes and its Institute for the Economic and Social Study of the Middle Classes, the Federal Republic of Germany its Institute for Middle Class Research, and the Netherlands its Middle Class Bank. All of these organizations were primarily devoted to the interests of small businessmen. In Great Britain little government legislation to assist or protect small business was adopted. In 1921 the minister of agriculture created the Rural Industries' Bureau to help rural area firms that employed 20 or fewer skilled workers, but this step was taken not so much to help small enterprise as to promote the welfare of the rural population as a whole.

In Asia, Africa, and South America an ambivalent attitude toward small business is sometimes found. The fact that small-scale production requires more labour per unit of output than does large-scale manufacturing causes some policy makers to look with favour on the former as a means of providing employment. Small business is also viewed as a source of entrepreneurship and of middle-class stability. However, the desire to secure the cost advantages of modern productive techniques, often available only in large-scale units, leads in some cases to conflicting programs.

In socialist countries varying policies toward small business have been followed. In the Soviet Union and in Communist China, after an initial period of toleration or even encouragement of small businesses while large enterprises were being nationalized and a socialist economy was being organized, small firms generally were either nationalized or incorporated in cooperatives. On the other hand, in Yugoslavia artisan firms with less than five employees are permitted to function as private enterprises.

The number of small businesses in the nonsocialist countries tended to increase as population and income increased. In the United States the business population rose from about 3,000,000 firms in 1929 to 4,700,000 in 1960. Most of these firms were very small, about 75% of them consisting of the operator and, at most, three paid assistants. Nearly half the firms with no more than three employees were engaged in retail trade and one-fifth were in the service industries. In Europe the numerical persistence of small producers and tradesmen received much attention in the arguments over the revisionist thesis of Eduard Bernstein at the turn of the century. The Marxist theory that capitalist development tended to polarize society into the two classes of capitalists and workers was interpreted by Bernstein to imply a prophecy of rapid decline in the number of small producers and tradesmen, although Marx quite clearly anticipated the emergence of new groups of the petty bourgeoisie. In much of Europe large sectors of the economy, particularly retail trade and the service industries, long seemed immune to the process of concentration and centralization that development of large-scale organization in these industries brought about in the United States and Great Britain. However, the rapid economic expansion in western Europe during the 1950s was accompanied by the introduction of new forms of large-scale enterprise and the growth of older enterprises. At the same time full employment made available alternative sources of livelihood. As a consequence, many artisans and small shopkeepers took jobs as employees. Although the number of small enterprises did not decline, there was a decline in the percentage of nonagricultural self-employed persons in the labour force of most western European countries.

The long-run future for small business in the nonsocialist countries is governed primarily by two opposing forces: the growth of their economies tends to increase opportunities for both small and large firms; the process of concentration of economic activity in large units of business tends to eliminate small firms. These two forces are closely interrelated. Economic growth has been associated with the rise and consolidation of large-scale enterprise and has involved both growth of population and increasing per capita incomes. It has been accompanied by expanding commercialization of economic activity and increasing division of labour. These developments have contributed to the expansion of the market that makes feasible large-scale enterprise, while the economies made possible by large-scale organization of economic activity have been a decisive factor in economic growth. Historically, the interplay of these forces has been evident, on the one hand, in the "squeezing out" of small firms from certain industries as a result of the competition from large firms and, on the other hand, in the emergence of new opportunities for small firms as a result of the expansion of existing industries and the development of new ones whose markets are too limited to permit large-scale organizations to function effectively. Thus the increased volume of products turned out by large manufacturing enterprises greatly expanded the opportunities for small firms in

retailing. One line of retailing after another has been invaded by large units of business as markets have been extended by growth of population, incomes, and mobility, and as technical and organizational problems have been solved. In the United States the corner grocery store, that epitome of small business, has given way in striking fashion to the supermarket; the number of retail grocery outlets has been greatly reduced.

Whether the proliferation of small business industries as a result of the development of new products and services will continue to occur at a pace that offsets the process of concentration into larger units of business remains to be seen. The growing concentration of population in metropolitan areas provides a greater base for large-scale enterprise in lines that have long been strongholds of small business. Furthermore, the struggle for diversification on the part of large firms, which reflects both a desire to spread risks and the process of capital accumulation out of retained profits, has led more and more of them to explore opportunities in industries traditionally considered the preserve of small business. However, the increasing share of the service industries in consumer expenditures as per capita incomes rise expands economic activities that are often not suitable to large-scale organization. In any case, it is evident that small business will long continue to provide a livelihood for millions of owners and employees and to produce a significant fraction of the goods and services demanded by consumers.

BIBLIOGRAPHY.—A. D. H. Kaplan, *Small Business: Its Place and Problems* (1948); Joseph D. Phillips, *Little Business in the American Economy* (1958), *The Self-Employed in the United States* (1962); Joseph Steindl, *Small and Big Business* (1946); Harold G. Vatter, *Small Enterprise and Oligopoly* (1955); U.S. Small Business Administration, *Annual and Semiannual Reports* (1953–); U.S. Temporary National Economic Committee, *Problems of Small Business*, Monograph No. 17 (1941); James B. Jefferys and Derek Knee, *Retailing in Europe: Present Structure and Future Trends* (1962); A. Gutersohn, *Das Gewerbe in der freien Marktwirtschaft*, vol. 1 (1954), vol. 2 (1962); Herbert Gross, *Small Business im grossen Markt* (1958); Camera dei Deputati, XII Commissione Permanente dell'Industria, *Problemi della piccola e media industria* (1963); Donald R. Liggett, *Small Industry Development Organizations: a Worldwide Directory* (1959); Marian C. Alexander-Frutschi, *Small Industry: an International Annotated Bibliography* (1960). (J. D. P.)

SMALL HOLDINGS. The term small holdings covers several categories of relatively small farm tracts. Throughout the world, the essential characteristics of small holdings include recognized rights to operate the land, a high ratio of labour to land with the labour furnished mainly by the operator and his family, and relatively low levels of farm production and income. In most parts of the world, the labour is mainly by hand with the assistance of one or two horses, oxen or cows and simple tools.

This article outlines the characteristics and traces the history of small holdings after the breakdown of feudalism and indicates their prevalence in the second half of the 20th century. Separate sections deal with farm ownership in the United States and discuss some of the economic aspects of small holdings. *See* also AGRICULTURE; FARM MANAGEMENT; FARM TENANCY; LAND REFORM; LAND TENURE: ECONOMIC AND AGRARIAN ASPECTS.

Included in the commonly accepted meaning are the subsistence tracts of the peasant farmer and garden tracts and part-time farms as well as small family farms. The mechanized family farm, of substantial size in commercial agricultural areas, usually is excluded. Certainly excluded are the plantation, the estate, the collective farm and the factory-in-the-field farm. Many ranches, even if family operated, would not commonly be thought of as small holdings, for the size is an important, though not by any means the only, determinant. The United Kingdom Small Holdings act of 1892 and the Small Holdings and Allotment act of 1926 provided examples of legal applications of the term. In other countries somewhat the same results were accomplished by legislation not so exactly named.

Very much of the world's population lives on farms and perhaps three-fourths of the world's farmers live on small holdings. The farm population constitutes about 1,500,000,000 people or about [illegible] of the estimated total world population. Small holdings are most prevalent in Asia and other underdeveloped parts of the world but are also numerous in other countries. Population pressure on the land is sufficient in much of the world to reduce the average size of farms well within the small-holding limits. In most countries small holdings also are used by labourers and other part-time farmers for housing, subsistence and incidental agricultural production.

History of the Small Holding.—In Denmark, England and other countries of western Europe, governmental interest in small holdings was aroused by the breakdown of feudalism. The object of governmental programs was to provide holdings for labourers and tenants who were being dispossessed. As early as 1682 20 years before serfdom was abolished, it was decreed in Denmark that no farms that were large enough to support a family should be combined with other farms. This was an anticonsolidation rule aimed at landlords who had discovered that the prevailing three-field system was not very efficient and might be profitably replaced by improved rotation methods, thereby consolidating some of the peasant-cultivated land. But that would have meant less taxes and fewer soldiers for the crown—hence the rule that meant in practice that no tenant could be dispossessed. In the same country, as early as 1781–82, many of the agricultural labourers were by law given ownership of four to six acres of land to keep the individual available to work the lord's land. Comprehensive legislation after 1899 furthered the establishment of small holdings averaging five to seven acres, the government lending the prospective holder nine-tenths of the purchase price of land and buildings.

Later emphasis was on having a farm large enough to keep the farmer fully employed and provide a living for the family; hence 12½ ac. of land were set as a minimum goal for all farms. Still later government measures were directed at obtaining land for small holdings or to add to small holdings, as equitably as possible, from entailed estates and from the higher valued, more extensive holdings.

In England small independent holdings existed under the feudal system, their numbers being greater in some manors than in others. For centuries, through the modification and disappearance of the feudal system, the number of small holdings or family farms increased until they became the predominant form of land occupancy. This trend continued up to the last great enclosure movement beginning in the latter part of the 18th century. (*See* ENCLOSURE.)

The enclosure movement contributed greatly to a reduction in the number of small holdings and to an increase in the number of farm labourers.

The success of large numbers of small holdings was dependent on the common for pasture. When the common was enclosed an allotment was made to the commoner in proportion to his interest. These allotments were of little value to many small holders, however, because they were often far from the allottee's cottage and because he was obliged to fence them. Consequently many small holders sold their land to larger landowners and became labourers.

With the reaction to the materialism of the Industrial Revolution, the social and economic consequences of the decline of small holdings aroused the attention of the English government. The weight of arguments for small holdings resulted in reform legislation beginning with the Allotments acts of 1887 and 1890. The Small Holdings act of 1892 authorized county councils to create small holdings on a self-supporting basis where a need was proved to exist. Small holdings were legally defined in the Small Holdings act of 1892 as holdings of less than 50 ac. in extent or under £50 in annual rental value. In 1926 the limit of annual rental value was raised to £100. If an applicant could not purchase land he could obtain a holding on lease. The results of the act of 1892 were negligible largely because it was only permissive, and it was amended by the Small Holdings act of 1908, a much more comprehensive enactment, particularly in introducing the principle of central initiative and authority. This act instructed county councils to provide small holdings in response to the demand. In case the council failed to carry out its obligations the ministry of agriculture was given power to develop a plan. If a loss was actually incurred the ministry was authorized to bear one-half the loss.

Operations under the act of 1908 were suspended during World War I. Legislation following World War I had the main objective

of providing small holdings under particularly favourable terms for ex-servicemen.

The Small Holdings and Allotments act, 1926, renewed the powers of county councils to provide small holdings. In the event that holdings could not be provided without loss, the ministry of agriculture would contribute 75% of the estimated annual loss. Under this series of enactments a total of 28,700 small holdings were established from 467,000 ac. of land. Only a small part of this land was held by councils on lease.

In 1947 small-holding policy was made the subject of part iv of the 1947 Agricultural act. The objective emphasized in this act was the use of small holdings as a step in the agricultural "ladder" for farm families with little capital. Permitted sizes of holdings were increased to 75 ac. or £150 rental value. In 1958 a Small Farmers scheme was introduced providing grants to eligible small farmers for the purpose of carrying out approved plans for improving the farm business.

During the early 1900s small-holdings programs were established in many other countries, including Germany, Rumania, Hungary and parts of Russia immediately following the Revolution. The collectivization of agriculture greatly reduced numbers of small holdings in many Communist countries. However, in 1953 Yugoslavia instituted a policy permitting decollectivization of agriculture. Beginning about 1930, attention in France, Switzerland, Germany and other countries was focused on the consolidation of fragmented holdings and enlargement of excessively small holdings. A major share of the Mexican land reform was undertaken during the 1930s. After World War II land reforms favourable to small holdings were undertaken in Japan, Italy, Egypt, Formosa (Taiwan), Bolivia and other countries. (*See* also LAND REFORM.)

Prevalence of Small Holdings.—In England and Wales in the second half of the 20th century, 22% of the holdings of over 1 ac. were from 1 to 5 ac., 24% were from 5 to 20 ac. and only 36% of the holdings were of 50 ac. or more. The number of units under 50 ac. provides a rough measure of the importance of small holdings in Great Britain. They vary considerably in size, in returns and in types of products grown. Many of the holdings under five acres are cottage gardens or allotments. Some occupy small acreages with specialization in intensive enterprises such as growing vegetables or producing eggs. Others are larger and more diversified. Probably most of the holdings from 20 to 50 ac. are commercial units producing mainly for sale. In Ireland, long after the period commonly considered as the reform period, very small holdings continued to increase both in number and in percent of total, though average size remained fairly constant. Three-fourths of the holdings in Ireland are under 50 ac. In Germany and other countries of western Europe considerable permanent hired labour is used on farms of 20 ha. (hectare = 2.47 ac.) and over.

Data from selected countries show that the size distribution of holdings is highly variable from country to country, but small holdings greatly predominate. Moreover, allotments made as a part of land reform, as in Italy, are small—only one to five hectares per holder. The numbers of holdings reporting less than one hectare vary widely, largely because of the varying farm definitions used in various countries.

In eastern Europe small units were prevalent prior to collectivization. In Bulgaria in 1934 only 1% of the holdings were larger than 49 ac. Less than 3% of those of Yugoslavia in 1931 were more than 49 ac. Even in Hungary, with its large pre-World War II estates, 70% of all holdings in 1935 were in tracts of 7 ac. or less. It may be significant that in some of these countries members of the new state cooperative farms were permitted to retain small plots—of 0.5 to 1.2 ac. in Bulgaria and Rumania, up to 2.5 ac. in Yugoslavia and about 2 ac. in Hungary.

In Egypt, where 95% of those who own land have 5 ac. or less, expropriated land was to be distributed in units varying in size, according to quality, from a minimum of 2 feddans to a maximum of 5 feddans (about 2 to 5 ac.). In Iran units of 14.5 to 21 ac. each were distributed. In Iraq settlement after 1946 appeared to recognize that tracts could be too small; the Dujail project involved allocation of 62½ ac. per family from a tract of 75,000 ac. of newly irrigated, formerly desert land.

In the orient, holdings are extremely small and tend toward further reduction under the pressure of population. More than 90% of Japanese farmers work less than 2.0 ha. After World War II Japan instituted a comprehensive system of land reform. Under these laws strict upper limits were set on farm size. Numbers of small farms increased substantially but net productivity of labour apparently rose with increased savings and capital formulation. Very small farms are also characteristic of India and China. More than half the farms in India are under 2.0 ha. In the late 1950s considerable emphasis was placed on the formation of cooperative and collective farms in China.

In Africa small native holdings greatly predominate even in areas such as Nigeria, producing export crops of cocoa and peanuts for the world market. Ninety per cent of the non-European farms in Rhodesia also are in holdings of under 14 ac. Cultivation in these countries is by plow or hand hoe.

There is appreciable variation in the size of holdings in Central and South America. Chile, with 2.5 ac. of cropland per capita, has about 20 cultivated acres per person engaged in agriculture. On the other hand, in Guatemala more than one-fifth of all agricultural holdings are less than 0.7 ha. and more than 75% are less than 3.5 ha. Over half of the holdings in Puerto Rico are under 3.5 ha. in size. Even in Uruguay, which is still accepting settlers, 42% of the holdings are less than 20 ha. in size.

United States.—In the United States, even during the period of abundant land, relatively small holdings were generally favoured. In 1820 the minimum unit of sale of government land was lowered to 80 ac. Limitation of acreage of public land that could be acquired by one individual was early recognized. The Pre-emption act of 1841, as amended, provided that a settler could legally enter upon public land and establish a claim for not more than 160 ac.; he could not make more than one such entry, and such a claim could not be made by anyone owning as much as 320 ac. The Homestead act of 1862, under which something like 285,000,000 ac. were patented, provided that a qualified individual might acquire title to land, not to exceed 160 ac., by five years of residence thereon with improvements, or by six months of residence plus payment of $1.25 per acre.

The 1954 United States census of agriculture classified 4,782,000 holdings as farms. A postenumerative survey indicated that about 419,000 additional places that would have qualified were missed by census enumerators; most of the latter were small holdings. About 35% of the 4,782,000 farms enumerated operated less than 50 ac., and more than 50% were under 100 ac. In 1959 the U.S. census definition of a farm was changed to exclude a considerable number of the small holdings enumerated in earlier years.

Because of the wide variation in productivity, acreage is not a very satisfactory measure of size of farms in the United States. In terms of value of products for sale, nearly one-third of the farms are part-time and residential units, most of them classed as small holdings, with farm production for sale of less than $2,500. About one-fifth of the farms in the United States can be classified as small-scale units primarily dependent on farming for a living. While the number of small-scale farms was decreasing, they remained a substantial segment of U.S. agriculture, characterized by low levels of mechanization, relatively small acreages and underemployment of operator and family labour. A substantial number of the operators were 55 years of age or over.

Beginning in the 1930s the problem of the small farmer was approached in a modest way by the Farmers Home administration of the U.S. department of agriculture with a program of supervised credit and individual guidance for the development and enlargement of land resources and working capital. In 1955 a rural development program was initiated by the U.S. department of agriculture in several counties in areas where farm incomes were low, emphasizing the development of both farm and nonfarm resources. In the 1960s this program was substantially expanded.

Some Economic Aspects of Small Holdings.—Developments in machinery, equipment and modern aids and techniques focused

attention on the fragmentation of holdings in some areas and on the uneconomic sizes of fields and lack of the opportunity to make efficient use of modern agricultural machinery.

This situation suggested consolidation of small and scattered holdings and the injection of capital to mechanize the consolidated units to give at least part of the reorganized population more income per hour worked and turn part of the population to other work. The difficulty of attaining efficient production with a small land unit (except for a few types of intensive specialization) led to attempts to overturn the small-holding pattern in some countries.

Another aspect of the problem is what is called the underemployment of farm labour. It has been argued that many small holders value leisure and the satisfactions of self-employment more highly than the advantages they might gain from fuller or more productive employment. Furthermore, a high percentage of families on small holdings have some off-the-farm employment. In these cases the small tract can serve to provide a form of security plus some food and space for family living. But for many families more productive use of labour involves farm enlargement, occupational adjustments and the development of nearby nonfarm industries. *See* also Agricultural Economics; Farm Management; Land Tenure: Economic and Agrarian Aspects.

Bibliography.—J. Ackerman and M. Harris (eds.), *Family Farm Policy* (1947); Waldorf Astor, 2nd Viscount Astor, and B. S. Rowntree (eds.), *Smallholdings Studies* (1938); W. V. Ball, *Smallholdings for Pleasure and Profit* (1947); B. O. Binns, *The Consolidation of Fragmented Agricultural Holdings* (1950); Kenneth H. Parsons *et al.* (eds.), *Land Tenure* (1956); *Development of Agriculture's Human Resources*, prepared for the Secretary of Agriculture, U.S. Department of Agriculture (1955); A. W. Menzies-Kitchin, *Land Settlement; A Report Prepared for the Carnegie United Kingdom Trustees* (1935); "Assistance to Small Farmers," presented to Parliament by the Home Secretary, the Secretary of State for Scotland and the Minister of Agriculture, Fisheries and Food (1958). (J. K. R.; K. L. B.)

SMALL ISLES, a parish of islands of the Inner Hebrides, Inverness County, Scot. (Pop. (1961) 170. It consists of the islands of Canna, Sanday, Rum, Eigg and Muck, lying in the order named, like a crescent with a trend from northwest to southeast, Canna being the most northerly and Muck the most southerly. They are separated from Skye by Cuillin Sound and from the mainland by the Sound of Ardnamurchan. The surface is moorland, pasture and mountain (exceeding 2,600 ft. on Rum). The islands are rich in seafowl and Rum contains red deer and wildcats. The fisheries include cod, ling and herring. Steamers call at Eigg, Rum and Canna.

Antiquarian remains at Canna include a weather-worn sculptured stone cross and the ruins of a chapel of St. Columba. On the northeast coast of Eigg is a cave with a narrow mouth, opening into a hollow 255 ft. (78 m.) long. In it Macleod of Skye, toward the end of the 16th century, ordered 200 Macdonalds to be suffocated; their bones were found long afterward.

SMALLPOX (Variola), one of the world's most dreaded plagues, is an acute infectious disease characterized by fever and, beginning about two days later, an eruption which, after passing through the stages of papule, vesicle, and pustule (*see* Skin, Diseases of: *Types of Skin Lesion*), dries up, leaving more or less distinct scars.

The characteristic eruption may be so profuse as to be confluent, especially on the face, or so scanty that the lesions are missed altogether. The lesions are commonly more deeply seated in the skin than are the lesions of chicken pox and the scars are therefore more permanent than those of chicken pox, but mild forms of smallpox may also have comparatively superficial pocks, with correspondingly less tendency toward pustulation and less scarring.

These modifications, both toward fewer lesions and toward their being more superficial, may occur either naturally or because of vaccination that was not recent enough to give complete protection against contracting the disease. With nearly complete vaccinal protection, few lesions will appear; but even if vaccination had been effected many years before, smaller and more superficial lesions are the rule where an unvaccinated person would have a severe attack. Such superficial lesions are also characteristic of the naturally occurring mild strains of the disease (variola minor). Even with strains of full severity (variola major) there may be much variation in severity among the unprotected.

Besides the characteristic focal eruptions (papule to vesicle to pustule to scab to scar) there is sometimes a toxic eruption during the initial fever. These toxic rashes may be diffuse blushes on trunk or limbs somewhat suggestive of scarlet fever or measles; they may on the other hand be deeper red, with small hemorrhages like fleabites or larger blotches in the skin. They are unusual enough to be of not much importance except for the fact that the deeper red eruptions have often disguised the diagnosis so that smallpox is not thought of until the contagion has spread over a hospital or farther. Cases with such a hemorrhagic eruption are likely to be the most severe, even fatal before the characteristic focal eruption appears, and can carry an intense contagion.

Diagnosis.—Prompt diagnosis is of the highest importance in combating smallpox. Though general vaccination helps, there is no community anywhere so efficiently vaccinated as to prevent smallpox from occurring occasionally. The mechanical refrigerator is largely responsible for the scarcity of the disease in the United States and Europe: if a supply of refrigerated vaccine is available constantly, a screen of recently and successfully vaccinated persons can be thrown promptly around any cases that may occur. Better refrigeration, quicker transportation, safer and more regularly potent vaccine, and more enlightened sanitary practices also have caused marked reductions elsewhere in the world. Vaccine dried while frozen has proved advantageous for vaccinating large numbers of persons in regions where ordinary liquid vaccine would lose its potency due to temperature and transportation difficulties. Smallpox rates and infant mortality rates are two of the best indexes of living standards.

The scarcity of the disease has brought about a considerable dulling of a former keenness in diagnosis. Except for laboratory confirmation of the virus (*see* below), which should be sought at once, the surest criterion for diagnosis is the course of a typical case. It begins with fever 7–21 days after effective exposure, followed by papules appearing after 1–5 days of fever and becoming vesicles in 1–4 days, pustules in another 1–4 days, and crusting over (scabbing) in 2–6 days. The scabs fall off in 10–40 days after the beginning of the eruption, leaving pink scars which become white after months or years. Unfortunately this long-term diagnosis is too slow for the preventive measures demanded by the disease. Complete isolation of the patient, search for persons who may have been in contact with him, and vaccination of those who might be exposed should be begun within hours, not days, after the first trace of a case. Therefore an early diagnosis based on the distribution of the eruption, which can be used at any stage after it appears, is imperative.

The eruption of smallpox, whether scanty or profuse, has more lesions on the parts of the body exposed to irritation than elsewhere, more on the face than on the abdomen and chest, more on back of the arms and hands than on the front, more on the limbs than on the trunk, and more on the back than on the abdomen. Protected places such as the armpits, the depression of the eyelid, and the back of the ear (unless irritated by wearing glasses) are relatively spared. However, unusual and extremely severe forms of smallpox may be evidenced only by prostration, deep redness of the skin with hemorrhages, and sometimes a fatal outcome occurring before the delayed eruption of pocks with its characteristic distribution becomes apparent. The presence of vaccination scars should not influence the diagnosis of smallpox; immunity diminishes in some persons more rapidly than in others.

Of laboratory methods of diagnosis, the most trustworthy is the identification of the lesions produced on the chorioallantoic membrane of the chick embryo. This requires a minimum of three days and a laboratory with 12-day incubated fertile eggs, a skilled inoculator, and a pathologist trained to recognize the lesions. Microscopic examination, agar-gel precipitation, hemagglutination inhibition, and complement-fixation tests are quicker and helpful, but less definite in interpretation.

Cause.—Smallpox is caused by a virus of the poxvirus group whose particles, ovoid or rounded-quadrangular in shape, are relatively large. The virus can be cultivated in a developing chick

embryo and on some other living animal tissue.

Some virologists tend to regard the viruses of variola major, variola minor, cowpox, and vaccinia as distinct, on account of differences which breed true in laboratory passage and, by common observation, in natural passage. On broader grounds, however, it seems reasonable to regard them all as derived from the original smallpox virus. Each immunizes against all the others. Charles Armstrong has shown that by selective transfer a strain of vaccinia can be made to become much more virulent.

Method of Spread.—Each case of smallpox arises from contact, direct or indirect, with another case of the disease. There are no natural animal carriers or natural propagation of the virus outside the human body, and the virus ordinarily does not live very long outside the body nor is it ordinarily carried more than a few feet through the air. Nevertheless all articles which may have been contaminated from a smallpox patient should be thoroughly disinfected by heat before they come in contact with unprotected persons. Smallpox is one of the few diseases in which public-health regulations specify terminal disinfection (as by formaldehyde) of living quarters after removal or death of the patient. The virus may be carried passively by a third person, himself immune, from a sick person or from a corpse. The disease is presumably contracted by inhalation of particles bearing the virus; therefore persons having necessary contact with the sick should be thoroughly masked as well as gowned. Isolation must be much more rigid than with the other contagious diseases.

Treatment.—No specific treatment is yet known for smallpox. Prevention is by means of isolation and vaccination (*q.v.*). Of all widely prevalent diseases, this is the most completely preventable by public health measures.

History and Prevalence.—The Persian physician Rhazes about A.D. 900 wrote the first known account separating smallpox from measles, although Eusebius mentioned a Syrian epidemic in A.D. 302 and the term variola ("pox") was used by Bishop Marius of the disease in France and Italy in 570. About 1200 Gilbert of England first referred to smallpox as a contagious disease, but even this was continually contradicted—even by Thomas Sydenham (1675), who described the disease better than anyone before him.

Smallpox was pandemic in Europe in 1614 and epidemic in England during 1666–75, and there were scattered outbreaks in New England all through the 17th century. On account of their habits during sickness, the American Indians were severely stricken, and students have assigned smallpox as one of the chief reasons for the conquest of the land by the white men. Of course there were many mistakes in diagnosis, and smallpox was confused with the great pox (syphilis). William Heberden in 1767 first clearly distinguished chicken pox from smallpox, but even in the 19th century the great skin physician Ferdinand von Hebra of Vienna denied the difference. After the turn of the 20th century the predominance in some localities of very mild types gave rise to the thought that the mild type was a disease different from smallpox, but tests for immunity indicated that these are merely different forms of the same disease, each form generally breeding true to its type.

Statistics as to prevalence of smallpox must be accepted with some caution. Reporting is likely to be imperfect in the very regions where vaccination is imperfect. Asia and to a less extent Africa have been breeding grounds for outbreaks of the severe form of the disease in Europe and the United States. Not infrequent outbreaks arose from travelers from Mexico before that country undertook a thorough campaign of eradication.

Mild smallpox was once well established in the central and western United States, but the reduction has been so phenomenal that the elimination of the disease in the U.S., as far as indigenous cases are concerned, may be considered accomplished. Rapid air travel, however, emphasized the need for fresh vaccination of all international travelers.

See also references under "Smallpox" in the Index.

BIBLIOGRAPHY.—C. W. Dixon, *Smallpox* (1962); J. P. Leake, *Questions and Answers on Smallpox and Vaccination*, U.S. Pub. Health Serv. reprint no. 1137, rev. (1946). For current data on the reported prevalence of smallpox *see* publications of the World Health Organization. (J. P. L.)

SMART, CHRISTOPHER (1722–1771), English poet whose reputation perhaps rests partly on his strange personality. Browning saw him as a mediocre poet to whom a single flash of sublime poetic inspiration was granted in the *Song to David;* and an eminent surgeon has diagnosed from his poetry a classic case of cyclothymia, with its characteristic alternation of elation and depression. Neither, however, had a complete knowledge of his writings. The most significant addition to them was *Rejoice in the Lamb* (composed *c.* 1763; published in 1939), an antiphonal poem full of cryptic autobiographical references. This points directly forward to Smart's later and best poetry: *A Song to David* (1763), the *Hymns . . . for the Fasts and Festivals . . .* (1765), and the *Hymns for the Amusement of Children* (1770).

Smart was born on April 11, 1722, at Shipbourne Fairlawn, Kent, where his father was steward to William, Viscount Vane. After her husband's death in 1733, his mother moved to Durham, where the Barnard branch of the Vane family lived. Christopher's holidays from Durham Grammar School were spent at Raby Castle with the Barnards, whose patronage he enjoyed for most of his life. He was intellectually precocious, and at 17 was sent to Pembroke Hall, Cambridge, where, during a brilliant but erratic career, he became a fellow (1745), and held academic office. At about 27 he plunged into the literary life of London, though he continued to compete successfully for the Seatonian Prize, a yearly university award for a poem on the attributes of the Supreme Being. His five Seatonian poems held promise of his later religious verse.

In London Smart was associated as a hackwriter with the bookseller John Newbery (whose stepdaughter he married). Chronic extravagance and inebriety made his life unhappy. At least three times he was confined for madness—a mild religious mania. But he must have had an engaging personality, for among friends who helped him were Johnson, Garrick, Goldsmith, and Dr. Burney; and Fanny Burney speaks of him in her diary with pity and admiration. He died in a London debtor's prison on May 21, 1771.

Most of Smart's early poetry shows talent but little depth. His later religious poems were treated with kindly indifference by his contemporaries, but a change in taste has led later readers to perceive in them unique flashes of child-like penetration and vivid visual imagination. In some ways they foreshadow the poetry of Blake and John Clare.

BIBLIOGRAPHY.—*Collected Poems*, ed. by N. Callan, 2 vol. (1949), with bibliography; *Poems*, a selection ed. by R. Brittain (1950). *See* also C. Devlin, *Poor Kit Smart* (1961). (N. C.)

SMART, JOHN (*c.* 1742–1811), one of the foremost English miniature painters, the place and exact date of whose birth are not known. In his youth he won a series of prizes from the Society of Arts. He exhibited with the Society of Artists from 1762 and became president of that body in 1778. Smart spent ten successful years in India from 1785 till 1795, becoming miniature painter to the nawab of Arcot. After his return to England he exhibited fairly regularly with the Royal Academy. Smart's first wife left him for William Pars, the watercolour painter, and he married once, and possibly twice, again. He died in London on May 1, 1811.

In an age prolific in miniature painters of the front rank, Smart's achievement is regarded as highly as that of his rival and contemporary, Richard Cosway. He employed entirely different methods. His brushwork was meticulous and his finish highly wrought; he used body colour to render the rich costume of the late 18th century. Though not acute analyses of character, his miniatures have great finesse and charm. Throughout his life Smart made a practice of dating as well as signing his works; on those he painted in India he added the letter I to his signature J. S.

See B. S. Long, *British Miniaturists* (1929); A. Jaffé, *The Art Quarterly*, vol. xvii, pp. 242–254 (1954). (G. Rs.)

SMEATON, JOHN (1724–1792), English civil engineer, builder of Eddystone lighthouse and founder of the civil engineering profession in Britain, was born at Austhorpe, Yorkshire, on June 8, 1724. He learned mathematical instrument making in Lon-

don and, through contributing scientific papers to the Royal society, was elected a fellow in 1753. A tour of the Low Countries during 1755, studying canals, harbours and mills, was the turning point in his career. In 1756–59 he built the third lighthouse on Eddystone reef, opposite Plymouth, with dovetailed blocks of portland stone, and this established his reputation (*see* LIGHTHOUSES AND OTHER AIDS TO NAVIGATION). In 1882 the upper portion was re-erected on Plymouth Hoe as a memorial. He also constructed the Forth and Clyde canal, built arched bridges at Perth, Banff and Coldstream, and completed the harbour at Ramsgate.

In 1754 Smeaton introduced cast-iron shafts and gearing into wind- and water mills, and received the Royal society's Copley medal for *An Experimental Enquiry Concerning the Natural Powers of Water and Wind to Turn Mills* (1759). He constructed a boring mill at Carron ironworks and, due to his improved designs, the Newcomen engine achieved its maximum performance. He designed large atmospheric pumping engines for Long Benton colliery in Northumberland, Chacewater mine in Cornwall, and the docks of Kronstadt in Russia. Smeaton was an original member of the Society of Engineers, founded in 1771, and in 1791 wrote his *Narrative of the Building . . . of the Eddystone Lighthouse*. He died at Austhorpe on Oct. 28, 1792, and was buried in Whitkirk parish church.

BIBLIOGRAPHY.—S. Smiles, *Lives of the Engineers*, vol. ii, rev. ed. (1874); Smeaton's *Reports*, 3 vol. (1812), *Papers* (1814); *John Smeaton's Diary of His Journey to the Low Countries, 1755*, with introduction by Arthur Titley (1938). (AR. S.)

SMEDEREVO, a town in the Socialist Republic of Serbia, Yugos., and a Danube port, lies 24 mi. (39 km.) ESE of Belgrade. Pop. (1961) 27,182. It is a walled town with 20 towers. It manufactures railway trucks and there is a steel mill. In the neighbourhood of Smederevo is believed to be the site of the Roman settlement Mons Aureus. In the 15th century when the Serbian despot George Brankovich became lord of Tokay (Tokaj) in Hungary, he planted vines from Smederevo on his estates there, and from these came the famous white wine of Tokay. Smederevo became the capital of Serbia about 1430 and was taken by the Turks in 1459. *See* also SERBIA. (V. DE.)

SMELL AND TASTE. Food flavour is man's commonest and most familiar smell (olfactory) and taste (gustatory) experience. Commonly misnamed taste, flavour is a composite of pressure, temperature, pain, and odour, as well as taste. Seeing and hearing also contribute to delight the gourmet. The olfactory component in flavour can be demonstrated by pinching shut the nostrils: a slice of raw potato then tastes the same as an apple; cinnamon yields only mild sweetness (a true taste) in addition to texture. Perfumes, on the other hand, primarily excite the sense of smell, although some may induce slight gustatory or tactual experiences.

The sensation (*q.v.*) that an entirely new synthetic chemical will arouse is not yet fully predictable before it has been tasted or smelled. The chemistry of odour and taste remains incompletely known. Many known relations concern the presence of specific atom types in chemical compounds. For example, mercaptans (*q.v.*) structurally resemble alcohols, but have sulfur atoms in place of the oxygen in alcohol molecules. Such sulfur substitution often leads to a markedly disagreeable change in odour; *e.g.*, the scent of a skunk is traced to *n*-butyl mercaptan; in water H_2O, sulfur substitution gives the odour of rotten eggs in hydrogen sulfide H_2S (called stinkdamp by miners); the analogue of odourless carbon dioxide CO_2 is carbon disulfide CS_2, with a distinctive smell that becomes especially nauseating in the presence of organic sulfide impurities.

Another approach explains odour in terms of molecular shape regardless of atomic constituents (J. E. Amoore, J. W. Johnston, Jr., M. Rubin). Thus camphor, cyclooctane, thiophosphoric acid dichloride ethylamide, and hexachlorethane (all with similar odours) have widely different chemical formulas, but their molecules are of similar shape and size. This interesting stereochemical theory awaited fuller empirical verification in the 1960s.

The common chemical sense, often classed with taste and smell (also chemical senses), depends on undifferentiated free nerve endings in the moist mucous membranes of the mouth, upper respiratory passages, and other body orifices. The common chemical sense mediates the mild pain of pepper or other irritant spices, whereas gustatory sensory qualities are characterized as salty, sour, bitter, and sweet; and odour as foul, fruity, burnt, fragrant, and so on. The common chemical sense, taste, and smell form a hierarchy in terms of increasing complexity of anatomical structure, sensitivity, and differentiation in sensory quality. In lower forms, especially aquatic organisms, these distinctions may not always be obvious because irritant, olfactory, and gustatory stimuli are all transported in the fluid medium; on land, odorous or irritating vapours are largely air-borne, whereas taste stimuli are in solution and stimulate by direct contact between solution and taste bud. In fishes olfactory pits or taste buds on whiskerlike feelers (barbels) and on other parts of the body surface indicate specific sensory functions. Taste and smell, being less informative and involving relatively simpler physical structures than those that mediate seeing or hearing, are often classified as lower senses. Among lower animals the chemical senses have particular significance in the selection and acceptance of foods, the avoidance of noxious agents, and the detection and attraction of sexual mates. Civilized man, more dependent on his higher senses, finds in his lower senses the flavour and spice of life.

SENSE OF TASTE

The sense organs for taste are goblet-shaped clusters of cells (taste buds) that open by a small pore. The buds contain slender gustatory cells with hairlike terminal microvilli projecting into the taste pore (fig. 1). In man and other mammals, taste buds are primarily found in the fungiform, foliate, and circumvallate papillae of the tongue (*q.v.*), and in adjacent structures of the throat. Sensory nerve fibres enter the taste buds, entwine, and contact one or more taste cells. When nerves to the tongue are cut, these fibres and the whole taste bud degenerate. If the nerve fibres regenerate, the taste buds are reconstituted from surrounding cells of the epithelium.

ADAPTED FROM A. J. D. DE LORENZO, "ULTRASTRUCTURE AND HISTOPHYSIOLOGY OF MEMBRANES" IN "OLFACTION AND TASTE," ED. Y. ZOTTERMAN; PERGAMON PRESS, 1963

FIG. 1.—MICROSCOPIC SECTION OF TASTE BUDS OF CIRCUMVALLATE PAPILLA

In invertebrates, the clearest distinction between taste and smell can be made in insects; there is also ample evidence that worms, gastropods, and other forms reject bitter substances but accept sugars added to their food. In insects, contact chemoreceptors are located on hairs of the legs and mouth parts. Localized stimulation of these hairs with sugar solution induces extension of the proboscis. This response may be inhibited by adding salt to the sugar solution. In flies, the labellar hairs contain four nerve fibres, only three of which appear to extend to the chemoreceptive tip (*see* FLY); electrical discharges in these fibres are found to follow chemical stimulation. Similar receptors occasionally are seen on the proboscis or antennae, but the latter more commonly function as olfactory organs.

Innervation.—There is no single sensory nerve for taste in vertebrates. In man, the anterior two-thirds of the tongue is supplied by the lingual nerve, the back of the tongue by the glossopharyngeal nerve, and the throat and larynx by branches of the vagus nerve, all of which mediate touch, temperature, and pain sensitivity as well as taste. The taste fibres of the anterior tongue leave the lingual nerve to form a slender branch (the chorda tympani) that traverses the eardrum en route to the brain stem. When the chorda tympani is damaged, as in removal of the eardrum, taste buds disappear and taste sensitivity is lost on the anterior two-thirds of the tongue on the same side. The taste fibres ascend to the medulla oblongata; at this and all levels of

the brain, taste fibres seem closely associated with touch and temperature neurons for the tongue. From the medulla the fibres ascend to a small cluster of cells in the medial part of the sensory nucleus of the thalamus. From here fibres proceed to the anterior cerebral cortex close to the sensory-receiving area for the mouth (*see* BRAIN).

The tongue surface is not uniformly sensitive. The middle surface is insensitive to taste, but there is sensitivity to salt around all edges, sensitivity to sweet primarily at the tip, sour at the sides, and bitter at the back. For many years it was believed that there were only four basic types of taste receptor, one each for salt, sour, bitter, and sweet. However, electrical impulses in the gustatory nerves of animals contradict this simple schema. Individual nerve fibres from the tongue show mixed sensitivity, in many cases responding to more than one taste stimulus; *e.g.*, acid plus salt, acid plus salt plus sugar. Records from individual taste cells show the same multiple sensitivity. The taste buds, therefore, seem to have different clusters or patterns of sensitivity. In some organisms certain taste receptors may be activated by water but inhibited by sodium chloride. Thus nerve impulses to the brain may be increased or decreased by taste stimulation. Gustatory perception depends on the processing of such impulses by the brain.

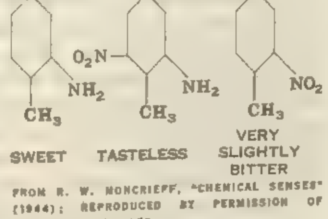
Chemical Factors.—There is no simple relation between chemical composition and taste quality except in the case of acids. The qualities of inorganic salts are complex, only sodium chloride yielding a purely saline taste. Sweet and bitter tastes occur in many different chemical classes.

Sour.—The hydrogen ions of acids are largely responsible for a sour taste, and sourness increases with hydrogen ion concentration; but this factor alone does not determine sourness. Weak organic acids such as the acetic acid of vinegar are more sour than would be predicted from hydrogen ion concentration alone (*see* ACIDS AND BASES; HYDROGEN IONS).

Salt.—Although a salty taste is often associated with soluble salts, most chemical salts, except sodium chloride, have such complex tastes as bitter-salt or sour-salt. Salts of low molecular weight are predominantly salty while those of higher molecular weight are bitter. The salts of heavy metals such as mercury have a metallic taste, although some salts of lead and beryllium are sweet. Both anion and cation (*qq.v.*) seem to influence taste quality and stimulating efficiency. In man the degree of saltiness decreases in this order: salts of ammonium, potassium, calcium, sodium, lithium, magnesium. The order appears different for other animals. (*See* also SALT.)

Sweet.—Except in the case of some inorganic salts of lead or beryllium, a sweet taste is associated largely with such organic compounds as alcohols, glycols, sugars, and sugar derivatives. The complex relation between chemical structure and sweet taste was not readily explained in the 1960s. This is particularly apparent for such synthetic sweeteners as saccharin and Dulcin. Sensitivity for these substances is especially remarkable; for sweetness to be just perceptible, saccharin need have only $\frac{1}{700}$ the concentration required for cane sugar. Spatial arrangement of atoms in the molecule is strikingly important; slight changes in arrangement within a sweet molecule (fig. 2) will make it bitter or tasteless. (*See* also CARBOHYDRATES.)

Bitter.—Bitter also is elicited by many chemical classes and is often found in association with sweet and other taste qualities. An increase in molecular weight of inorganic salts or an increase in length of the carbon chain in organic molecules may be associated with increased bitterness. The best-known bitter substances are the alkaloids (*q.v.*), which are often toxic, including quinine, caffeine, and strychnine (*see* POISONOUS PLANTS). Some of the lowest taste thresholds on



FROM R. W. MONCRIEFF, "CHEMICAL SENSES" (1944); REPRODUCED BY PERMISSION OF LEONARD HILL, LTD

FIG. 2.—EFFECT ON TASTE OF SLIGHT CHANGES IN MOLECULAR ARRANGEMENT

record are found in this class of agents (*see* PSYCHOPHYSICAL METHODS).

Some people cannot detect the normally bitter taste of such chemicals as PTC (phenylthiocarbamide) that have the thiocarbamide group

$$-\mathrm{N}-\overset{|}{\underset{\underset{\mathrm{S}}{\|}}{\mathrm{C}}}-$$

in the molecule. This form of taste blindness appears to be inherited as a Mendelian recessive (*see* GENETICS). About one-third of Caucasians cannot taste PTC; among Negro groups this proportion is about one-fifth. Evidence for such taste blindness has been observed only in primates, the gene responsible for the defect apparently having appeared at a relatively late stage of evolution. Taste blindness for thiocarbamides is not correlated with insensitivity to other bitter stimuli.

Factors Affecting Taste Sensitivity.—It has long been known that temporary taste insensitivity may arise at extreme temperatures, especially when a fluid is very cold. The optimum for taste generally seems to be near or slightly below body temperature. The effect is often difficult to assess because of changes in temperature as a solution is put in the mouth. When the tongue and mouth are first adapted to the temperature of the solution, sugar sensitivity increases with temperature, salt and quinine sensitivity decrease, and acid sensitivity remains relatively unchanged.

Taste adaptation (*i.e.*, partial or complete disappearance of taste) may occur if a solution is held in the mouth for a time. In some cases, cross adaptation will occur; *e.g.*, adaptation to one acid may carry over for other acids. Adaptation also may be followed by contrast; *e.g.*, distilled water tastes sweet following exposure to weak acid. Taste masking is an everyday experience; for example, the bitterness of tea and coffee or the sourness of lemonade are masked by sugar or saccharin.

The gustatory difference threshold (or percentage increase in concentration required for a just noticeable difference) is approximately one-fifth (or 20%). At very weak intensities, however, as much as a 100% increase in concentration may be required for a detectable difference.

Food Choice.—Sucking, one of the earliest reflex responses of the infant, can be controlled by gustatory stimuli. Sweet solutions are sucked more readily than is plain water; bitter, salty, or sour stimuli can stop the reflex. Among insects a very specific feeding reaction (proboscis extension) is so automatic that it is widely used as an index of gustatory stimulation. If a fly is held relatively immobile in wax, a drop of sugar solution will elicit proboscis extension when applied to the legs or mouth parts. A fly that has been starved will respond to a weak sugar solution that ordinarily would have no effect. Addition of salt or acid to the solution inhibits this response.

Lower mammals such as laboratory rats, when free to take pure carbohydrates, proteins, vitamins, and minerals from separate containers, will show a consistent pattern of choice that may be modified experimentally. A rat made salt-deficient by removal of the adrenal glands, for example, will increase its sodium chloride intake; normally, adrenalectomy is fatal without salt replacement. Similar effects have been reported in human beings; a dramatic case was that of a child with adrenal pathology who survived for a number of years by spontaneously satisfying his intense craving for salt. However, adult eating habits are much influenced by past experience and may even run counter to physiological requirements. Food habits and fads, cultural conditioning, social usage, and other complex psychological factors play a significant role in eating behaviour (*see* MALNUTRITION).

Toxic substances often are unpalatably bitter, but not all of them have a taste. The rat poison alpha-naphthylthiourea (ANTU) was developed from a relatively insoluble and therefore tasteless form of highly toxic chemical; soluble forms had been rejected by the animals. Lead acetate, sometimes called sugar of lead, is quite poisonous, but was once used as a sweetening agent with disastrous results. Other palatable stimuli also have toxic effects, so that taste alone is not a sure guide to safety.

SENSE OF SMELL

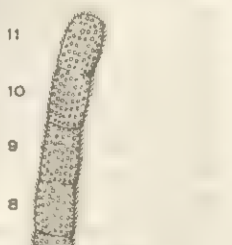
FROM G. H. PARKER, "SMELL, TASTE AND ALLIED SENSES"; REPRODUCED BY PERMISSION OF J. B. LIPPINCOTT CO.

FIG. 3.—SEPARATED SECTION OF OLFACTORY MUCOSA SHOWING SENSE CELLS WITH OLFACTORY HAIRS AND SUPPORTING CELLS

In mammals the olfactory receptors are located in the upper part of the nasal cavity. (*See* also OLFACTORY SYSTEM.) In man, the yellow-pigmented olfactory membrane measures about 2.5 cm.^{2} in each nasal cavity. The olfactory sense cells are long and thin, ending in several delicate hairs that project into and through the mucus covering the nasal epithelium (fig. 3). Electron microscope photographs show from 6 to 12 olfactory hairs per cell. The other end of each cell narrows to a fine nerve fibre that enters the olfactory bulb of the brain through a narrow channel in the bony roof of the nasal cavity. The olfactory area of the rabbit contains about 100,000,000 receptor cells which, with their olfactory hairs, provide a surface equal to the total skin area of the young rabbit. Free nerve endings of the trigeminal supply that mediates common chemical sensitivity are distributed throughout the nasal cavity and olfactory region. They respond to such mild odorants as orange oil as well as such irritants as ammonia.

Odorous molecules are carried to the olfactory mucosa by slight eddy currents during quiet respiration; vigorous sniffing produces a surge into the region. Sensitivity may be impaired by blocked passages, as when nasal membranes are congested.

It is generally agreed that in insects the antennae are the principal olfactory sites, but such appendages as the mouth parts may also participate (fig. 4). The olfactory structures of the honeybee appear to be what are called pore plates (fig. 5), but in some beetles, housefly larvae, and *Drosophila* they are peglike hair derivatives. The stimulus is separated from the neural elements of the peg by a nonliving cuticle with no detectable mucous or other secretion. This is cited as evidence that odour molecules stimulate directly without prior solution in a liquid covering. However, some authorities believe that in vertebrates olfactory stimulation occurs only after the molecules are dissolved in the mucous covering. In spite of differences, a number of studies suggest that odour sensitivity of man and insect are very similar.

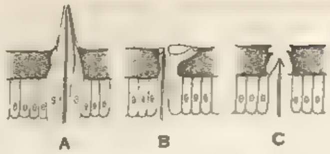


AFTER K. VON FRISCH

FIG. 4.—FEELER OF WORKER BEE: JOINTS 4 THROUGH 11 BEAR OLFACTORY ORGANS

In vertebrates nerve fibres enter the olfactory bulb to end in a series of intricate basketlike clusters called glomeruli. Each glomerulus receives impulses from about 26,000 receptor cells and sends them through 24 mitral cells and 68 tufted cells. Most mitral-cell axons pass into the lateral olfactory tract en route to the olfactory centres at the base of the brain. Most tufted cells cross over to the olfactory bulb of the opposite side. Electrical activity in the primary olfactory nerves, the olfactory bulb, lateral olfactory tracts, and pyriform lobe of the brain have been recorded following odour stimulation. When the bulb is removed, ability to discriminate odours is lost; details of higher olfactory centres were unclear in the 1960s.

Olfactory Qualities.—Most of the great variety of olfactory qualities have been given the names of substances; attempts to classify them have had limited success. One of the best known is that



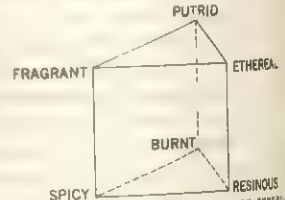
AFTER K. VON FRISCH

FIG. 5.—CROSS SECTION OF OLFACTORY ORGANS ON FEELERS OF INSECTS: (A) OLFACTORY PEG, (B) PORE PLATE, (C) PIT PEG

of H. Henning, who tested more than 400 different scents on people, concluding that there were six main odour qualities: ethereal (fruity), fragrant (flowery), resinous, spicy, putrid (foul), and burnt. Their interrelations could be systematized by a smell prism (fig. 6) based on sensory qualities rather than odorants or basic odour receptors. Electrical activity can be detected readily by wire electrodes inserted into the olfactory bulb. A wavelike response may be recorded from the surface of the bulb and impulse activity is observed that reflects discharges in the mitral cells following stimulation. Those portions of the bulb toward the anterior (or oral) region in the rabbit are most sensitive to water-soluble substances, the more posterior parts being most responsive to fat-soluble odorants.

Use of very fine electrodes shows that specific groups of mitral cells are sensitive to different chemicals. However, these findings fail to show the existence of only a few primary receptors since there is a variety of different combinations of sensitivity. Recordings from the primary receptor nerve fibres show similar patterns of sensitivity. Electrical recording also shows that discharges in the sympathetic nerves to the nasal mucosa instigated by a painful stimulus (*e.g.*, a pinch on the foot of the animal) enhance olfactory sensitivity. This appears to be a reflex that heightens sensory function when the animal is in the presence of danger.

Odorous Substances.—To be smelled, a substance must be volatile enough for its molecules to be carried by air currents. Although chemicals that have both water and fat solubility are often strong odorants, many that are soluble only in water or only in fat lack odour.



S. S. STEVENS (ED.), "HANDBOOK OF EXPERIMENTAL PSYCHOLOGY" (1951); REPRODUCED BY PERMISSION OF JOHN WILEY & SONS INC.

FIG. 6.—H. HENNING'S SMELL PRISM

Only seven of the elements are odorous: fluorine, chlorine, bromine, iodine, oxygen as ozone, phosphorus, and arsenic. Most odorants are organic compounds in which molecular arrangement, structure, and chemical formula are related to odour. Stereoisomers (*i.e.*, different arrangements of the same molecular components) may have different odours; yet a series of benzene derivatives all have a similar odour. The earliest known benzene derivatives (*e.g.*, oil of wintergreen, oil of anise) prompted chemists to call the whole class aromatic compounds (*see* BENZENE; ISOMERISM).

The scent of flowers and roots depends on minute quantities of highly odorous essential oils. Although the major odorant constituents can be identified by chemical analysis, botanical essences are so complex that their odours can be duplicated only by adding them in small amounts to synthetic formulations.

Odour Sensitivity.—In spite of the relative inaccessibility of the olfactory end organs, they are estimated to be 10,000 times more sensitive than taste buds. For example, ethyl mercaptan is reported detectable at a concentration of 4×10^{-8} mg. per litre of air. A just noticeable difference in intensity may appear with a 20% increase in odorant strength; at low concentrations as much as 100% increase may be required. Temperature and humidity influence the strength of an odour by affecting volatility. Hunting dogs follow a trail better when humidity is high since evaporation and dissipation of the odour are retarded. In perfumes (*q.v.*) fixatives are used to retard evaporation of the more volatile constituents. The temporary anosmia (lack of olfactory function) following colds in the nose may be complete or partial; in the latter case only the odours of certain substances are involved. Paranosmia (change in perceived odour quality) also may occur. Changes in sensitivity are said to occur during the menstrual cycle, particularly with odorants related to steroids and sex hormones. Also, olfactory acuity is said to become more acute during hunger.

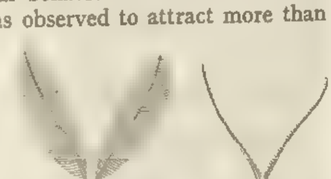
The stench of the slaughterhouse or chemical laboratory ceases to be noticeable after a few minutes. Such olfactory adaptation, as measured by a rise in threshold, is greater with stronger odours. Cross adaptation also is observed; thus eucalyptol may be difficult to detect after adaptation to camphor. Adaptation long was regarded as the result of changes in the receptor; however, peripheral receptors do not seem to adapt. Rhythmic electrical discharges in the olfactory bulb persist long after the experimenter fails to detect the odour that is stimulating the experimental animal. Apparently olfactory adaptation is mediated partially in the brain.

Odour Blending and Flavour.—The ancient art of perfumery and the modern techniques of odour control rest on mixing and blending. Two different odours presented at once may be readily identifiable. The more they resemble each other, the greater will be the tendency to blend, yet trained workers usually can discriminate the components of a successfully blended perfume. The substantially greater intensity of one odour may mask another. Masking and chemical neutralization are the commonest bases for odour control; air may be deodorized by passing it through activated charcoal (*see* DEODORIZER).

The distinctive flavours of food are known largely through the sense of smell. Flavour is the composite of experiences from many senses, but the aroma of roast beef and the delicate bouquet of wine are olfactory in origin. Flavour technology assumes an ever-increasing place in the food industry; it is common for flavour panels of several trained members to judge the flavour of new food products. Using psychophysical methods with adequate statistical control, such panels are highly reliable and can detect qualities so subtle that they defy the methods of physics and chemistry. In using large, untrained consumer panels the emphasis is on acceptability, often a hedonic or emotional reaction to the product. Such judgments are unstable and may vary among individuals because of idiosyncrasy or differences in experience. The strong cheese odour so palatable to the gourmet can produce revulsion in the uninitiated (*see* FLAVOUR; FOOD PREPARATION).

Effects on Behaviour.—Recognition of friend or foe by social insects may depend on olfactory cues: certain ants attack their own kind furiously if deprived of the sense of smell by amputation of their antennae; bees entering a strange hive are put to death because the scent of a foreign hive clings to them. Bees also have an organ on the end of the abdomen that deposits a scent to mark newly discovered food sources, and guide foraging workers. The scents of flowers attest to the evolutionary importance of odour; those flowers that attract insects most efficiently are the likeliest to be pollinated and to continue the botanical species (*see* SCENTED PLANTS).

The effect of odours on sexual behaviour of invertebrates is most striking; a female moth was observed to attract more than 100 males during relatively brief observation periods of 6½ hours. The remarkable difference in antennal development between the male and the female moth is shown in fig. 7. Mammals in the wild state appear to use their odour glands for sexual attraction. Laboratory rats show preference for the arm of a maze scented with the odour of a sexually receptive female. It is possible that some rudiments of these effects operate in man. The most sexually provocative perfumes have a high proportion of musk or musklike odour. Genuine musk is derived from glands of the musk deer and is chemically related to human sex hormones; recall also that odour sensitivity in humans varies with the menstrual cycle. Further study of these effects in man is needed to understand many aspects of smell and taste familiar in everyday life.



(A) AFTER SPIEGEL, (B) AFTER MARCUS

FIG. 7.—OLFACTORY ORGANS OF CRUSTACEA; FIRST ANTENNA OF (LEFT) CRANGON (MALE): (RIGHT) UROLYCHUS NITIDUS (FEMALE)

BIBLIOGRAPHY.—E. C. Crocker, *Flavor* (1945); V. G. Dethier and L. E. Chadwick, "Chemoreception in Insects," *Physiol. Rev.*, vol. 28 (1948); F. A. Geldard, *The Human Senses* (1953); R. W. Moncrieff, *The Chemical Senses* (1946); G. H. Parker, *Smell, Taste, and Allied Senses in the Vertebrates* (1922); C. Pfaffmann, "Taste and Smell" in S. S. Stevens (ed.), *Handbook of Experimental Psychology* (1951); Y. Zotterman (ed.), *Olfaction and Taste: a Symposium* (1963); J. E. Amoore *et al.*, "The Stereochemical Theory of Odor," Sci. Amer., vol. 210 (Feb. 1964); W. Gorman, *Flavor, Taste, and the Psychology of Smell* (1964); E. L. Wick, "Chemical and Sensory Aspects of the Identification of Odor Constituents in Food . . . ," *Food Tech.*, vol. 19 (May 1965).
(CA. P.)

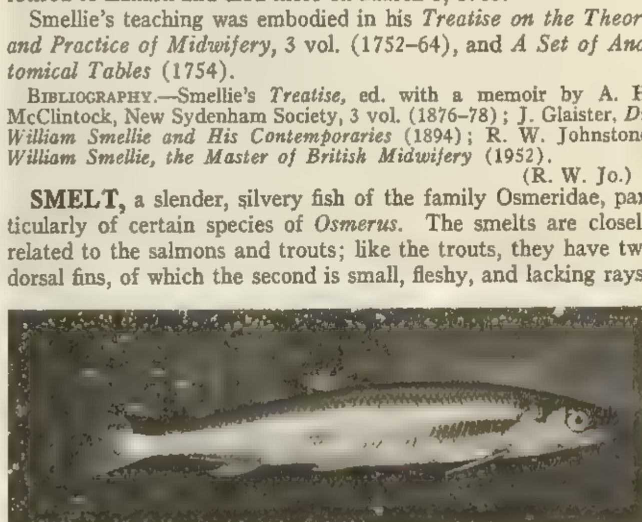
SMELLIE, WILLIAM (1697–1763), "the master of British midwifery," was born at Lanark in 1697. In 1739 he went to London, where in 1741 he began to teach midwifery by lecture-demonstrations for midwives and medical students. One reason for his success was that he attended poor women free of charge on condition that his students were allowed to watch. He thus broke the female monopoly in midwifery, and medical men (men midwives) began to be called in to attend childbirth.

Smellie invented (simultaneously with A. Levret of Paris) the "long" obstetric forceps, but his immortal contribution was his discovery and lucid description of "the mechanism of labour" (*i.e.*, the way in which the child's head adapts itself to the changes in shape and dimensions of the bony pelvic canal during birth). He retired to Lanark and died there on March 5, 1763.

Smellie's teaching was embodied in his *Treatise on the Theory and Practice of Midwifery*, 3 vol. (1752–64), and *A Set of Anatomical Tables* (1754).

BIBLIOGRAPHY.—Smellie's *Treatise*, ed. with a memoir by A. H. McClintock, New Sydenham Society, 3 vol. (1876–78); J. Glaister, *Dr. William Smellie and His Contemporaries* (1894); R. W. Johnstone, *William Smellie, the Master of British Midwifery* (1952).
(R. W. JO.)

SMELT, a slender, silvery fish of the family Osmeridae, particularly of certain species of *Osmerus*. The smelts are closely related to the salmons and trouts; like the trouts, they have two dorsal fins, of which the second is small, fleshy, and lacking rays;



W. T. DAVIDSON FROM NATIONAL AUDUBON SOCIETY

AMERICAN SMELT (OSMERUS MORDAX)

the ventral fins are attached below the first dorsal fins. They are chiefly cold-water fishes confined to the northern hemisphere.

The American smelt (*Osmerus mordax*) ranges along the shores from Virginia to eastern Labrador and has become naturally landlocked in many lakes and ponds of the northeastern U.S. and Canada. They were introduced in 1912 into Crystal Lake, Mich., from where they spread throughout the Great Lakes, becoming an important commercial species. Whether smelt are marine or landlocked, they come in to shore in late winter and spring and run for short distances upstream to spawn. The small, sticky eggs cluster and adhere to any object which they touch; they hatch in 8 to 30 days. The American smelt may grow to be 14 to 15 in. long, but most adults are usually smaller. The European smelt (*O. eperlanus*), ranging from Scandinavia to the English Channel, is similar in appearance and habits to the American form.

Several closely related species occur in the North Pacific region. These include the rainbow herring (*Osmerus dentex*), the surf smelt (*Hypomesus pretiosus*), the Kodiak smelt (*Thaleichthys albatrossis*), and the eulachon or candlefish (*Thaleichthys pacificus*). A number of unrelated small fishes somewhat resembling osmerids externally are also called smelt in various parts of the world.

In the United States commercial fishermen take almost 11,000,000 lb. of smelts annually, and amateurs take an unknown but large additional quantity. More than 70% of the total is from the Great Lakes, a large part of the remainder from the Pacific Coast, and about 1% from New England.

See also FISH: *Survey of the Bony Fishes*. (L. A. WD.)

SMELTING: *see* METALLURGY.

SMERDIS (6th century B.C.), the generally accepted Greek form of name originally given by Herodotus to Bardiya, son of Cyrus the Great of Persia and brother of Cambyses (*q.v.*). According to Xenophon, Cyrus on his deathbed appointed him governor of Media, Armenia, and Cadusia (Gilan). Murder and intrigue have made the true facts of Smerdis' life extremely obscure. According to one theory he was murdered by Cambyses (before Cambyses went to Egypt), presumably because he feared Smerdis might rebel in his absence. The murder was kept secret and Smerdis was successfully impersonated by Gaumata, a Magian, who seized the throne in March 522 B.C. while Cambyses was still in Egypt. He was acknowledged as king in Persia, Media, and the other provinces. In September, after the death of Cambyses, Darius I (*q.v.*), with the help of six associates, killed Gaumata and seized the throne.

This account, given by Darius in his inscription at Bisitun, is in the main followed by Herodotus, Ctesias, and other Greek authors. Contemporary Aeschylus, however, believed that King Smerdis (whom he calls Mardos) really was the legitimate heir (*Persae*, 774–776). This belief is shared by some modern scholars who take the view that Darius, not Cambyses, murdered Smerdis and invented the story of Gaumata to justify his action.

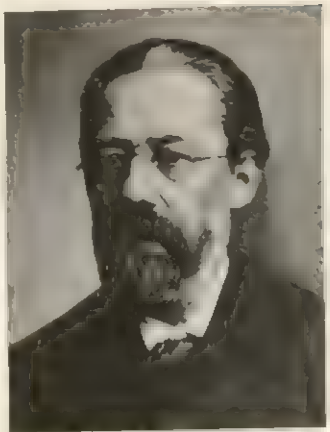
Whoever King Smerdis may have been, he did, according to Herodotus, confer great benefits on his subjects, granting them freedom from military service and tribute for three years, so that his death was greatly mourned. Darius also acknowledges his hold on the people but ascribes it to fear of his tyranny and accuses him of the destruction of temples and wholesale confiscation of property. His death was annually celebrated in Persia by a feast called "The Killing of the Magian," at which no Magian was allowed to appear.

In 521 B.C. a certain Vahyazdata, taking advantage of the confusion, rose against Darius and also claimed to be Smerdis the son of Cyrus. He gained considerable support and sent an army into Arachosia (Kandahar province), but was finally defeated and slain. (J. M. M.-R.)

SMET, PIERRE JEAN DE (1801–1873), Belgian Jesuit, known throughout the western United States as the Indians' truest friend, was born at Termonde, Belg., in 1801. He sailed to the United States in 1821 to enter the Society of Jesus, and was ordained priest in 1827 at St. Louis University in Missouri, his headquarters for life. During 40 years he crossed the Atlantic 19 times, visiting European cities and courts to beg funds, equipment, and recruits for the university and Indian missions. He founded his first mission in Iowa (1838). In negotiating peace between the Sioux and Potawatomi his unusual power over Indians was revealed. For 30 years, having their complete confidence, he was welcomed among all tribes. He helped pacify the Plains Indians in 1841; in 1863, when he went alone to avert a Sioux attack; and in 1868, when he induced Sitting Bull to negotiate. He arranged peace in the Pacific northwest in 1851 and, as United States commissioner with Gen. William S. Harney, again in 1858. Ever outspoken for Indian rights and against punitive expeditions, De Smet won ready consent to establish Christian missions and send teachers. Between sea voyages and western travels totaling 180,000 mi., he published three books in four languages. His vigorous and winning personality made friends of statesmen, pioneers, traders, and soldiers. He died in St. Louis, May 23, 1873.

See Hiram Martin Chittenden and Alfred Talbot Richardson, *Life, Letters, and Travels of Father Pierre-Jean De Smet, S.J.*, 4 vol. (1905). (J. V. J.)

SMETANA, BEDŘICH (1824–1884), Czech composer of operas and symphonic poems, founder of the Czech national school of music. Born at Leitomischl (Litomyšl) in eastern Bohemia on March 2, 1824, he studied the piano in his youth with Josef Proksch and later became music teacher to the family of Count Leopold Thun. Encouraged by Liszt, he opened a piano school in Prague in 1848 and the following year married the pianist Katharina Kolař. In 1856 he wrote his first symphonic poems and in the same year was appointed conductor of the philharmonic society of Göteborg (Sweden), where he remained until 1861. He then returned to Prague where he played the leading part in the establishment of the national opera house.



ARCHIV FÜR KUNST UND GESCHICHTE, BERLIN

BEDŘICH SMETANA

Smetana's first opera, *Braniboři v Čechách* ("The Brandenburgers in Bohemia"), was produced in Prague in 1866. This was followed by the production on May 30, 1866, of his second opera, *Prodaná nevěsta* ("The Bartered Bride"), which later established Smetana's reputation as a distinctively Czech composer throughout Europe and the U.S. His later operas were less successful. *Dalibor*, written under the influence of Wagner, was given in 1868. *Libuše*, named after a legendary figure in the history of Prague and intended to celebrate the projected coronation of the emperor Francis Joseph I as king of Bohemia, was not given until 1881, when its production marked the opening of a new opera house at Prague. In 1874 Smetana resigned his conductorship of the Prague Opera and later in the year became totally deaf. Between 1874 and 1879, however, he wrote some of his finest works, notably the cycle of six symphonic poems bearing the collective title *Má vlast* ("My Country") and the string quartet to which he gave the title *Zmého života* ("From My Life"). *Hubička* ("The Kiss") was successfully produced in 1876, but *Čertova stěna* ("The Devil's Wall"), given in 1882, shows signs of declining powers. Following attacks of depression and symptoms of mental instability, Smetana entered an asylum at Prague and died there on May 12, 1884. The Smetana Society, founded in Prague in 1931, maintains a museum containing the composer's manuscripts and other documents, and sponsors the publication and performance of his works.

BIBLIOGRAPHY.—F. Bartos (ed.), *Smetana: Letters and Reminiscences*, Eng. trans. (1955); Z. Nejedlý, *B. Smetana*, 7 vol. (1950–54); R. Gorer "B. Smetana and A. Dvořak," in *The Heritage of Music*, vol. iii, ed. by H. J. Foss (1951).

SMETHWICK, a municipal county, and parliamentary borough, Staffordshire, Eng., adjoining Birmingham 3½ mi. NW. Pop. (1961) 68,390. Area 3.9 sq.mi. The borough was incorporated in 1899, became a county borough in 1907, and a parliamentary borough in 1918. It returns one member. It was mentioned in Domesday Book as "Smedewich." The historic inventions of Matthew Boulton, James Watt, and William Murdock (*qq.v.*) had their origin in the Soho foundry at Handsworth which was partly removed to Smethwick in 1848. The town has since become known for its glass works and there are manufactures of lighthouse apparatus, iron and steel, and a variety of metal hardware.

SMETONA, ANTANAS (1874–1944), Lithuanian statesman and journalist who in 1919 became the first president of Lithuania and later returned to power as an authoritarian head of state for the last 13 years of his country's independence. Born on Aug. 10, 1874, in the Ukmergė district, under Russian rule, Smetona studied law at St. Petersburg, graduating in 1902. He edited the first Lithuanian daily, *Vilniaus Žinios*, and the Democratic Party's organ, *Lietuvos Ūkininkas*. In 1905 he was elected to the presidium of the Vilnius diet, which proclaimed Lithuanian autonomy. From 1907 to 1913 he edited the journal *Viltis*; and in 1913 he founded *Vairas*, later the organ of the Nationalist Party.

On the convocation of the *Lietuvos Taryba*, or Council of Lithuania, during the German military occupation of the country in World War I, Smetona was unanimously elected its president in September 1917; and from April 1919 to June 1920 he was president of the newly proclaimed state. In 1921 he served as chairman of the Lithuanian delegation at Riga for the settlement of the Latvian-Lithuanian boundary dispute; and early in 1923 he was authorized to negotiate with the Allies during the crisis over Memel (Klaipėda). In 1923 also he became lecturer on philosophy in the newly established university in Kaunas.

After the military *coup d'état* of Dec. 16–17, 1926, organized

by a Nationalist group and backed by the Christian Democrats, Smetona was elected president of the Lithuanian republic. He was reelected in 1931 and in 1938, but on June 15, 1940, when Lithuania was occupied by Soviet forces, he fled to Germany and thence, in March 1941, to the United States. He died in a fire at Cleveland, O., on Jan. 9, 1944. *See* also LITHUANIA: *History*. (MA. G.)

SMIBERT (SMYBERT), **JOHN** (1688–1751), British-American portrait painter, one of the first well-trained artists to work in the American colonies, was born in Edinburgh, March 24, 1688, and apprenticed to a house painter. He went to London about 1709 to study art and worked there, except for a three-year visit to Italy, until 1728. In that year he accompanied George Berkeley (*q.v.*) to Rhode Island as prospective professor of fine arts in the college which Berkeley hoped to establish in Bermuda. When this project failed, he went in 1730 to Boston, Mass., where he married and spent the rest of his life. He painted many prominent Bostonians, but had to supplement his income by keeping a shop, where he sold art supplies and prints. His collection of copies from the "old masters" was unique in colonial America and influenced the development of several later artists, notably Copley and Allston. In 1742 he drew the plans for Faneuil Hall, Boston. He died in Boston, April 2, 1751. His best-known work is "Bishop Berkeley, Family and Friends" (1729, Yale University), which contains a self-portrait of the artist. Many of his single portraits are exhibited in American museums and historical societies. (D. H. W.)

SMILES, SAMUEL (1812–1904), Scottish author best known for his didactic work *Self-Help* (1859), which with its successors, *Character* (1871), *Thrift* (1875), and *Duty* (1880), enshrined the basic Victorian values associated with the "gospel of work." He was born at Haddington, Berwickshire, on Dec. 23, 1812. One of 11 children left fatherless in 1832, he early learned the meaning of self-reliance. Although he qualified in medicine at Edinburgh in 1832, he soon abandoned medical practice for journalism, moving to Leeds, where from 1838 to 1842 he edited the radical *Leeds Times*. His radicalism had much in common with that of Ebenezer Elliott and Joseph Hume (*qq.v.*), and, like them, he was a zealous advocate of material progress based on individual enterprise and free trade. From 1845 to 1866 he was engaged in railway administration, and in 1857 he published a life of George Stephenson. He followed this with *Self-Help: With Illustrations of Character and Conduct*, the outcome of a series of lectures on self-improvement given to young men in Leeds; 250,000 copies had been sold by the end of the century and it was widely translated. Smiles wrote many other books, including *Lives of the Engineers* (three volumes, 1861–62; five-volume enlarged edition, 1874), a pioneer study in economic history; and an *Autobiography* (edited by T. Mackay, 1905). He died in Kensington, London, on April 16, 1904.

See Self-Help, with introduction by Asa Briggs (1958); T. B. Green, *The Life and Work of Dr. Samuel Smiles* (1904). (A. BRI.)

SMILLIE, the name of a family of U.S. painters, engravers, and etchers.

JAMES SMILLIE (1807–1885), engraver, was born in Edinburgh, Scot., on Nov. 23, 1807, and emigrated to New York in 1829. He was a landscape line engraver of some distinction, the best known of his numerous works being a series of large plates after Thomas Cole's "The Voyage of Life." After 1861 he devoted his time to the engraving of bank notes. He died on Dec. 4, 1885.

JAMES DAVID SMILLIE (1833–1909), his eldest son, an engraver, etcher, and painter, was born on Jan. 16, 1833, in New York City. He studied with his father and at the National Academy of Design. His work includes engraved steel vignettes for bank notes and some illustrations, notably F. O. C. Darley's pictures for Dickens' and Cooper's novels. He was elected an associate of the National Academy in 1865—the year after he first began painting—and an academician in 1876, and was a founder (1866) of the American Water Color Society and of the New York Etching Club. Smillie's most significant works are his etchings, dry points, and aquatints, which show complete technical mastery. He died on Sept. 14, 1909.

GEORGE HENRY SMILLIE (1840–1921), younger son of James Smillie, was a landscape painter. Like his brother, he first studied with his father; later he was a pupil of James MacDougal Hart. He, too, was a member of the National Academy of Design and of the American Water Color Society. Smillie's best-known work probably is the "Lake in the Woods," first shown at the National Academy in 1872. He died Nov. 10, 1921.

SMITH, ADAM (1723–1790), Scottish political economist and philosopher, whose book *The Wealth of Nations* exerted the greatest influence on subsequent economic and political theory and practice, was the son of the comptroller of the customs at Kirkcaldy, Fife, Scot. The exact date of his birth is unknown, but he was baptized at Kirkcaldy on June 5, 1723, his father having died some six months previously. In 1737 he proceeded to Glasgow university, studying moral philosophy under "the never-to-be-forgotten" Francis Hutcheson (as Smith called him). In 1740 he entered Balliol college, Oxford, but as William Robert Scott has said, "the Oxford of his time gave little if any help towards what was to be his lifework," and he withdrew from the college in 1746. In 1748 he began delivering public lectures in Edinburgh under the patronage of Lord Kames. Some of these dealt with rhetoric and belles-lettres, but later he took up the subject of "the progress of opulence," and it was then, in his middle or late 20s, that he first expounded the economic philosophy of "the obvious and simple system of natural liberty" which he was later to proclaim to the world in his *Inquiry into the Nature and Causes of the Wealth of Nations*. About 1750 he met David Hume, who became one of the closest of his many friends.

In 1751 Smith was appointed professor of logic at Glasgow university, transferring in 1752 to the chair of moral philosophy. His lectures covered the field of ethics, rhetoric, jurisprudence and political economy, or "police and revenue." In 1759 he published his *Theory of Moral Sentiments*, embodying some of his Glasgow lectures. This work, which established Smith's reputation in his own day, is concerned with the explanation of moral approval and disapproval. His capacity for fluent, persuasive, if rather rhetorical argument is much in evidence. He bases his explanation, not as the third Lord Shaftesbury and Hutcheson had done, on a special "moral sense," nor, like Hume, to any decisive extent on utility, but on sympathy. There has been considerable controversy as to how far there is contradiction or contrast between Smith's emphasis in the *Moral Sentiments* on sympathy as a fundamental human motive, and, on the other hand, the key role of self-interest in *The Wealth of Nations*. In the former he seems to put more emphasis on the general harmony of human motives and activities under a beneficent Providence, while in the latter, in spite of the general theme of "the invisible hand" promoting the harmony of interests, Smith finds many more occasions for pointing out cases of conflict and of the narrow selfishness of human motives.

Smith now began to give more attention to jurisprudence and political economy in his lectures and less to his theories of morals. An impression can be obtained as to the development of his ideas on political economy from the notes of his lectures taken down by a student in about 1763 which were later edited by E. Cannan (*Lectures on Justice, Police, Revenue and Arms*, 1896), and from what Scott, its discoverer and publisher, describes as "An Early Draft of Part of *The Wealth of Nations*," which he dates about 1763.

BY COURTESY OF THE NATIONAL GALLERIES OF SCOTLAND

ADAM SMITH, PASTE MEDALLION BY JAMES TASSIE, 1787

At the end of 1763 Smith obtained a lucrative post as tutor to the young duke of Buccleuch and resigned his professorship. From 1764–66 he traveled with his pupil, mostly in France, where he came to know such intellectual

leaders as Turgot, D'Alembert, André Morellet, Helvétius and, in particular, François Quesnay, the head of the Physiocratic school (*q.v.*) whose work he much respected. On returning home to Kirkcaldy he devoted much of the next ten years to his magnum opus, which appeared in 1776. In 1778 he was appointed to a comfortable post as commissioner of customs in Scotland and went to live with his mother in Edinburgh. He died there on July 17, 1790, after a painful illness. He had apparently devoted a considerable part of his income to numerous secret acts of charity.

Shortly before his death Smith had nearly all his manuscripts destroyed. In his last years he seems to have been planning two major treatises, one on the theory and history of law and one on the sciences and arts. The posthumously published *Essays on Philosophical Subjects* (1795) probably contain parts of what would have been the latter treatise.

The Wealth of Nations.—It is on *The Wealth of Nations* that Smith's renown securely rests. This work, in the first place, with those of Sir William Petty, Richard Cantillon, Quesnay and Turgot, did most to create the subject of political economy and develop it into an autonomous systematic discipline; and secondly, in the western world, it is the most influential book on the subject ever published. In the plan of Smith's lifework, however, it seems that *The Wealth of Nations* had represented only one section of a comprehensive system of moral or social philosophy, in which he had intended to survey and explain the progress of society in all its aspects.

Controversial views have been expressed as to the extent of Smith's originality in *The Wealth of Nations*. However, it seems justifiable to regard him neither as the original and almost unique creator or founder of political economy, nor, on the other hand, as a completely unoriginal systematizer, deeply indebted in particular to Turgot and Quesnay. A survey of the influences on, or ingredients of, Smith's work would almost amount to a history of the economic thought of the century or more before his time. In his general philosophical approach Smith continued the tradition of the "natural law" school of Hugo Grotius and Samuel von Pufendorf, and in the broad layout of his ideas started from his Glasgow predecessors Gershom Carmichael and, above all, Francis Hutcheson. Hume must have helped in strengthening and developing his liberal ideas, and Montesquieu his historical approach. His English predecessors, Petty, John Locke, Bernard Mandeville, Joseph Harris and Sir James Steuart, must also have been of some help.

The Wealth of Nations is a work of much wider sweep than its 19th-century successors, much wider even than *The Principles of Political Economy* by John Stuart Mill. Above all, it is marvelously rich in learning about earlier periods as well as acute in its observation of the contemporary scene. It is also a great work of political philosophy, most eloquently and persuasively delivered. As regards the economic analysis and theorizing, this provides a fairly loose and slender thread, not usually at all rigorously worked out and occurring mostly in the first two books. Nevertheless, *The Wealth of Nations*, rambling and prolix though it often is, did systematize its subject, with enormous gains to its public influence. Its five books are surveyed below:

Book I: The Division of Labour, Value and Distribution.—"The annual labour of every nation is the fund which originally supplies it with all the necessaries and conveniences of life" is Smith's opening sentence, and it enunciates, with its primary emphasis on the "real" productivity of labour, one of the leading themes of the work. Increasing the productivity of labour requires, above all, its specialization or "division," on which Smith places extreme emphasis as the key to economic progress. The division of labour is limited by the extent of the market, but, as exchanges multiply and ramify, money as a medium of exchange is required, and there arise the problems of the values and prices of goods and services. In these opening five chapters Smith pursues a now well-worn sequence of ideas, than which there is no more lucid or logical way of leading into the central problems of economics or social economy.

Smith begins his treatment of value with the distinction between value in exchange and value in use, or "utility," and drives this distinction so far as to proclaim that value in use is not necessary even for "the greatest value in exchange." After setting out, without keeping clearly separate, the concepts of the labour commanded by a good, the toil and trouble of commanding a good, and the labour contained in it, and having laid down labour as a measure or standard of value, he settles for an explanation of "natural" prices in terms of cost of production, made up of wages, profit and rent. Smith's distinction between "natural" and "market" prices established a pattern of analysis followed out by the main 19th-century theories of value and price. Natural price is the "long-run" competitive equilibrium price determined by cost of production, while market price is the "short-run" price, determined by supply and demand. Natural price consists of wages, rent and profits at their "natural" levels. As regards wages, these will tend in a stationary economy to a subsistence level, though they will remain above subsistence in a progressive economy. Smith emphasizes the stronger bargaining position of combinations of employers as against the workers, and also suggests the idea behind the later "wages fund" doctrine that wages are an advance by the employer to maintain the worker while his product is ripening. He distinguishes profit as a separate income by marking it off from wages of "inspection and direction," and he notes the element of compensation for risk. He argues that through increasing competition the rate of profit will tend to fall as an economy progresses toward the stationary state. On rent Smith's position is rather ambiguous, and his chapter on the subject includes a vast digression on variations in the value of silver. His explanation mentions the element of monopoly, as well as the productivity of land, and he argues that rent enters into price "in a different way from wages and profit," since "high or low wages and profit are the causes of high or low price; high or low rent is the effect of it"—thus anticipating David Ricardo's theory. Book I ends with an analysis of the effects of economic progress on wages, profits and rents—the incomes, that is, of the "three great, original and constituent orders of every civilized society, from whose revenue that of every other order is ultimately derived." Smith's analysis provided the pattern for subsequent important attempts on this subject, though the precise logic of his model is not entirely clear.

Book II: Capital, Saving and Investment.—The title of the second book is "Of the Nature, Accumulation, and Employment of Stock," the term "stock" including consumers' capital, as well as producers' capital, which latter is capital in Smith's sense. He classifies capital into fixed and circulating, each with four subbranches. He includes money as a part of the circulating capital of society, but not explicitly the means of subsistence advanced to productive labourers. In a long chapter on money, "the great wheel of circulation, the great instrument of commerce," he explains the saving to the economy which arises from the development of banking and paper money in place of the precious metals.

Smith comes next to his definitions of "productive" and "unproductive" labour, in which he is attempting to adapt the peculiar definition of the Physiocrats according to which only agricultural labour was "productive." Smith's distinction seems little more helpful for modern analysis, since he defines services—though not retailing—as unproductive, "how honourable, how useful, or how necessary soever" they may be. There follows Smith's theory of saving and investment, or that saving *is* investment. This lays unqualified emphasis on the universal and inevitable beneficence of "parsimony," or private voluntary saving, as "the immediate cause of the increase of capital," since, for any income-receiver, "that portion which he annually saves, as, for the sake of the profit, it is immediately employed as a capital, is consumed in the same manner" as what he spends on his consumption. Hence "every frugal man is a public benefactor," since, apparently, for every individual decision to save there is a corresponding decision to invest.

Joseph Schumpeter has pointed out the immense influence over the next century and a half of this analysis. It is very similar to that of Turgot's published ten years previously in his *Réflexions*, though Smith elaborates its implications, especially for policy, much more emphatically. The Turgot-Smith theory swept away

the rudimentary analysis of effective demand of their predecessors as well as Quesnay's reservations regarding the inevitable beneficence of saving.

Smith's analysis of saving and investment, implying the utilization of all income for consumption or investment spending, formed the basis for important parts of his attack (notably in the first chapter of book iv) on what, following Physiocratic usage, he called "the mercantile system," for its preoccupation with monetary policies and "the popular notion that wealth consists in money." There Smith argues that "it is not for its own sake that men demand money" which "can serve no other purpose besides purchasing goods." Complaints of a general scarcity of money are wrongheaded, for "though a particular merchant, with abundance of goods in his warehouse, may sometimes be ruined by not being able to sell them in time, a nation or country is not liable to the same accident . . . It might, indeed, suffer some loss and inconveniency, and be forced upon some of those expedients which are necessary for supplying the place of money. The annual produce of its land and labour would be the same or very nearly the same."

Smith's theory of saving and investment is also an essential part of his general campaign for reliance on the invisible hand of the individualist profit motive as against state action. For, as he argues when discussing tariffs, "no regulation of commerce can increase the quantity of industry in any society beyond what its capital can maintain. It can only direct a part of it into a direction which it might not have gone." The same system of ideas pervades, also, his final chapter of book iv, "Of Public Debts," with its underlying assumption that all government expenditure is "unproductive." However, perhaps in contrast with his extremely emphatic and influential analysis of saving and investment, Smith favoured a legal maximum rate of interest rather than complete freedom. Nevertheless, what John Maynard Keynes was to call "the classical system of ideas" first emerged fully developed in *The Wealth of Nations*. The "law of markets" of Mill and Jean Baptiste Say, and what came to be known in Great Britain in the 1920s and 1930s as "the Treasury view," that government expenditure could do little or nothing to raise the level of employment, are traceable, above all, to Adam Smith.

Book III: "Of the Different Progress of Opulence in Different Nations."—This is probably the oldest part of the work, deriving from Smith's Edinburgh lectures. It forms an interlude on the course of economic development as it proceeds by the exchange of goods and services between town and country or manufacturing and agriculture, and it is illustrated by a selective survey of European economic history. Smith argues for a "natural course of things," according to which "the capital of every growing society is, first, directed to agriculture, then to manufacture, and lastly to foreign commerce." Any other order is "unnatural and retrograde." Indeed, it is in agriculture that the investment of a given capital is most advantageous to society. Over much of Europe, though fortunately not to the same extent in Britain, pernicious legislation with regard to land tenure and primogeniture, and internal obstructions and export prohibitions on the corn trade, have distorted the natural order and prevented the invisible hand from harmonizing the private and social gains from investment.

Book IV: "Of Systems of Political Economy".—This is devoted to Smith's attack on "the mercantile system" in the framework of a general review of commercial policies. It concludes with a critical but friendly chapter on "the agricultural system" of the Physiocrats. Smith's most definitely identifiable target was the existing mass of restrictive regulations and monopolistic institutions, practices and privileges, in fact, all those "systems either of preference or of restraint" which "being thus completely taken away, the obvious and simple system of natural liberty establishes itself of its own accord." He has already argued for free choice of occupations by the removal of apprenticeship restrictions and settlement laws, and for free trade in land by the abolition of primogeniture and entails. In book iv he calls, of course, for internal free trade, which he recognizes as a vital factor in the economic progress of Britain as contrasted with France. Most important are his weighty and detailed case for freedom in international trade and his attacks on duties, bounties and the monopolistic privileges of the chartered companies. However, he makes certain concessions regarding restrictions in special circumstances on the export of corn, a moderate export tax on wool and moderate import duties on manufactures where these might yield "a considerable advantage" to a country's own industries, though he will give no protection to infant industries as such. Smith argues for gradualness or "a good deal of reserve and circumspection" in the removal of restrictions, where considerable unemployment might be caused, and says that to expect that complete free trade should ever be established in Great Britain is Utopian.

It is in book iv that Smith introduces several of his most famous passages advocating the free play of self-interest: "As every individual, therefore, endeavours as much as he can both to employ his capital in the support of domestic industry, and so to direct that industry that its produce may be of the greatest value; every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it."

In his very striking chapter "Of Colonies" Smith campaigns against all nationalistic and monopolistic restrictions on colonial trade, claiming that these benefit only a small minority and not the broad masses of the nation. Because of the burden of defense "under the present system of management, therefore, Great Britain derives nothing but loss from the dominion which she assumes over her colonies." After a searching discussion of political relationships he suggests the vision of a free-trade association of politically independent partners, in which the mother country and its colonies would become "faithful, affectionate, and generous allies."

Book V: "Of the Revenue of the Sovereign or Commonwealth."—This deals with the expenditure (or functions) of the state and with taxation and public debts. Smith confines the role of the state to three heads: (1) defense (which in a famous aside approving the Navigation acts he had declared to be "much more important than . . . opulence"); (2) justice and civil government, particularly "the defence of the rich against the poor"; and (3) that of "erecting and maintaining those public institutions and those public works, which, though they may be in the highest degree advantageous to a great society, are, however, of such a nature, that the profit could never repay the expense to any individual or small number of individuals." This third heading could, of course, be a most capacious one. Smith specifically mentions the provision of roads, bridges, canals, harbours and, to some extent, education as duties of the state. He discusses further possibilities such as banks, his main criterion being that any state enterprise should pay its way. He emphasizes that what the state may beneficially take on is relative to the efficiency and standards of the government of the day, and he expresses a low view of the capacities of the contemporary British government; however, he has an equally unfavourable opinion of the efficiency of most joint stock companies. Previously he had enunciated as a principle (which might legitimatize an almost indefinite extension of governmental intervention) that "those exertions of the natural liberty of a few individuals which might endanger the security of the whole society, are, and ought to be, restrained by the laws of all governments." As regards taxation, Smith favours proportionality, along with certainty, convenience and economy, though later he claims that "it is not very unreasonable that the rich should contribute to the public expense, not only in proportion to their revenue, but something more than in that proportion."

Though Smith's main treatment of the functions of government comes in book v, various suggestions for beneficent state action are scattered throughout *The Wealth of Nations*. As J. Viner has

pointed out, Smith never collected his full program together and, indeed, probably omitted to mention some kinds of state activity which he would heartily have favoured, such as some sort of provision for the destitute or unemployed. So state action emerges as an exception, which is to be assumed guilty until proved innocent or beneficent, while private initiative is to be assumed beneficent until proved guilty. Taken together, his considerable exceptions might build up into a very large role for the state, though, unlike his predecessors and his successors since Keynes, he held rigidly that the state neither could nor should seek to control the aggregate level of activity and employment. Indeed it was the sweep and force of Smith's advocacy of the simple system of natural liberty which gave *The Wealth of Nations* its revolutionary influence and which constituted its great message for succeeding generations: "All systems either of preference or of restraint, therefore, being thus completely taken away, the obvious and simple system of natural liberty establishes itself of its own accord. Every man, as long as he does not violate the laws of justice, is left perfectly free to pursue his own interest his own way, and to bring both his industry and capital into competition with those of any other man, or order of men. The sovereign is completely discharged from a duty, in the attempting to perform which he must always be exposed to innumerable delusions, and for the proper performance of which no human wisdom or knowledge could ever be sufficient; the duty of superintending the industry of private people, and of directing it towards the employments most suitable to the interest of the society." At the same time *The Wealth of Nations* provided a first great survey of economic life in all its interdependence, and a framework for the development of political economy as a separate discipline.

For a modern edition of *The Wealth of Nations* see that of E. Cannan, 2 vol., 6th ed. (1950). *See* also *The Theory of Moral Sentiments*, 6th ed. (1790); and *Essays on Philosophical Subjects*, new ed. (1872).

See also references under "Smith, Adam" in the Index.

BIBLIOGRAPHY.—Dugald Stewart, *Biographical Memoirs of Adam Smith*, etc. (1811); J. Rae, *Life of Adam Smith* (1895); W. R. Scott, *Adam Smith as Student and Professor* (1937). For commentaries on Smith's political economy *see* the prefaces to J. R. McCulloch's, J. E. T. Rogers', J. S. Nicholson's and E. Cannan's editions of *The Wealth of Nations* (1863, 1869, 1901 and 1950); *Adam Smith 1776–1926*, including notably an essay by J. Viner, "Adam Smith and Laissez-faire" (1928); J. A. Schumpeter, *History of Economic Analysis*, pp. 181–194 (1954). For Smith's moral philosophy *see* Sir Leslie Stephen, *History of English Thought in the 18th Century*, vol. ii, pp. 70–77 (1876); W. R. Sorley, *History of English Philosophy*, new ed., pp. 156–194 (1937).
(T. W. H.)

SMITH, ALFRED EMANUEL (1873–1944), U.S. politician and governor of New York state, was born in New York city on Dec. 30, 1873, in the Old Fourth ward under the Brooklyn bridge, of poor but respectable parents. Upon the death of his father, who had been a rather unsuccessful truckman, in 1886, Smith's education at St. James's parochial school, near his home, was interrupted. Forced to help support his mother and sister, the boy held a variety of jobs and finally became a checker at the Fulton fish market, where he remained for seven years. Even as a small boy Smith had given indications of talent in elocution and amateur dramatics, and, despite long hours at work, he continued his interest in theatricals and at one time considered a stage career.

CULVER PICTURES, INC.

ALFRED E. SMITH

Smith's political career began in 1895 when the Tammany district leader made him an investigator in the office of the city commissioner of jurors, with a salary of $800 a year. In the autumn of 1903 he was elected to the New York state assembly on the Democratic ticket. He went to Albany, N.Y., with little formal education but with an alert mind and an intimate knowledge of the city district he represented. He served continuously from 1904 to 1915. By 1913 he had become speaker of the assembly, an office second only to that of governor in power. During his first ten years in the legislature Smith gave small indication of the political independence he was to show later as governor. Industrious and intelligent, he was nevertheless a routine politician who obeyed the orders of Charles F. Murphy the Tammany Hall "boss." After 1911, however, his viewpoint was broadened through his appointment to a commission which investigated factory conditions. In 1915 he was sent as a delegate to the constitutional convention held at Albany to revise the state's fundamental law.

Tammany Hall rewarded Smith in 1915 by making him sheriff of New York county. In 1917 he was elected president of the board of aldermen of Greater New York. The following year he resigned to run for governor against Charles S. Whitman. Although it was generally believed that he did not have a chance, he won by the narrow margin of 14,000 votes. He served four terms as governor, the first man in the history of the state to have this honour.

"Al" Smith, as he was affectionately known, showed extraordinary ability as a vote-getter. In 1920, when he ran against Nathan L. Miller, he was defeated in the Republican presidential landslide, but polled 1,000,000 more votes in his state than did the Democratic presidential nominee. He then became head of the United States Trucking corporation, but in 1922 was drafted to run for governor again. This time he defeated Miller, seeking re-election, 1,397,633 to 1,011,725—at that time the largest majority ever given a gubernatorial candidate in New York. In 1924 he defeated Theodore Roosevelt, Jr., and in 1926 defeated Ogden L. Mills. As governor, Smith fought for adequate housing, improved factory laws, proper care of the insane, child welfare and state parks. He effected a reorganization of the state government on a consolidated, businesslike basis. He repeatedly demonstrated his gubernatorial leadership by forcing Republican legislatures to accept his recommendations.

Smith was the first Roman Catholic to receive serious consideration as a presidential candidate in the United States. His religion combined with his long record as an opponent of prohibition, resulted in a prolonged deadlock at the Democratic national convention of 1924. His nomination was opposed by the Protestant "dry" faction led by William G. McAdoo. On the 103rd ballot a compromise candidate, John W. Davis, was nominated, but lost the election to Pres. Calvin Coolidge.

In 1928 Smith was the leading candidate for the Democratic presidential nomination. At the Houston (Tex.) convention of the party on June 28 he was again placed in nomination by Franklin D. Roosevelt. The first ballot gave him the necessary two-thirds vote of the delegates, and Sen. Joseph T. Robinson was later nominated for vice-president. Though the western and southern Democratic delegates managed to bury the hatchet at the convention they could not persuade the voters to do so in the election. The opposition to Smith in the rural districts of closely contested states in the west and south contributed to his defeat in the election.

Smith carried on an aggressive campaign, making several extended speaking tours into the south and west. Hailed by Roosevelt as the "happy warrior," he was a picturesque figure who made his brown derby, cigar and colourful speech his trademarks and "The Sidewalks of New York" his theme song. But when the electoral votes were counted it was found that he had lost to Herbert Hoover, 444 to 87. Smith lost his own state as well as five southern states that had not gone Republican since the reconstruction era. The popular vote was much closer, 21,391,000 to 15,016,000. Smith polled the largest popular vote ever given to a Democrat up to that time.

The remaining years of Smith's life were embittered by the financial difficulties of the Empire State building, of which he became president, and by a bitter conflict with Franklin D. Roosevelt. Smith lost contact with his old following and supported the

Republican presidential candidates in 1936 and 1940. He died on Oct. 4, 1944.

Bibliography.—Oscar Handlin, *Al Smith and His America* (1958); Alfred E. Smith, *Up to Now* (1929); Henry F. Pringle, *Alfred E. Smith, a Critical Study* (1927). (H. F. Pr.; Os. Ha.)

SMITH, CHARLOTTE (1749–1806), English novelist and poet, once famous for her sonnets and her romances, was born in London, May 4, 1749, the daughter of Nicholas Turner, a man of property. Her husband, Benjamin Smith, whom she married in 1765, was imprisoned for debt from December 1783 to July 1784 and then fled to France to avoid creditors. There his wife joined him until, by her efforts, he could return to England (1785). Two years later she left him and turned to writing to support her children (she had 12, of whom 6 survived her). She had already published *Elegiac Sonnets* (1784) but, although these were well received, her free translations of *Manon Lescaut* (1785) and part of *Les Causes Célèbres* (*The Romance of Real Life*, 1786) promised higher financial reward for novels: hence *Emmeline, or the Orphan of the Castle* (1788), *Ethelinde, or the Recluse of the Lake* (1789), which Fanny Burney enjoyed, etc. *Desmond* (1792), based on the innocent love of a man for a married woman, was accused of immorality, and its political ideas, inspired by the French Revolution, also caused some talk. Mrs. Smith began *The Old Manor House* (1793), generally considered her best novel, when, with Cowper and Romney, she was the guest of William Hayley (*q.v.*); she read a chapter to the company each evening. Scott praised "her invention . . . . her knowledge of the human bosom, her power of natural description, her wit and her satire"; but in spite of her Cowperlike attitude to nature and her radical ideas her novels, with their rather stilted dialogue and occasional Gothic effects, belong essentially to the large derivative group of 18th-century women novelists.

Toward the end of her life she turned to palatably instructive books for children, the best being *Conversations Introducing Poetry* (1804). She died at Tilford, near Farnham, Surrey, on Oct. 28, 1806.

See F. M. A. Hilbish, *Charlotte Smith* (1941).

SMITH, GEORGE (1840–1876), English Assyriologist who deciphered (1872) a cuneiform tablet in the British Museum that bore a Chaldean account of the flood (*see* Flood [in Religion and Myth]). He was born in Chelsea, London, on March 26, 1840, and apprenticed to an engraver of bank notes at 14. Through Sir Henry Creswicke Rawlinson, Smith was permitted to study inscriptions in the British Museum and in 1867 became an assistant in the department of Assyriology. His translation of one inscription established the date of a total eclipse of the sun in the month Sivan, in May 763 B.C., and another fixed the date of an invasion of Babylonia by the Elamites in 2280 B.C. In 1873 the *Daily Telegraph* gave 1,000 guineas to Smith for excavations in Mesopotamia that unearthed tablets at Nineveh (*q.v.*) with missing portions of the account of the flood. The British Museum sent Smith in 1874 and 1876 to continue excavations of the library of Ashurbanipal (*q.v.*); he died at Aleppo, Syr., on Aug. 19, 1876. (J. M. Wi.)

SMITH, SIR GEORGE ADAM (1856–1942), Scottish preacher and Semitic scholar who helped to make generally acceptable the higher criticism of the Old Testament, was born in Calcutta, India, on Oct. 19, 1856. Educated in Edinburgh (at the Royal High School, the university and New College), with vacation study at Tübingen and Leipzig, he taught at the Free Church College, Aberdeen (1880–82), during the suspension of W. Robertson Smith (*q.v.*). After ordination he made his reputation as a preacher at Queen's Cross Free Church, Aberdeen (1882–92), uniting sound scholarship with a vivid sense of the relevance of God's Word to the listening congregation. These qualities were apparent also in *The Book of Isaiah* (2 vol. 1888–90; revised 1929). During his tenure of the Old Testament professorship at the Free Church College, Glasgow (1892–1909), he made several lecture tours in the U.S. and published *The Historical Geography of the Holy Land* (1894; revised 1931), the outcome of detailed observation and investigation made in Palestine; it proved invaluable to Gen. Sir William Allenby in the Palestine campaign of 1917. At Glasgow Smith also produced *The Book of the Twelve Prophets* (two volumes, 1896–98; revised 1928), a life of his friend Henry Drummond (1898), *Modern Criticism and the Preaching of the Old Testament* (1901), and *Jerusalem* (two volumes 1907–08). The advanced views of *Modern Criticism* (lectures delivered at Yale) nearly led to a process for heresy against him at home. As principal of the University of Aberdeen (1909–35) he wrote *The Early Poetry of Israel* (1913) and *Jeremiah* (1923), was knighted in 1916, and was moderator of the General Assembly of the United Free Church of Scotland, 1916–17. He died at Balerno, near Edinburgh, on March 3, 1942.

See S. A. Cook in *Proceedings of the British Academy*, vol. 28 (1942), and the life by his widow, L. A. Smith (1943).

SMITH, SIR HARRY GEORGE WAKELYN, Bart. (1787–1860), British general, governor of Cape Colony and high commissioner in South Africa from 1847 to 1852, was born at Whittlesey, Cambridgeshire, on June 28, 1787, the son of John Smith, a surgeon. Commissioned as an ensign in the army in 1805, he saw service in South America (1807) and then for two periods during the Peninsular War. He took part in the retreat to, and battle of, Corunna in January 1809, and was at the storming of Badajoz (April 1812). Here two strikingly beautiful Spanish sisters, one of them, the 14-year-old Juana María de los Dolores de León, sought protection in the British lines. Smith later married Juana, who accompanied him constantly on his military and subsequent civil duties. From 1814 to 1815 Smith, now a captain, served as assistant adjutant general to the British force in the war between Britain and the United States, and was present at the capture and burning of Washington (August 1814). On his return to England he was appointed assistant quartermaster general to a division of the duke of Wellington's army in the Low Countries, in which capacity he served in the Waterloo Campaign of June 1815.

After a number of military postings in Britain, Smith went to Jamaica in 1826 as deputy quartermaster general of the forces there and in 1828 was transferred to the Cape Colony in the same capacity. On the outbreak of the sixth Kaffir War (*see* Kaffir Wars) in December 1834, Sir Benjamin D'Urban, governor of the Cape, appointed Smith, with the rank of colonel, commandant of the regular and the burgher forces, and second in command of the colony. Smith at once made his famous ride from Cape Town to Grahamstown (January 1835), covering the 600 mi. in 6 days, in order to allay panic in the eastern districts of the colony which were exposed to the depredations of the Amaxosa, a branch of the Bantu people. His soldierly qualities and vigorous conduct of the campaign earned him the admiration of the Boer farmers. At the end of the war Smith administered the frontier territory between the Keiskamma and Great Kei rivers (which was named the province of Queen Adelaide) by means of a benevolent despotism until the territory was returned to the tribes at the end of 1836.

After a brief period in Cape Town, Smith was appointed adjutant general of the army in India in March 1840. With the honorary rank of major general he distinguished himself in a number of engagements in 1845 and 1846, winning special fame as the hero of the victory over the Sikhs at Aliwal (Jan. 28, 1846), when he led the final charge which routed the enemy. For this achievement he was created a baronet and in November was promoted to be major general. He returned to the Cape in December 1847 as governor and also held the recently created office of high commissioner in South Africa. Soon after his arrival he ended the seventh Kaffir War and reannexed the old province of Queen Adelaide to the crown as British Kaffraria, where he instituted, as part of a civilizing policy, a number of far-reaching changes which reduced the authority of the chiefs.

In February 1848 he visited Natal, where he tried, with little success, to check a new emigration of Boers from that district. In the same month, he annexed to the crown, as the Orange River sovereignty, the territory north of the Cape Colony between the Orange and the Vaal rivers. Smith was faced with hostile demonstrations of the colonists in the Cape in 1848 and 1849 against the attempt of the colonial secretary, Earl Grey, to land convicts at the Cape. He faithfully transmitted their objections to London,

and in September 1849 refused to allow any of the convicts to come ashore. The precipitate nature of Smith's changes in British Kaffraria was probably the principal cause of the eighth Kaffir War (1850–53). The British government's dissatisfaction with his handling of the war, which was long and costly, caused his recall in March 1852. He subsequently held various military posts in Britain and was appointed lieutenant general in 1854. He died in London on Oct. 12, 1860. Harrismith (Orange Free State), Ladysmith (Natal), and Ladismith (Cape) are South African towns named after Smith and his wife. His autobiography, edited by G. C. Moore Smith, was published in two volumes in 1901. (L. M. Y.)

SMITH, HENRY JOHN STEPHEN (1826–1883), British mathematician who specialized in the theory of numbers, was born in Dublin, Ire., on Nov. 2, 1826. When he was two years old his father died and his mother left Ireland for England. After being privately educated he entered Rugby in 1841 and Balliol College, Oxford, in 1844. He was elected a fellow of Balliol in 1850 and Savilian professor of geometry in 1861. He was elected a fellow of the Royal Society in 1861, served on various royal commissions, and from 1877 was chairman of the managing body of the Meteorological Office. He died at Oxford on Feb. 9, 1883.

Smith published a few papers on geometry and then began a study of the existing work on the theory of numbers. The results of his researches are contained in his *Report on the Theory of Numbers*, which appeared in the British association volumes from 1859 to 1865. His further original researches on the subject were communicated to the Royal Society in two memoirs, "Systems of Linear Indeterminate Equations and Congruences" and the "Orders and Genera of Ternary Quadratic Forms" (*Phil. Trans.*, 1861 and 1867). After 1864 he devoted himself chiefly to elliptic functions. His *Collected Papers*, prefaced by several biographical notices, were edited by J. W. L. Glaisher (1894). (O. Oe.)

SMITH, JEDEDIAH STRONG (1799–1831), an explorer of the American West who deserves to rank with Lewis and Clark (*see* Lewis and Clark Expedition); he was the most notable single personality of the fur trade era. Born Jan. 6, 1799, in Jericho (now Bainbridge), N.Y., he first ascended the Missouri River with the first fur trade venture of William H. Ashley and Andrew Henry in the spring of 1822. Smith quickly rose to prominence, for he was intelligent, tough, and adaptable, though "a mild man and a Christian." In the fall of 1823 he led a party of fur hunters to Wind River, and in March 1824, made the effective discovery of South Pass (in present-day Wyoming), gateway to the American West. (The pass had been traversed a dozen years earlier by Robert Stuart, but the discovery was not followed up.) After a year as Ashley's partner, in 1826 Smith formed a new partnership with David E. Jackson and William L. Sublette and promptly led to southern California the first expedition to reach that area from the Great Salt Lake country. The following spring Smith became the first man to cross the Sierra Nevada and traverse the Great Basin from west to east. On a second expedition in 1827–28 he made the first exploration by land up the California-Oregon coast. All these explorations were financed by beaver catches. In 1830 Smith, Jackson and Sublette sold out to the new Rocky Mountain Fur Company, transporting their furs to St. Louis, Mo., in the first wagons ever brought to the northern Rockies.

On May 27, 1831, Smith was killed by Comanches on the Cimarron River, while en route to Santa Fe, N.M. This early death delayed the publication of his journals and maps, which would have signally advanced geographic knowledge.

Bibliography.—Jedediah Smith, *Travels*, ed. by Maurice S. Sullivan (1934); Dale L. Morgan, *Jedediah Smith and the Opening of the West* (1953); Dale L. Morgan and Carl I. Wheat, *Jedediah Smith and His Maps of the American West* (1954). (D. L. M.)

SMITH, JOHN (*c.* 1580–1631), soldier, explorer, author, and one of the principal founders of the first permanent English colony in North America, was born into a farming family at Willoughby, Lincolnshire, Eng. At the age of 15 he left school and was apprenticed to a wealthy merchant. Soon after his father's death he went to fight in the Netherlands. At the age of 20, having returned to England and studied military horsemanship, he went off to the wars against the Turks in Hungary. His exploits there once questioned, have been substantiated by Hungarian records. His acts included deluding the enemy by the use of fireworks, for which he was made a captain, and killing three opponents in individual combat. Left for dead on the battlefield, Smith was taken as a captive into Asiatic Turkey from which he escaped into Russia and eventually returned to England in 1604.

ENGRAVING IN THE CORNER OF A MAP OF NEW ENGLAND BY S. DE PASSE, 1616; BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

JOHN SMITH

He soon attached himself to a group that was preparing to establish an English colony in North America. In 1606 the way was cleared with the granting of a royal charter to the Virginia Company. When the colonists (105 in number according to Smith) set sail on Dec. 20, Smith was one of their acknowledged leaders.

On April 26, 1607, the three little ships ("Susan Constant," "Godspeed," and "Discovery") sailed into Chesapeake Bay. According to instructions given in London, the colonists then opened a sealed box that contained the names of seven men who were to form the governing council of the colony. Smith, though named, was not allowed to take his seat because he had been charged with concealing a mutiny. A few weeks later, having proved his courage and resourcefulness as explorer and soldier, he was admitted to the council. Meanwhile, on May 14, the colonists settled at Jamestown, Va. Soon after the "Susan Constant" returned to England for badly needed supplies, the colony suffered a virulent disease that nearly killed Smith. Food was running short; no adequate houses had been built; and discord was rising among the settlers. The council, reduced to four members, deposed their president, Edward Maria Wingfield. Smith, taking charge of outdoor affairs, soon became the dominant figure. He directed the building of houses, traded with the Indians, and began the series of river voyages that enabled him to draw a remarkably fine map of Virginia. While exploring the Chickahominy River in December 1607, he was ambushed and captured by Indians. Taken to the great chief Powhatan, he was about to be executed when, according to Smith's account, the chief's daughter Pocahontas, a girl of about 13, threw herself between him and his executioners. Thereafter Smith was initiated into the tribe and Pocahontas became his friend, coming often to Jamestown with gifts and on one occasion warning Smith of a planned attack.

When Smith returned to Jamestown, Gabriel Archer, a member of the council, tried to have him hanged for the loss of the men who had gone with him, and would have succeeded but for the return of Captain Christopher Newport from England with fresh supplies. He found only 38 of the original settlers still alive.

Smith, after further exploration of Chesapeake Bay and its rivers, became president of the colony on Sept. 10, 1608. He expanded the fort to form a five-sided enclosure, tightened up on military training and guarding, sent off boats to trade for food, went with Newport to crown Powhatan according to orders received from London, and continued to get corn from the Indians where others often failed. But the situation remained desperate. After two years the English still depended on England and the Indians for their food. They had failed to find any gold or silver or to return any profits to the investors in London. They had not found the hoped-for passage to the Pacific and were in constant danger from the Indians. Under Smith's direction, however, small quantities of tar, pitch, and soap ashes were made, a well was dug, 20 houses were built, fishing was put on a regular basis, crops were planted, and outlying forts were built.

In August 1609, four supply ships arrived, bringing several settlers who had formerly been sent home and were bitter enemies of Smith. Gabriel Archer and John Ratcliffe demanded that Smith give up the government because, under a new charter granted by King James, a new lieutenant governor, Sir Thomas Gates, and an admiral, Sir George Somers, were on their way with five more ships, with Lord De La Warr to follow as governor. Since the new officers and their commissions had not arrived, however, Smith insisted he must govern until they came, and refused to admit Archer and Ratcliffe to their former council seats, though he accepted John Martin, another former member. Hoping to improve a bad situation by dispersion, Smith sent Francis West (De La Warr's brother) up to the falls of the river (the site of Richmond) with 120 men. Martin went down the river with an equal number to settle near the Nansemond Indians. In September, returning from an attempt to straighten out difficulties among West's men, Smith was so seriously injured when his powder bag caught fire in his boat that he was impelled to go aboard a ship that was soon to leave for England. The measure of Smith's ability is found in what happened after he left. Ratcliffe and a whole company were slaughtered by Indians; West sailed off with another group to England; all the domestic animals were killed and eaten; guns were traded to the Indians for food; and still the colony starved. Men fed upon roots and acorns, and even upon their dead comrades. Of the 500 Smith left behind, only 60 were alive when the new leaders, shipwrecked in the Bermudas, finally reached Jamestown. Within a few days they decided to abandon the settlement and return to England, and were on their way when they met De La Warr coming with new supplies.

Accusations against Smith had been sent to the Virginia Company in London, but the disposition of his case is not known, as the records have disappeared. Time, however, vindicated many of his policies—the necessity of dealing firmly with the Indians, the wisdom of spreading out the plantations, the folly of trying to produce naval stores, silk, glass, or gold.

Having quarreled with the Virginia Company, but still eager to explore and settle in America, Smith made contact with the Plymouth Company, which had a charter for colonizing the northern coast from 38° to 45°. In 1614, having arranged for the financing of two vessels, he sailed to the area he named New England, carefully mapping the coast from Penobscot Bay to Cape Cod. Though he failed to find gold and whales, he did catch some fish and trade for furs. He saw that to make fishing profitable, settlers should be established on land to dry and pack the catch. In time each port of old England could have its counterpart in New England.

In 1615 Smith set out again, this time to establish a permanent settlement in New England. A ship of 200 tons and a smaller one of 50 sailed from Plymouth, but within 200 miles they were separated by a gale. The larger lost its masts and had to return to port. In a 60-ton bark Smith set out again, but was soon captured by French pirates. Smith's ship was eventually freed, but he himself (through the treachery of his own captain) was left aboard one of the pirate ships. When they attacked Spanish ships, Smith joined in the fighting on promise of a share in the spoils. During his three months with the pirates he wrote his *Description of New England*, a counterpart to his *Description of Virginia* published in 1612. In the end the pirates threatened to accuse him as the destroyer of French settlements in America unless he renounced his share. So he escaped from the vessel at night, survived a storm in a rowboat, and reached shore only to learn that the pirates had sunk in the storm with all their spoils.

Returning penniless to England, Smith met his friend Pocahontas, now on a visit with her husband John Rolfe, and in 1617 assembled three ships for another colonizing attempt, but was windbound for three months and never got off. He lived, wrote, and talked New England for the rest of his life, attempting to go first with the Pilgrims and then with the settlers of Boston, both of whom benefited from his books and maps. But he never saw America again. Much of his time was now devoted to writing. His works include the beautifully printed *Generall Historie of Virginia, New England and the Summer Isles* (1624), *The Seaman's Grammar* (1627), and *The True Travels, Adventures, and Observations of Captaine John Smith in Europe, Asia, Africa and America* (1630), which tells of his early exploits. A set of first editions of his works, if they could be assembled, would cost over $25,000.

As a pioneer in Virginia, explorer of the Chesapeake and other rivers, map maker, namer of New England (to which his press agentry attracted the Puritan settlements), Smith might well say that the English colonies were "pigs of his sow." He had a faith in America and an understanding of its greatness as a cradle of freedom which few men of his time approached. He died in June, 1631, and was buried in St. Sepulchre's Church, London, where his epitaph may still be seen.

Smith's works were edited by Edward Arber (Birmingham, 1884, and Edinburgh, 1910). They have often been printed in part in collections of voyages. *Captain John Smith, His Life & Legend* (1953), by Bradford Smith, is based on evidence collected from contemporary sources gathered for the first time, most of which corroborate Smith's own accounts. (B. Sm.)

SMITH, JOSEPH (1805–1844), U.S. founder of the Mormon religion, was born at Sharon, Vt., Dec. 23, 1805, the third son of Joseph and Lucy Mack Smith. His early career is shrouded in controversy. Neighbours in Palmyra, N.Y., described him as a romancer and diviner who dug for buried treasure, and their stories can be largely verified by court records, early newspaper accounts, and the autobiography of his mother. In his own history, written after 1838, Smith described himself as an unschooled, pious boy who had a vision of God and Jesus Christ at the age of 14. At 17, he said, an angel directed him to buried golden plates containing a history of the American Indians, which described them as descendants of Hebrews who had centuries earlier sailed to America by way of the Pacific. This *Book of Mormon* he translated from "reformed Egyptian" with the aid of magic stones he called the Urim and Thummim.

By most objective scholars the book is regarded as a potpourri of local Indian origin legends, fragments of autobiography, and the current religious and political controversies, particularly Anti-masonry, all transformed with remarkable ingenuity into a religious saga of quasi-biblical authority. After its publication in 1830, Smith offered it as scientific evidence of his divine calling, and claimed that his new church, organized April 6, 1830, was a restoration of the ancient, primitive Christian religion.

Converts in ever-increasing numbers followed Smith from New York to Ohio, then to Missouri, and on to Illinois, despite great suffering. The young prophet governed his people with the help of revelations, which dealt with such divergent matters as the anthropomorphic nature of God, the communistic United Order, and the size, location, and financing of a hotel. He taught that the heavens are peopled with many gods, and that every man can aspire in time to be a god himself.

Joseph Smith was a man of commanding presence, athletic grace, and prodigious personal charm. A genius at improvisation, he managed to combine Jewish and Christian mysticism with the goal of perpetual prosperity, and to establish the tradition of Mormonism as a complete "way of life." Although his revelation on polygamy was not made public (by Brigham Young) until 1852, eight years after Smith's death, and though it is not supported by *The Book of Mormon*, there is evidence that he may have married as many as 50 wives. He publicly acknowledged only his first, Emma Hale Smith, who bore him nine children.

When a handful of dissenting Mormons attacked him in their newspaper, Smith had their press destroyed. For this he was imprisoned, with his brother Hyrum. On June 27, 1844, an anti-Mormon mob stormed the jail at Carthage, Ill., and shot both men. The church split into pieces, but the majority followed Brigham Young (*q.v.*), a sagacious administrator and colonizer, who led the Mormon exodus to Great Salt Lake. A smaller group, led by the eldest of Smith's four surviving sons, who denied that their father had practised polygamy, settled in Independence, Mo. (*See* also LATTER-DAY SAINTS, CHURCH OF JESUS CHRIST OF.)

Joseph Smith's writings include, in addition to *The Book of Mormon* (1830), *A Book of Commandments for the Government*

of the Church of Christ (1833); *Doctrine and Covenants of the Church of The Latter Day Saints* (1835).

BIBLIOGRAPHY.—Fawn M. Brodie, *No Man Knows My History: the Life of Joseph Smith, the Mormon Prophet* (1945); Thomas O'Dea, *The Mormons* (1957); Lucy Mack Smith, *Biographical Sketches of Joseph Smith the Prophet and His Progenitors for Many Generations* (1853); B. H. Roberts (ed.), *History of the Church of Jesus Christ of Latter-Day Saints*, 6 vol. (1902–12). (F. M. B.)

SMITH, SIR KEITH MACPHERSON (1890–) and **SIR ROSS MACPHERSON** (1892–1922), brothers, Australian pioneer aviators, were born at Adelaide, S. Austr., on Dec. 20, 1890, and Dec. 4, 1892, respectively. They were both educated at Queen's school, Adelaide, and at Warriston school, Moffat, Scotland. During World War I, Keith served in the Royal Flying Corps and Royal Air Force as a pilot, 1917–19, while Ross served with the Australian Light Horse in Gallipoli and Sinai. In 1916 he learned to fly in Egypt and spent the last two years of the war in the Australian flying corps in Palestine. He made the first flight from Cairo to Calcutta in 1918. In Nov.–Dec. 1919, Ross as first and Keith as second pilot, in a Vickers Vimy twin-engined biplane with Rolls Royce 375 h.p. engines, accompanied by Sgts. J. M. Bennett and W. H. Shiers as mechanics, made the first flight from England to Australia, landing at Darwin, Northern territory, on Dec 10, 1919. For this flight the brothers received knighthoods and the prize of £10,000. Three years later they began preparations for a pioneer round-the-world flight in a Vickers pusher-amphibian aircraft, but Ross was killed, together with Bennett, while testing the aircraft at Brooklands, Eng., on April 13, 1922. Keith went into business in Sydney, Austr. (D. CR.)

SMITH, SEBA (MAJOR JACK DOWNING) (1792–1868), U.S. editor and humorist, creator of Major Jack Downing, a rustic Yankee whose horse sense and wise saws as the purported counselor and friend of Andrew Jackson brought laughs that swept the country, was born in Buckfield, Me., on Sept. 14, 1792. A graduate of Bowdoin college, Brunswick, Me., Smith in 1829 founded the *Portland Courier*, where the Major's fictional letters first appeared in Jan. 1830, continuing later in the *National Intelligencer* until July 1853. Major Jack was a common man magnified as oracle; yet there was irony in his role of threadbare office seeker exposing follies in a mobocracy. As adviser to President Jackson, Jack was in the heyday of his popularity; this led to shameless counterfeiting of Smith's invention and to his collection of the letters in book form, the last volume being published in 1859 under the title of *My Thirty Years Out of the Senate*. In the Downing letters Smith is intent on the portrayal of New England character, an aim more completely carried out in his *Way Down East* (1854). Smith's title to distinction comes from his hitting upon something new in humour that has become a part of American tradition. Practically forgotten himself, he numbers among his followers in satire James Russell Lowell's Hosea Biglow, Artemus Ward and Will Rogers.

Smith died July 28, 1868, in Patchogue, N.Y.

See Mary A. Wyman, *Two American Pioneers, Seba Smith and Elizabeth Oakes Smith* (1927); Constance Rourke, *American Humor* (1931). (M. A. WN.)

SMITH, SYDNEY (1771–1845), English man of letters and canon of St. Paul's (1831–45), was not only one of the foremost preachers of his day but, as Macaulay said, was, in controversy, "the greatest master of ridicule since Swift," and for wit in conversation "beat all the men I ever met" (Thomas Moore). The intellectual versatility and warmth of feeling beneath his superficial frivolity are best seen in his letters.

The second son of a wealthy and eccentric landowner, Sydney Smith was born at Woodford, Essex, on June 3, 1771. At Winchester, where he went in 1782, he acquired a lifelong hatred of public schools. In 1789 he went up to New College, Oxford, with a scholarship, exchanging it in 1791 for a fellowship of £100 a year. His father refused to let him study law, and, after taking his degree in 1792, he was ordained in 1794 and went as curate to Netheravon, near Amesbury, on Salisbury Plain. In 1797 the squire of Netheravon asked him to accompany his son to Germany as tutor; war intervened, and they went to Edinburgh, where Smith attended lectures in moral philosophy (under Dugald Stewart), chemistry, and medicine. He made many friends, among them Henry Brougham and Francis Jeffery (*qq.v.*), with whom, in 1802, he founded the *Edinburgh Review*. The suggestion was Smith's, and, although never officially editor, he acted as one for the first three numbers. He continued to write for it for 25 years, and his trenchant articles were a main element in its success. In 1800 he married and published his first book. In 1803 he moved to London.

There he quickly made friends, and was appointed alternate Sunday evening preacher at the Foundling Hospital, and morning preacher at Berkeley and Fitzroy Chapels, Mayfair. In 1804 he gave a course of lectures on moral philosophy at the Royal Institution. Their blend of sense and wit attracted attention and people flocked to hear him: the second and third series (1805 and 1806) were also successful.

Had he been on the powerful side in politics, or willing to modify his opinions, Smith would soon have been a bishop. As it was, all that his friends Lord and Lady Holland could achieve, when the predominantly Whig "ministry of all the talents" took office in 1806, was presentation to the living of Foston-le-Clay, Yorkshire. At first he discharged his duties through a curate, but enforcement of the Clergy Residence Act put an end to his pleasant metropolitan life, and in June 1809 he left London.

Meanwhile, in March 1807 the "ministry of all the talents" had been forced to resign on the issue of Catholic emancipation. Smith supported emancipation on liberal and rational grounds. In the summer of 1807 he published the first of his famous *Letters of Peter Plymley to my Brother Abraham who lives in the Country*, attacking Protestant ignorance, obscurantism, and bigotry. Its success was immediate, and it was followed by four more in 1807, and five in 1808, in which year they were collected. Inspired by "a passionate love for common justice, and for common sense," and armed with the weapon of ridicule, Smith did more than anyone else to change public opinion. The letters were published anonymously and were never openly acknowledged, but rumours of Peter Plymley's identity soon got abroad, and Lord Holland wrote to Smith, expressing his view, and Lord Grenville's, that there had been "nothing like it since Swift," but pointing out that, for this, his wittiest performance, Smith had lost a bishopric.

In Yorkshire, although he hated the country, Smith soon showed his characteristic energy and cheerfulness. He planned and built his own house, learned to manage a farm, bred horses, set up a dispensary, and won his parishioners' affection. He continued to support Catholic emancipation, in 1826 publishing *A Letter to the Electors, upon the Catholic Question*, one of his most effective polemics. At last, in 1828, he gained preferment. Just before going out of office, the chancellor, Lord Lyndhurst, presented him with a prebendal stall at Bristol Cathedral, and in 1829, having exchanged his living at Foston for one at Combe Florey, he moved to Somerset.

When he became a canon, Smith stopped writing for the *Edinburgh Review*, thinking it more fitting that a dignitary of the church should sign what he wrote. It was expected that, when the Whigs came to power, he would be made a bishop, but opposition was strong. He only succeeded (September 1831) in exchanging his Bristol stall for a residentiary canonry at St. Paul's Cathedral, which necessitated his living in London three months of the year. He attributed his failure to be made a bishop to his being a "high-spirited, honest, uncompromising man, whom all the bench of bishops could not turn upon vital questions."

He was a champion of parliamentary reform and one of his best fighting speeches was at Taunton (Oct. 11, 1831) when he made the well-known comparison of the House of Lords, which had just thrown out the Reform Bill, with Dame Partington, with mop and pattens, "vigorously pushing away the Atlantic Ocean."

Despite disappointment, and lasting grief at the death in 1829 of his eldest son, Smith retained his high spirits, wit, practical energy, and powers of argumentative ridicule to the last. His *Letter to Archdeacon Singleton, on the Ecclesiastical Commission* (1837), and the other *Singleton Letters* (1838–39); and his *Petition* and *Letters* (1843, to the *Morning Chronicle*; reprinted 1844) on Pennsylvania's repudiation of its debts are as vigorous and spirited as the best of his early essays for the *Edinburgh Review*.

On his brother Courtenay's death in 1839, Smith inherited £50,000, and bought a house in London; and there, on Feb. 22, 1845, he died.

BIBLIOGRAPHY.—*Editions, Correspondence, etc.: The Works,* 4 vol. (1839–40), 3 vol. (1845; 1854), 2 vol. (1859, etc.); *Essays, Social and Political, 1802–25,* 2 vol. (1874, enlarged 1877); *Selections,* ed. by E. Rhys (1892); *Selected Writings,* ed. by W. H. Auden (1956); *The Letters of Peter Plymley, with Other Selected Writings, Sermons and Speeches,* ed. by G. C. Heseltine (1929); *The Letters of Sydney Smith,* ed. by Nowell C. Smith, 2 vol. (1953).

Biography and Criticism: A Memoir of Sydney Smith by his Daughter, Lady Holland. With a Selection from his Letters, ed. by Mrs. [Sarah] Austin, 2 vol. (1855); G. W. E. Russell, *Sydney Smith,* English Men of Letters series (1905); O. Burdett, *The Rev. Smith, Sydney* (1934); Hesketh Pearson, *The Smith of Smiths* (1934); G. Bullett, *Sydney Smith: a Biography and a Selection* (1951).

SMITH, THEOBALD (1859–1934), U.S. pathologist, a pioneer in the comparative etiology and immunology of infectious and parasitic diseases, was born in Albany, N.Y., July 31, 1859. He earned baccalaureate and medical degrees at Cornell University, Ithaca, N.Y. (1881), and Albany Medical College (1883). From 1884 to 1895 Smith was director of the pathological laboratory of the Bureau of Animal Industry in Washington, D.C., where he was also professor of bacteriology at Columbian University (later George Washington University). He served also as director of the pathological laboratory of the Massachusetts State Board of Health (1895–1915); as professor of comparative pathology at Harvard University (1896–1915); and as director of the department of animal pathology of the Rockefeller Institute for Medical Research (1915–29). Some of his most valuable research included the role of ticks in transmitting Texas or southern cattle fever (1888); his differentiation of bovine from human tubercle bacilli (1898), which was of value to Robert Koch in his later work; and the demonstration that killed cultures of bacteria may produce immunity. He died on Dec. 10, 1934.

See Hans Zinsser, "Biographical Memoir of T. Smith," *Biographical Memoirs of National Academy of Science* (1937–38), with bibliography. (W. F. N.)

SMITH, WALTER BEDELL (1895–1961), U.S. Army officer and diplomat, was born in Indianapolis, Ind., Oct. 5, 1895. He began his military career as an enlisted man in the Indiana National Guard, 1910–15, and in 1917 was commissioned second lieutenant of infantry in the U.S. Army. In February 1942, during World War II, Smith was appointed secretary of the joint chiefs of staff and U.S. secretary of the combined chiefs of staff, with the rank of brigadier general. The following September he became chief of staff, European theatre of operations, and chief of staff to Gen. Dwight D. Eisenhower, serving in those posts until Eisenhower's departure from Europe after the war. Smith negotiated and accepted for the Allies the surrender of Italy in 1943 and that of Germany in 1945.

On his return to the United States in 1945 Smith became chief of the Operations and Planning Division of the War Department general staff. Shortly thereafter he was appointed U.S. ambassador to the Soviet Union, holding the post from 1946 to 1949. Later he commanded the U.S. 1st Army, 1949–50, was director of the Central Intelligence Agency, 1950–53 (having been promoted full general in 1951), and on Jan. 31, 1953, retired from active military service to become undersecretary of state. In October 1954 he resigned from government service and entered private business. He died in Washington, D.C., on Aug. 9, 1961. General Smith was the author of *My Three Years in Moscow* (1950) and *Eisenhower's Six Great Decisions* (1956). (F. C. PE.)

SMITH, WILLIAM (1769–1839), English geologist, called the father of English geology and known as "Strata Smith," was born at Churchill in Oxfordshire on March 23, 1769. At 18 he became assistant to a surveyor, Edward Webb, of Stow-on-the-Wold, and traversed the oölitic lands of Oxfordshire and Gloucestershire, the Lias (Lower Jurassic) clays and red marls of Warwickshire and other districts, studying their varieties of strata and soils. In 1791 his observations at Stowey and High Littleton in Somersetshire first impressed him with the regularity of the strata. In 1793 he executed the surveys and levelings for the line of the Somerset coal canal, in the course of which he confirmed his earlier supposition, that the strata above the coal were not horizontal, but inclined in one direction—to the east—so as to terminate successively at the surface.

On being appointed engineer to the canal in 1794 he made a tour with regard to inland navigation. He carefully examined the geological structure of England, and corroborated his generalization of a settled order of succession in the strata. In 1794 he made his first geological map—that of the vicinity of Bath—showing the ranges of the different strata in the area.

In 1799 Smith dictated his first table of British Strata (presented to the Geological society of London in 1831). It was headed *Order of the Strata, and Their Imbedded Organic Remains, in the Neighbourhood of Bath; Examined and Proved Prior to 1799.* In 1813 Joseph Townsend, a clergyman who was also a geologist, published, with acknowledgment, much information on the English strata communicated by William Smith, in a work entitled *The Character of Moses Established for Veracity as an Historian, Recording Events From the Creation to the Deluge.* Meanwhile Smith was completing and arranging the data for his large *Geological Map of England and Wales, With Part of Scotland* (15 sheets, 1815). The map—of great significance in the development of modern geology—was reduced to smaller form in 1819, and from this date to 1822, 21 separate county geological maps and several sheets of sections were published in successive years, constituting a *Geological Atlas of England and Wales.*

Smith's collection of fossils was purchased in 1816–18 by the British museum. In 1817 a portion of the descriptive catalogue, *Stratigraphical System of Organised Fossils,* was published. In 1816 he had commenced the publication of *Strata Identified by Organised Fossils,* with figures printed on paper to correspond in some degree with the natural hue of the strata. In this work (of which only four parts were published, 1816–19) is exemplified the principle he established of the identification of strata by their included organic remains, which changed the whole conception of paleontology. In 1831 the Geological society of London conferred on Smith the first Wollaston medal; and from the government he received a life pension of £100 per annum. The last years of his life were spent at Hackness (of which he made a good geological map), near Scarborough; and in the latter town. He died at Northampton on Aug. 28, 1839.

Smith's *Memoirs,* edited by his nephew, John Phillips, appeared in 1844.

See C. L. and M. A. Fenton, *Giants of Geology,* rev. ed., pp. 70–83 (1952).

SMITH, WILLIAM ROBERTSON (1846–1894), Scottish Semitic scholar and encyclopaedist, whose introduction in the pages of *Encyclopædia Britannica* of the new higher criticism of the Old Testament from Germany led to his being accused of heresy, and who later did fruitful research into comparative religion and social anthropology. He was born on Nov. 8, 1846, at Keig in Aberdeenshire. Educated at home by his father, he had a brilliant career at Aberdeen University (1861–66) and studied for the ministry at New College, Edinburgh (1866–70), for half the time being concurrently assistant to P. G. Tait, professor of natural philosophy at Edinburgh University. In 1867 he spent the summer studying at Bonn and in 1869 at Göttingen. He was ordained in 1870 on his appointment as professor of Oriental languages and Old Testament exegesis at the Free Church College of Aberdeen. Two years later he again visited Göttingen, studying Arabic. Few others in Britain were at that time familiar with the higher critical work of Old Testament scholars on the continent of Europe, and when Robertson Smith's articles on biblical subjects, notably that on the Bible itself (1875), appeared in the ninth edition of *Encyclopædia Britannica* the authorities of the Free Church took strong exception to them.

In 1877 a church committee reported that Smith's advanced views were incompatible with his being a teacher of candidates for the ministry. The assembly in June 1877 suspended him from his teaching duties. He demanded a formal trial and conducted his defense in the protracted case (1878–80) with conspicuous ability and courage, showing that his views did not conflict with the formulation of his church's belief in the inspiration of Scripture. Mean-

while he continued his scholarly work, including that of a reviser of the Authorized (King James) Version of the Old Testament, a task to which he had been appointed in 1875. He also perfected his Arabic by residence and travel in Egypt and other Middle Eastern countries during the winters of 1878 and 1879. In May 1880 the assembly dropped the indictment against him. A second attack on his opinions was launched after the publication in June of his article "Hebrew Language and Literature" in *Encyclopædia Britannica*. He was again suspended, and in May 1881 the assembly voted no confidence in him as a professor and he was removed from his chair (*see* further SCOTLAND, FREE CHURCH OF: *Theological Developments*).

The next month he was appointed joint editor of *Encyclopædia Britannica*. He moved to Edinburgh, and two crowded courses of lectures which he delivered there and in Glasgow respectively in 1881 and 1882 were published as *The Old Testament in the Jewish Church* (1881) and *The Prophets of Israel* (1882). In 1883 he was appointed Lord Almoner's Professor of Arabic at Cambridge and took up residence in Trinity College, moving to Christ's College in 1885 on his election as a professorial fellow there. His Oriental books are preserved at Christ's College. He remained editor in chief of *Encyclopædia Britannica* until the ninth edition was completed in 1888, though much of the work now devolved on a resident editor in Edinburgh, and continued to contribute many articles to it. His article "Sacrifice" (1886) together with his earlier book *Kinship and Marriage in Early Arabia* (1885) are important landmarks in the study of comparative religion, for which he had long had an interest through the influence of his close friend the ethnologist J. F. McLennan.

In 1886 Robertson Smith became university librarian at Cambridge, a post which he relinquished in 1889 on his appointment to succeed William Wright as Sir Thomas Adams's Professor of Arabic. In the same year appeared his most original work, *Lectures on the Religion of the Semites*, which had been delivered in Aberdeen. His health broke down in 1890 and remained precarious until his death at Cambridge on March 31, 1894. One of the most remarkable men of his time, a generous friend, and a brilliant conversationalist, he was regarded by continental scholars and by orientalists generally as the foremost Semitic scholar in Britain.

BIBLIOGRAPHY.—*Life*, with list of his publications, by J. S. Black and G. W. Chrystal (1912). *See* also T. K. Cheyne, *Founders of Old Testament Criticism* (1893); James Bryce, *Studies in Contemporary Biography* (1903); C. E. Raven and S. A. Cook, *William Robertson Smith* (1951). (D. W. TH.)

SMITH COLLEGE, an institution for the higher education of women at Northampton, Mass., was founded under the will of Sophia Smith (1796–1870) of Hatfield, who chose the neighbouring town of Northampton as the site of the college and selected the first trustees. The college was chartered in 1871 and opened in 1875. It became the largest independent college for women in the United States. *See* NORTHAMPTON (Mass.).

SMITHFIELD, an area in the north of the City of London, Eng., known chiefly for its meat market. The market, now administered by the Corporation of London, existed in A.D. 1183. The site also was used for jousting and military training, as a place of execution during the Reformation and Counter-Reformation and for the ancient Bartholomew Fair which is regarded as the forerunner of the famous Caledonian market developed at Islington after the live cattle trade was transferred there from Smithfield in 1855. The several sections of the market, which have been progressively established since 1860, handle fresh and frozen meat, poultry, provisions, fish, and fruit; and 15 mi. of meat rails can accommodate more than 9,000 tons daily with ample cold storage facilities. Supplies are received mainly from Great Britain, the Commonwealth, Europe, and South America. The church of St. Bartholomew the Great is a fragment of the church of the Augustinian priory of St. Bartholomew founded by Rahere in 1123. St. Bartholomew's hospital, founded in 1137, one of the great teaching hospitals of London, was rebuilt beginning in 1730. (H. G. B.)

SMITHSON, JAMES (1765–1829), English scientist and founder of the Smithsonian Institution at Washington, D.C., was the natural son of Hugh Smithson Percy, first duke of Northumberland of the third creation. The duke's family name was Smithson. James Smithson's mother, Elizabeth Keate Macie, was a gentlewoman of large fortune, who had been twice married before James Smithson was born. She was a lineal descendant of King Henry VII. Most of Smithson's large fortune came from his mother's family but some also came from his father's relatives. He was educated at Pembroke College, Oxford, and is said to have been the best chemist and mineralogist of his year. He later became well known for his important work in these fields. The carbonate of zinc is named smithsonite (*q.v.*) in his honour. At the age of 22 he was admitted into membership in the Royal Society on the recommendation of the great physicist Henry Cavendish, among others. He had a great regard for original research and publication. He wrote: "Every man is a valuable member of society who by his observations, researches, and experiments procures knowledge for men."

Smithson never married. Much of his life was spent on the continent, where he was well acquainted with the scientific leaders of the period. He died in Genoa, Italy, on June 27, 1829. The founding of the Smithsonian Institution was the result of a provision of Smithson's will. During his life Smithson never visited America, but in 1904 his remains were brought to the U.S. and are now interred in the original Smithsonian building. *See* also SMITHSONIAN INSTITUTION. (L. CAR.)

SMITHSONIAN INSTITUTION, an institution of learning in Washington, D.C., founded by the bequest of an English scientist, James Smithson (*q.v.*). His estate was left to a nephew, Henry James Hungerford, with the stipulation that, should Hungerford die without issue, or should his issue die intestate or under 21, the whole estate should go "to the United States of America, to found at Washington, under the name of the Smithsonian Institution, an establishment for the increase and diffusion of knowledge among men." Hungerford died without issue in 1835. There was much opposition in the U.S. to the acceptance of Smithson's bequest, especially by John C. Calhoun and others, who held that Congress had no power under the Constitution to accept such a gift, but it was accepted, largely through the efforts of John Quincy Adams.

Establishment.—In September 1838, £104,960 in gold sovereigns was delivered from the clipper "Mediator" to the Philadelphia mint, where it was recoined into U.S. money to the amount of \$508,318.46. In 1867, after the death of Hungerford's mother, a residuary legacy of \$26,210 was received and the fund then amounted to \$650,000. By savings of interest and by other gifts the fund was increased.

After ten years of debate, Congress in 1846 accepted the trust and created by enactment an "establishment" called the Smithsonian Institution, consisting of the president of the United States the vice-president, the chief justice, and the members of the president's cabinet. Smithson's money, a great fortune in that day, was lent to the U.S. Treasury, the government agreeing to pay perpetually 6% interest upon it. The act of Congress that created the institution provided that it should be governed by a board of regents composed of the vice-president and chief justice, three members of the Senate, three members of the House of Representatives, and six private citizens. The institution has a secretary who is its executive officer. The act also provided for a library and for a museum to contain "objects of art and of foreign and curious research, and objects of natural history, etc.," belonging to the United States. The museum was later designated the United States National Museum but remains under the direction of the Smithsonian Institution.

The regents met for the first time on Sept. 7, 1846, and in the autumn of the same year they elected as secretary Joseph Henry (*q.v.*), then a professor at Princeton University, known for his experiments on the electromagnet and other subjects relating to electricity. Under his guidance the institution took shape. Henry seized the unique opportunity offered by the opening up of the great western areas of the United States to make collections of vegetation and wildlife and to study the Indian tribes. A principal feature of his administration was the establishment of international exchanges of scientific literature.

The diffusion of knowledge was promoted by publishing a series of periodical reports on the progress of different branches of knowledge, occasional separate treatises on subjects of general interest, and monographs on special subjects investigated by experts. The Smithsonian was probably the first establishment in the United States to have a staff of full-time research scientists in a broad series of fields. In 1846 a plan was presented for the unification and systematization of weather observation under the institution. In December 1847 an appropriation was made by the board for such meteorological research; in 1849 telegraphic transmission of meteorological intelligence collected by the institution was begun; weather maps were successfully made in 1856. In 1870 the meteorological work of the institution was incorporated with the U.S. Army Signal Corps. Still later the U.S. Weather Bureau was established by act of Congress. After 1854 Henry's annual reports contained a general appendix with reports of lectures, such as were held under the auspices of the institution until 1865, summaries of correspondence, special papers, etc. On the grounds of the Smithsonian is a bronze statue of Joseph Henry by W. W. Story.

National Museum.—Spencer F. Baird (*q.v.*), Henry's successor, incorporated in the general appendix annual reports on the progress of the sciences. He also perfected Henry's system of international exchanges under which the institution, through agents in the principal cities of the world, exchanges its own publications, those of other departments of the U.S. government, and those of learned societies for foreign publications. Baird had been at the head of the U.S. National Museum before he became secretary of the institution, and the museum particularly was developed during his administration. It was built up around the collections of the U.S. Patent Office, which were turned over to it in 1858, and those of the National Institute, transferred to the Smithsonian in 1861 when the institute was dissolved. A part of the collection (including Smithson's collection) was destroyed by fire in 1865. The museum gained much valuable archaeological and ethnological material from the exploring parties sent out under John Wesley Powell (*q.v.*). It obtained excellent ichthyological specimens through Baird's position as U.S. fish commissioner, and acquired general collections from the Centennial Exhibition of 1876. Its great collection of plants is known as the National Herbarium. The museum had a phenomenal growth.

The museum comprises two units: the Museum of Natural History and the Museum of History and Technology. In the 1960s it occupied parts of four buildings and its catalogued objects totaled approximately 60,000,000, distributed among its various departments (anthropology, botany, zoology, entomology, paleobiology, mineral sciences, science and technology, arts and manufactures, civil history, and armed forces history).

The Bureau of American Ethnology was established as a branch of the institution in 1879, when the various organizations doing survey work in the west united as the U.S. Geological Survey, and anthropological and ethnological research under the direction of John W. Powell, was transferred to the Smithsonian Institution. The bureau carried on studies of the American Indians and published monographs and reports of its researches. The river basin surveys became an active program of the bureau in 1945 when a unit was organized to salvage and preserve archaeological materials in the path of national flood-control, irrigation, hydroelectric, and navigation projects. In 1965 the bureau was amalgamated with the National Museum's department of anthropology to become the Smithsonian office of anthropology under the Museum of Natural History.

Astrophysical Observatory.—In 1887 Samuel P. Langley (*q.v.*) was appointed as assistant secretary of the institution, and succeeded as secretary upon Baird's death in the same year. In 1890 Langley established the Smithsonian Astrophysical Observatory to carry on research on the physical aspects of the heavenly bodies, particularly the sun. Its early research concerned especially the measurement of the intensity and characteristics of the infrared solar rays. Langley developed the bolometer, an instrument capable of recording one-millionth of a degree of temperature. After 1900 the research concentrated on measuring the quantity and quality of the sun's radiation as it is found in free space at the earth's mean diameter. Field stations for solar observation were operated at several points in Africa, South America, and the United States. In 1955 Fred L. Whipple became director and the observatory moved its scientific headquarters to Cambridge, Mass. Its activities encompass the study of meteorites, the optical tracking of artificial earth satellites, the planetary and lunar sciences, exobiology, cometary science, and solar and stellar observations. In 1965 the observatory's division of radiation and organisms was accorded full Smithsonian bureau status, becoming the radiation biology laboratory.

National Zoological Park.—By acts of Congress of March 2, 1889, and April 30, 1890, the National Zoological Park was established under the institution. In a wooded area of 175 ac. in the valley of Rock Creek in Washington, D.C., the institution installed a small collection of animals that later grew to be one of the foremost collections of its kind in the United States.

Art Collections.—Upon her death in 1903, Mrs. Harriet Lane Johnston left her art collection to a national gallery of art, when such a gallery should be established. In 1906 the Supreme Court of the District of Columbia decreed that the art collection of the Smithsonian Institution was a national gallery, and turned the bequeathed collection over to it. It was later renamed the National Collection of Fine Arts. Housed in the Smithsonian's natural history building, the gallery attracted gifts of great interest and value, including the Ralph Cross Johnson collection of old masters, the William T. Evans collection of paintings of American artists, and the John Gellatly collection of paintings, glass, jewels, and antiques. The gallery also directs a Smithsonian traveling exhibition service.

In 1907 Charles Doolittle Walcott, eminent geologist and paleontologist, was elected secretary. During his administration of almost exactly 20 years the outstanding event for the Smithsonian was the receipt of the gift by Charles L. Freer of Detroit, Mich., of the Freer Gallery of Art, together with a large testamentary endowment. The Freer gift comprised more than 9,000 pieces, including works of American artists, especially James A. McNeill Whistler, Dwight W. Tryon, Abbott H. Thayer, and T. W. Dewing, and of Japanese and Chinese masters, including precious screens, ceramics, and bronzes.

A major event of the 1930s was the establishment of the National Gallery of Art, the gift of Andrew W. Mellon (*q.v.*) to the American people, as a bureau of the Smithsonian.

Further Developments.—In 1928 Charles Greeley Abbot, astrophysicist and assistant secretary of the Smithsonian from 1918, was elected secretary. He continued the researches begun by Langley on the variation of the sun's radiation, improving instruments and methods and establishing three solar observatories in widely separated regions of the earth.

Abbot was succeeded in 1945 by Alexander Wetmore, eminent zoologist, whose administration was marked by the addition of two bureaus to the Smithsonian organization, both in 1946—the National Air Museum, created by Congress to collect, preserve, and display aeronautical equipment of historical interest and significance; and the Canal Zone Biological Area, a 3,000-ac. jungle-covered reserve on Barro Colorado Island, C.Z., where a wide variety of biological studies are conducted.

Leonard Carmichael, psychologist, became secretary in 1953. Under his leadership two Smithsonian programs were notably advanced—an extensive project of renovating and modernizing museum exhibits; and a long-term program to provide new buildings for the growing institution.

During this period also, two additional Smithsonian bureaus were authorized by Congress: the National Cultural Center (1958), later renamed the John F. Kennedy Center for the Performing Arts, which, like the National Gallery of Art, operates under a separate board of trustees; and the National Portrait Gallery (1962).

In 1964 S. Dillon Ripley, ornithologist, became the eighth secretary of the Smithsonian.

The annual total number of visitors to the Smithsonian, including all its buildings, approximates 20,000,000.

The Smithsonian Park occupies an area in Washington, D.C., equivalent to about 12 city blocks. The oldest building, that of

the institution proper, was erected in 1847–55. It is of Seneca brownstone in a mingled Gothic and Romanesque style, designed by James Renwick (*q.v.*), and is located in the southwestern part of the grounds. Southeast of it is the arts and industries building, erected in 1881; on the north side of the park is the natural history building (1911). Southwest of the original Smithsonian building is the Freer Gallery of Art (1921). The monumental National Gallery of Art building (1941) faces on the Mall and Constitution Avenue, between Fourth and Seventh Streets, NW.

Early in 1964 a monumental new building for the Museum of History and Technology, erected directly west of the natural history building, was opened to the public.

In 1965 the Institution celebrated the 200th anniversary of the birth of James Smithson, the founder, with appropriate programs and ceremonies.

Publications.—The institution conducts an extensive publication program and has issued and distributed throughout the world more than 10,000 scientific books and monographs, in the following series: *Annual Reports* (1846 *et seq.*), in which the *Reports* of the National Museum were included until 1884, after which the museum *Reports* appeared as separate volumes; *Smithsonian Contributions to Knowledge* (1848–1916); *Smithsonian Miscellaneous Collections* (1862 *et seq.*); *Proceedings of the United States National Museum* (1878 *et seq.*); *Bulletins of the United States National Museum* (1875 *et seq.*), containing larger monographs than those printed in the *Proceedings*, and including *Contributions From the United States National Herbarium* (1890) and *Contributions from the Museum of History and Technology* (1959); *Annual Reports of the Bureau of American Ethnology* (1880 *et seq.*); *Bulletins of the Bureau of American Ethnology* (1886 *et seq.*); *Publications of the Institute of Social Anthropology* (1944–1953); *Annals of the Astrophysical Observatory* (1900 *et seq.*); *Smithsonian Contributions to Astrophysics* (1956 *et seq.*); *Catalogues of the National Collection of Fine Arts* (1922 *et seq.*); *Smithsonian Annals of Flight* (1964 *et seq.*); *Oriental Studies of the Freer Gallery of Art* (1933 *et seq.*); *Freer Gallery of Art Occasional Papers* (1951 *et seq.*); *Ars Orientalis* (1954 *et seq.*), sponsored jointly by the Smithsonian Institution and the University of Michigan; and many special publications.

Bibliography.—*The Smithsonian Institution, 1846–1896: the History of Its First Half-Century* (1897); William J. Rhees, *James Smithson and His Bequest*, in the "Smithsonian Miscellaneous Collections Series," vol. xxi (1879); William J. Rhees (comp. and ed.), *The Smithsonian Institution: Documents Relative to Its Origin and History, 1835–1899*, 2 vol. (1901); W. P. True, *The Smithsonian Institution*, vol. i in the "Smithsonian Scientific Series" (1929), *The First Hundred Years of the Smithsonian Institution, 1846–1946* (1946); "Smithsonian Institution Centennial Issue," *Science*, vol. 63, no. 5 (Aug. 9, 1946); Paul H. Oehser, *Sons of Science: the Story of the Smithsonian Institution and Its Leaders* (1949); Bessie Z. Jones, *Lighthouse of the Skies* (a history of the Smithsonian Astrophysical Observatory) (1965); Walter Karp, *The Smithsonian Institution* (1965).

(C. G. A.; L. Car.; P. H. O.)

SMITHSONITE, a mineral consisting of zinc carbonate, an ore of zinc which was formerly, until sphalerite (*q.v.*) began to be used in the 1880s, the principal source of all zinc produced (*see* also Zinc). In England smithsonite is known as calamine (*q.v.*). The formula of smithsonite is $ZnCO_3$. It crystallizes in the hexagonal system and is a member of the calcite group. Crystals are rare; it occurs usually in crusts or botryoidal masses. A porous variety is called dry-bone ore. The hardness is 4 to 4.5, and the specific gravity is 4.43 for the pure material, but varies with isomorphous replacement of the zinc by iron, manganese, calcium, and magnesium. The colour of the pure mineral is white, but it is more often brownish and sometimes green or blue; a bright yellow variety containing some cadmium sulfide has been found in Arkansas and is known as "turkey-fat" ore. It is a secondary mineral, formed by the action of ground water containing carbon dioxide on primary zinc minerals. It is frequently associated with hemimorphite, sphalerite, and galena.

Smithsonite occurs in Silesia and the Rhineland (Germany); Sardinia (Italy); Laurium, Greece (a translucent blue-green variety); Santander, Spain; Missouri, New Mexico, Utah, Colorado, Montana, and other localities in the United States; and in Algeria, Tunisia, Zambia (Northern Rhodesia), and Tsumeb, South West Africa. The mineral was named after James Smithson, the founder of the Smithsonian Institution. (L. S. Rl.)

SMOKE: IN WARFARE. While smoke has been one of the symbols and signs of warfare from earliest times, it was usually fire—rather than its by-product, smoke—that was dreaded by ancient armies in the field. Among primitive warriors, the American Indians had learned to employ smoke as an adjunct to military activity. They developed a visual signaling system of ascending smoke puffs that could be seen and interpreted by friendly tribesmen at considerable distances. Up until World War I, however, the use of smoke for tactical purposes was confined to special situations and was produced by improvised methods. It was during World War I that smoke and other chemical munitions had their inception on a standardized basis on the battlefields of Belgium and France. Later the employment of smoke munitions in warfare became recognized doctrine in all armies. In military operations smoke is used to advantage by ground, air and naval forces either on the offensive or on the defensive. Its main purposes are screening and signaling.

Screening Smokes.—For tactical purposes, smoke clouds are formed by firing smoke shells, dropping smoke bombs, burning smoke pots or operating mechanical smoke generators. These clouds are made up of masses of tiny particles that remain suspended in the air and hence produce an obscuring curtain. The behaviour of the smoke is dependent on wind and other natural factors, but the mechanical agencies that are used to create the clouds give them certain initial characteristics. Smoke created by a bursting shell, for instance, is released instantaneously and quickly forms into an expanding, rising puff. Smoke pouring from a continuing source, such as smoke pots, generators and burning-type shells, on the other hand, rises in a long, unbroken stream and remains effective for a much longer time. In deciding on the use of smoke under a given tactical situation, the commander must be thoroughly familiar with the prevailing natural conditions of the area in which the target is located and know the possibilities and limitations of the smoke munitions immediately available to him. Otherwise, the employment of smoke might revert to the advantage of the hostile forces.

While the direction and the velocity of prevailing winds exert the major effect on released smoke, humidity and the nature of the terrain also have a bearing on its behaviour. As might be expected, the obscuring power and effective duration of smoke is enhanced under conditions of high humidity. Ordinary ranges of temperature encountered in the field, however, have no perceptible effect. Level and unbroken terrain is reflected in a more evenly spread obscuring cloud above it, whereas smoke over hilly and broken terrain is less predictable because of the prevalence of inconsistent and erratic wind directions and speeds in such areas. In wooded regions smoke clouds persist longer and tend to cling to the treetops. Ordinarily smoke will follow ravines and defiles that crease the terrain in target sectors. Chemical shells produce smoke of varying properties that help overcome, or at least minimize, many of the screening handicaps inherent in the prevailing natural conditions of the battleground. Whenever smoke is used there is need for adequate observation and close co-ordination so that gaps and bare spaces that occur in the curtain can be readily repaired by directing artillery to fire smoke-producing munitions to the proper sectors of the target.

Individual soldiers and military units of all sizes have some means at their disposal for producing smoke. Smoke-producing munitions include the hand grenade and rifle grenade designed for the use of the individual soldier. Shells are made for firing from mortars, howitzers and guns. In the U.S. army, they range in size from those fired by 4.2 in. mortars up to the big 155 mm. guns and howitzers (and even larger calibres) used by the ground forces. Airplanes use either smoke generators or bombs when assigned to screening missions. Generator-produced smoke establishes long, low-hanging, persistent curtains and can be used over water as well as land areas. Naval vessels are equipped with smoke generators to provide smoke curtains for self-concealment.

The most widely used chemical agent in smoke-producing artil-

lery shells is white phosphorus. This chemical is inert when sealed in airtight jackets but burns and produces billows of dense white smoke when the shell bursts and the contents are exposed to the atmosphere. Special petroleum oils, heated and mixed with steam, are the smoke-producing agents used in mechanical smoke generators.

In tactical situations, smoke has many purposes that result in reducing casualties, saving matériel and speeding up operations in favour of the using forces. Its most important functions are to (1) blind hostile observation; (2) reduce effectiveness of enemy fire power; (3) hamper and confuse enemy operations; and (4) deceive the foe regarding the user's tactical preparations. Field tests and combat experience prove that the fire from troops enveloped by smoke is tremendously less effective than that from troops aiming into it. As a consequence, an attacking force can develop an advantage by firing smoke shells directly on hostile defensive positions both before and during an assault. This will not only minimize enemy fire power but will cause confusion among his troops. At the same time smoke conceals from enemy observation the size, deployment pattern, armament and movement of the advancing forces so that they achieve a certain degree of surprise in the attack. This is particularly important in assaults on strong points such as caves, dugouts, bunkers, pillboxes and machine gun emplacements where success is dependent upon getting close enough to the target so that flame throwers or hand grenades can be used effectively. Before the introduction of modern smoke munitions, attacking forces were obliged to rely on the natural concealment of night and early dawn to create similar advantages.

In the rear areas as well as in the combat zones smoke screens are used to conceal vital targets from observation and air attack. Troop concentrations, supply dumps, ports, communication centres, towns and industrial areas can be given this screening security in both daylight and darkness. If the strategic situation indicates the need, smoke curtains can be maintained continuously in friendly locations over periods of weeks or even months by the use of mechanical generators. Smoke screens are sometimes used for deceptive purposes. Spread over strategically valueless areas, they divert enemy attention from real targets and tend to draw wasted fire or useless bombing.

Signal Smokes.—Smoke munitions used for signaling or marking purposes contain organic dyes mixed with the basic burning chemicals which produce coloured smoke clouds. These clouds are visible for miles in good weather. Smoke-signal munitions come in a variety of colours, each having its own significance. By using different colours or combinations of colours, almost any type of information concerning objectives can be conveyed to friendly forces deployed over wide areas. Coloured smokes are most frequently used to mark targets, to signal the need of specialized attack on specific objectives, to announce the time for attack and to call for artillery fire.

Artillerymen often use signal-smoke shells for registering fire on targets. Smoke signals are also used to indicate rallying points and assist members of units in getting oriented in difficult terrain during the heat of battle. In joint air and ground operations, troops use coloured smoke signals to identify themselves and their positions to friendly aircraft. For the use of the individual soldier, coloured smoke chemicals are contained in hand and rifle grenades. They are also used in various sizes of artillery shells as well as in smoke pots.

(M. B. H.)

See also CHEMICAL WARFARE.

SMOKELESS POWDER: *see* EXPLOSIVES.

SMOKE PRECIPITATION: *see* PRECIPITATION, ELECTROSTATIC.

SMOLENSK (SMOLENSKAYA OBLAST'), an *oblast* of the Russian Soviet Federated Socialist Republic, U.S.S.R., was formed in 1937. Area 19,266 sq.mi. (49,900 sq.km.). Pop. (1959) 1,142,969. The greater part of the *oblast* lies in the basin of the upper Dnieper (Dnepr), which rises in the north. The eastern part drains to the Volga, the northeast directly, the east by the Ugra and Oka. The northwestern part drains to the Zapadnaya (Western) Dvina, which crosses the extreme northwest tip of the *oblast*. This close conjunction of three major river basins, flowing to three different seas, and the easy portages between them have made the area a focus of routes since earliest times. The main relief feature is the Smolensk-Moscow Upland, a ridge of terminal moraines, which crosses the *oblast* from east to west, just south of the Dnieper. It rises to 1,050 ft. (320 m.) northwest of Vyaz'ma. Between the confused morainic hills and ridges are broad valleys and depressions, often occupied by peat bog or grass marsh. The climate is continental, but much modified by Atlantic influences brought in by depressions from the west. The January average temperature is −7.7° C (18° F) and the July average 17.2° (63°). Rainfall is 22–26 in. (560–660 mm.) a year, with a summer maximum. The whole *oblast* lies in the zone of mixed forest, with a natural vegetation of pine, spruce, oak, and birch. The area has long been settled by Russians and the forest, once covering the whole surface apart from the swamps and natural meadows along the rivers, has been greatly reduced and now occupies about 18% of the *oblast*. Timber cutting is now restricted. Soils are mostly sod-podsols, weakly or moderately podsolized. Where sandy glacial soils occur, they are usually in pine forest.

Of the 1959 population, the low proportion of 32% was urban. Apart from Smolensk (*q.v.*; 147,196), the 14 towns and 7 urban settlements are small. The most significant are the rail junctions of Vyaz'ma in the east and Roslavl' in the south. In these three towns most of the industry, especially engineering, is concentrated. Many of the industries of the *oblast* are long-established, and have grown from small-scale enterprises: linen making, timberworking, glassmaking (mostly in the south), and food processing, especially cheese and butter making. Some lignite is mined at Safonovo. Much peat is dug for fuel. Agriculture is dominated by flax growing, for which Smolensk is the second most important *oblast* of the R.S.F.S.R. Rye and oats are the chief grains, although some wheat and maize (corn) are also grown. Potatoes are important and used for human and animal consumption, distilling, and starch making. Dairy farming is highly developed, helped by the abundant natural pasture, although sown grasses occupy over a quarter of the arable land.

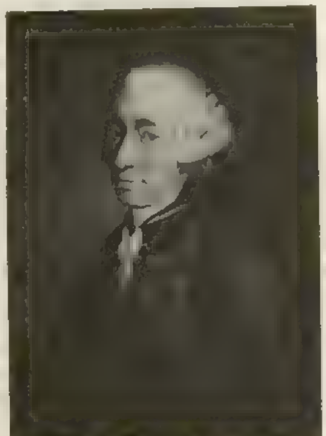
(R. A. F.)

SMOLENSK, a town and *oblast* centre of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on both banks of the Dnieper (Dnepr), 260 mi. (418 km.) WSW of Moscow. Pop. (1959) 147,196. Smolensk is one of the oldest and most historic of Russian towns, dating back to the 9th century, but the ravages of war (particularly World War II) have left little of the many ancient churches and other buildings. Among those surviving and now restored are the 12th-century churches of SS. Peter and Paul, the Svirskaya (1191–94), and St. John the Divine (1173–76). Modern Smolensk is a key junction on the main Moscow-Warsaw railway and highway, on the navigable Dnieper, and on other railways to Vitebsk and the northwest, to Bryansk and the south, and to Sukhinichi and the southeast. It has engineering industries, concerned with vehicle and tractor repair equipment and textile machinery, and also linen making. Other industries include timber working; the manufacture of glass, bricks, clothing, knitwear, and shoes; flour milling; and other food industries. There are medical, veterinary, pedagogic, and physical training institutes.

The first documentary mention of Smolensk is in 882, but it has been suggested that the adjacent burial ground of Gnezdovo, with nearly 4,000 tumuli, may be even earlier. Certainly Smolensk at a very early date was one of the leading Russian towns, deriving its importance from its control of the key portages between the Dnieper and the Zapadnaya (Western) Dvina. The town was thus the main centre of the middle section of the Baltic-Byzantium "Water Road." At first under Kiev, in 1054 the town was given to the fifth son of Yaroslav the Wise and later chose its princes by the *veche*, or town assembly. Trade with the west increased, especially in the 12th and 13th centuries, and in 1229 a trade treaty was signed with the Germans, which refers to a "Latin" church in the town. As usual in old Russian towns, there was a citadel, or kremlin, and a commercial town, both originally on the left bank. The Tatar invasion of 1238–40 sacked the town, and it subsequently passed to the Grand Duchy of Lithuania. For over four centuries Smolensk was the key in a bitter struggle between

Muscovy and the west. In 1340 it fell after a siege to the Muscovites; in 1408 another siege led to its capture by the Lithuanians. The year 1514 saw it recaptured by Moscow, and in 1596–1602 its kremlin was rebuilt by the architect Fedor Kon; but in 1610, in the "time of troubles," it fell to the Poles, after a 21-month siege. Only in 1654 was Smolensk finally captured by Russia (and formally ceded in 1686), but its position on the main route from Moscow to Warsaw and western Europe gave the town no peace. One of the major engagements of Napoleon's advance to Moscow in 1812 took place at Smolensk on Aug. 17–18 and the town was burned. During the retreat Napoleon stayed there for four days in a vain attempt to rally his army. From July to September 1941 one of the bitterest battles of World War II was fought in and around Smolensk, preventing the Germans from occupying Moscow. In October 1943 another great battle forced the Germans to abandon the town. (R. A. F.)

SMOLLETT, TOBIAS GEORGE (1721–1771), Scottish novelist who combines realistic description of the life of his age with satire on the folly, selfishness, and cruelty of mankind, was born in March 1721, near Cardross in Dunbarton. His father, who died soon after Tobias' birth, belonged to a distinguished family notable for its pride, its resourcefulness, and its independence. The Smolletts were Whigs in politics and Presbyterians in religion, and had a strong traditional connection with the law. Smollett was sent to Dumbarton Grammar School, memories of which no doubt lie behind some of the graphic descriptions of schooldays in *Roderick Random*—perhaps the ferocity of the punishments, for example, and the violence of the practical jokes. He was then apprenticed to two well-known Glasgow surgeons and attended classes at the university, before setting out for London at the age of 18, burning with desire to see his tragedy, *The Regicide* (publ. 1749), produced in the metropolis—a desire destined never to be fulfilled.



PORTRAIT BY AN UNKNOWN ITALIAN ARTIST; PHOTO, BY COURTESY OF THE NATIONAL PORTRAIT GALLERY

TOBIAS SMOLLETT, ABOUT 1770

It is unfortunate that little is known about the next few years, since Smollett anticipated later "realistic" writers in often using his own experiences in his books with relatively little modification; but it is clear that he gained experience of life in London, in the navy, and in the West Indies. The description of the tragic expedition to Carthagena in *Roderick Random* is comparable with Goya's etchings of scenes of warfare. In 1742 or 1743 he married a Jamaican heiress, Anne Lassells, who was to outlive him by 20 years; and soon after he took up his residence in London.

It is significant that Smollett's first considerable publications, apart from journalism and the fine lyric "The Tears of Scotland" (written after the defeat of Charles Edward at Culloden on April 16, 1746), should have been *Advice* (1746) and *Reproof* (printed with *Advice*, 1748), two Juvenalian satires; for there is a marked element of satire in all Smollett's work—not the satire of an unfeeling man (in spite of its harshness), but rather that of a sensitive observer whose nerves are shocked by the brutality and selfishness of human life. That Smollett should then have turned to the novel as a form of expression is not surprising: this was the decade following Defoe's death, the decade that saw the first work of Samuel Richardson and Henry Fielding. There was a strong tide running in favour of prose fiction, a tide which carried *The Adventures of Roderick Random* (1748) to immediate success. The preface makes it clear that Smollett's conception of the novel was much less lucid and classical than that of Fielding: the three main terms which Smollett uses—"satire," "romance," and "comedy"—are not clearly related to each other. What the book gives us is simply a narrative in the first person describing a number of vivid and characteristic scenes of contemporary life, tempered by some severe satire on the folly and cruelty of mankind. Except in *The Expedition of Humphry Clinker* (1771) there is no real unity in Smollett's books: he uses as his protagonist the *picaro* of Spanish tradition, a young man as resilient as a rubber ball who is unencumbered with principles or ideals. The value of *Roderick Random* lies in the powerful truth to life of individual scenes, particularly those describing the violent and squalid life of the 18th-century navy.

Smollett translated Le Sage's *Gil Blas* in 1748 and made a journey to France in 1750, probably to collect material for *The Adventures of Peregrine Pickle* (1751), which is written in the third person but which resembles *Roderick Random* in the power and humour of its individual scenes as it does in its lack of general structure. The description, in the first volume, of Peregrine's colourless father and his remarkable friend Commodore Trunnion, with his faithful Lieutenant Hatchway, is an incomparable piece of humorous writing which gave a hint to Sterne for Uncle Toby and his "garrison" in *Tristram Shandy*. The protagonist, for all the "modest merit" attributed to him, is an even greater scoundrel than Roderick Random: he deserves a long spell in prison rather than the heiress with whom he is rewarded at the end of the book.

In his remaining 20 years Smollett wrote a great deal, much of it hackwork of little interest today. His writings include *A Complete History of England, Deduced from the Descent of Julius Caesar to the Treaty of Aix La Chapelle, 1748* (four volumes, 1757–58), the *Continuation of the Complete History of England* (five volumes, 1760–65), and numerous translations, as well as the *Travels Through France and Italy* (1766) which annoyed Sterne and helped to produce *A Sentimental Journey*. He wrote an immense amount of journalism and edited *The Critical Review* from 1756 to *c.* 1762. Three minor works of fiction deserve mention: *The Adventures of Ferdinand Count Fathom* (1753), *The Adventures of Sir Launcelot Greaves* (serialized 1760–61; 1762), and *The History and Adventures of an Atom* (1769). Derivative as these books are, here and there in them Smollett anticipates later developments in popular prose fiction. But *Humphry Clinker* is his masterpiece. It consists of letters written by the various members of a family who are making a tour of Britain: an idea that Smollett probably took from Christopher Anstey's *New Bath Guide* (1766—a series of poetical letters sent by an imaginary family to friends in the country). In the early chapters (particularly) of *Humphry Clinker* humorous effect is gained from the contrasting impressions made by Bath on a romantic young girl and a testy valetudinarian; but later Smollett concentrates on describing the places which his travelers visit and on the opportunities that arise for satirical comment on human life. The romantic plot is no more than a thread on which to hang the "humours" of old Matthew Bramble, who is well described by his nephew when he says that he "affects misanthropy, in order to conceal the sensibility of a heart, which is tender, even to a degree of weakness." Through the *persona* of Bramble (a Welshman) Smollett is able to pay tribute to his native Scotland, from which he had been exiled for most of his life but which he had revisited as an ailing man in 1766.

In December 1769 he went to Italy in search of health, and in 1770 settled near Leghorn, where he died on Sept. 17, 1771—the same year that saw the birth of a greater Scottish novelist, Sir Walter Scott, who was destined to be both the admirer and the biographer of Smollett.

BIBLIOGRAPHY.—*Works:* No edition exists which includes reliable texts of all his writings, and there is no collection of most of his journalism. Of "collected" editions, *see* those by R. Anderson, 6 vol. (1796; with enlarged *Memoir*, 1800); J. Moore, 8 vol. (1797); W. E. Henley and T. Seccombe, 12 vol. (1899–1901); and G. M. Maynadier, 12 vol. (1902). There is an edition of the novels by G. Saintsbury, 12 vol. (1895); and an Oxford edition, 11 vol. (1925–26). There is an unexpurgated edition of *Peregrine Pickle*, by J. L. Clifford (1964). *Humphry Clinker* was ed. by L. M. Knapp (1966). *Travels Through France and Italy* was ed. by T. Seccombe (1907). The *Plays and Poems* were collected in 1777. The *Letters* were ed. by E. S. Noyes (1926).

Biography and Criticism: Sir Walter Scott, in *Lives of the Novelists*, 4 vol. (1821–24; ed. by G. Saintsbury, Everyman's Series, 1910); H. S. Buck, *A Study in Smollett, chiefly "Peregrine Pickle"* (1925); G. M. Kahrl, *Tobias Smollett, Traveler-Novelist* (1945); L. M. Knapp, *Tobias Smollett, Doctor of Men and Manners* (1949); R. Alter, *Rogue's Progress: Studies in the Picaresque Novel* (1964). (I. J.)

SMUGGLING, the conveyance of things by stealth, is particularly applied to the clandestine movement of goods to evade customs duties or import or export restrictions, but may include the secret movement of persons.

Smuggling flourishes wherever there are high revenue duties (*e.g.*, on tea, spirits, and silks in 18th-century England, coffee in many European countries, and tobacco almost everywhere) or prohibitions on importation (narcotics) or on exportation (arms and currency). The most convenient commodities to smuggle are high-value goods of small bulk such as watches or drugs. Some geographical conditions favour smuggling, particularly a long coastline or remote land frontiers, or the proximity of a fiscally favoured territory as a base. The Isle of Man was such a threat to England until 1765, as was Gibraltar to 19th-century Spain; Hong Kong, Macao, Goa, Andorra, and Tangier are other examples of free or low-duty enclaves which have endangered the revenue security of their neighbours.

History.—Smuggling is probably as old as the first tax or regulation on trade. England, which was the first modern nation to develop a national system of finance and a professional customs force, has been particularly prone to both commodity and currency smuggling from the end of the 13th century onward. Other European nations retained the seigneurial system much later than England, their monarchs being financed by domain fines or fees, with little reliance on customs taxation. In America, however, the customs system grew up with the Colonies and was adapted to benefit the nation after Independence.

In the late 17th century smuggling was greatly encouraged by a technical development in the rigging of cutters, which enabled vessels to sail in and out on more or less any wind, and by wars on the European continent whose cost forced up the rates of customs duties.

In the early 1700s England secured the right to send one ship a year to the Spanish-American colonies. This was the pretext for much smuggling. Actions between Spanish coastguardsmen and English traders led to the "War of Jenkins' Ear" in 1739 (*see* JENKINS, ROBERT). The 18th century became the heyday of the smuggler, and smuggling in England became so serious as to threaten the structure of finance and government. Wool continued to be smuggled to the European continent, and tea, tobacco, spices, silks, and spirits were smuggled into England in quantities exceeding the legitimate trade. English popular opinion sided with the "free traders."

In France smuggling against the tobacco monopoly and the exorbitant gabelle (*q.v.*) on salt became widespread. For the protection of the farmers-general, severe penalties were imposed which sent 3,000–4,000 persons every year to prison, the galleys, or the gibbet. During the Napoleonic Wars, English smugglers were paid by the French to break the British blockade. After Waterloo, when the European armies were discharged, smuggling grew to alarming proportions. The English army and navy were called in to maintain an expensive blockade until, after much violence, the illicit trade was suppressed, while the fiscal policy of free trade helped to prevent its immediate recurrence.

Britain could not enforce its policy of requiring its colonies to trade with the rest of the world only through the mother country: by 1744, more than 40 vessels from American colonies were trading direct at Hispaniola. Many convicted English smugglers were transported to British colonies where they found opportunities to ply their trade. Spain's trade with her colonies had been strictly regulated, but was interrupted by wars, and Mexican and Peruvian merchants began to rely upon smugglers. Jamaica became the centre of the Caribbean smuggling trade. In spite of the Molasses Act of 1733, American merchants got their supplies direct from French and Dutch possessions in the vicinity. The burden of the British revenue and mercantile laws fostered independence; from the large-scale smuggling of tea it was only a short step to the Boston Tea Party, which repudiated the whole customs system. After Independence, smugglers such as Jean Laffite (*q.v.*) were active on the Louisiana coast, but during Jefferson's Embargo and Non-Intercourse acts the centre of illicit activities shifted to the North.

Attempts by the Chinese government to stop the smuggling of opium led to the opium war of the 1840s and the ceding of Hong Kong to Britain. Despite its illegality Englishmen and Americans continued to smuggle opium from India into China in exchange for tea and silks. When the trade was legalized, the import duty was so high that smuggling continued into the 20th century. British India in the 19th century suffered smuggling of salt between the different states with varying tax rates, while smuggling of all kinds of dutiable goods occurred between Goa and India and between Gibraltar and Spain.

Throughout the 19th century a contraband traffic in African slaves persisted. In the 1830s, after most civilized nations had declared the trade illegal, as many as 135,000 Africans were being enslaved each year, and mostly carried to Brazil, the United States, and Arabia. The British navy endeavoured to suppress this traffic around African coasts. The Atlantic slave trade reached a peak in its last years before the American Civil War. "Blackbirders" ran Negroes into the South; the Underground Railroad (*q.v.*) spirited some slaves northward to freedom. In the Middle East the illegal slave trade continued into the 20th century. (*See* also SLAVERY: *International Efforts to Suppress Slavery Throughout the World.*)

In the latter half of the 19th century smuggling developed in Africa, particularly of spirits from the Portuguese colonies into the Boer states and from French colonies into the Gold Coast and Nigeria. Illegal supplies of arms and ammunition were also poured into Africa and New Zealand. Another problem of the period was the white-slave traffic, or smuggling of girls abroad for prostitution, which was checked only by international agreements, leading to control by the United Nations in the 20th century. Fortunes were also made in the traffic in obscene publications during this period, and this trade revived in the decades after World War II.

20th Century.—In the 20th century smuggling against revenue duties has been overshadowed by smuggling of prohibited goods. International syndicates trade in narcotics, arms, diamonds, bullion, and illicit currency. The 13 years of prohibition (*q.v.*) of liquor in the United States from 1920 presented such opportunities to smugglers that violations threatened to overwhelm the authorities in conditions reminiscent of 18th-century England. Fleets of craft laden with liquor left Europe and the West Indies for the Atlantic coast, while truckloads were run all along the Canadian frontier (*see* also BOOTLEGGING). Prohibition was also tried in Finland, Norway, and Sweden; spirits were prohibited in Belgium. The smuggling syndicates were so successful that in Finland, for instance, there was never any real shortage of liquor. Smuggling scandals in Canada involving members of the customs department precipitated the resignation of the prime minister, William Mackenzie King, in 1926. In the 1920s and 1930s guns were smuggled into Ireland, Spain, and the Middle East.

In the period following World War II, there was an abundance of money but great shortages of goods varying between one country and another. Currency restrictions prevented free purchase and encouraged a return to barter in Europe, coffee and cigarettes becoming a common form of currency. As after the Napoleonic Wars, there were numbers of adventurous ex-servicemen and former resistance fighters well trained in movement by stealth: fast ex-naval motorboats could be had cheaply. Perhaps the worst smuggling was into Italy by sea and land, and over the land frontiers between France and Belgium and Germany. The principal goods smuggled were tobacco (particularly into Italy where there is a state tobacco monopoly), watches into all countries from Switzerland, nylon stockings, and other portable manufactured goods. In spite of decreasing tariffs between the Common Market countries, butter smuggling from Holland to Belgium was still a problem in the 1960s, and cigarette smuggling continued into Italy, while Denmark was troubled by the illegal importation of cigarettes and cigarette papers. Shortages of consumer goods still occurred in the 1960s in some European countries through quota restrictions, and there were always smugglers ready to make good the deficiency.

Stringent currency restrictions imposed in the era after World

War II provided huge profits for international smuggling gangs. There was a great West-East movement of gold, as peasants of Asia would buy gold for hoarding at well above world prices. Licensed users of gold in the United States and other western countries evaded official restrictions and provided the source of the bullion smuggled into India and Pakistan. Currency was smuggled out of India to pay for the traffic. Valuable paintings (permission to export which might have been withheld) were smuggled from Italy and also from Britain, usually for sale in the United States.

Even if tariffs, quotas, and currency controls were abolished there would still remain some prohibitions on the free movement of goods. Many western countries prohibit the export of potential war material to unapproved destinations, but gunrunning to unstable countries continued to occur in the 1950s and 1960s, and some strategic material was diverted to the Soviet bloc or Communist China. Most countries restrict the import of narcotic drugs, but this enormously profitable traffic is most difficult to eradicate. In the United States, gangs which had flourished during Prohibition turned to dope smuggling, and in spite of international cooperation and some brilliant successes by the federal authorities the problem remained serious. In December 1964 New York customs officials seized cocaine worth over $1,000,000 coming from South America, the biggest single quantity of drugs till then intercepted in the U.S. since before World War II. The traffic existed also in Europe, where Italy had been shocked by its extent. (*See* also NARCOTICS, LAWS RELATING TO.)

Stolen diamonds were being smuggled out of the diamond fields of Africa to evade the monopoly of the Diamond Selection Trust and continued through illicit channels to Europe, America, and Russia. The mining Diamond Corporation created the Independent Diamond Security Organisation to combat the leakage.

Methods of smuggling change little; all are variants on two themes: the undetected "running" of cargoes across frontiers, or concealment in unlikely places on ships or cars or on the person or in baggage or cargo.

Countermeasures.—Countermeasures have a similar pattern in most countries. Current basic acts are: in the U.S., the Tariff Act of 1930, and the Anti-Smuggling Act of 1935; and in Britain, the Customs and Excise Act, 1952. Usually the law prohibits the import or export of goods except at approved ports or frontier crossings, where permanent customs staffs can control traffic. This must be backed by efficient patrol systems using fast craft and modern equipment such as radio or, in some countries, helicopters. There are laws against hovering off the coast or signaling to smugglers. Vessels must stop when challenged and may be fired upon if they fail to do so. The captain of a ship or aircraft must present a cargo list without delay. All this must be reinforced by legal powers of search of ships, vehicles, persons, and premises, powers to arrest offenders and to seize contraband goods and vehicles or craft carrying them. Severe legal penalties have been set as deterrents. In English-speaking countries, persons suspected of violating customs laws have fewer rights than common-law offenders. Passengers and crews must submit to questioning and may be searched without special warrant. A claimant seeking to recover seized goods carries the burden of proof that duty has been paid. (*See* also COAST GUARD: *United States*.)

Because successful smuggling is undetected smuggling, no firm estimate is possible of the extent of evasion. The profits, however, are such as to tempt men into smuggling. An increase in the number of seizures may point to an increase in smuggling, but it may be due to improved methods and efficiency of the administration. Any comparison of the number of convictions secured by different countries must take account of the extent of legitimate trade. For instance, U.S. Treasury Reports for the decade after World War II showed a yearly average of about 400 convictions for revenue offenses, but at the same time more than 22,000,000 ships, aircraft, and other vehicles and almost 90,000,000 persons entered the United States in an average year. The figure of convictions is but a tiny fraction of the traffic. A European table of major customs convictions showed Britain at the head with 93 in the years 1955–64, followed by Austria with 85. Whereas the British figures are fairly evenly spread over the ten years, the major convictions for some countries (including Spain, Portugal, and France) all occurred in a single year.

BIBLIOGRAPHY.—The best sources are the treasury (or appropriate ministry) reports of the countries concerned, together with the *Bulletins* of the Customs Co-operation Council, Brussels.

History: The best general survey is N. Williams, *Contraband Cargoes: seven Centuries of Smuggling* (1959); H. Atton and H. H. Holland, *The King's Customs*, 2 vol. (1908–10); G. D. Ramsay, "The Smugglers' Trade: a Neglected Aspect of English Commercial Development," *Transactions of the Royal Historical Society* (1951); F. D. Arnold-Forster, *At War with the Smugglers: the Career of Dr. Arnold's Father* (1936); E. K. Chatterton, *King's Cutters and Smugglers, 1700–1855* (1912); T. C. Barker, "Smuggling in the XVII Century: the Evidence of the Scottish Tobacco Trade," *Virginia Magazine of History and Biography*, vol. 62, no. 4 (Oct. 1954); F. Funck-Brentano, *Mandrin, Capitaine Général des Contrebandiers de France* (1908), L. Saxon, *Lafitte the Pirate* (1930); Lord Teignmouth and C. G. Harper, *The Smugglers: picturesque Chapters in the History of Contraband*, 2 vol. (1924).

Historical Background: Sir Reginald Coupland, *The Exploitation of East Africa, 1856–1890: the Slave Trade and the Scramble* (1939); C. H. Haring, *Trade and Navigation Between Spain and the Indies in the Time of the Hapsburgs* (1918); L. A. Harper, *The English Navigation Laws* (1939); E. Hughes, *Studies in Administration and Finance, 1558–1825* (1934); W. Mohrhenn, *Helgoland zur Zeit der Kontinentalsperre* (1928).

Special Aspects: H. J. Anslinger and W. F. Tompkins, *The Traffic in Narcotics* (1953); H. S. Dennison and A. E. Sawyer, *Federal Prohibition Enforcement*, rev. ed. (1930); United Nations, Economic and Social Council, *Study on Traffic in Persons and Prostitution* (1959), United Nations, Secretary-General, Economic and Social Council, *The Suppression of Slavery*, XIV, 2 (1951).

Modern Accounts: I. Fleming, *The Diamond Smugglers* (1957); B. Herrington, *Smugglers Ahoy!* (1957); W. J. Makin, *Smugglers of Today* (1937), A. Moray, *Diary of a Rum Runner* (1929); G. Pinnock, *Dark Paths: the Story of Modern Contraband Running in Europe* (1938); R. Ellis, *Coast Guard Law Enforcement* (1943); D. E. Walker, *Modern Smugglers* (1960).
(ME. N.)

SMUT AND BUNT, parasitic fungus diseases of various flowering plants, especially cereals and other grasses. Smuts and bunts compose the order Ustilaginales of the class Basidiomycetes. The name smut refers to the black or brownish powderlike masses of fungus spores that commonly appear in the flowers or heads, but in some cases on the stems, leaves or even roots, of host plants. The smut fungi usually give little if any evidence of their presence in the host prior to spore formation. In some cases the fungus mycelium causes hyperplasia (increase in cell division) or hypertrophy (abnormal enlargement of existing cells) of the host, resulting in galls or tumours as exemplified in corn smut. The smuts in general are best typified by the genus *Ustilago*, species of which attack oats, barley, wheat, corn, etc. The loose and covered smuts (in the former the membrane enclosing the infected grain ruptures; in the latter it does not) which destroy the inflorescence of cereals are among the best known of the *Ustilago* types.

Bunt indicates a special group of the inflorescence smuts, caused by species of *Tilletia*. The bunts, or stinking smuts, so called because of their characteristic and unpleasant odour of decaying fish, are best exemplified by the bunt of wheat (*Tilletia caries* and *T. foetida*), a classical disease of plant pathology. The bunt fungi invade the host when the latter is in the seedling stage. Their mycelia ramify and grow harmoniously with the host plant to its maturity. At this point the presence of the fungus is detected as it sporulates in the wheat head and completely replaces the grain with powdery masses of black spores.

The principal means for controlling most smuts is by seed treatment, either with chemicals or heat. Soil treatment, the use of resistant varieties and the adoption of certain cultural practices are other important control measures. Seedling-infecting smuts usually are controlled by various chemical seed treatments. The embryo-infecting loose smuts of wheat and barley may be checked by the hot-water or water-soak treatment of seed or by the use of certified smut-free seed.

The economic importance of smut fungi becomes apparent when it is realized that cereals constitute 73% of the total food consumed by man, and that smuts are one of the most important and widely studied groups of cereal diseases. In the Pacific northwest region of the U.S., where wheat bunt has always been a serious problem, the average annual loss attributed to bunt in the five-

year period 1951–55 was $4,500,000. The loose and covered smuts of barley, wheat and oats constantly threaten these cereal crops, especially where they are grown in the more humid or irrigated regions of the world. Annual losses attributable to corn smut in the U.S. have run as high as 55,000,000 bu.

See FUNGI: *Basidiomycetes* (*Club Fungi*).

A bibliography pertaining to the smut fungi may be found in G. W. Fischer and C. S. Holton, *Biology and Control of the Smut Fungi* (1957). (E. L. KE.)

SMUTS, JAN CHRISTIAAN (CHRISTIAN) (1870–1950), South African statesman, soldier and philosopher, was prime minister of the Union of South Africa from 1919 to 1924 and again from 1938 to 1948. His greatness lay in his continuous pursuit of Anglo-Afrikaner unity in South Africa, his reconciliation of Dominion autonomy with Commonwealth cohesion, his contributions to international order and his vigorous leadership in World Wars I and II. He was born on the farm Bovenplaats, near Riebeeck West, Cape colony (Cape of Good Hope), on May 24, 1870. Both parents were Afrikaners, his father, Jacobus Abraham Smuts, being a farmer and a member of the Cape parliament. At the Riebeeck West school, where he was for a time a fellow student with his later political adversary Daniel François Malan, at the Victoria college, Stellenbosch, and at Christ's college, Cambridge, Smuts stood out as a student of great ability, with a mind that was both broad and deep. In 1894 he was top of the first class in both parts of the Cambridge law tripos: and in England he also wrote a psychological study based on the life and work of Walt Whitman, but failed to find a publisher for it. He took practically no part in the social life of his fellow students.

© KARSH, OTTAWA

JAN CHRISTIAAN SMUTS, ABOUT 1943

Early Career.—On his return to South Africa in 1895 he practised at the Cape Town bar and supplemented his income by working as a political journalist. At first he supported Cecil Rhodes, prime minister of Cape colony, whose declared aim was Anglo-Afrikaner co-operation and the development of South Africa under Cape leadership. The raid of L. S. Jameson (*q.v.;* Dec. 1895), however, destroyed his confidence in Rhodes and in 1897 he moved to Johannesburg, where a few months later he took his bride, Sybella Margaretha Krige, whom he had met while at college at Stellenbosch. In June 1898 Smuts became state attorney of the South African republic, hoping that, as other young Afrikaners came to the fore, the regime would become less corrupt and more progressive. But Joseph Chamberlain, the British colonial secretary, and Sir Alfred (afterward Viscount) Milner, the high commissioner for South Africa, were impatient to assert British supremacy throughout South Africa and to use the grievances of the Uitlanders to that end. (*See* SOUTH AFRICA, REPUBLIC OF: *History*.)

Smuts worked for a peaceful solution to the Uitlander question, persuading Pres. Paul Kruger to offer considerable concessions to Milner at the Bloemfontein conference in June 1899 and himself offering further concessions to Conyngham Greene, the British agent at Pretoria, in August. When the negotiations collapsed and war broke out in October (*see* SOUTH AFRICAN WAR), Smuts devoted all his energy to the struggle for the maintenance of the independence of the South African republic and its ally, the Orange Free State, producing a plan for the overthrow of the British empire and publishing *A Century of Wrong* (1899), a vitriolic attack upon British policy in South Africa. After the fall (June 1900) of Pretoria, the Transvaal capital, Smuts himself became a guerrilla fighter, leading a commando deep into the Cape colony to harass the British and foment rebellion. This experience demonstrated his leadership ability and won him the lifelong allegiance of those who served under him. April 1902 found him besieging O'okiep in Namaqualand, whence he was summoned to the conference of Boer representatives at Vereeniging (May). When he had heard reports on the military situation, and when Lord Kitchener, the British commander in chief, had told him that it was likely that the Liberal party would soon be in power in Great Britain and that it would give the former republics a form of self-government, he realistically concluded that it would be wise to come to terms with the British; and he was a negotiator and signatory of the treaty of Vereeniging (May 31, 1902).

Soon afterward Smuts entered into close co-operation with Louis Botha, his wartime commander in chief. They denounced the English language educational policy of Milner's crown colony government in the Transvaal and Orange River colony and the decision in 1904 to import Chinese labourers to the Witwatersrand gold mines, and in 1905 they founded a political organization *Het Volk* ("The People"). As soon as Sir Henry Campbell-Bannerman had formed a Liberal ministry (Dec. 1905) in Britain, Smuts hurried to London to urge the cabinet to give the Transvaal immediate self-government, with an electoral system favourable to the Boers. He succeeded. In March 1907 the Transvaal became a self-governing colony, with Botha prime minister and Smuts colonial secretary (later also minister of education).

Work for Political Union.—The conduct of the Liberal government confirmed Botha and Smuts in their belief that Britain was no longer an enemy but a friend. Thereafter they strove to put an end to the feud between Boer and Briton in South Africa by welding them into a new nation which, like Canada, Australia and New Zealand, was to be a willing partner in the evolving British Commonwealth. Smuts therefore framed an education act (1907) for the Transvaal which permitted Afrikaner children to be taught in the Dutch language to a much greater extent than under the crown colony regime, but was mild enough to reassure the Transvaal British. He also prepared the ground for the unification of the four self-governing British South African colonies—the Transvaal, the Cape colony, Natal and the Orange River colony. He used the breakdown of the railway and customs agreements between the colonies as the occasion for calling in Oct. 1908 a national convention to consider some form of political union. He then wrote the outline of a draft constitution for South Africa, sent copies of it to leading statesmen in the Cape colony and the Orange River colony, and amended it in consequence of their comments. Next he discussed it with members of the opposition as well as his colleagues in the Transvaal and, after further alterations, gained their approval. Smuts's modified plan was the basis of the document which was unanimously adopted by the national convention (May 1909), enacted without amendment by the British parliament, and brought into force as the constitution of the Union of South Africa (May 31, 1910).

In the first Union cabinet, as previously in the Transvaal cabinet, Botha was prime minister and Smuts his right-hand man. At the outset their South African National party (formed in 1910) had the support of nearly all the Afrikaner voters, but it soon suffered a serious setback. J. B. M. Hertzog, minister of justice, considered that Botha and Smuts were placing the interests of Britain and the empire above the interests of South Africa, and that their policy would result in the denationalization of the Afrikaner people. In 1912 Hertzog made a series of speeches criticizing their policy and Botha formed a new government without him. Two years later Hertzog founded the Afrikaner Nationalist party, with the object of preserving the identity of the Afrikaners as a nation within a dualistic white South African nation. The result was that as Hertzog weaned Afrikaner voters away from Botha's party, now named the South African party, Botha and Smuts came increasingly to depend upon the support of British South Africans.

World War I.—This trend was promoted by World War I. In Sept. 1914 the Botha government decided to act on a British request that South African troops should occupy German South West Africa. Many Afrikaners resented this decision and some of them rebelled (October). When the rebellion had been suppressed (December), Smuts took part in the conquest of South

West Africa and in 1916 he became commander in chief of the imperial forces in East Africa. He had nearly completed the conquest of Tanganyika when, in March 1917, he went to attend the imperial war conference in London. He remained there as a member of the imperial war cabinet and the British war cabinet, performing many valuable services for the British government and the allies. His speeches at that time stressed the importance of maintaining the cohesion of the Commonwealth on the basis of Dominion autonomy; and in Dec. 1918 he wrote a memorandum for the imperial war cabinet outlining the project of a League of Nations, which was a major contribution to the origin of that body. At the peace conference at Versailles, recalling his own position at Vereeniging, he opposed the imposition of severe reparations on Germany and it was with great reluctance that he signed the treaty of Versailles (June 28, 1919).

Prime Minister.—Botha died in Aug. 1919 and Smuts succeeded him as prime minister. By that time his international prestige was high. In 1921, besides attending the imperial conference in London, he persuaded some of the Irish leaders to enter into the negotiations which led to the Anglo-Irish treaty of Dec. 1921 and the establishment of the Irish Free State. But his political position in South Africa was difficult. The Nationalists gained ground in the 1920 elections and in the next year, rather than unite with them, Smuts's party joined forces with the Unionists, the party of most of the English-speaking South Africans, and gained an increased majority at the election of 1921. The effect of this amalgamation was reduced by industrial unrest on the Witwatersrand. As early as 1914 Smuts had suppressed a general strike of white mineworkers by proclaiming martial law and arresting and deporting the leaders. In 1922 another such strike developed into a large-scale rebellion; the rebels gained control of most of the Witwatersrand, and there were 700 casualties before Smuts in person restored order. He then tried to improve his political prospects by offering the voters of Southern Rhodesia favourable terms for incorporation in the Union, but when a referendum was taken in Oct. 1922 most of them preferred that Southern Rhodesia should become a self-governing colony, separate from the Union. In 1923 the Nationalists came to an agreement with the Labour party, whose main strength was among the Witwatersrand mineworkers, and in 1924, after a government defeat in a by-election, Smuts went to the country and was defeated by the Nationalists and their Labour allies. His exclusion from office enabled Smuts to complete a philosophical treatise, *Holism and Evolution*, which he had started many years earlier. Its publication in 1926 attracted the attention of philosophers throughout the world.

For some years the Nationalist government was secure, for the Afrikaners were a majority of the voters and they were assisted by an electoral system which favoured the rural areas. During the world depression which began in 1929, however, the government encountered difficulties through trying to remain on the gold standard after Britain had left it in 1931, and in 1933 Hertzog and Smuts formed a coalition government, with Hertzog prime minister and Smuts deputy prime minister. In 1934 the Nationalist and South African parties fused into the United party.

World War II.—Hertzog and Smuts co-operated successfully in internal matters, such as the enactment of segregation laws for Africans, and in completing the legal autonomy of the Union by the Status act of 1934; but they differed on how the Union should exercise its autonomy in foreign affairs. This difference came to a head when Germany invaded Poland in 1939. Hertzog considered that South Africa should remain neutral, but Smuts, sensitive to the broader implications of Nazi expansionism, disagreed. The cabinet split into two irreconcilable groups and the issue was debated in the house of assembly, which decided for Smuts by 80 votes to 67. On Sept. 5, 1939, Smuts formed a government with the support of the Labour party, the Dominion party, and a majority of the United Party, to wage war against Nazi Germany.

Notwithstanding widespread opposition to participation in the war, and a certain amount of sabotage, the South African war effort was impressive. South African troops played an important part in the Ethiopian, north African and Italian campaigns and South African industry was geared to the production of war materials. As prime minister, Smuts actively supervised the administration of the Union, and as commander in chief of the South African armed forces he paid several visits to the African and European theatres of war. He also kept in touch with other allied leaders, especially Winston Churchill, who set a high value on his judgment. In 1945 he attended the San Francisco conference, playing a major part in drafting the United Nations charter.

Postwar Years.—Smuts and his colleagues planned a progressive postwar program for South Africa, involving white immigration, industrial expansion and the improvement of agriculture; but after World War II, as after World War I, political forces in the Union were against him. In the 1943 election he had won a good majority, partly because the opposition had been splintered into several factions; but by 1948 Malan had consolidated most of the factions into his new Nationalist party, which exploited discontents caused by wartime controls and fears that white supremacy was being undermined. Continuous erosion of the tribal system by economic forces and the democratic ideology of the allies in World War II had promoted among the African people of the Union demands for the relaxation of segregationist measures. Smuts, without making any significant concessions to them, had adopted a conciliatory attitude. Malan and his party accused Smuts of endangering white supremacy and propounded the theory of *apartheid*, according to which the races were to be separated from one another and white supremacy was to be assured. The 1948 election resulted in a narrow but decisive victory for Malan and his Afrikaner party allies. Smuts was defeated in his own constituency, but a safe seat was soon found for him elsewhere. He was leader of the opposition when he died, on Sept. 11, 1950, at his home at Irene near Pretoria.

Smuts held many honours. He was a privy councilor (1917), a Companion of Honour (1917), and a member of the Order of Merit (1947); rector of the University of St. Andrews (1931–34) and chancellor of the University of Cape Town (1936) and of Cambridge university (1948); a British field marshal (1941); and he was awarded many honorary degrees and the freedom of many cities. Smuts's shortcomings were those of his background. To him the South African nation was essentially a white nation; he was never clear as to the role of the African, the Asian and the Coloured peoples in South African society. To him the Commonwealth was essentially an association of white nations: he could not fully reconcile himself to the admission of India, Pakistan and Ceylon as equals. To him the world order was to be based upon the continuing leadership of white nations.

His setbacks in South Africa arose out of his failure to carry his own Afrikaner people with him into the wider groupings of nation and Commonwealth. He might have been more successful if he had shown a keener insight into the mind of the ordinary Afrikaner farmer, and if he had been interested in Afrikaans literature. As it was, he was regarded by many Afrikaners as completely denationalized—an agent of the British South African and of British imperialism.

BIBLIOGRAPHY.—S. G. Millin, *General Smuts*, 2 vol. (1936); H. C. Armstrong, *Grey Steel, J. C. Smuts* (1937); F. S. Crafford, *Jan Smuts* (1943); B. Williams, *Botha, Smuts and South Africa* (1946); J. C. Smuts, *Jan Christian Smuts* (1952); W. K. Hancock, *Smuts*, vol. 1, *The Sanguine Years 1870–1919* (1962). (L. M. T.)

SMYRNA (ZMYRNA) was in later antiquity one of the principal cities of Asia Minor. Known in modern times by its Turkish name of IZMIR (*q.v.*), it is situated in the *il* (province) of Izmir in Turkey at the head of a deep gulf of the eastern Aegean, with the Hermus (modern Gediz) discharging on its northern shore. The little plain of Smyrna extends nine miles inland from the coast. The prehistoric and early Greek city of Smyrna lay at the northwest corner of this plain; after the time of Alexander the Great the city of Smyrna was built 3 mi. (5 km.) farther south at the southwest corner of the plain.

Old Smyrna.—The oldest settlement was on a low hillock (probably a peninsula in antiquity) by the modern suburb of Bayrakli. Anglo-Turkish excavations in 1948–51 revealed numerous strata of prehistoric habitation with remains of houses of

crude brick on stone foundations, the earliest occupation found being contemporary with that of the first city of Troy at the beginning of the Bronze Age (*c.* 3000 B.C.). In Greek tradition the pre-Greek city of Smyrna was ascribed to an Amazon foundress of the same name; to the Leleges; and to the legendary king Tantalus. It is not known whether Hittite dominion ever reached this far, but the weather-god Teshub seems to be portrayed in rock-relief in the Karabel pass 10 mi. S.E. of the plain of Smyrna. Greek settlement is first clearly attested by the presence of pottery of the style called Protogeometric, dating perhaps from *c.* 1000 B.C. The Greek city of Smyrna is said by Herodotus to have been founded by Aeolians but seized by Ionic refugees from Colophon who barred the gates when the Aeolic inhabitants were out at a festival. Excavation has shown that from modest beginnings (single-room horseshoe-shaped huts with thatched roofs) Greek Smyrna grew into a stately city in the 7th century B.C. with massive fortifications and axially planned blocks of two-storied houses. Attacked by Gyges of Lydia in the early 7th century, Smyrna is said to have repelled the invasion; its citizens won fame for their exploits against the Lydian horsemen. But later the city fell to the Lydian king Alyattes; it may well have been this emergency that evoked the martial elegiacs that the poet Mimnermus addressed to the Smyrnaeans. After the Lydian capture Smyrna ceased to exist as a city and (as the geographer Strabo records) was inhabited in village fashion. It did not join the Delian league in the 5th century. The old town was not entirely depopulated, and in the 4th century it was again densely inhabited. Smyrna is mentioned as a wheat mart in a decree of 387 B.C. and is commemorated on a silver coin minted at Colophon about the same time. Above all it is remembered as the reputed birthplace of the poet Homer. Two ruined fortifications on the mountain above the old city have been thought to be relics of a remote epoch, but may rather be the mansion and outpost of a landowner or local despot of this time.

The New City.—Strabo says that Antigonus I and Lysimachus reestablished Smyrna (before and after 301 B.C.), but local tradition made Alexander himself conceive the idea of refounding the city on a visit there in 334 B.C., and the evidence of excavation at Old Smyrna seems to favour a date in his lifetime for the removal. The new city had its citadel on the flat-topped hill called Pagus, from which the two arms of the city wall descended to enclose the city and its inner harbour on the north; this closed harbour, which was partly filled in when Timur (Tamerlane) sacked Smyrna in 1402, is now wholly dry land and occupied by the bazaars. The visible ruins of the citadel are medieval but are built on ancient foundations in the southwest part. On the upper slope of the Pagus was the stadium and a theatre of Roman date. Excavations conducted by Turkish archaeologists have uncovered much of the agora (forum) at the foot of the slope; rebuilt (with the bounty of Marcus Aurelius) after the disastrous earthquake of A.D. 178, it was surrounded by two-storied colonnades, and it contained numerous statues and an altar with reliefs of the 12 gods. It is said that an avenue, called the Golden, traversed the city from east to west, running round the lower slopes of the Pagus. Built on these slopes and at their foot, with paved streets laid out on a varied rectilinear plan, ancient Smyrna was renowned for its beauty; the one fault attributed to its architects was that they did not lay down drains in the streets, so that the roadways were awash with filth in heavy rain.

Smyrna was for a short time known as Eurydicea in honour of Lysimachus' daughter. After Lysimachus' fall (281 B.C.) the city espoused the cause of the Seleucids and received Magnesia ad Sipylum as a reward for its services in 244 B.C., but it transferred its loyalty to Pergamum soon afterward and as early as 195 B.C. set up a temple dedicated to the city of Rome. It was a favourite place of exile for condemned Roman governors and legates. After Caesar's murder Trebonius was beheaded by Dolabella in Smyrna (43 B.C.), and the city suffered some damage; and Caesar's murderers, Brutus and Cassius, met there for their council of war. Smyrna was the centre of a civil diocese in the Roman province of Asia, and vied with Ephesus and Pergamum for the title "first [city] of Asia." It was chosen by the Roman senate for the honour of erecting a temple to the emperor Tiberius in A.D. 26. It was visited by Hadrian and Marcus Aurelius, and the sophist Apollonius of Tyana resided there. Smyrna had a considerable Jewish colony and was one of the early seats of Christianity; its church endured the tribulation predicted of it in the Apocalypse (Rev. ii, 8–11) by the martyrdom of its bishop Polycarp (A.D. 156 or 168) and of Pionius (A.D. 250).

The stream Meles, beginning and ending its brief course near the walls of Smyrna, was associated in local tradition with Homer, who was reputed to have been born by its banks and to have composed his poems in a cave near its source; the stream was celebrated in local literature and worshiped for its healing powers. Its identity has long been a subject of controversy; yet there can be no doubt that it is the stream that issues at Halkapinar (the so-called "baths of Diana") and forms a fishpond at the shore on the northeast of the modern city. Homer and Mimnermus were far the greatest poets associated with Smyrna; but the pastoral poet Bion and the late epic writer Quintus Smyrnaeus also deserve mention. The city was celebrated for its library, its school of medicine in which Galen studied, and a rhetorical tradition dignified by the names of Polemon and Publius Aelius Aristides. Its specialties included prawns, lettuce, vines that produced a Pramnian wine, malachite and white lead. It possessed thermal baths (probably those at Balcova southwest of the city) named after Agamemnon, who was said to have halted there to treat his wounded men. The tutelary deity of ancient Smyrna was the Sipylene Mother.

Smyrna continued in Byzantine times as a metropolitan see, and was made the capital of the naval theme of Samos. Despite Arab attacks in the 7th century and its capture by the Turkish raider Tsacha (Tzachas) in 1090, Smyrna was still a great city when the Norse pilgrim Saevulf put in there in 1103. Its ruined citadel was rebuilt by the emperor John III Vatatzes *c.* 1225. By the treaty of Nymphaeum in 1261 Michael VIII Palaeologus ceded the flourishing port of Smyrna to the Genoese, reserving only Greek ecclesiastical autonomy. As the result of Byzantine military and naval disarmament in the east, Turkish ghazis with their hordes broke through the Seljuk state of Rum (Iconium) and established principalities in western Asia Minor. Perhaps as early as 1317 the amir Mohammed of the Aydin dynasty captured the citadel of Smyrna; and despite the prolonged resistance of Martino Zaccharia, Mohammed's son Omer became master of the city in 1329 and established his naval station there in that year.

It may in conclusion be remarked that, after Athens, Smyrna is second to none of the cities of the Mediterranean world in its continuous historical importance during the last 5,000 years.

For Smyrna since the Turkish conquest *see* IZMIR. *See* also references under "Smyrna" in the Index.

BIBLIOGRAPHY.—C. J. Cadoux, *Ancient Smyrna* (1938), which supersedes earlier works; J. M. Cook, "Old Smyrna" in *Annual of the British School at Athens*, vol. liii (1958). For the Turkish conquest *see* P. Lemerle, *l'Émirat d'Aydin* (1957). For the ancient coinage *see* British Museum, *Catalogue of Greek Coins of Ionia* (1892). (Jo. M. Co.)

SMYTH, JOHN (d. 1612), a pioneer in the demand for liberty of conscience, who "may be regarded as the father and founder of the organized Baptists of England" (A. C. Underwood); he also influenced the Nottinghamshire Separatist group of which some later emigrated to New England as the "Pilgrim Fathers." His origins are obscure. He matriculated in 1586 at Christ's College, Cambridge, where he was a fellow during 1594–98. From 1600 to 1602 he was city preacher at Lincoln, but in 1606 he renounced his episcopal ordination and became minister to a group meeting in separation from the Church of England at Gainsborough, Lincolnshire. For two years he and John Robinson (*q.v.*), of Scrooby, Nottinghamshire, "provided the necessary leadership and theological equipment" (R. A. Marchant) for organized Separatism in this district. In 1608 both men, with their followers, left England for Amsterdam, where they soon parted. Smyth adopted Baptist principles and baptized (by affusion, or pouring water over the head, not immersion) not only Thomas Helwys (who later became a Baptist pioneer in London) and other members of the congregation but, first, himself. As he received fresh light, Smyth continually altered his convictions, and unashamedly: "wee are inconstant in erroer." This inevitably caused

further division, and he died in late August 1612, excommunicated by his own people and unacceptable to the Dutch Mennonites, with whom he had sought membership, after adopting their Arminian tenets. Rejecting the dogma of original sin, he held that "the new creature . . . needeth not the outwoard scriptures." The magistrate should "leaue Christian religion free, to euery mans conscience." *See* also BAPTISTS: *History*.

BIBLIOGRAPHY.—*Works* ed. by W. T. Whitley with introduction, 2 vol. (1915). *See* also W. H. Burgess, *John Smyth* (1911); A. C. Underwood, *A History of the English Baptists* (1947); T. Lyon, *The Theory of Religious Liberty in England, 1603–39* (1937); R. A. Marchant, *The Puritans and the Church Courts in the Diocese of York, 1560–1642* (1960). (G. F. N.)

SMYTHE (SMITH), **SIR THOMAS** (1558?–1625), English entrepreneur, was pre-eminently instrumental in launching, financing and administering pioneer English joint-stock enterprises in overseas trade, colonial settlement and exploration during the first two decades of the 17th century. A member of the London Haberdashers' and Skinners' companies from 1580, he accumulated a considerable fortune from commerce. Besides holding other official posts, he was incorporator and for some years governor of the East India company, treasurer of the Virginia and Somers Islands (Bermuda) companies, governor of the Muscovy and French companies and ambassador to the tsar (1604–05). He was a prime initiator of voyages to discover a Northwest passage. Smythe's efforts and a good part of his fortune were consistently and selflessly devoted to experiments that underlay England's subsequent imperial and commercial greatness. Although there was some controversy as to his management of the Virginia company, his fame as the entrepreneur and supporter of outstanding ventures is secure. Smythe died at Sutton-at-Hone, Kent, on Sept. 4, 1625.

BIBLIOGRAPHY.—Alexander Brown (ed.), *Genesis of the United States*, vol. ii, pp. 1012–18 (1890); W. F. Craven, *Dissolution of the Virginia Company* (1932); C. M. Andrews, *The Colonial Period of American History*, vol. i (1934). (B. E. SU.)

SNAIL, a name generally applied to any member of the class Gastropoda, the largest of the six classes of the phylum Mollusca (*see* MOLLUSK). The gastropods include land, fresh-water and marine forms, among them such well-known shelled groups as whelks, periwinkles, conchs, cones, limpets, abalones, land snails and pond snails. Land gastropods lacking shells commonly are known as slugs, and shell-less marine forms as nudibranchs or sea slugs (*see* SLUG). There is, so far as is known, only a single species of shell-less fresh-water snail.

This article on snails is divided into the following sections and subsections:

I. GENERAL FEATURES

Gastropods are primarily distinguished from other mollusks by the shell, which, when present, is a single structure (except in the Juliidae, which have a bivalved shell), spirally coiled, at least in the larval stage, and typically sufficiently large so that the animal may retract completely within for protection. In many forms, however, the shell is very much modified in the adult. It may be nearly flat, cup-shaped or tubular. In some forms the mantle may completely cover the shell and in others the shell may be entirely absent. Gastropods also are distinguished by having a well-defined

CY LA TOUR

FIG. 1.—COMMON GARDEN SNAIL (HELIX ASPERSA)

head region, which possesses one or two pairs of tentacles, eyes and a mouth with jaws, liplike folds and a rasping tongue, or radula. The foot usually is developed as a creeping organ, and locomotion is accomplished by continuing waves of muscular contractions. Respiration is carried on either by means of gills, a pallial outgrowth, a pulmonary sac or, as in a few groups, the entire body surface. In all snails, however, the entire body surface may function in respiration under adverse conditions.

Gastropods are further distinguished from other mollusks by their asymmetry. The vital organs are not paired nor are they lined up on a median axis with an anterior mouth and posterior anus as is the case with the classes Monoplacophora (*q.v.*) and Amphineura (*q.v.*; chitons). In the gastropods the visceral mass has been doubled in a loop so that the anus is brought forward close to the mouth. This is not a simple flexure, however, for there is also a 180° twisting of the loop posterior to the head, which causes the organs of the left-hand side of the animal to lie on the right side, places the anus above the mouth and twists the nervous system to form a figure 8. Associated with this radical twisting known as torsion, is the loss of the organs of the normal right side of a symmetrical animal (the post-torsional left side).

II. DISTRIBUTION AND HABITAT

Gastropods are among the most widely distributed of all groups of animals and the only mollusks that have adapted to life on land. The majority, however, occur in marine, brackish and fresh-water habitats. They are found from the depths of the ocean to the tops of high mountains in the tropics, from deserts to tropical rain forests and from small springs and ditches to large lakes and rivers. Some marine snails may live most of their lives well above the strand line on rocks, trees and even the sides of houses, returning to the water only to breed. Others, such as the littorines, are found in the intertidal zone, where they are covered by water much of the time but are exposed during low tide. The majority of the littoral, or shore, species, however, are found at or just below the low-tide mark, for they cannot tolerate long periods outside the water. Such species, however, often may be found in tide pools. Many species found at considerable depths can be obtained only by dredging.

1. Marine and Brackish-Water Snails.—Water temperature, salinity and the type of substrate are the most important factors controlling the distribution of marine snails. Some species found in the intertidal zone in cold waters often can be obtained by dredging in adjacent warmer waters, the snails moving out into cooler, deeper water. Species requiring full marine conditions will not be found near the mouths of rivers pouring large amounts of fresh water into the sea. A stable brackish-water area has a unique gastropod fauna not found in other areas, but an area that fluctuates between brackish- and salt- or fresh-water conditions is usually barren. The type of bottom is an important factor controlling distribution, for some species require a sandy or muddy bottom, while others are restricted to rocky shores. A long

stretch of sandy shore may completely block the distribution of rock-loving forms that might otherwise be able to live there. Pelagic forms are widely distributed, their distribution being controlled mainly by temperature. Many species of parasitic marine snails are restricted by their feeding habits, which may limit them to a single host species.

Most families of marine snails have a wide range of distribution, but some genera and many species may be very restricted. Families such as the Cypraeidae (cowries) are found in nearly all tropical and warm temperate seas but reach their greatest numbers in the vast Indo-Pacific area. The four great oceanic regions—the eastern Atlantic, the western Atlantic, the eastern Pacific and the Indo-Pacific—have been divided into numerous subregions, each with gastropods peculiar to it, the controlling factor being temperature, usually that which is suitable for the breeding and survival of the young. Certain intertidal areas, such as the coast of the United States from Virginia to South Carolina, may have a rather poor molluscan fauna, for the range of temperature from the summer high, which may reach 85° F. (about 29° C.), to freezing winter temperatures is too great for most species. Snails with a free-swimming larval stage are generally more widely distributed and the populations more uniform than those that emerge from the egg capsule as young, crawling snails.

2. Land Snails.—In the cooler regions of the world, land snails are usually rather inconspicuous, brownish and fairly small. As a rule they live on the ground or on low shrubs or grass, feeding mainly on detritus (organic debris). They burrow into the ground during the winter for protection and generally lay their eggs there. In desert areas snails are usually light coloured and heavier shelled. Most species burrow into the ground during the dry season for protection, but some cement themselves to the trees or rocks by a limy mucus secretion. They become very active when the rains begin, feeding voraciously and breeding. In the warm temperate and tropical regions, where there is sufficient moisture, the snails will remain active throughout the year. Land snails are most abundant in areas where there is plenty of limestone; some species, in fact, may be found only on limestone rocks, where they feed on lichens. In general, insular areas are richer in land snails than are the continents, perhaps a reflection of the reduction of predators on islands. Probably the richest land snail area in the world is the island of Cuba, but the largest land snails are found in Africa (*Achatina*) and in South America (*Strophocheilus*). In general, tree snails such as *Liguus* in Florida and Cuba and *Helicostyla* in the Philippines are highly coloured, while ground-living forms are drab. The distribution of land snails on islands is often fortuitous, the snails being carried as eggs or young snails by man, birds, hurricanes or by rafting.

3. Fresh-Water Snails.—Fresh-water snails are found in nearly all permanent bodies of water, regardless of size, though they are rare and small in areas without limestone and in fast-flowing streams in the colder portions of the world. Swift streams in warmer regions may be very rich in mollusks, particularly in areas of abundant limestone. The southeastern United States is perhaps the richest area in the world for fresh-water snails. This area is old geologically and was not glaciated; thus, the snails have had time to evolve many genera and species peculiar to the rivers of this area, the Coosa river of Alabama being the most remarkable. Lakes are generally transient and seldom have species peculiar to them. There are, however, several very old lakes that support remarkable snail faunas. Most famous of these is Lake Tanganyika in Africa, which has a large number of endemic genera and species of mollusks. Because the physical characteristics of this large, deep lake approximate those of the ocean and because some of the snails resemble marine forms, Lake Tanganyika was long thought to have had a marine origin. Most of the snails, however, belong to the local fresh-water family Thiaridae (or Melaniidae). Fresh-water snails may be carried from one stream or pond to another as eggs or young snails on the feet of water birds; they may be blown with leaves or carried by floods. Most fresh-water snails cannot withstand submergence in salt water, so migration from one stream to another via the sea is virtually impossible.

III. IMPORTANCE

1. Economic Importance.—Gastropods are not as important a source of food for man as are the bivalves, but a number of species are eaten. The Roman land snails or *escargots* (*Helix pomatia* and *H. aspersa*) are considered a delicacy, particularly in France; as in Roman days, these snails often are raised in special gardens called cochlearia. Many species of marine snails are eaten in various parts of the world, particularly by the poorer classes. Among those more commonly eaten are the periwinkle (*Littorina littorea*) in Europe, the top shell (*Livona pica*) and the conch (*Strombus gigas*) in the West Indies. The flesh of conchs is used to make a delicious chowder. In California and Australia the large foot of species of abalone (*Haliotis*) is tenderized and served as a steak. Abalone is often canned and can be purchased in the markets. In some portions of Africa the large land snail *Achatina* is an important item of food for the natives. *Achatina fulica* was introduced into many of the Pacific islands by the Japanese.

Tremendous numbers of larger gastropods such as the whelks are used as bait by fishing fleets, particularly on the long lines, for the tough, muscular foot is not easily torn from the hook.

To the paleontologist and the oil geologist snails are important as index fossils, and a study of the distribution and habits of living forms aids in the interpretation of the events of the past.

Gastropods long have played an important role in the jewelry market. The beautiful, iridescent, highly coloured shells of the abalone are used in inlay work and in the manufacture of rings, pins, earrings and buttons. The Japanese collect *Trochus* shells of the Indo-Pacific by the ton for use in the button trade. *Cassis rufa* of the Red sea is the shell most commonly used in the manufacture of cameos. Here the great difference in the colour of the two shell layers is utilized, the figure being cut in the light outer layer while the dark-reddish inner layer serves as a background. Pearls occasionally found in gastropods, though they may be very beautiful, generally are irregular and not of great economic importance. Those found in the lower gastropods with iridescent shells have a lasting pearly lustre, but the beautiful pink pearls

(ABOVE) D. P. WILSON, (RIGHT) RALPH BUCHSBAUM

FIG. 2.—SNAILS OF DIFFERENT HABITATS

(Above) The eyeless purple snail (*Janthina janthina*) drifts on the warm ocean surface buoyed by a "raft" of air cells and mucus; (right) the tree snail (*Liguus fasciatus*), common to Cuba and parts of Florida, glides on a mucous film it secretes on the bark of trees

of *Strombus* eventually fade and lose much of their beauty.

Among native tribes various species of snails have been used as money, the most famous being the money cowrie (*Cypraea moneta*) of the Indo-Pacific. Shells, particularly cowries, are used for adornment by the natives of New Guinea, the Fijis and other Pacific islands. The rare and beautiful golden cowrie (*C. aurantium*) is a badge of office and may be worn only by chieftains. Snails also are found in religious ceremonies; the triton trumpet (*Charonia tritonis*) is used by the Shinto priests of Japan to call the people to worship, and in India the left-handed chank shell (*Xancus rapa*) is considered sacred. The Tyrian or royal purple dye of the Phoenicians was a product of the common *Murex* species of the Mediterranean and was a most important item in the economy of the people. The Mexican natives dye cloth with the exudate of the snail *Thais patula pansa*. (*See* also SHELLS AND SHELL COLLECTING.)

2. Ecological Importance.—The place of gastropods in nature is very important. As free-swimming larvae they are eaten by larger planktonic animals such as shrimp and small fish. Many of the bottom-feeding marine fishes and a few fresh-water fishes feed almost entirely upon mollusks (in fact, fish stomachs are a good source of shells not generally found in the intertidal zone). Many ducks and other water birds also feed on snails.

From a negative point of view snails are far more important, for many of them, especially fresh-water forms, are intermediate hosts of trematodes (flatworms or flukes), which are parasitic in man, cattle, sheep and various other mammals as well as in birds. In man the most important of these trematode diseases is schistosomiasis (*q.v.*), the blood fluke disease, which in various forms occurs in Africa, the near east, China, Japan, the Philippines, Brazil and in certain of the West Indies, where it has been introduced. Probably most snails are the intermediate hosts of some fluke, but fortunately most of them do not affect man directly. Eliminating the snail host is still the best means of controlling these diseases. In areas where the poorer classes use the streams, lakes and ponds for washing clothes, bathing and in religious rites, the education of the people concerning the disease and its spread is of some help. (*See* also FLUKE.)

Many snails are predators, feeding particularly on bivalves; they may be very destructive to the commercially important clam and oyster beds. The slipper limpet (*Crepidula fornicata*) was accidentally introduced into Europe, where it has caused considerable damage by smothering the oyster beds.

The cone shells (*Conus*) are not normally harmful to man but, if improperly handled, can inflict a fatal wound. The radula ribbon in this group is highly modified and is associated with a poison gland. Cones are sluggish snails and their dartlike teeth and poison are normally used in subduing prey.

IV. STRUCTURE AND FUNCTION

1. Basic Body Plan.—Though the anatomy of the various groups of gastropods appears superficially very different, in reality the many patterns are only variations of a basic theme. Once the stylized gastropod (fig. 3) is understood it is possible to follow the arrangement through with its increasing elaborations, reductions and changes from the generalized to the highly specialized forms.

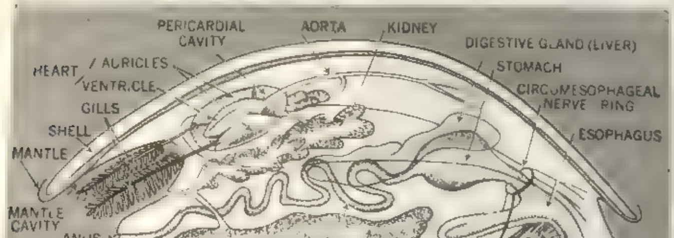


FIG. 3.—HYPOTHETICAL PRE-GASTROPOD SHOWING PROBABLE POSITION OF ORGANS BEFORE FLEXURE AND TORSION

It is impossible, of course, to know the complete anatomy of the first gastropods, for they are known only as fossils. However, it was probably similar to that of a monoplacophoran and so had a caplike shell, a large, flat foot developed for creeping or clinging to the substratum, a head region with a mouth and radula but probably lacking eyes and tentacles. The mantle was probably a simple "cape" hanging over the dorsal surface of the animal and extending just beyond the edge of the shell. The muscle attaching the animal to the shell was probably U-shaped, open in front. Internally the earliest gastropod was probably bilaterally symmetrical; however, unlike a monoplacophoran, it must have had only a single pair of each of the vital organs such as gills, kidneys (nephridia) and reproductive organs. The nervous system undoubtedly had two sets of parallel nerves—a dorsal ring serving the visceral mass and a ventral one with a ladderlike arrangement in the foot. The heart was at the posterior end of the animal; the ventricle in the centre with an auricle on each side (fig. 4).

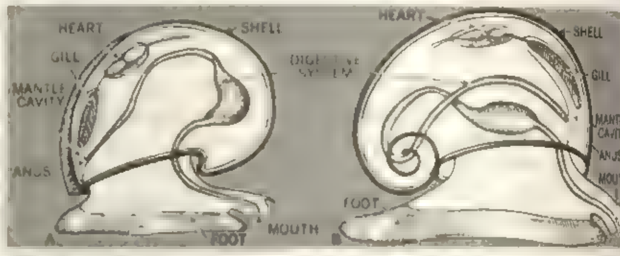


FIG. 4.—(A) POSSIBLE ARRANGEMENT OF AN EARLY GASTROPOD BEFORE TORSION; (B) PROSOBRANCH GASTROPOD SHOWING TORSION

The gills were located in a groove or cavity posterior to the heart and between the mantle and the foot. The paired kidneys and reproductive organs had common openings near the base of the gills.

The sexes were separate but could only be distinguished by the colour and products of the gonads. The coils of the long intestine were probably interwoven with the gonads.

If this hypothetical, bilaterally symmetrical animal is doubled into a loop and twisted in 180°, the result is the basic anatomy of the gastropod. Torsion takes place during the veliger stage of development, observable in the embryo (fig. 5). The earliest

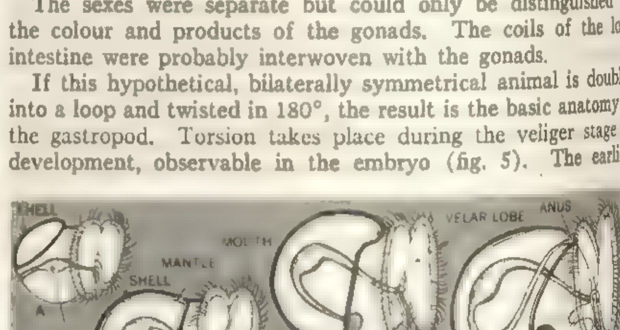


FIG. 5.—FLEXURE AND TORSION OF VELIGER: (A) BEFORE BEGINNING OF FLEXURE; (B) AT TIME OF FLEXURE AND BEGINNING OF DEVELOPMENT OF FOOT; (C) JUST BEFORE TORSION; (D) AFTER TORSION

gastropods soon gave rise to forms with coiled shells, which in the process of evolution were first bilaterally symmetrical and so planispiral (coiled in a single plane), and later asymmetrical with a protruding spire. This type of shell was very successful and has persisted; it is compact and strong and at the same time portable affording mobility and protection.

2. Shell.—The shell is carried dorsally and is usually large enough so that the animal may retreat within it completely. Like that of bivalves, the shell is typically composed of three layers. The inner layer is further laminated; in some it may be nacreous or iridescent. It is laid down by the entire surface of the mantle. The outer prismatic layer is produced at the growing edge of the shell (the lip) and is laid down by cells at the outer edge of the mantle. Prolongations of the mantle at the lip may extend out to produce spines, nodules or other sculptural features.

The outer surface of the shell in most species is covered by a protective coating of periostracum, a chitinlike material that ranges from a thin, smooth and nearly transparent layer to a thick and hairy covering. The periostracum is produced in a special fold in the outer edge of the mantle. The ability to produce

colour is invested in all three layers. The laminated layer in genera such as *Haliotis, Patella, Cypraea* and *Cassis* may produce various colours, and in *Haliotis* and *Patella*, which are nacreous, additional iridescent colour effects are produced by the structural arrangement of the component layers. In most groups colour is produced in the prismatic layer uniformly or in the form of strips or bands, either spiral or axial, in spots, zigzag lines or irregular blotches. The periostracum in most groups is some shade of yellow or brown, sometimes being nearly black. In a few forms, such as the highly coloured tree snails, other hues, particularly green, rare in mollusks, may be produced in the periostracum. Very little is known about the physiology of colour production in mollusks, and nothing is known concerning the mechanism by which the complicated colour patterns of some species are made. In addition to the shell many forms have a calcareous or horny plate, the operculum, which is produced on the dorsal surface of the foot and which closes the aperture when the animal retreats within the shell. Pulmonates—lunged snails of fresh water and land—do not have an operculum, the aperture of some species usually being closed by a limy secretion from glands in the foot. The animal is attached to the shell by means of a large, powerful muscle attached to the central pillar or columella of the shell. The various shapes as well as sculptural and colour patterns produced by snails are almost infinite; it is these characters on which species, most genera and some families are based. However, since the same general forms may appear in widely separated families, the shells cannot be used in the classification of higher categories.

3. Digestive System and Feeding.—Snails feed in a number of different ways and on a variety of food. All snails, except the adults of some parasitic forms, have a radula, which is used in procuring food. It is ribbonlike and on the outer surface armed with numerous rows of teeth (fig. 6). Worked by muscles, it functions as a rasping organ and a conveyor belt to carry the food back into the mouth and esophagus. Many groups also have horny jaws with which to cut off small particles.

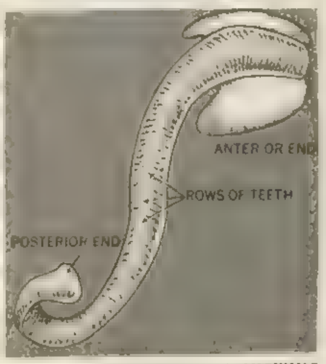


FIG. 6.—RADULA RIBBON SHOWING TRANSVERSE ROWS OF TEETH

The earliest gastropods were probably very unselective in their feeding, merely scraping in bottom detritus containing organic material. Among the least specialized snails of the subclass Prosobranchia are the archeogastropods (order Archeogastropoda), most of which are plant feeders, browsing on the algal film on rocks or feeding on larger algae. In their browsing, however, they also may scrape up some small encrusting bryozoans and sponges. Land forms (family Helicinidae) feed on algae and lichens that form a film on the trees and rocks. The order of mesogastropods includes detritus feeders, plant feeders, scavengers, predators and parasitic forms. Some are filter feeders and use the gills in a manner similar to that of bivalves. By means of cilia and mucus on the gills these species filter the minute particles of food from the water brought into the mantle cavity. The food is collected in mucous strings that are carried by means of cilia to the mouth and are pulled in by the radula. Some attached snails such as the Vermetidae use mucous traps to catch their food. Thin sheets of mucus are extended into the water where they wave about and collect minute organisms and floating debris. The radula pulls the mucous sheet back into the mouth and it is eaten.

Most of the large mesogastropods, such as the cowries, conch shells (Strombidae), cameos (Cassidae) and cask shells (Tonnidae), feed mainly on echinoids and bivalves. Cowrielike *Trivia* and *Erato* have developed a proboscis that can be extended into the mouths of the encrusting colonial tunicates on which they feed. Some cowries have a varied diet and one species has been observed feeding on algae, sea anemones, sponges, dead and decaying mollusks or fish and even the eggs of other snails.

Some mesogastropods are scavengers feeding on dead animal matter, while others (*e.g., Hipponyx*) have formed a commensal relationship, living on the shells of turbinids and feeding on their fecal pellets.

The scavengers and browsing forms that fed on encrusting animals probably led the way toward the carnivorous habit. Among the mesogastropods the Naticidae (moon shells) and Pterotracheidae (heteropods) are true carnivores that actively hunt their prey. The moon shells drill neat, round holes in the shells of bivalves, insert their proboscides and devour the soft parts within. The heteropods can swim and actively catch the small fish and other animals on which they feed.

All neogastropods (order Neogastropoda) are flesh eaters, either scavengers or carnivores, and many combine both habits. Scavengers such as *Nassarius* often may be seen in numbers approaching a dead fish, which they can detect from some distance, gathering in large numbers to feed at a common table. Baiting such forms by staking out a dead fish is often a successful way to collect them.

Opisthobranchs also show the complete range of the feeding habits found in the prosobranchs; the lower types are generally plant and detritus feeders, while the higher forms are scavengers and carnivores. Notes on their feeding habits are included under the section on the major groups, below.

The subclass Pulmonata includes mainly plant and detritus feeders, though *Euglandina, Streptaxis* and some of the slugs are carnivorous. They feed on worms, insect larvae and other snails and their eggs. Ground-living pulmonates such as *Mesodon* (Polygyridae) often are found feeding on mushrooms. Some species of helices are capable of digesting cellulose by means of an enzyme produced by certain intestinal bacteria.

The few species of gastropods that are completely parasitic are quite small. Many species in the Eulimidae are parasitic on sea urchins, starfish and sea cucumbers. The tips of their white shells often may be seen protruding slightly from the surface of their host. The young eulimid has a typical radula with which it bores into its host, but the radula is soon lost and the snail feeds by sucking the body juices of the host. Most species in the Pyramidellidae are now thought to be parasitic; this group has developed elaborate mouth parts adapted for puncturing the host and sucking the body juices. The most highly specialized of the parasitic snails is *Entoconcha*, which lives in the gonads of the sea cucumber. These snails have lost the radula, jaws and shell and look more like worms than mollusks, but the young have a well-developed shell and even an operculum, features that are definitely gastropodan.

V. REPRODUCTION AND DEVELOPMENT

1. Breeding Habits.—Fertilization in a few opisthobranchs and in many of the lower groups of gastropods may be external, but in most forms it is internal and there is usually a male intromittent organ. The embryonic stages include trochophore and veliger larvae, which may be free-swimming or enclosed within a gelatinous covering or egg capsule, or the eggs may be retained within the body of the parent until the young are fully developed and emerge as diminutive crawling snails.

2. Embryology and Development.—In all gastropods, cleavage of the egg is complete. In groups lacking yolk it is possible to follow the early cleavage stages of living embryos with the aid of a microscope. The first larval stage, the trochophore, usually lasts only a few hours; it is variable but commonly has a ring of cilia around the broad portion with a tuft of cilia at either end. During the second larval stage, the veliger, the embryonic shell is developed. Generally those species having a long, free-swimming larval life have large velar lobes with strong cilia that are used in feeding. Species with such larvae usually are widely distributed. In species with a short, free-swimming larval life, usually those that produce their eggs in gelatinous strings or masses, the velar lobes are smaller and not so elaborately developed. Such forms use the larval stage mainly as a means of dispersal rather than feeding.

LYNWOOD M. CHACE

FIG. 7.—SLUG LIFE CYCLE

Stages in the life cycle of the black slug (*Limax*), a hermaphroditic pulmonate: (left) pair of slugs, having lowered themselves from a branch by means of a mutually produced mucous string, intertwine and mate while suspended; (above) the shiny, gelatinous eggs are laid in clusters in sheltered places; (right) eggs and recently hatched slugs, tiny replicas of the adults

Many snails produce their eggs in capsules within which the young may spend their entire larval life. Eggs of such forms may have a large yolk supply, or the capsules may contain nurse eggs on which the developing young feed. The shell of the veliger, which usually differs markedly from the adult shell both in structure and sculpture, often may be seen at the apex of older specimens. Generally speaking, if the veliger shell is composed of several whorls, the larval stage has been a long one. Those species with a short larval life have shells with few whorls, usually only one or one and a half. Larvae that are retained within the egg capsule apparently use their velar lobes to create currents in the capsular fluid as an aid in respiration and to keep themselves distributed throughout the capsule. Some groups of land snails have calcareous egg shells. A few fresh-water forms (the Viviparidae) retain the eggs within the body of the parent until the time of hatching and the young emerge as tiny snails; such species are referred to as ovoviviparous.

VI. EVOLUTION AND RELATIONSHIPS

On the basis of comparative anatomy and embryology, definite natural groups of gastropods have been established and the relationship of most of the groups determined. The difficulty in studying gastropod evolution is in relating the recent to the fossil forms. As has been shown, the shell is not a reliable indicator of classification above the generic level. As a consequence paleontologists have never attempted to work out a classification of gastropods based on shell characters, as has been done with bivalves. Instead they have fitted gastropod families, genera and species into higher categories provided by the students of recent mollusks.

The recent discovery of *Neopilina,* a living monoplacophoran (*see* MONOPLACOPHORA), has shown that all limpetlike fossil shells cannot be classed as gastropods. However, *Helcionella,* a limpetlike genus of the Cambrian period, generally is considered to be a primitive gastropod that had undergone torsion but had a caplike shell. The bellerophontids (*Owenella*) of the Upper Cambrian and Lower Triassic were perhaps the next step in evolution; the shell, though bilaterally symmetrical, is planispiral. The slit at the aperture of these shells is very similar to that of the Pleurotomariidae, and most paleontologists today believe that it functioned as an anal slit, as it does in the recent forms; therefore, it is quite reasonable to believe that the animals had undergone torsion.

The first asymmetrically coiled shells probably arose from the bellerophontids as a result of the protrusion of the early whorls to the right or left to form the apex of the spire. In most forms the protrusion was to the right, producing a dextral, or orthostrophic, shell; those that protruded to the left are sinistral, or hyperstrophic. In Upper Cambrian times a group arose that can be traced in a direct line to the present-day Pleurotomariidae; these were undoubtedly the ancestors of present-day prosobranchs.

It is generally believed that the Opisthobranchia arose from some archeogastropod stock such as the trochoids in which the right pallial organs have been lost. The opisthobranchs are believed to be monophyletic (derived from a single line) because all forms have a sinistral larval shell even though the adult shell, when present, is dextral. The Pyramidellidae and Acteonidae are more nearly allied to the prosobranchs than are other opisthobranchs, but they clearly belong in this subclass because of the sinistral larval shells, the anatomy of the digestive tract and the occurrence of hermaphroditism. The ancestors of these families, which were probably transition forms between the prosobranchs and the opisthobranchs, are known at least as far back as the early Carboniferous. The pteropods appear to be opisthobranchs that have retained the pelagic habit throughout life. The nudibranchs probably arose at a much later period from more than one group of opisthobranchs.

The pulmonates are generally considered to be close to the opisthobranchs and probably arose from some early opisthobranch stock or from some archeogastropod stock close to that from which the opisthobranchs arose. The earliest reliable pulmonate record is from the Carboniferous. There is still some question about the placement of the order Systellommatophora. Some authorities place it in the Opisthobranchia and others in the Pulmonata.

There is no question that they are transition forms between the two subclasses, but since two of the three families included in the order appear to be closer to the Pulmonata, they are included in that order here. However, a great deal more work needs to be done on this obscure group before definite placement can be made.

VII. SURVEY AND CLASSIFICATION

Many factors are taken into consideration when attempting to arrange gastropods into natural groups. No one character can be used alone. Consequently it is little wonder that there is no one classification universally agreed upon by all workers. No attempt will be made here to discuss the various theories or to give a detailed classification of the gastropods but only to outline the major groups which seem to be most universally accepted. For those interested in pursuing the subject further, the references listed in the bibliography will be of help.

The class Gastropoda is divided into three subclasses on the basis of the nervous, respiratory, circulatory and reproductive systems. The nervous system becomes progressively more concentrated, the reproductive system more complex and the respiratory system more specialized, beginning with the oldest known group,

the archeogastropods (subclass Prosobranchia) which are known from the Cambrian, and progressing upward through the opisthobranchs (subclass Opisthobranchia) to the Stylommatophora (subclass Pulmonata), which first made their appearance in the Upper Cretaceous.

Subclass Prosobranchia
  Order Archeogastropoda
    " Mesogastropoda
    " Neogastropoda
Subclass Opisthobranchia
  Order Cephalaspidea
    " Anaspidea
    " Pteropoda
    " Sacoglossa
    " Acochlidacea
    " Notaspidea
    " Nudibranchia
Subclass Pulmonata
  Order Systellommatophora
    " Basommatophora
    " Stylommatophora

1. **Prosobranchia.**—In the first subclass, the Prosobranchia (*proso*, "forward"; *branchia*, "gills"), the gills, when present, are located in front of the heart. Where the gills are absent, respiration is carried on by a special outgrowth of the mantle (Patellidae), or a simple pulmonary chamber in land species (Helicinidae). All prosobranchs show evidence of torsion, the visceral mass and nervous system being twisted to form a figure 8; hence this group is often referred to as the Streptoneura. An operculum is present except in those groups with caplike shells or shells with very large apertures as in abalones (Haliotidae). There is only one pair of tentacles with the eyes at the base (fig. 8). The foot is usually large, developed for creeping or holding fast to the substrate. The radula, present in all but the parasitic forms, is varied and can be used in grouping the prosobranchs into orders. The sexes are separate, and in higher forms the males have a large copulatory organ.

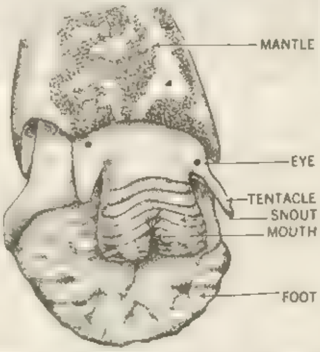


FIG. 8.—ANTERIOR END OF PROSOBRANCH SNAIL (TRUNCATELLIDAE) EMERGING FROM SHELL WITH FOOT STILL FOLDED

This subclass contains mainly marine species; of the approximately 90 families of prosobranchs only three are restricted to land and four to fresh water. Some families, such as the Bulimidae, are found in both marine and fresh-water habitats, and the Truncatellidae are found near the high-water mark at the edge of the sea (*Truncatella*) or on land (*Geomelania* and *Taheitia*). The Neritidae, though mainly marine, may be found also in fresh water and on land. The subclass is divided into three orders on the basis of the shell and radula and the nervous, respiratory and reproductive systems.

Archeogastropoda.—This order contains the oldest and least specialized of the gastropods (fig. 9). Other than the families Neritidae and Helicinidae, which are highly specialized, the archeogastropods are entirely marine and are characterized, in most families, by their pearly shells. The shell may vary in shape from typically top-shaped, spirally coiled shells, as in the Pleurotomariidae, Trochidae and Turbinidae, to a flattened coil, as in the Haliotidae, or a volcano- or cap-shaped shell that has lost all evidence of its early coiling and again become outwardly bilaterally symmetrical, as in the Fissurellidae, Acmaeidae and Patellidae. The foot is large and the operculum, in most of the trochoid forms, may be calcareous or horny. Respiration is by means of single or paired bipectinate gills, free at the anterior end, by a pallial outgrowth (Patellidae) or by a modified pulmonary sac (Helicinidae).

The archeogastropods, mainly plant feeders, have specialized radulae. The rhipidoglossate radula found in most families is a long, wide ribbon with strong lateral teeth for scraping and numerous long, weak marginal teeth for carrying the minute particles into the mouth. The docoglossate radulae found in the Acmaeidae and Patellidae are close to those found in the chitons; both these families feed in much the same manner as do the chitons.

FIG. 9.—SUBCLASS PROSOBRANCHIA, ORDER ARCHEOGASTROPODA, SHOWING REPRESENTATIVE GENERA AND FAMILIES AND CHARACTERISTIC DENTITION

The twisted nervous system is much less concentrated than in other prosobranchs. There are no large ganglia in the head region. In most species the ladderlike arrangement of nerves in the foot is similar to that found in the chitons. The eyes are open, being only retinal pits lacking both cornea and lens, though in a few forms the opening is small and the pit is filled with a vitreous body bathed with water. Many archeogastropods retain much of the original bilateral symmetry, having two pectinate gills, two kidneys and a heart with two auricles.

The sexes in all groups of archeogastropods are separate, and in all except the Neritidae and Helicinidae there is no true penis or oviduct, the reproductive products being passed out through the modified right kidney. The sexes in these groups can be distinguished only by the colour of the sex organs or their products. The male organs are usually cream coloured, while the female organs are more brightly coloured. In many archeogastropods, as in some species of *Haliotis* and *Patella*, the eggs and sperm are shed into the water and fertilization is external. In such groups the eggs lack a protective covering, or have only a thin, gelatinous covering. Other groups deposit eggs in gelatinous ribbons that are attached to the substrate. The life span of the free-swimming larva varies greatly: in some forms the trochophore and veliger stages are passed within the egg; others have only a free-swimming veliger; and for a few the trochophore and veliger are planktonic.

Mesogastropoda.—This is a varied order that contains land, fresh-water and marine forms (fig. 10). Some species cement their shells to the substrate; others are completely pelagic. There are filter feeders, plant feeders, carnivores and parasites. Some are only a few millimetres in length when fully adult, while the giant *Megalatractus*, the false trumpet shell of Australia, reaches over two feet in length. Some are completely smooth (Naticidae) and others have an elaborate sculpture of spines (Muricidae). The cowries (Cypraeidae) cover the shell with the mantle and have highly polished shells, while the tritons (Cymatiidae) have a heavy, hairy outer coat of periostracum. Some are restricted to the intertidal zone; others are found only in the abyssal depths. Many are intermediate and have a considerable range.

Though most families have spirally coiled shells, in a few groups such as the Capulidae (the cap shells) and the Calyptraeidae (the

slipper limpets) the shell has become flattened and limpetlike or disc-shaped. The shell is never pearly and there is usually an operculum, which may be calcareous or horny.

In this order the nervous system is more concentrated around the esophagus and there is only a single, large pedal ganglion; however, the visceral loop is still in the form of a figure 8. The eyes are always closed and there is a well-developed cornea. There is no evidence of bilateral symmetry, the organs of the pretorsional right half of the animal having disappeared completely. Respiration is generally by means of a single, large monopectinate gill, which is attached throughout its length. In terrestrial families such as the Cyclophoridae and the Pomatiasidae the pallial or mantle cavity is transformed into a lung, while in the fresh-water family Ampullariidae both a gill and pulmonary sac are present, separated by an incomplete septum. There is a single kidney usually opening through a slitlike aperture. The heart has only a single auricle.

The sexes are separate, except in the Cerithiacea, and the male has a well-developed and often large copulatory organ. The genital opening is separate, the reproductive products no longer passing out through the kidney. Fertilization is internal and the eggs may be deposited singly, in gelatinous masses or strings, or in a variety of egg capsules.

The radula is usually taenioglossate, consisting of a large central tooth, two lateral teeth and four marginal teeth, but in the Epitoniidae and Janthinidae it is ptenoglossate, being composed of more or less uniform teeth, the central tooth lacking. The shorter, narrower, more strongly toothed radulae of the mesogastropods foreshadow the powerful radula of the carnivorous Neogastropoda. Some families of mesogastropods, such as the Naticidae (the moon shells), the Pterotracheidae (the heteropods), some of the cowries (Cypraeidae) and the tonnas (Tonnidae) are carnivorous. The heteropods are pelagic, fast-swimming animals that catch and feed on small fish, copepods, jellyfish and other planktonic animals. Most mesogastropods, however, are rather sluggish, grazing animals feeding on detritus, plants or encrusting animals such as sponges and bryozoans. In parasitic species the radula may be absent in the adult.

Neogastropoda.—The third order of the prosobranchs, the Neogastropoda, is entirely marine (fig. 11). All have well-developed, solid, spirally coiled shells, though a few, with very large apertures,

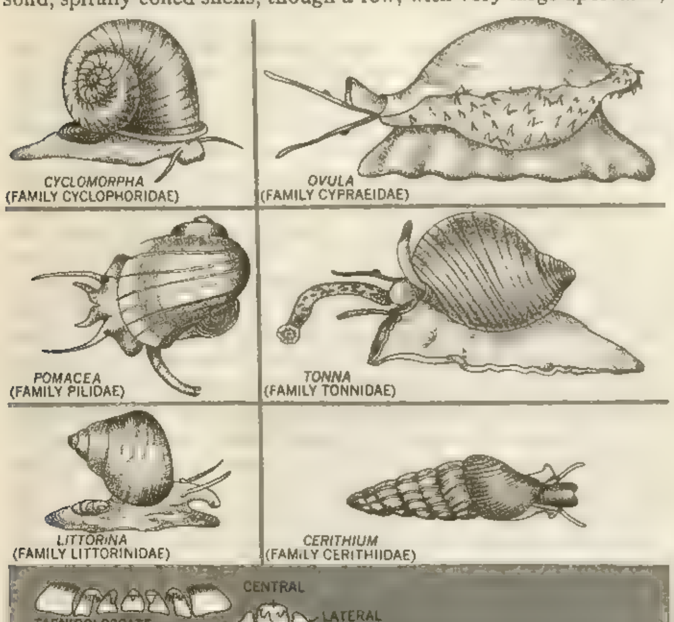


FIG. 10.—SUBCLASS PROSOBRANCHIA, ORDER MESOGASTROPODA, SHOWING REPRESENTATIVE GENERA AND FAMILIES AND CHARACTERISTIC DENTITION

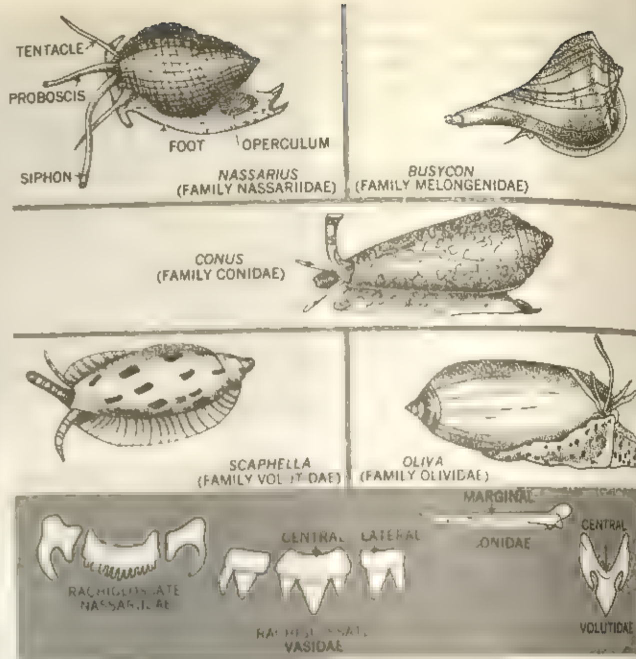


FIG. 11.—SUBCLASS PROSOBRANCHIA, ORDER NEOGASTROPODA, SHOWING REPRESENTATIVE GENERA AND FAMILIES AND CHARACTERISTIC DENTITION

have become somewhat flattened. The operculum, when present, is always horny and has few spirals. The neogastropods are all carnivorous. In most forms there is a well-developed proboscis that can be everted into the aperture of other gastropods, into the substrate to feed on worms or into the testes of echinoderms, etc. The radula is a short, narrow ribbon usually with only three large strong teeth per row—a central tooth and two lateral teeth. In some forms such as the Volutidae only the central tooth remains, and in the Toxoglossa (the Terebridae and Conidae), considered the highest group of the prosobranchs, the central and lateral teeth are lacking and the remaining marginal teeth are highly specialized. The teeth are needlelike and hollow, forming a minute hypodermic, usually with a barb at the end. In the Conidae the associated poison gland produces a powerful neurotoxin used to immobilize prey. The nervous system in the neogastropods is increasingly more complex; the visceral loop is still in the form of a figure 8, but the two intestinal ganglia come to lie close to the pleural ganglia in the head region. In this order the osphradium, a chemical sense organ, which lies near the gill, is usually double and highly developed. The sexes are separate, and the males have a large copulatory organ. Fertilization is internal; the eggs may be protected by a variety of egg cases or capsules, and the young may emerge in the late veliger stage or as young crawling snails.

2. Opisthobranchia.—In the second subclass, the Opisthobranchia (*opisthen*, "behind"; *branchia*, "gill"), the gills are behind the heart, though in many forms the gills are lacking and respiration is carried on over the entire surface of the body (fig. 12). The visceral loop and nervous system are not twisted into a figure 8, except in the Acteonidae (this group has undergone a secondary detorsion and so is often referred to as the marine Euthyneura). The nervous system is more concentrated around the esophagus, though there are still visceral and pedal ganglia that are connected by long commissures. The shell is greatly reduced or lacking, except in the Acteonidae, Bullidae and Pyramidellidae, though it is present in all forms in the larval stage. It is often flattened and platelike or ear-shaped and it may be partly or entirely internal. An operculum is present only in the Acteonidae and Pyramidellidae. All opisthobranchs are hermaphroditic.

There are usually two pairs of tentacles, though the posterior pair are more properly referred to as rhinophores, their function being olfactory. The foot is variously developed in the several orders as a creeping or swimming organ. As most opisthobranchs have little or no shell, the fossil record is limited, but they are known at least as far back as the Carboniferous and may go back

as far as the Upper Cambrian. Included here are the sea hares and their relatives, the nudibranchs, or sea slugs, and the pteropods, or sea butterflies.

The opisthobranchs are the least-known gastropods and their classification the most disputed. The number of orders recognized may vary from three to eight, depending upon the authority, but the general order in which the approximately 70 families are arranged, except in a few cases, is much the same. The classification most generally followed is that of Odhner.

Cephalaspidea.—Members of the order Cephalaspidea are characterized by usually having a well-developed external shell, though it may be internal or even absent. The head has a large shield or disc used for digging or plowing in the mud or sand in which the animals live. There are no tentacles, the eyes being sessile. The primitive Cephalaspidea such as *Actaeon* are mainly detritus feeders, but the more advanced forms have a radula developed for seizing prey and a muscular gizzard with horny or calcareous plates for crushing the prey, which is swallowed whole. The larger species may feed on bivalves and other gastropods; the smaller ones on foraminifera, worms, etc. This order includes the bullas or bubble shells, which have large apertures but lack opercula. Only in the Acteonidae, the most primitive of the group, is the operculum present. The nervous system is still twisted in a figure 8. In all other families in this order the visceral mass has become detorted, but the visceral loop is large and the nerves have not become highly concentrated around the esophagus.

Anaspidea.—Anaspidea lack the cephalic shield or disc; the shell is reduced, becoming largely or completely internal or lacking. The head has two pairs of tentacles; the large posterior pair, rhinophores, are olfactory organs; the short anterior pair are tactile organs. The eyes are sessile and are located just in front of the rhinophores. The foot is large, extending forward, anterior to the head and posteriorly beyond the visceral mass. Parapodia or swimming lobes, about midway on the dorsal surface of the foot, are fleshy, mobile flaps that cover the visceral hump when the animal is resting or crawling but can be moved in an undulatory way when the animal is swimming. Members of this group are known as sea hares, their large rhinophores giving them a rabbit-like appearance. Many species in this group emit a purple dye when disturbed. They feed mainly on seaweeds such as *Ulva* and *Fucus* and are capable of adapting their colour to match the seaweed on which they are feeding.

Pteropoda.—The order Pteropoda includes the families of small pelagic opisthobranchs that have the foot modified into a pair of large parapodia used for swimming. They have transparent shells and bodies so that it is possible to see the large jaws, radula and other organs without dissection. This order is divided into two orders by many workers—the Thecosomata, the shelled forms, and the Gymnosomata, those without shells.

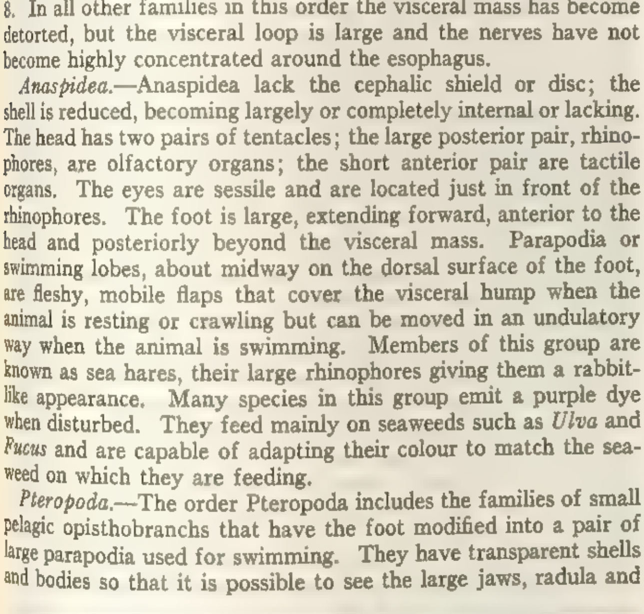


FIG. 12.—SUBCLASS OPISTHOBRANCHIA SHOWING REPRESENTATIVE ORDERS, FAMILIES AND GENERA

The Pteropoda often are referred to as sea butterflies because of their flying motion in the water and their delicate, beautiful colouring. The shelled pteropods, the Thecosomata, are ciliary feeders, collecting microscopic plants and animals by means of cilia on the parapodia or wings and entangling them in mucus from the pallial mucous gland. The muscular gizzard is used to crush diatoms. The Gymnosomata are predators, actively hunting their prey and feeding on minute planktonic animals.

Sacoglossa.—The Sacoglossa is a small order of very specialized opisthobranchs adapted for feeding on the fluid within the cells of algae. They have a radula composed of a single series of strong teeth specialized for piercing and a buccal pump for sucking up the plant juice. Only one tooth is used at a time, and, when it is worn out, it is stored in a sac below the mouth. Members of this order may have thin bulla-like shells or they may be naked and look much like nudibranchs. It is a small but varied group containing five to seven families and probably not more than 50 species. However, since they are small and not easily seen, they really may be more abundant. They are world-wide in distribution, living on seaweeds.

The only known bivalved gastropods (Juliidae) belong to this order. As veliger larvae they have a typical small, coiled shell, but in the adult they are bivalved, the coiled embryonic shell remaining on the left valve. The shells have been known from the fossil record of the Oligocene of Europe. Studies made of the living animal, discovered by S. Kawaguti in 1959, proved that the Juliidae are gastropods rather than bivalves of the class Pelecypoda, as was earlier believed. Species in two genera, *Julia* and *Berthelinia*, have now been found extant. Once it was known that these small and remarkable animals were restricted to feeding on the seaweed *Caulerpa*, workers readily found them in waters off Japan, Australia, Hawaii, Jamaica and the west coast of Mexico.

Acochlidacea.—The order Acochlidacea is a very small group of minute and highly specialized opisthobranchs that live in sand and mud and feed on detritus. They are adapted for burrowing, the anterior end being narrow and extensible. The visceral sac is free from the foot and extends out posteriorly. Much of the body may be covered with spicules that probably aid in burrowing. The shell is lacking in this order as well as the gill and parapodia. (This group was included along with the Notaspidea in the order Aceola by Thiele.) They are probably world-wide in distribution, though little is known about them. They are often found inside minute gastropod shells where they probably go both for protection and to feed on the remains of dead animal matter.

Notaspidea.—The order Notaspidea contains two families of large opisthobranchs, usually with a flat, disclike shell that may sit like an umbrella on the top of the animal (Umbraculidae), may be covered by the mantle or may be absent (Pleurobranchidae). The foot is large and flat, adapted for creeping and adhering to rocks; there are no parapodia. A single large ctenidium or gill is located on the right side in the groove between the mantle and the foot. The tentacles may be small or large and, in pleurobranchs, are fused together, forming a broad veil that extends over the mouth. The rhinophores are usually large. These are bottom-living detritus and plant feeders; they have a broad radula with many rather uniform teeth typical for this type of feeding.

Species in these families are world-wide in distribution in tropical and subtropical seas, a few forms reaching temperate regions. In Australia they are commonly referred to as umbrella shells and side-gilled sea slugs. They may be found from just below the low-tide line out into somewhat deeper water in areas where seaweed is abundant, providing surfaces to which these species usually attach their strings of gelatinous eggs. Although the shells of these forms are not attractive, the animals may be brilliantly coloured and must be seen alive to be appreciated.

Nudibranchia.—This is the largest order of the Opisthobranchia, comprising more than 30 families and containing those forms usually referred to as sea slugs. Nudibranchs are among the most beautiful of all mollusks, though generally they are known only by

the specialists, for most of the species are small. Their beautiful colouring and varied form can be appreciated only when the animal is alive and active, for these soft-bodied mollusks, often with elaborate outgrowths on the dorsal surface, collapse when removed from the water.

Externally they are bilaterally symmetrical; they are shell-less in the adult stage and lack a true gill and mantle cavity. Detorsion is complete in this group, and the nervous system is more highly concentrated in the head region than in other opisthobranchs. The club-shaped rhinophores may be elaborately developed and accessory gills may be present in some forms, while others use the entire body surface for respiration. Dorsal outgrowths (cerata), which may be club-shaped or branching, are common. One group of nudibranchs, the aeolids, feed on coelenterates and store the nematocysts (the stinging cells) of corals in their dorsal cerata; they may serve to protect the animal. Other nudibranchs feed on sponges; the pelagic forms feed on jellyfish and siphonophores.

Nudibranchs are world-wide in distribution, though they are far more abundant in tropical seas. The largest and most beautiful sea slugs are to be found on the Barrier Reef of Australia; one species there reaches 12 in. (30 cm.) in length. They may be found living in tide pools or below low-tide line on rocky shores, among attached seaweed, on coral, hydroids, sponges or barnacles. Some species live only in floating seaweed, such as the sargassum sea slug (*Scyllaea pelagica*), while others, such as the sea lizard (*Glaucus eucharis*), live in swarms in the open sea, where they may form a violet-blue carpet on surface waters.

3. Pulmonata.—The third subclass, the Pulmonata (*pulmo*, "lung"), includes only land and fresh-water species, with the exception of the families Siphonariidae, Oncidiidae, Ellobiidae and Amphibolidae (fig. 13). The gills are lacking and respiration takes place in a pulmonary cavity well supplied with blood vessels and having a contractile orifice. This subclass is often referred to as the land Euthyneura, for most specialists believe they, like the marine Euthyneura, also have undergone detorsion. The nervous system is highly concentrated in the head region.

All pulmonates are hermaphroditic, and in most forms the reproductive system is highly complicated and specialized. Fertilization is internal, and the eggs may be deposited in gelatinous masses or they may have a calcareous shell. The radula is broad, relatively short, with a central tooth and numerous rather similar laterals and marginals. Most species are plant and lichen feeders, though a few, such as members in the Streptaxidae, Haplotrematidae and

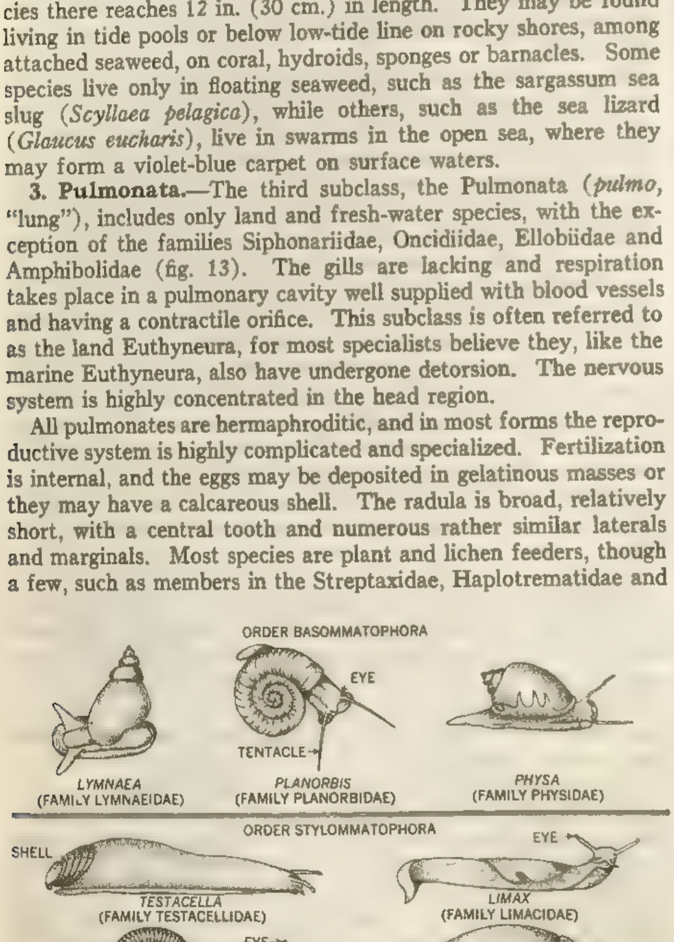


FIG. 13.—SUBCLASS PULMONATA SHOWING REPRESENTATIVE ORDERS, FAMILIES AND GENERA AND CHARACTERISTIC DENTITION

Oleacinidae, are carnivorous. Most families have well-developed spiral shells, though in the Ancyclidae, Siphonariidae, Lancidae and in one genus of the Planorbidae the shells are limpetlike. Some of the land slugs are shell-less, while others have the shell greatly reduced.

The fossil record of the Pulmonata is very limited, for most land and fresh-water snails do not live and die in situations favourable to fossil formation. The best records are from relatively recent loess formations. They are known at least as far back as the Pennsylvanian. The three orders in this subclass are easily recognized and are agreed upon by all workers, though there is some disagreement about the placement of the Systellommatophora, some regarding this group as belonging to the Opisthobranchia.

Systellommatophora.—This is a small order of primitive, sluglike, pulmonate snails that have no mantle cavity and have the lung, nephridial opening and anus at the posterior end of the body. The eye stalks are contractile but cannot be inverted as in the Stylommatophora. They are hermaphroditic and the openings of the male and female systems are separate, as in the Basommatophora. The male opening is on the right side of the head, the female opening farther back on the right side. Three families are included: the Oncidiidae are intertidal but move about and feed only during low tide when they are out of water; the Veronicellidae and Rathouisiidae are land forms restricted to humid areas of the tropics and near tropics. The common Cuban species *Veronicella floridana*, also found in southern Florida, feeds only at night and may be quite destructive in gardens.

Basommatophora.—Members of this order are characterized by having only a single pair of short, broad, noncontractile tentacles, with the eyes at the base. The shell may be extended and spirally coiled, flattened and planospirally coiled or cap-shaped. The openings of the male and female reproductive systems are separate and lie close together on the right side near the eye. This order includes mainly fresh-water species, though there are a few brackish-water and marine forms.

The Ellobiidae are heavy-shelled, spirally coiled marine and brackish-water gastropods often found in salt marshes; the Siphonariidae are limpetlike marine species found on rocky shores and on roots of mangroves. The Physidae and Lymnaeidae include the common thin-shelled, fresh-water snails often referred to as pond snails, though they are also found in streams and large lakes. Although some lymnaeids come to the surface to fill the lung with air, other deep-lake forms are entirely aquatic, filling the pulmonary cavity with water. These two large families are easily distinguished: in the Physidae the aperture is on the left (sinistral) and in the Lymnaeidae on the right (dextral). The Planorbidae have flattened shells, generally coiled in a single plane. The Ancyclidae are limpetlike fresh-water snails. Only the Amphibolidae have an operculum; this is a small family of rather primitive pulmonates found in the upper intertidal salt and brackish-water marshes of Australia and New Zealand.

The reproductive system in the Basommatophora is very much like that of many opisthobranchs; only Ellobiidae (the salt marsh snails) have free-swimming young. In the Amphibolidae and Siphonariidae the eggs are produced in a jellylike mass or in coiled strings similar to those of opisthobranchs, without a separate capsule for each egg. The remainder of the Basommatophora produce the eggs in a jelly mass but each egg has its own capsule and sufficient yolk so that it passes through the trochophore and veliger stages within the capsule, emerging as miniatures of the adult.

Stylommatophora.—Families in the order Stylommatophora contain members characterized by having two pairs of tentacles that are retractile and can be invaginated. The eyes are located at the tip of the large tentacles. The openings of the male and female reproductive systems are united. The shell, which is present in most forms, is helicoid and generally one of three basic forms: lenticular, globose or attenuate. In land slugs the shell is rudimentary and internal or absent.

The reproductive system in the Stylommatophora is the most highly developed in the Mollusca and many groups of land snails have an involved courtship behaviour. The eggs of land snails are generally fewer in number, larger, with more yolk and with a much

better protective covering than is the case in marine and freshwater forms, adaptations to the hazards of living on land. Many land snails that lay their eggs in damp ground have a tough, gelatinous coating around the egg; others lay eggs with calcareous shells; while still others retain the eggs within the body of the parent until they are ready to hatch (Sagdidae).

Most land snails are plant and detritus feeders, though a few are carnivorous, feeding mainly on other snails, snail eggs, worms or insect larvae. Most pulmonate snails belong in this order, which contains about 50 families (many more according to some workers) and perhaps 8,000 species.

See also references under "Snail" in the Index.

Bibliography.—The following list contains only a very small portion of the multitude of books and papers written on this subject; most of them, however, have extensive bibliographies that will lead to greater details on the various groups of gastropods.

B. C. Cotton, *South Australian Mollusca, Part 3, Agchaeogastropoda* (1959); D. R. Crofts, *"Haliotis," Liverpool Marine Biology Series 29* (1929); L. R. Cox, "Thoughts on the Classification of the Gastropoda," *Proc. Malac. Soc. Lond.*, vol. 33, no. 6, pp. 239–261 (1960); J. B. Knight, "Primitive Fossil Gastropods and Their Bearing on Gastropod Classification," *Smithson. Misc. Coll.*, vol. 117, no. 13 (1952); A. B. Leonard, "Handbook of Gastropods in Kansas," *Publ. Mus. Nat. Hist. Univ. Kans.*, (1959); Ernst Marcus, "On Western Atlantic Opisthobranchiate Gastropods," *Amer. Mus. Novit. no. 1906* (1958); T. E. Thompson and D. J. Slinn, "On the Biology of the Opisthobranch Pleurobranchus membranaceus," *Journal of the Biological Association of the United Kingdom*, vol. 38, pp. 507–524 (1959); V. Fretter and A. Graham, *British Prosobranch Molluscs* (1962); A. R. Mead, *Giant African Snail: A Problem in Economic Malacology* (1961); N. B. Eales, "Torsion in Gastropoda," *Proc. Malac. Soc. Lond.*, vol. 28, pp. 53–61 (1950); H. J. Van Cleave, "Snails and Their Relations to the Soil," *Biologist*, vol. 34, no. 1, 2 (Sept.–Dec. 1951); F. C. Baker, "Fieldbook of Illinois Land Snails," *Ill. State Nat. Hist. Surv. Manual 2*, pp. 1–166 (1939); A. M. Keen and A. G. Smith, "West American Species of the Bivalved Gastropod Genus *Berthelina*," *Proc. Calif. Acad. Sci.*, vol. 30, no. 2, pp. 47–66 (1961). (R. D. T.)

SNAKE, common name for reptiles whose most distinctive features—reduction or loss of limbs, reduction or loss of one lung, and lengthening of internal organs—are associated with extreme elongation of the body. Among the smallest is a thread snake of Syria, about the size of a knitting needle; the giants are tropical pythons reported to be more than 30 ft. long. Snakes, suborder Ophidia or Serpentes, are closely related to the lizards, and with them constitute the dominant order of modern scaly-skinned reptiles, the Squamata. They may have originated from some monitorlike lizard stock of which the Bornean *Lanthanotus* may be a unique survivor; fossils identifiable as snakes are known from the Cretaceous Period.

Snakes have always inspired mankind with feelings of fear and revulsion, but also with awe. Their pictorial representations, carved in antlers, are found as early as the Paleolithic Age. Serpents figure prominently in the religion and mythology of all primitive peoples; their worship, still widespread, may be the greatest of all animal cults. The economic importance of snakes is considerable. The major item on the debit side is the mortality arising from snakebites, estimated at more than 30,000 deaths yearly, with the highest incidence in Burma. There are also losses due to predation on domestic and stock animals, and more seriously, destruction of animals that serve as natural controls of injurious insects. These ill effects, however, are offset by the beneficial activities of the rodent-eating snakes. Snakes are also used for food in many countries, and their skins produce ornamental leathers for which there is a demand. The venom has some limited pharmacological applications. To satisfy these demands there have been attempts to "farm" snakes. The various snake parks, mainly associated with the preparation of antivenins, are usually dependent upon captured specimens.

Distribution.—Snakes are cosmopolitan in distribution, with the exception of the Arctic and Antarctic regions, New Zealand, Ireland, and some of the more recent oceanic islands. A few species reach or cross the Arctic circle (*e.g.*, the adder in Europe and a ribbon snake in Canada), but the areas of permanently frozen subsoil limit snake distribution both geographically and topographically. Upper montane limits are 9,800 ft. (3,000 m.) in the European Alps (adder), about 13,000 ft. (4,000 m.) in Mexico (a garter snake and a rattlesnake), and 15,700 ft. (4,800 m.) in the Himalayas (a pit viper). Within these limits snakes occur in almost every kind of environment, arid or humid, subterranean, terrestrial, arboreal, freshwater, and marine, with the greatest numbers being found in the damp tropics.

Characteristic Features.—The following characters, typical of snakes, are also to be found, though seldom in combination together, in various lizards, notably in burrowing snakelike forms such as the Amphisbaenidae.

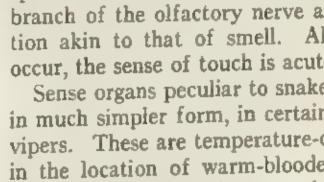
Snakes typically are devoid of limbs, though vestiges of the hind pair occur in a few families. Associated with the attenuated body form are increase in the number of vertebrae and ribs, reduction or suppression of the left lung, staggering of the other paired internal structures, and closure of the front end of the braincase. Although the mouth is narrow, certain anatomical features allow the passage of bulky prey: bones supporting the jaws are reduced in number and are loosely articulated; the rami of the lower jaw are connected by a very elastic ligament; and the glottis is protrusible to permit respiration while the mouth is occluded by prey.

The flexibility and strength of the vertebral column necessary to withstand the stresses set up by the methods of locomotion are ensured by additional articular surfaces between the vertebrae (zygantra and zygosphenes). There are no movable eyelids; the cornea is protected by a fixed transparent scale called the brille. Sight is one of the more important senses in most terrestrial and arboreal snakes; probably there is colour vision. There is no external ear or tympanum and the middle-ear bone abuts laterally against the quadrate bone of the skull so that vibrations of the substratum are picked up; only low-frequency sound waves, between 100 and 700 cycles, are perceived. The tongue, elongate and retractile into a basal sheath, has a forked tip whose points can be inserted into two depressions on the roof of the mouth—the openings of the chemoreceptor organ of Jacobson.

There are adaptive modifications associated with the different environments. For example, in burrowing forms the skull bones are more firmly articulated, eyes are smaller and may be sunken, and scales are close fitting and highly polished. Nocturnal and twilight-active species have elliptically contractile pupils. A few arboreal species may have binocular vision; this is achieved, not as in other vertebrates by a forward rotation of the orbits but by the pupils being long, horizontal slits situated excentrically near the anterior margins of the lateral orbits.

The senses of smell and taste are supplemented by the tongue operating in conjunction with Jacobson's organ. The repeated protrusion and flickering of the tongue, so characteristic of snakes, is a chemical sampling of the environment. Airborne particles adhering to the tips of the tongue are transferred to the sensory epithelium lining Jacobson's organ; this organ is innervated by a branch of the olfactory nerve and presumably produces a sensation akin to that of smell. Although no special tactile organs occur, the sense of touch is acute.

Sense organs peculiar to snakes are found in the pit vipers and, in much simpler form, in certain boids, and perhaps also in some vipers. These are temperature-differential receptors and are used in the location of warm-blooded prey in darkness. In the pit vipers, the organ lies between the eye and the nostril, and consists of two chambers separated by a diaphragm. One of the chambers is wide open to the exterior, exposing the diaphragm, which is richly supplied with nerves. The diaphragm is very sensitive to infrared radiation: an object can be accurately located when its heat radiation produces on the membrane a temperature only 0.36° F higher than that of the surrounding air. In the boas the heat-receptors are a series of simple pits along the lips and the suspected receptors of the pit vipers are sacs associated with the nostrils.



FROM J. Z. YOUNG, "THE LIFE OF VERTEBRATES," AFTER GADOW, 1950. THE CLARENDON PRESS, OXFORD

FIG. 1.—HEAD OF PIT VIPER, SHOWING POISON AND SENSE MECHANISMS

There are three main types of dentition in venomous snakes. In the Opisthoglypha (many Colubridae), the fangs, one to three in number, lie at the posterior end of the upper jaw; they

are enlarged teeth with a lateral groove that acts as a channel for the venom. In the other two types, the fangs lie at the front of the upper jaw and the channel for the venom is an enclosed canal opening at the fang tip. In the Proteroglypha (Elapidae, Hydrophidae) the bone bearing the fangs is relatively long and immovable; it often bears some simple teeth behind the fangs. In the third type, the Solenoglypha (Viperidae), there are no teeth immediately behind the fangs, and the short bones on which the fangs are inserted are rotatable about their transverse axis through about 90°; this allows the fangs to be folded down flat, tips directed posteriorly, when the mouth is closed. The harmless colubrine snakes with solid conical teeth are sometimes called the Aglypha.

BY COURTESY OF THE AMERICAN MUSEUM OF NATURAL HISTORY, NEW YORK

FIG. 2.—SKULLS OF (LEFT) NONPOISONOUS AND (RIGHT) POISONOUS SNAKES

(Left) Royal python skull showing teeth adapted for seizing and swallowing prey; (right) Gaboon viper skull showing fangs

Salivary glands are present, two of which in certain snakes are modified into venom sacs. Food is well coated with saliva and is then smoothly worked into the esophagus by muscular action and the gripping power of the teeth (see *Feeding*, below). The esophagus bears longitudinal folds that permit expansion, another adaptation to allow passage of large prey. Products of digestion pass into the slightly coiled small intestine, while indigestible matter are regurgitated. A valve forms the junction of the small and large intestines; at this point is a primitive colic caecum or pouch. The short, straight large intestine opens into the cloaca, where sex products of the paired testes and the wastes of the paired kidneys also empty; there is no urinary bladder.

Locomotion.—Locomotion is accomplished by several methods, here broken down into four main types: serpentine (undulatory), rectilinear (straight crawling), thrust creeping and sidewinding.

The commonest means of locomotion is by serpentine movement, in which sinuous lateral undulations that originate anteriorly travel toward the tail; the pressure of these waves against objects produces a thrust that drives the animal forward.

Terrestrial and arboreal snakes often exhibit rectilinear locomotion. The belly has a series of transversely enlarged overlapping plates that provide for forward movement in a straight line. When some of the muscles of the body wall are contracted and relaxed, the ventral plates act like a series of ratchets. Their anterior edges, being overlapped by the preceding plate, slide freely forward even over a rough surface, but any backward movement is prevented because the projecting posterior edges catch on the least irregularities. In consequence, contractions of the muscles produce a series of small thrusts forward. Since increased length increases efficiency for both undulatory and rectilinear locomotion, slender species are generally swifter and more active than thickset ones. In completely aquatic forms, swimming is by lateral undulations and there are no enlarged ventral scales; efficiency is increased by lateral compression of the body, and, particularly, of the tail, which is paddle-shaped in the sea snakes.

Since serpentine movement is not possible in the confines of a narrow burrow, there is provision for thrust creeping by means of short, blunt, muscular tails, often armed with tubercular or spinose scales. The body is first thrown into a number of sinuous curves of the restricted amplitude allowed by the burrow and then straightened while the tail is held fast by firm pressure against the sides of the burrow.

One of the most extraordinary means of snake progression, well suited for traveling over loose sand, is sidewinding. The most notable example is found in the North American Desert, or horned, rattlesnake (*Crotalus cerastes*), often called sidewinder. This snake repeatedly throws a loop of its body forward and to one side, leaving a series of more or less parallel oblique tracks in the sand.

Feeding.—Snakes are insectivorous or carnivorous, and their prey is swallowed whole. The teeth, thornlike, with their tips recurved backward, are arranged in single rows along the jaws with, frequently, two additional rows on the roof of the mouth. The prey, which in most instances must be alive to interest a snake, is seized in the mouth by a rapid strike; then the process of overpowering and swallowing begins. Nonvenomous species swallow prey alive if it is small. A few species can subdue large and struggling prey by constriction; *i.e.*, coils of the snake's body are thrown around it and gradually tightened. The prey is worked around until its head is in the snake's mouth and swallowing begins.

Due to the loose articulations of the bones supporting the jaws and the elastic ligament between the halves of the lower jaw, the potential gape of the mouth is enormous. The hooked teeth permit food to be pushed easily into the mouth but not easily withdrawn. While the teeth of one side of the mouth are engaged and act as a fulcrum, the jaws of the opposite side are pushed forward for a short distance; then their teeth in turn act as a fulcrum, and so on. During this process teeth are frequently broken off, but there is a series of replacement teeth in different stages of development, and a lost tooth is quickly replaced from this successional series; even when no injury occurs, the functional teeth are periodically replaced, alternate teeth at a time. Venomous snakes usually release their victim immediately after injecting venom with the first bite. The prey is then followed either by sight or by scent, and is eaten only when the poison has taken effect.

The majority of snakes have restricted food preferences, which may involve some physiological or morphological adaptations. For instance, the American king snakes, though themselves nonvenomous, attack and eat venomous snakes with impunity. The African *Lycophidion* preys largely on hard-scaled lizards (skinks) and has enlarged caninelike teeth on both jaws. Some snakes (Dasypeltinae) subsist almost exclusively on birds' eggs and the thirst snakes (Dipsadinae) eat snails. The egg-eating snakes have a reduced, almost functionless, dentition but a mouth so distensible that a large egg can be swallowed whole. Their cervical vertebrae, however, have enlarged lower processes (hypapophyses), some of which perforate the esophagus to form a median row of toothlike structures on which the egg is crushed; the contents of the egg pass on into the stomach but the shell is regurgitated. The Dipsadinae deal with their hard-shelled prey differently. Here, although the upper jaw teeth are reduced in number, the lower jaw, with enlarged teeth in front, can be swung forward a long way; it is thrust into the shell, its teeth engaged in the soft tissues, and then retracted, drawing the snail out.

Life History.—Snakes are essentially nongregarious, though congregations of individuals may be attracted by favourable environmental conditions. The burrows of other animals may be attractive for shelter (*e.g.*, for hibernation) or for the food they offer (*e.g.*, termite galleries); or many females may gather at the same spot for egg laying. There are, however, no known instances of social organization or community life. In consequence, except

where mass hibernation is the rule (*e.g.*, among rattlesnakes) and mating occurs with the spring awakening, the meeting of the sexes for reproduction may not be easy. Females are probably located by scent, which in some instances arises from an epidermal secretion but in others from the secretion of a pair of saclike glands that open just inside the cloaca. There is evidence that certain snakes (*e.g.*, *Coluber*, *Naja*) mate with only one partner in a season, and in others the female is able to store viable sperm, sometimes for as long as five years (*Leptoderia annulata*), as a safeguard against failure to find a mate.

Courtship usually consists of pushing and weaving movements initiated by the male, during which the chin is rubbed against the female's neck. In a few instances, for example, in the European Aesculapian snake, there is a nuptial dance that may last for more than an hour; in this species there is mutual pursuit, the two snakes coiling together, each raising the head and neck and the two forming a lyre-shaped figure. Behaviour similar to this occurs in many other snakes (adders, rattlesnakes, cobras, etc.), but in these instances the participants are both males and the "dance" appears to be a ritual displav of aggression.

Fertilization is internal. The male terminates courtship by inserting one of a pair of intromittent organs, collectively called hemipenes, into the cloaca of the female; sperm passes along a groove on the hemipenis, enters the cloaca, and travels up the oviduct where it fertilizes the eggs. At rest, each hemipenis lies in a fold of the skin at the base of the tail, near the cloaca. Diminutive hemipenes of the female are of doubtful function. (*See* also REPTILE: *Urogenital System* and *Reproduction and Development*.)

Most snakes lay elongate eggs with a parchmentlike shell; they are usually buried in warm, damp soil. In some groups ovoviviparity, or hatching of eggs in the female's body, is the rule; in a few instances there is true viviparity, the developing young being nourished through a placenta. The number of eggs or young varies from one or two to more than a hundred, and the incubation or gestation period depends upon the prevailing temperature. In some ovoviviparous species in cold climates (*e.g.*, the adder), the young may not be born until the summer following that of conception. Few snakes exhibit parental care. The king cobra, or hamadryad, scoops together a pile of detritus in which the eggs are laid and the female coils herself upon it. Some other cobras, and also some kraits and pythons, guard their eggs. Pythons coil around the clutch; it has been claimed that the python can raise its body temperature to assist incubation. Growth is greatest in the earliest stages but probably never ceases entirely. Associated with growth is sloughing, whereby the skin is cast off and renewed periodically, every one to three months on the average; beginning at the lips, the old skin is turned back on itself, inside out, in one piece.

Although a few snakes have hornlike processes on the snout or above the eyes, these do not appear to be weapons. The

FRITZ GORO "LIFE," © 1942, TIME, INC

FIG. 3.—COLUBRINE SNAKE LAYING EGGS

ISABELLE HUNT CONANT

FIG. 4.—SHEDDING OF SKIN BY RED-BELLIED WATER SNAKE (NATRIX ERYTHROGASTER ERYTHROGASTER)

teeth and the venom apparatus are primarily offensive, in connection with the acquisition of food, and only secondarily defensive, except in the cases of the black-necked and spitting cobras; these cobras have the ability to eject a highly irritating venom into the eyes of an enemy. Other defensive devices include pungent glandular secretions. Despite all these methods, defense is achieved, in the main, by behaviour. Explosive hissing and threats of biting are common practices, often accompanied by intimidatory posturing such as inflation of the neck or the spreading of a "hood," as in some cobras. Shamming dead and rolling into a compact ball with the head concealed in the middle are also common defense measures. The North American rubber boa (*Charina*) rolls into a ball and also directs attention to a nonvital part by waving its tail in the air. Many other snakes divert attention even more effectively; the tail is often distinctively coloured and makes stabbing movements like a striking head. A few snakes, like the desert-dwelling vipers *Aspis* and *Echis*, give audible warning by rubbing together certain scales on their flanks or, like the rattlesnakes, by means of a rattle at the tip of the tail.

Venom.—Venom is secreted by venom sacs, situated laterally behind the eyes and connected by ducts to the base of the fangs. The venom contains certain poisonous proteins present in varying amounts according to the species of snake. The poisons belong to two principal types: histolytic substances that destroy the body tissues, including blood cells and the lining of blood vessels, and neurotoxins, which attack the nervous system and cause paralysis, especially of the respiratory apparatus. When neurotoxins predominate, paralysis, general prostration, and difficulty in breathing are the most dangerous effects; if the victim survives this stage, recovery is rapid and there are few serious local symptoms at the bite. When the other substances predominate, there is no paralysis, but there are severe constitutional symptoms, with vomiting and prostration, local and general hemorrhage, and much painful swelling of the bitten parts, which may become gangrenous.

Since the poison apparatus is primitively and primarily a device for killing food, the toxicity of the venom is related to the normal prey; for example, the venom of the crab-eating, oriental *Fordonia* is highly toxic to crustaceans but not to fish or amphibians, and that of the fish-eating sea snakes is toxic to fish but much less so to warm-blooded animals. In the aglyphous, "harmless" Colubridae, which lack any venom canals on their teeth, the saliva may be toxic, at least to the animals on which they prey.

First aid in cases of snakebite on the extremities consists of the immediate application of a light ligature, on the side of the wound closest to the body, and suction of the wound. Exertion and excitement should be avoided to prevent an increase of the pulse rate; for the same reason, stimulants should be avoided. The application to the wound of substances such as potassium permanganate is likely to be harmful rather than helpful.

Antivenins, manufactured in many countries, are prepared by the immunization of animals against particular snake venoms. The immunized serum is effective only against the venom with which it was prepared. To ensure treatment with an appropriate antivenin, it is therefore important that the snake be identified. It is highly desirable that antivenins be administered under medical supervision since there may be allergic reactions. In cases where the main toxin destroys blood cells, blood transfusions may be indicated. (*See* also VENOM.)

Classification and Survey.—Snakes number about 3,000 extant species. The generally recognized families may be grouped as follows, under a conservative taxonomic system.

Typhlopidae (*Blind Snakes*).—Small, harmless, burrowing snakes with bluntly rounded heads and tails, without enlarged ventral scutes and usually with vestiges of pelvic bones; teeth are present on the upper jaw, but absent on the lower; eyes are minute. Blind snakes occur in most tropical countries. Their food consists of small invertebrates.

Anomalepidae.—A small family of South and Central American snakes that resembles the blind snakes in appearance and habits but lacks any vestiges of hind limbs or pelvis. They have teeth in the upper jaw and a single, small tooth in the lower jaw. They are unique among snakes in possessing, in common with lizards, a jugal bone bordering the orbit of the eye.

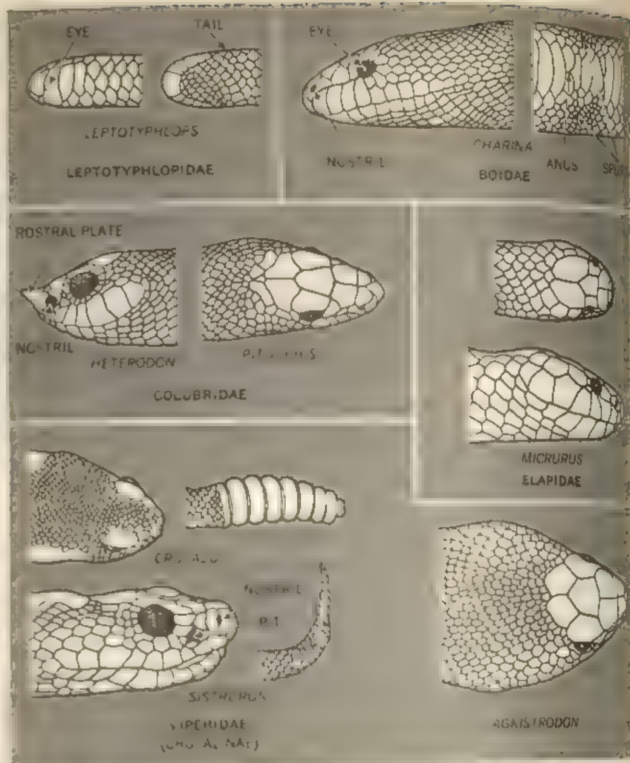
Leptotyphlopidae (*Thread Snakes*).—Although superficially similar to the preceding, thread snakes have teeth only on the lower jaw. They inhabit Africa, southwestern Asia, and America. Most species are uniformly brown or black, but the commonest species in tropical America, *Leptotyphlops albifrons*, has white patches on the head and tip of the tail, a pattern which enhances the similarity of these extremities to the possible confusion of an aggressor. Some species have the unique habit of sucking out the contents of termites' abdomens.

Aniliidae.—This family of harmless burrowing snakes of South America and southeastern Asia is characterized by teeth on both upper and lower jaws, clawlike vestiges of hind limbs, and slightly enlarged ventral scales. They are larger than most burrowing snakes, reaching two and one-half to three feet. Their food consists principally of other snakes and eels. The South American *Anilius scytale* resembles some coral snakes in its colour pattern of bright coral red with black rings. The Indo-Malayan *Cylindrophis rufus* is usually crossbanded in brown or black and white, but the underside of the tail is vermilion; if the snake is disturbed it raises the tail aloft, showing the bright colour, and makes darting movements with it. The same tail colour and reactions are characteristic of the venomous elapid snake *Maticora intestinalis*, which inhabits the same region.

Uropeltidae (*Shield Tails*).—These harmless burrowing snakes, similar to the preceding but without vestiges of hind limbs, have an enlarged spiny or rugose shield on the tip of the tail. Inhabiting Ceylon and southern India, they burrow in damp, forest earth, and feed mainly on earthworms.

Xenopeltidae (*Sunbeam Snakes*).—The unique species, *Xenopeltis unicolor*, of southeastern Asia is a harmless, handsome snake, black or chocolate brown in colour with highly iridescent scales; there are no vestiges of limbs, but a left lung half the size of the right is present. It is cryptozoic and feeds on other snakes, small rodents, and frogs.

Boidae (*Boas, Pythons, and Wood Snakes*).—This group of nonvenomous snakes, with a lineage extending back to at least Cretaceous times, is characterized by the retention of vestiges of hind limbs and, usually, of two lungs; teeth are present on both upper and lower jaws; and there are enlarged ventral scales. Four subfamilies may be recognized, the two largest being the circumtropical boas (Boinae) and the Old World pythons (Pythoninae). The boas mostly bear living young. The larger species are nearly all arboreal. The biggest of them, the South American anaconda, is arboreal and aquatic. The smallest boas, seldom exceeding three feet, are the sand boas (*Eryx*) of the semi-desert regions of northern Africa and southwestern Asia. The egg-laying pythons inhabit the tropics from western Africa to northern Australia, Melanesia, and Polynesia. The largest species are the Oriental reticulate python and the African rock python, both of which may exceed 30 ft. (9 m.) in length. *Loxocemus*, a Central American form, was at one time thought to be a python (the only one in the New World); but it has proved to be related in several points to the sunbeam snakes and is allotted an isolated position in the family as the sole member of the subfamily Loxoceminae. The Central American and Antillean wood snakes, at one time regarded as boas, are now known to have only one lung and to have a few unique features; they are placed in another subfamily, the Tropidophiinae. (*See* also BOA; PYTHON.)



FROM A. H. WRIGHT AND A. A. WRIGHT, "HANDBOOK OF SNAKES OF THE UNITED STATES AND CANADA," P. 33. © 1957 BY CORNELL UNIVERSITY. USED BY PERMISSION OF CORNELL UNIVERSITY PRESS

FIG. 5.—CERTAIN ANATOMICAL CHARACTERISTICS OF VARIOUS FAMILIES OF SNAKES

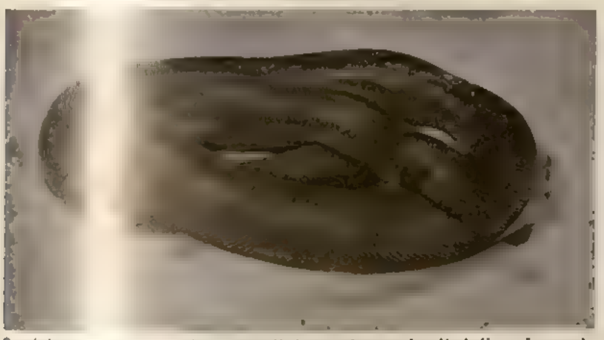
Bolyeridae.—This family of two genera, *Bolyeria* and *Casarea*, each with one species, occurs only on a tiny island, Round Island in the Indian Ocean; these and extinct remains of a third species found on nearby Mauritius, have some boalike features. But they have no vestiges of hind limbs or pelvis, their left lung is almost vestigial, they have some peculiarities of the vertebral column and the unique feature of a transversely divided upper jawbone. They may represent a little group that has been isolated since the Indian Ocean islands were formed, probably in Eocene times.

Acrochordidae (*Wart Snakes*).—This family contains only a single genus with two species. These are harmless aquatic snakes that live in the fresh waters and estuaries of the Indo-Australian region. They resemble the harmless members of the Colubridae (below) in lacking vestiges of limbs and in having one lung; but they have no enlarged ventral plates and are peculiar in having a very loose baggy skin with small, wartlike granules instead of overlapping scales. Their eyes are tiny, and the air passage from the nostrils can be closed to exclude water by a flap on the roof of the mouth. Sluggish in disposition and almost helpless if removed from the water, they have a diet consisting exclusively of fish. They produce large litters of young, up to more than 30.

Colubridae.—This family contains the great majority of typical snakes in which there is neither a left lung nor vestiges of limbs; generally the ventral scales are enlarged and the head is covered with relatively few enlarged plates. Teeth are present on both jaws, and the dentition is aglyphous or opisthoglyphous, those with the former condition being harmless to man. The opisthoglyphs are technically venomous, but the small size and position of the

Reticulated pyth… *…hon reticulatus*), one of the largest of all snakes, may reach a length of more than 30 ft. Found in s… …east Asia and the Philippines

Diamond snake or carpet python (*Morelia argus*) of Australia and New Guinea



Sand boa (… *…yx johni*), a small burrowing snake that lives in sandy regions fro… …orth and central Africa to India and central Asia

White Indian python (*Python molurus*), a rare colour freak, not albino, caused by domination of white pigment cells over those of normal coloration

Boa constrictor (*Constrictor constrictor*), widely distributed in South America and found north, along the coasts of Mexico, almost to the U.S. border

Emerald boa (*Boa canina*), an arboreal snake found in the tropical jungles of South America

PYTHONS AND BOAS

PHOTOGRAPHS, (TOP LEFT, BOTTOM LEFT) JOHN H. GERARD, (TOP RIGHT, CENTRE RIGHT) CY LA TOUR AND (TOP RIGHT) BROOKFIELD ZOO, (CENTRE LEFT) NATHAN W. COHEN, (BOTTOM RIGHT) EDWARD VAN BAERLE

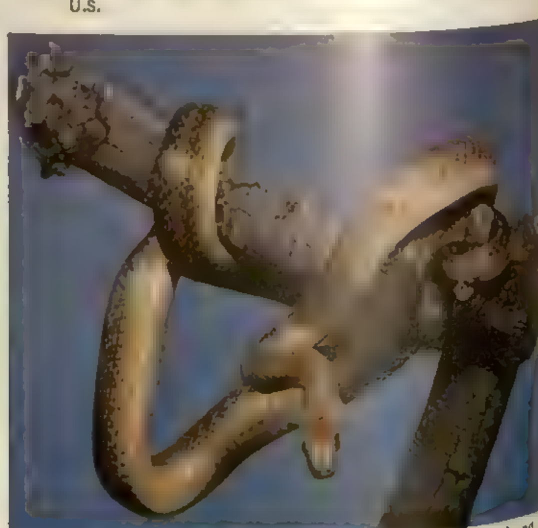
Eastern hognose snakes (*Heterodon platyrhinos*), showing two of the several colour variations that are characteristic of this species. Found in sandy areas of most of the eastern and central U.S.

Yellow rat, or chicken, snake (*E[illegible]he obsoleta quadrivittata*), an arboreal species of the [illegible]th Atlantic coastal U.S.

Texas rat snake (*Elaphe obsoleta lindheimeri*), found in Texas and western Louisiana. Like most other *obsoleta* varieties, it is a large snake and may reach a length of 7 ft.

California king snake (*Lamp[illegible] getulus californiae*) one of many *Lampropeltis* spec[illegible] dely distributed in the U.S.

Red milk snake (*Lampropeltis doliata syspila*). Banded pattern of the harmless milk snakes, or king snakes, is similar to that of some of the venomous coral snakes (Elapidae). Central U.S.



Aesculapian snake (*Elaphe longissima*), a chicken snake found in Europe and Asia Minor

KING, RAT AND HOGNOSE SNAKES

PHOTOGRAPHS, (TOP LEFT, TOP RIGHT, CENTRE LEFT, BOTTOM LEFT) JOHN H. GERARD, (CENTRE RIGHT) WILLIS PETERSON, (BOTTOM RIGHT) JOHN MARKHAM

Eastern yellow-be racer (*Coluber constrictor flaviventris*), an extremely fast-moving, smooth snake found chie Middle West, north of the Ohio valley to the Great Lakes

Rough green snake (*Opheodrys aestivus*), a long, slim arboreal species seen in south central and Atlantic coastal states

Texas long-nosed (*Rhinocheilus lecontei tessellatus*), a speckled, burrowing snake found in arid regions of the reme southwest and northern Mexico

Western worm snake (*Carphophis amoenus vermis*), a small terrestrial snake usually found hidden beneath moist foliage, rocks or ogs

Sonora gopher snake (*Pituophis catenifer affinis*), a large western bull snake known for the loud hissing it makes when attacked or frightened

Western ribbon snake (*Thamnophis sauritus proximus*), a slender garter snake found in most of the middle western states west of the Mississippi river

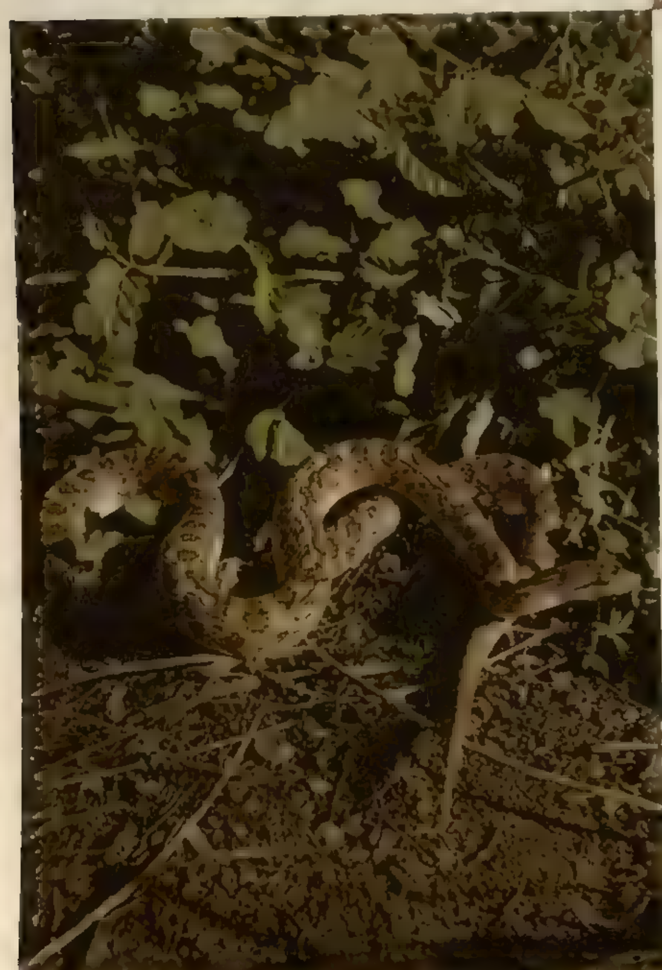
Eastern garter snake (*Thamnophis sirtalis sirtalis*), a very common snake of the entire eastern half of the U.S. and southeastern Canada

COLUBRIDS: HARMLESS SNAKES OF THE U.S.

PHOTOGRAPHS, JOHN H. GERARD EXCEPT (CENTRE RIGHT) CY LA TOUR

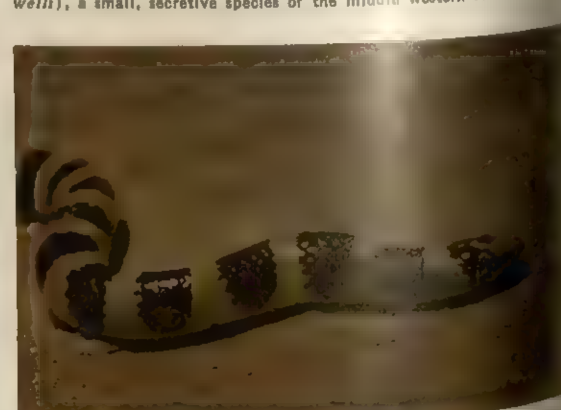
Sonora king snake (*Lampropeltis getulus splendida*), a largely nocturnal species found in the southwest and north central Mexico. It constricts its prey, often other snakes

Speckled king snake (*Lampropeltis getulus holbrooki*), another of the snake-eating snakes. Found in the prairie and south central states



Glossy snake (*Arizona elegans*), a burrowing snake usually found in sandy areas of the southwest and northern Mexico

Northern black-headed, or flat-headed, snake (*Tantilla gracilis ha[illegible]welli*), a small, secretive species of the middle western states

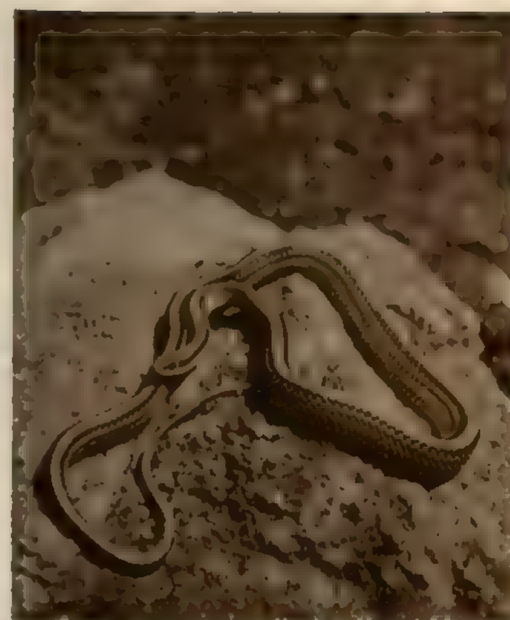


Banded sand snake (*Chilomeniscus cinctus*), a small snake whose [illegible] is adapted for rapid burrowing into sand. Found chiefly in Arizona and Baja California

COLUBRID SNAKES OF THE U.S.

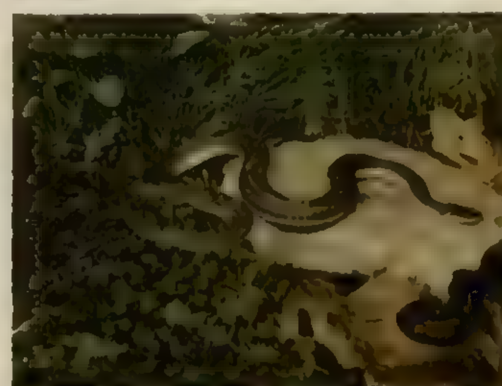
PHOTOGRAPHS, (TOP LEFT, CENTRE RIGHT) JOHN H. GERARD, (TOP RIGHT) JOHN MARKHAM, (BOTTOM LEFT) WILLIS PETERSON, (BOTTOM RIGHT) HYMEN MARX

Western ringneck snake (*Diadophis amabilis*) in its defense posture, coiled with the underside of its tail exposed. Found chiefly on the Pacific coast

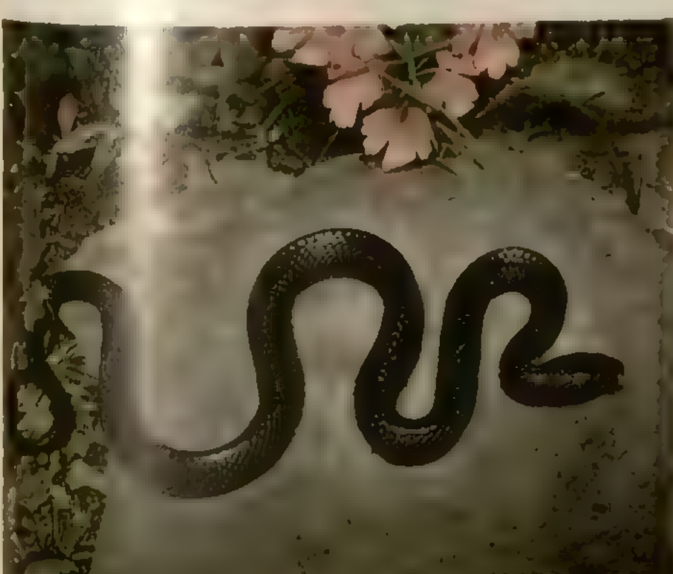


Texas patch-nosed snake (*Salvadora lineata*), a slender species found most often in rocky terrain in northern Mexico, Texas and Oklahoma

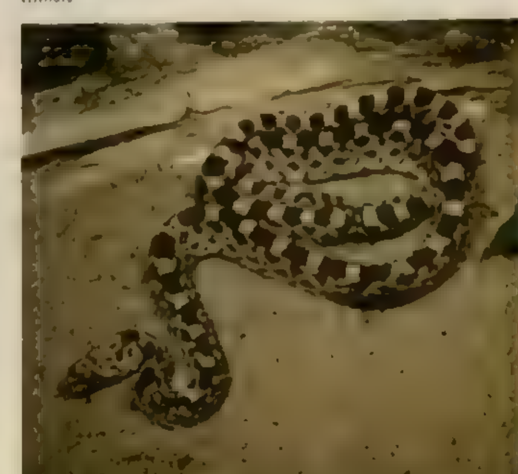
Pacific gopher snake (*Pituophis catenifer catenifer*), a constricting bull snake found on the west coast



Graham's water snake (*Natrix grahami*), found in streams and quiet water west of the Mississippi river to Texas and north to Illinois



Eastern indigo snake (*Drymarchon corais couperi*), a relative of the racers, may reach a length of about 8 ft. Found chiefly in Florida, Georgia and Alabama



Bull snake (*Pituophis melanoleucus sayi*), a large prairie dwelling species that will strike when alarmed but is not dangerous to man

HARMLESS SNAKES OF THE U.S.

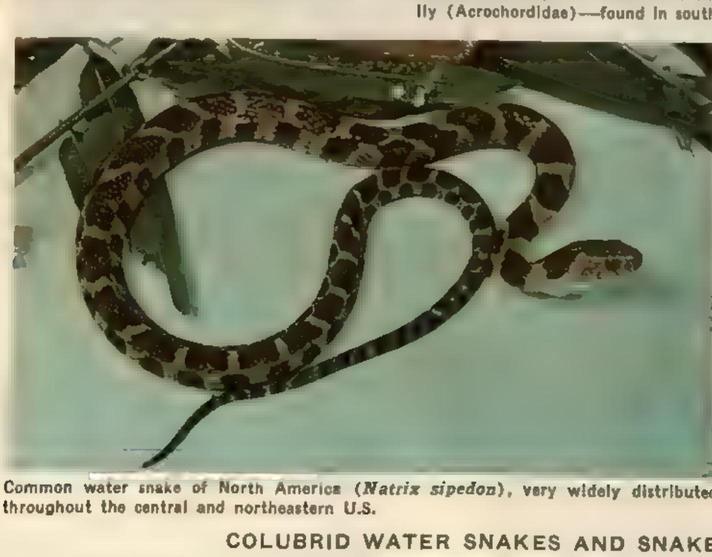
PHOTOGRAPHS, (TOP LEFT) NATHAN W. COHEN, (TOP RIGHT) WILLIS PETERSON, (OTHERS) JOHN H. GERARD

Mangrove snake (*Boiga dendrophilia*), a rear-fanged arboreal species found in tropical Asia

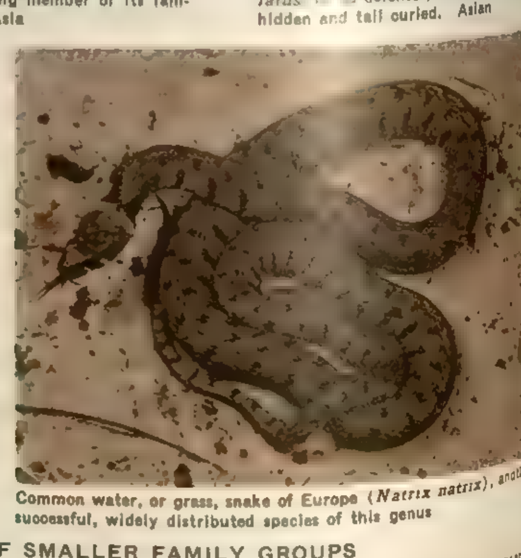
Blunt-headed tree snake (*Imantodes cenchoa*), native to Central and South America

Elephant trunk, or wart, snake (*Acrochordus javanicus*), a loose-skinned aquatic snake—the only living member of its family (Acrochordidae)—found in southeast Asia

An ar[illegible]ake, *Cylindrophis maculatus* [illegible] defense position w th head hidden and tail curled. Asian



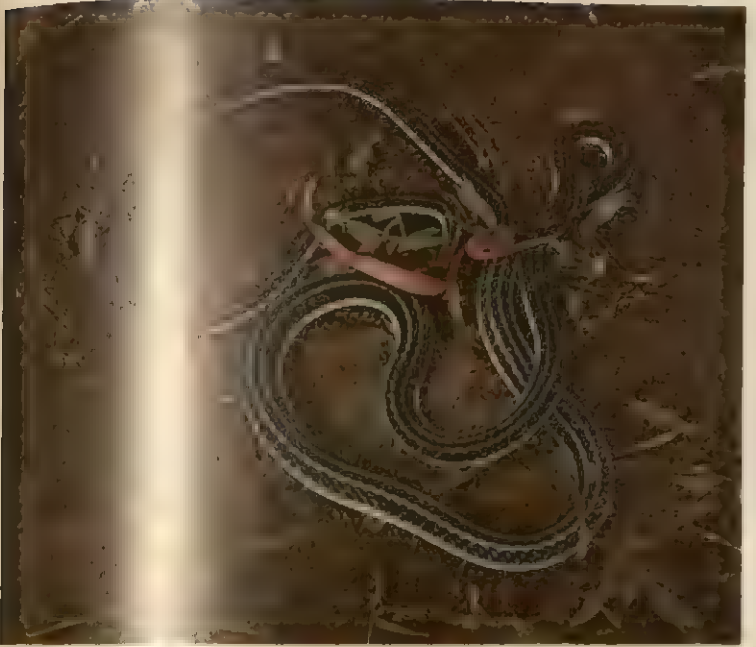
Common water snake of North America (*Natrix sipedon*), very widely distributed throughout the central and northeastern U.S.



Common water, or grass, snake of Europe (*Natrix natrix*), another successful, widely distributed species of this genus

COLUBRID WATER SNAKES AND SNAKES OF SMALLER FAMILY GROUPS

PHOTOGRAPHS, (TOP, BOTTOM RIGHT) JOHN MARKHAM, (CENTRE LEFT) CY LA TOUR, (CENTRE) HYMEN MARX, (CENTRE RIGHT) G. B. RABB—BROOKFIELD ZOO, (BOTTOM LEFT) JOHN H GERARD



East Indian coral s[illegible]acora (formerly *Doliophis*) *bivirgata*, shown exposing the vivid underside of its tail, a defensive maneuver

Indian cobra (*Naja naja*) in characteristic pose, with its neck muscles (hood) extended (front view). Found in Asia and Africa

Head of a black m[illegible]a (*Dendroaspis polylepis*), a large, aggressive and extremely venomous [illegible]frican elapid

Vermicella calonota, a small, weakly venomous cobra (harmless to man) of western Australia

Krait (*Bungarus fasciatus*), found in India and southeast Asia. Large and strongly venomous

Eastern coral snake (*Micrurus fulvius*), found in the southeastern section of the U.S.

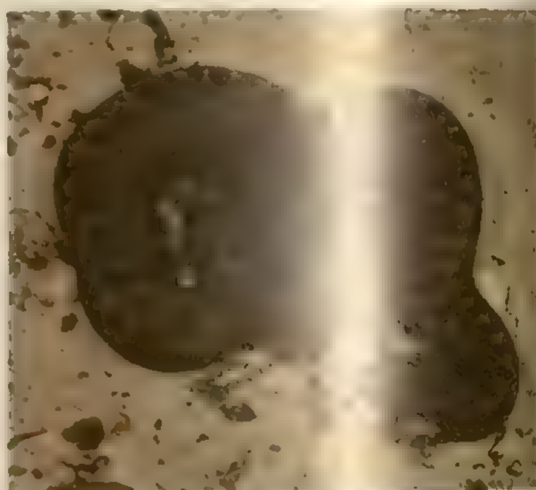
VENOMOUS SNAKES: COBRAS, CORALS AND OTHER ELAPIDS

PHOTOGRAPHS, (TOP LEFT) D. DWIGHT DAVIS, (TOP RIGHT, BOTTOM LEFT) CY LA TOUR AND (BOTTOM LEFT) DETROIT ZOO, (CENTRE LEFT) JOHN MARKHAM, (CENTRE RIGHT) R. MERTENS, (BOTTOM RIGHT) ALLEN D. CRUICKSHANK FROM THE NATIONAL AUDUBON SOCIETY

SNAKE

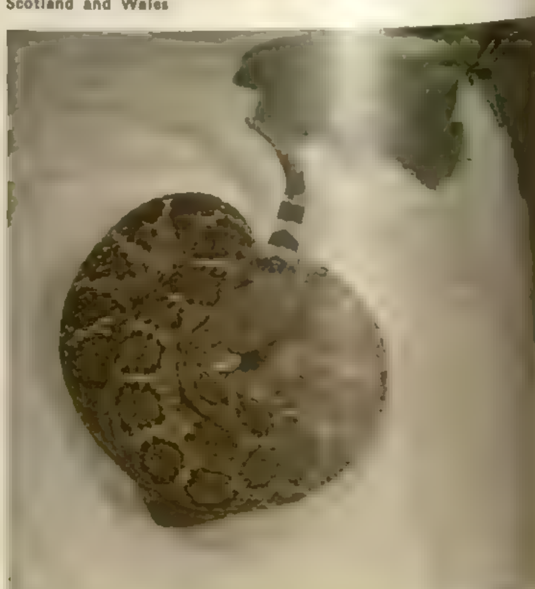
Gaboon viper (*Bitis gabonica*), strikingly coloured, is one of the largest of true vipers, with a length up to 6 ft. and a diameter of 6 in. It is unaggressive but extremely venomous. Found in Africa

Water moccasin or cottonmouth (*Agkistrodon* [or *Ancistrodon*] *piscivorous*), a dangerous aquatic pit viper of the southeastern U.S.



Common viper (*Vipera berus*), known as the adder in England, is the most widely distributed poisonous snake of Europe, England and Scotland and Wales

Copperhead (*Agkistrodon* [or *Ancistrodon*] *contortrix*), one of four varieties that are found in most of the eastern, southern and southwestern U.S. Its bite is rarely fatal to man



Western diamondback rattlesnake (*Crotalus atrox*) showing the distinctive tail of the rattlesnakes. The western diamondback is one of the largest and most dangerous of its type

VENOMOUS SNAKES: VIPERS AND PIT VIPERS

PHOTOGRAPHS, (TOP) CY LA TOUR AND THE ST. LOUIS ZOO, (CENTRE LEFT, BOTTOM LEFT, BOTTOM RIGHT) JOHN H. GERARD, (CENTRE RIGHT) JOHN MARKHAM

fangs make it difficult for them to inflict a dangerous bite on large objects. The larger species, however, can be dangerous, especially the African boomslang, whose venom is hemolytic, like that of the vipers. Several subfamilies can be recognized, though their limits and interrelationships are far from clear.

The Dipsadinae are aglyphous arboreal forms with a specialized diet of mollusks. These snakes lack the groove beneath the chin, present in most other colubrids and a part of the mechanism permitting the halves of the lower jaw to be widely separated. The group is confined to southeastern Asia (*Amblycephalus, Haplopeltura*) and to tropical America (*Dipsas* and allies).

The Homalopsinae, aquatic snakes from the fresh or saline waters of the same region as the Acrochordidae, differ from the latter in being opisthoglyphous, in closing the nasal passages by valvular nostrils, and in having larger imbricating scales of which those on the belly are slightly enlarged.

The Dasypeltinae comprise the egg-eating snakes of Africa (*Dasypeltis*) and northern India (*Elachistodon*), terrestrial snakes with enlarged ventral plates. *Elachistodon* is opisthoglyphous and *Dasypeltis*, aglyphous.

The Colubrinae are snakes with either aglyphous or opisthoglyphous dentition. This large subfamily (often given family rank as the Colubridae) is cosmopolitan in distribution, though poorly represented in Australia. There are many subaquatic forms, generally subsisting on fish and frogs; among them is the cosmopolitan genus *Natrix*, with such species as the European grass snake (*N. natrix*), the Indian keel back (*N. piscator*), and the North American common water snake (*N. sipedon*). Terrestrial forms prey mostly on small mammals, lizards, toads, and birds. Among the better-known forms are the chicken snakes, racers, and rat snakes (*Elaphe* and *Coluber*) of Europe and North America; the Indian rat snakes (*Ptyas*); the African house snakes (*Boaedon*); and the European smooth snake (*Coronella*), the rarest of the three indigenous snakes of Britain. The ophiophagous habits of American king snakes have been mentioned, but this genus, *Lampropeltis*, is noteworthy for another reason: some species, with a pattern of broad red bands edged with black and separated by yellow or buff interspaces, mimic the venomous coral snake.

Arboreal forms, many of which are opisthoglyphous, prey mostly on birds and lizards. Often they are greatly elongate and slender; many are protectively coloured green and brown. Some, such as the Indian green whipsnake (*Dryophis nasutus*) and the Malagasy *Langaha*, have leaflike dermal appendages on the tips of their snouts. Other well-known arboreal species are the dangerous boomslang (*Dispholidus typus*) and the Indian flying snake (*Chrysopelea ornata*). The latter is a handsome black, yellow-spotted snake, with a series of markings like red-centred, yellow flowers along its back; it can slip from a high branch, with its ribs pushed outward to produce a flat or concave lower surface, and execute a steep glide to earth.

Terrestrial genera are represented in Europe by *Malpolon*, the largest snake of that continent. In tropical America is the polymorphic *Erythrolamprus aesculapii*, some of whose many colour varieties mimic coral snakes. In Africa are numerous forms including the ubiquitous sand snakes (*Psammophis*). Some burrowing forms, like *Apostolepis* in South America and *Chilorhinophis* in eastern Africa, exhibit deflective coloration; the head and stumpy tail are alike in shape and colour, both being black with lighter dots.

Elapidae.—This family comprises snakes with fixed poison fangs at the front of the upper jaw, with enlarged ventral plates, and the tail not flattened. All elapids are venomous, the larger ones dangerously so; the venom is usually neurotoxic. They include the cobras (*Naja*) of southwest Asia and Africa; king cobra (*Ophiophagus*), the biggest venomous snake; spitting cobras (*Hemachatus*) and mambas (*Dendroaspis*) of Africa; Asiatic kraits (*Bungarus*); American coral snakes (*Micrurus* and allies); and several Australian genera. The Australian elapids constitute three-quarters of all the terrestrial snakes on that continent, the more important being the taipan (*Oxyuranus*), black snake (*Pseudechis*), tiger snake (*Notechis*), Australian copperhead (*Denisonia*), and death adder (*Acanthophis*). (*See* also COBRA.)

Hydrophidae (*Sea Snakes*).—This is a family of marine snakes that resemble the Elapidae in dentition and most other characters, but lack obvious ventral shields and have vertically flattened, paddlelike tails. Other modifications associated with a completely aquatic life are valvular nostrils and a reduction in the length of the tongue. In connection with their saltwater habitat, sea snakes have nasal glands believed to excrete excess salt. The majority of the species inhabit the littoral seas from the Persian Gulf to southeastern Japan southward to northern Australia and eastward to Samoa. Only one species, *Pelamis platurus*, is truly pelagic, ranging from coast to coast across the Indian and Pacific oceans. Although a few species may attain a length of about 8 ft. (2 m.), the majority are only 3 or 4 ft. (0.9 to 1 m.) long. The usual colours are browns or grays, with blackish crossbars, but *Pelamis* has a uniform brown or black back sharply delimited from a bright yellow belly.

There are two subfamilies, the Hydrophinae, which bear living young and are completely independent of the land, and the Laticaudinae, which lay eggs above the high-water mark.

Viperidae (*Vipers and Pit Vipers*).—These snakes have their fangs set on a rotatable bone (maxilla) that bears no other teeth. There are well-developed, enlarged ventral shields, but the head normally lacks the enlarged plates of the Colubridae and some other families. The venom is usually strongly histolytic.

Viperids are absent from the Papuan, Australian, and Polynesian areas but are widespread elsewhere. Two subfamilies may be recognized; the pit vipers (Crotalinae), from the Americas and across Asia to eastern Europe, and the true vipers (Viperinae), from Europe, Africa, and Asia. Both groups are often given family rank, the pit vipers then being separated as the Crotalidae.

Among the better-known viperine genera are *Vipera*, *Bitis*, *Atheris*, and *Cerastes*. In the first mentioned are such species as the adder (*V. berus*) and asp viper (*V. aspis*) of Europe, and the Indo-Malayan Russell's viper (*V. russelli*). The larger species of the African *Bitis* are the puff adder (*B. lachesis*), the Gaboon viper (*B. gabonica*), and the rhinoceros viper (*B. nasicornis*). *Atheris* is an arboreal African genus, often coloured a velvety emerald green, with a prehensile tail.

Cerastes includes the desert horned asp of Egypt and neighbouring countries, which has a thornlike spine above each eye. The crotalines include the moccasins and copperheads of North America and related species in Asia (*Agkistrodon*); the fer-de-lance group (*Trimeresurus* or *Bothrops*), also American and Asiatic; the notorious bushmaster (*Lachesis*); and many species of rattlesnakes (*Crotalus, Sistrurus*) in the Americas. (*See* also RATTLESNAKE; VIPER.)

See REPTILE; *see* also references under "Snake" in the Index.

BIBLIOGRAPHY.—V. Fitzsimons, *The Snakes of Southern Africa* (1962); J. R. Kinghorn, *The Snakes of Australia*, 2nd ed. (1957); L. M. Klauber, *Rattlesnakes: Their Habits, Life Histories and Influence on Mankind*, 2 vol. (1956); K. Klemmer *et al.*, *The Poisonous Snakes of the World, Activities and Antigenicity of Venoms, Treatment of Snakebite*, in German (1963); H. W. Parker, *Snakes* (1963), *Natural History of Snakes* (1965); C. H. Pope, *Snakes Alive, and How They Live* (1937), *The Giant Snakes* (1961); W. Rose, *Snakes: Mainly South African* (1955); M. A. Smith, *Fauna of British India, Reptilia and Amphibia, III, Serpentes* (1943); M. W. F. Tweedie, *The Snakes of Malaya* (1953); A. H. Wright and A. A. Wright, *Handbook of Snakes of the United States and Canada* (1957); M. Ludicke, "Serpentes," *Handbuch der Zoologie*, 7, fasc. 5 and 6, pp. 1–298 (1962, 1964). (H. W. P.)

SNAKEBIRD (DARTER OR ANHINGA), a bird belonging to the same order as the cormorant (*q.v.*), which it resembles in outward form and in habits. The snakebirds, however, are slenderer and have an elongated neck and tail. The pointed bill has its edges beset with backwardly directed "teeth." There is only one genus of snakebirds, *Anhinga*, with four species occurring (one each) in the Americas, Africa, southern Asia, and Australia. The male of the American species, *A. anhinga*, often called the water turkey, has black plumage, glossed with green, with a white line on the neck; the bare skin of the head is both green and orange, and there are white stripes and patches on the wings and tail. The bird haunts large rivers and lakes, frequently perching on overhanging branches. It feeds on fish, which it chases and spears

with its daggerlike bill; its agility under water is amazing. It often swims with only its head above water. The bird has been observed to play at catch with small twigs. The nest, built in a tree close to the water, is a large structure of sticks. The four eggs have white, chalky shells.

All four species are characterized by their serpentine necks; the vertebrae articulate in such a way to permit great flexibility and rapid forward thrust.

A. W. AMBLER FROM NATIONAL AUDUBON SOCIETY

AMERICAN SNAKEBIRD (ANHINGA ANHINGA), SHOWING CHARACTERISTIC WING PATCHES OF THE MALE

SNAKE RIVER, one of the most important streams in the Pacific northwest section of the United States, is the largest tributary of the Columbia River. The drainage basin of the Snake (109,000 sq.mi. [282,000 sq.km.]) is 30% greater than the total area of Idaho and contains one-half the entire area of the Columbia River Basin in the United States. Runoff from Wyoming, Utah, Nevada, Idaho, Oregon, and Washington combines in this stream. From elevations of 10,000 ft. (3,048 m.) the river descends to an elevation of 300 ft. (91 m.) and contributes 36,820,000 ac-ft. (45,417,470,000 cu.m.) of water annually to the Columbia.

The river rises in high, rugged mountains of the continental divide near the southeast corner of Yellowstone National Park in Wyoming. It flows south along the eastern base of the Teton Mountains, and swings northwest near the mouth of Greys River and enters Idaho. Near Heise, the river leaves the mountains and crosses southern Idaho in a huge, southwesterly curving arc that terminates near the junction of the Boise and Snake rivers on the western edge of Idaho. Turning north, it forms the Oregon-Idaho boundary for 216 mi. (348 km.). From the northeast corner of Oregon it forms the Washington-Idaho boundary to Lewiston, Ida., and then turns west to join the Columbia near Pasco, Wash. Its total length is 1,038 mi. (1,670 km.).

The upper Snake River, above King Hill, Ida., is used to irrigate over 1,500,000 ac. (607,050 ha.) of land and to generate electric energy. The main stem is regulated by Jackson Lake, Palisades, American Falls, Minidoka, and Milner reservoirs. Principal tributaries below Heise are Henrys Fork (the largest), Blackfoot, Portneuf, Raft, Little Wood, and Big Wood rivers. The last two and Henrys Fork enter the river from the north. Other north side streams sink into Snake River plains and become part of an immense underground reservoir. Numerous large springs—outpourings of the underground water body—emerge from the canyon walls in Hagerman Valley. Twin Falls and Shoshone Falls, with drops of 65 and 212 ft. (20 and 65 m.), respectively, are located downstream from Milner Dam.

The central Snake River, from King Hill to Weiser, Ida., is used primarily for hydroelectric generation. Over 1,000,000 ac. (404,700 ha.) of land are irrigated in this section, primarily from tributary streams. Principal reservoirs and their associated rivers are Lucky Peak, Arrowrock, and Anderson Ranch on the Boise; Cascade, Deadwood, and Black Canyon on the Payette; Owyhee on the Owyhee; and Warm Springs and Agency on the Malheur.

The lower Snake River, from Weiser to the mouth, flows through a 1-mi. (2-km.)-deep gorge known as Hells Canyon, one of the deepest river gorges in North America. Salmon River, largest tributary and most important wildlife area of the river system, joins the main stem near the downstream end of the canyon section. (G. V. Sk.)

SNAKEROOT. In most countries where snakes abound some root or herb is used by the natives as an antidote for the bites of venomous species, and many herbs have consequently received the name of snakeroot. Botanically speaking, the name properly belongs to *Ophiorrhiza mungos*, a plant of the family Rubiaceae, used in Indonesia for the purpose above indicated. In medicine, however, the roots of *Aristolochia serpentaria*, *Polygala senega*, and *Cimicifuga racemosa* were understood by this name being distinguished as the Virginian, seneca, and black snakeroots. The root of *Aristolochia reticulata* is known in the United States as Red River or Texas snakeroot.

The roots, or rhizomes, of *Liatris spicata*, *Eryngium aquaticum*, and *Eupatorium rugosum* have all been used in North America for snake bites, the first two species being known as button snakeroot and the last as white snakeroot. The rhizome of *Asarum canadense* passes under the name of Canadian snakeroot. All of these contain acrid or aromatic principles that, when given in a warm decoction, exercise a powerfully diaphoretic or, in some cases, diuretic, action, to which the benefit, if any, derived from their use may be attributed.

SNAPDRAGON, any herbaceous plant of the genus *Antirrhinum*, of the snapdragon or figwort family (Scrophulariaceae), of which there are over 40 species, scattered over the North Temperate Zone, especially in the western states of the U.S. The common garden snapdragon, *A. majus*, much grown for its handsome flowers, is a perennial from the Mediterranean region, usually grown as a summer annual; growers for the florist trade, however, force it for winter bloom. It is a stout plant 12–25 in. high, producing pouchlike, irregular flowers in a showy terminal cluster (raceme).

A. W. KERR

A VARIETY OF SNAPDRAGON, ANTIRRHINUM MAJUS

Another species is the beautiful chaparral snapdragon, *A. coulterianum* of California, with purple or white flowers that have conspicuous yellow hairs in the throat of the corolla.

Scores of showy garden varieties have replaced the typical *A. majus*. They are dwarfs (6–9 in. high), intermediates (10–20 in. high), and talls (20–40 in. high).

For outdoor culture, seeds are shallowly sown in late summer and wintered in cold frames for transplanting the following spring. In all cold regions this must be done annually, for snapdragons are not frost-hardy. When forced for winter bloom seed is sown in early summer, and the potted seedlings are transferred to a cool greenhouse in late summer.

Snapdragon is also the somewhat inappropriate name of *Galvesia speciosa*, closely related to *Antirrhinum*, an evergreen shrub with showy scarlet flowers, found on the islands off the coast of California. (N. Tr.)

SNELL (van Roigen), **WILLEBRORD** (1591–1626), commonly known as Snellius, Dutch astronomer and mathematician the discoverer of the law of refraction, was born at Leiden. In 1613 he succeeded his father, Rudolph Snell (1546–1613), as professor of mathematics in the University of Leiden. In his *Eratosthenes Batavus* (1617), he describes his method of measuring the earth, and gives as the result of his operations between Alkmaar and Bergen op Zoom a degree of the meridian equal to 55,100 toises (=117,449 yd.). His discovery (1621) of the law of refraction was of significance for the study of the nature of light (*see* Light: *History*).

Snell died at Leiden, Oct. 30, 1626.

SNELLMAN, JOHAN VILHELM (1806–1881), Finnish philosopher, statesman, and rouser of Finnish nationalism during

the period when Finland was under Russian rule, was born at Stockholm, Swed., on May 12, 1806. He studied at the University of Turku and received his B.A. degree in 1831 at Helsinki, where he became instructor in philosophy in 1835. In 1841 Snellman published a philosophical work, *Die Idee der Persönlichkeit*, and in 1842 appeared *Läran om staten* ("Political Science"), in which, leaning on Hegel, he showed that the essence of a state is a national spirit.

The publication of the *Kalevala* (*q.v.*) in 1835 had initiated an interest in developing a national literature in Finnish. Regarding this as essential to the growth of a national spirit, Snellman began his revolutionary influence as a stimulator of the national cultural life in 1844, when he started publication of the *Farmer's Friend* (*Maamiehen ystävä*) and the Swedish-language paper *Saima* (suppressed in 1846). Later, with Elias Lönnrot (*q.v.*), he edited at two different periods the paper *Litteraturblad för allmän medborgerlig bildning* ("Literary News for General Civic Culture"). In 1856 Snellman was named professor of "ethics and the system of sciences" at Helsinki, and from 1863 to 1868 he was a senator. He exerted a decisive influence on the promulgation of the 1863 statute extending the use of the Finnish language. He was also active in promoting the 1865 legislation that changed the basis of the monetary standard from rubles to marks. Snellman was knighted in 1866. He died July 4, 1881, at Kirkkonummi.

Snellman's *Collected Works* (*Kootut teokset*) were published in Finnish in 1928–33, the *Correspondence* (*Kirjeenvaihto*) with his wife in 1928. *See* also FINLAND: *History: Autonomy Under Russian Rule*; FINNISH LITERATURE. (F. P. M.)

SNIPE, a shore bird of the sandpiper family (Scolopacidae), much sought after as game. The common snipe (*Capella gallinago*) of Europe and America is 11 in. long with a long ($2\frac{1}{2}$–3 in.) bill, eye placed far back in the head, rather short legs, but long toes, a very short tail, and a complicated pattern of brown, buffs, and black on its upper parts. It lives in bogs and wet meadows where it probes for worms.

Secretive, it flushes from the grass with a hoarse cry and zigzags rapidly away. The birds, though usually solitary, sometimes feed and travel in loose flocks. They winter south to Africa, Java, and South America. In the spring the male gives a flight song, the pair lines a hollow in the ground for a nest; four spotted eggs are laid and incubated by the female; and the active, downy young are cared for by both parents.

There are other related species in Europe, Asia (some of which migrate to Australia), Africa, and South America. A more different species, the jacksnipe (*Lymnocryptes minimus*) of Europe, is only $7\frac{1}{2}$ inches long. In Europe, dowitchers are also called snipe.

The species of painted snipes, one of Africa and the area of southern Asia and Australia, the other of South America, form a related family, Rostratulidae, in which the birds are more brightly coloured; in this family the female is larger and brighter than the male, and takes the initiative in courtship, while the male incubates the eggs. (A. L. RD.)

SNOILSKY, CARL JOHAN GUSTAF, COUNT (1841–1903), Swedish poet, the most notable of a group of early poetic realists, was born in Stockholm on Sept. 8, 1841. At Uppsala University Snoilsky made himself known by his great poetic talent. He went to Italy in 1864 and wrote poems praising the joy and beauty of the Mediterranean landscape. After several years as a diplomatist and civil servant he gave up his career in 1879 and lived abroad, devoting himself to poetry. Deeply affected by social problems, he favoured the breaking down of class barriers and expressed his social anguish and humane liberalism in important poems. From these years also dates his well-known series on figures and episodes from Swedish history, *Svenska bilder* (1886). In 1890 Snoilsky returned to Stockholm, where he died on May 19, 1903. Snoilsky's poetry is clear and elegant and has, though rather traditional in form, a strong note of realism.

BIBLIOGRAPHY.—*Samlade dikter*, 5 vol. (1903–04); K. Warburg, *Carl Snoilsky* (1905); Henry Olsson, *Den unge Snoilsky* (1941).
(G. SDT.)

SNORRI STURLUSON (1179–1241), Icelandic historian and author of the *Prose Edda* (*see* EDDA). A descendant of the great poet Egill Skallagrímsson, he was born in 1179 in the west fjords, but from the age of three was brought up at Oddi in the south, at the home of Jón Loptsson, the most influential chieftain in Iceland. Jón was a grandson of Saemund Sigfússon (Saemundr the Wise), the first known Icelandic historical writer, and of Magnus II Barfot (Bareleg), of Norway. Like many chieftains, Jón was an ordained priest, and he was keenly interested in historical traditions. From him Snorri acquired both a deep knowledge of Icelandic tradition and a European breadth of outlook.

Jón died in 1197, and in 1199 Snorri married an heiress in Borgarfiord (Borgarfjordhur), and began to acquire lands and power. In 1206 he settled at Reykjarhóll, where most of his works were written. There he built a magnificent house, with a tiled bath, heated by volcanic springs. During 1215–18 (and again during 1222–32) he was *lǫgsǫgumaðr* ("lawspeaker" or president) of the Icelandic high court. In 1218 he was invited to the Norwegian court by King Haakon IV; he also visited noble families in Sweden, acquiring an unusual knowledge of Swedish tradition.

While in Norway, he began to intervene in politics. A quarrel had broken out between the Bergen merchants and Icelanders from Oddi, and the Norwegians were planning a punitive raid. Snorri persuaded Haakon that he could become king of Iceland by peaceful means, and promised to win the Icelanders over to him. He became Haakon's vassal (*lendrmaðr*), and on return to Iceland in 1220, sent Haakon his son as hostage. However, he did little to further Haakon's cause; and, in revenge, Haakon stirred up a feud between him and his kinsman, Sturla. Snorri was outlawed and fled to Norway (1237), there joining the faction of Earl Skúli. In 1239, hearing that Sturla had fallen in battle against Gissur Thorvaldsson (Snorri's son-in-law), Snorri returned to Iceland against Haakon's express command. In 1240 Haakon defeated Earl Skúli, and ordered Gissur to kill Snorri, or send him captive to Norway. Gissur attacked Snorri in Reykjarhóll, and killed him on Sept. 22, 1241.

Snorri's writings are remarkable both for their scope and for their formal assurance. In the *Prose Edda*, he succeeds in arranging the legends and drawing meaning from them, and in communicating his enjoyment of them as stories. He also explains the ornate diction of the ancient scaldic poets, telling the heroic legends to which their metaphors allude, and illustrates the great variety of poetic metres, often with his own verses.

Though he also composed court poems in honour of great men, his greatest gift was in prose. As well as the *Prose Edda*, he wrote a life of St. Oláf of Norway, included in the *Heimskringla*, a history of the Norwegian kings from their mythological origins as descendants of Oðinn down to Magnus Erlingsson (1184). For this, he used earlier histories, but he had gathered much fresh material. He particularly valued poems handed down orally from the time of the events themselves, and selected the traditions that seemed to him most authoritative and true both to contemporary politics and to human nature. To him, the writing of history meant the reconstruction of action, motives, and dialogue: his genius lay in his power to present with the immediacy of drama all that he perceived critically as a historian.

To Snorri is also attributed, almost certainly correctly, *Egils Saga*, one of the greatest of the Icelandic sagas.

The qualities of intelligence, warmth, and scholarly industry in Snorri's writing contrast sharply with the weak, shifty character that emerges from the account of his life by his nephew preserved in the *Sturlunga Saga*. Cold and miserly toward his family, opportunist in politics, he was ambitious for greatness but lacked the courage and willpower necessary to achieve it.

See also ICELANDIC LITERATURE.

BIBLIOGRAPHY.—The *Prose Edda of Snorri Sturluson*, trans. by J. I. Young (1964); *Heimskringla*, trans. by W. Morris and E. Magnússon, 4 vol., including a life of Snorri (1893–1905). *See* also S. Nordahl, *Snorri Sturluson* (1920; supplemented by an article in *Skirnir*, 1941).

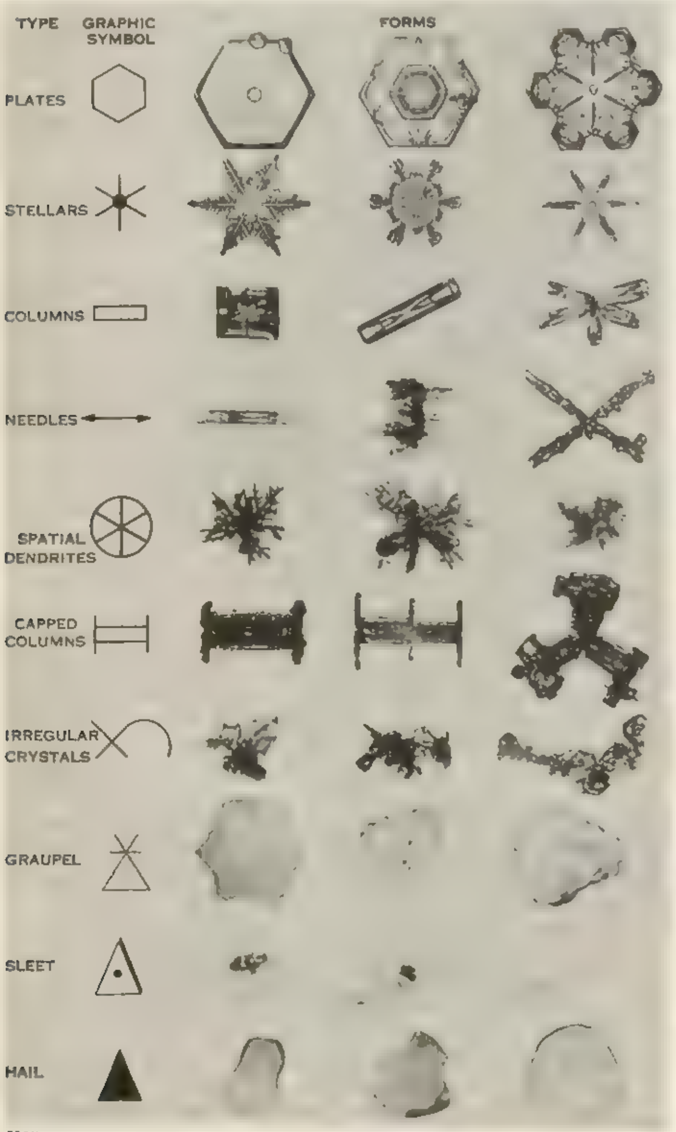
SNOW is the solid form of water which grows while floating, rising or falling in the free air of the atmosphere. It often takes the form of beautiful crystals generally having a hexagonal pattern. Snow crystals sometimes come from a cloudless sky since they may form spontaneously or in the presence of suitable sub-

limation nuclei before there is enough water vapour present in the air to form the common water droplet cloud.

Snow has attracted the attention and admiration of man from ancient times. The word crystal is of Greek origin and is derived from *kryllos*, the word for frost. One of the first indications of man's early awareness that there was more to snow than cold, hunger, trouble, and restricted travel is the biblical reference to the "storehouses of the snow" in the Book of Job 38:22.

One of the first recorded instances showing recognition of the hexagonal symmetry of snow crystals is ascribed to Olaus Magnus, archbishop of Uppsala in Sweden, who depicted the crystal form in a book on natural phenomena published in Rome in 1555. The development of the microscope increased this knowledge. Robert Hooke of England studied the snow crystal structure and published illustrations of some of the beauty and symmetry he saw in his fascinating book, *Micrographia*, published in 1665. During the middle of the 19th century, poetry and prose were written about snow and snow crystals. An extremely interesting book of this period entitled *Cloud Crystals* written by "A Lady" and published in 1864 contains scientific fact, poetic fancy and excellent wood cuts showing some of the basic forms of snow.

Snow Crystals.—It has been said that no two snow crystals are the same. This statement is true in the sense that it is quite unlikely that among all the countless myriads that fall from the sky there are ever two identical in shape, size and absolute number and distribution of water molecules.



FROM "ARTIFICIAL STIMULATION OF RAIN" (1957); REPRODUCED BY PERMISSION OF PERGAMON PRESS, INC.

CLASSIFICATION OF FROZEN PRECIPITATION

However, it is no problem during a particular storm to observe two crystals which to the unaided eye or even at low-power magnification are similar in outline and general structure. The main point to observe in this relation is that there is a bewildering variety of size, shape and pattern. They range from simple, solid triangles and hexagons to the exquisite fernlike dendrites, with all manner of complicated three-dimensional formations.

In an attempt to simplify the task of the research scientist concerned with reaching a better understanding of the fundamental science related to the formation of snow and ice, a classification of solid precipitation was developed in 1951 for international usage. This is shown in the accompanying figure which illustrates the typical forms of ten types of frozen precipitation. It is quite feasible to subdivide any one of these ten types into subtypes and for special purposes this is being done. For all practical purposes, however, the types illustrated can be used to describe the precipitation of snow, ice and hail storms.

A considerable amount of the popular interest in snow crystals stems from the pioneer and classic photographic work done by Wilson W. Bentley, farmer-meteorologist of Jericho, Vt. Bentley became enthralled by the beauty of snow crystals and spent many winters with his camera, recording the varieties and symmetry of the crystals that are such a common feature in the vicinity of his northern Vermont home. Of the more than 5,000 photomicrographs prepared by him, nearly half have been perpetuated in *Snow Crystals*, published in 1931. His persistent efforts, careful observations and beautiful photographs did much to arouse the interest of scientists in snow crystals

In Europe a similar effort was expended by A. B. Dobrowolska of Poland who wrote many scientific articles on this subject. In Japan, the classic work in the study of snow by U. Nakaya of the University of Hokkaido culminated in his book *Snow Crystals* (1954).

The relationship of snow to ice and the peculiar underwater snow storms caused by frazil (needle, or lolly) ice formations have been studied for many years and are of much economic importance.

The bewildering variations and modifications occurring in snow after it has reached the ground are a specialized study that still continues. The early studies have been well summarized by G. Seligman. The modern work increases in complexity with much of it in North America under the leadership of the Snow, Ice and Permafrost establishment of the U.S. army corps of engineers.

There are many factors controlling the size and shape of snow crystals. One of these—probably the most important—is the supply of water molecules in the air. If the air contains a rich supply of moisture, the snow crystals will grow fast and are likely to have dendritic or fernlike arms, will be large and may clump together with others to form a snowflake. Snowflakes may have 50 or more crystals interlocked with each other. If a stellar-type crystal forms or falls into a cloud containing a high density of subcooled cloud droplets, it may sweep up some of them which freeze upon contact and cover the crystal until it looks like a miniature snowball. Such particles are called soft hail or snow pellets but are most commonly called by the German name *graupel*.

The temperature of the air is another condition controlling the form of the snow crystal. If the air is colder than about −30° C. it contains very little water vapour so that any crystals developing grow very slowly and are likely to be tiny, dense and compact in shape and crystal form. This is the condition which produces the crystals of the filmy cirrus clouds that form in the high levels of the atmosphere often at altitudes of five to eight miles over the middle and low latitudes. In the polar regions, cirruslike clouds may be close to the ground due to the intense cold.

At temperatures close to −40° C. (−40° F.) which commonly occur in the polar regions and at levels in the sky above 30,000 ft., ice crystals form spontaneously whenever a sufficient concentration of water molecules exceeds the amount that can exist in vapour form. This is called homogeneous nucleation phenomenon. At all temperatures warmer than −40° and colder than 0° C. (32° F.), heterogeneous nucleation occurs. In this range special type microscopic or submicroscopic particles floating in the air serve as nuclei for ice crystal formation. Particles in the

size range of 0.01 to 10 μ (0.00000039 to 0.00039 in.) in diameter, consisting of such materials as volcanic dust, loess, certain types of clays, silver iodide and a special form of copper sulfide, serve as effective ice nuclei. Except for the silver iodide and copper sulfide, which become active at temperatures colder than −4° C., most ice nucleating substances are effective in the temperature range of −12° to −25° C. There is much still to be learned about the behaviour and characteristics of such particles.

Effects of Snow.—Like so many of the small things in nature, a snow crystal, when considered as a single particle, is at once beautiful or at least interesting in structural form and seemingly ineffectual as an element of force or destruction. It is only when they occur in quantity—nearly countless numbers—that their presence causes wonder, interest, concern or terror.

Floating in the rare air of the high atmosphere, the tiny, pointed or flat-ended hexagonal columns and thickened hexagonal plates form the coloured halos, pillars, arcs and circles surrounding the sun in orderly array. Formed in the upper parts of turbulent clouds, the reactions between crystals and subcooled water cause the development of electrical charges often manifested as cloud-to-cloud or cloud-to-ground lightning and radio static and lesser degrees of atmospheric electricity. Sweeping over and smashing on airplane wings and fuselage, the snow crystal causes electrification which may appear as St. Elmo's fire (*q.v.*) on leading surfaces and may illuminate the trailing edges of the wing tips.

Much of the moderate to heavy rain in the middle latitudes starts as snow in the high reaches of the sky and thus is responsible for the precipitation on which most life as we know it is primarily dependent. Snow packs (the annual accumulations of snow at higher elevations) represent a form of water storage that is very useful to man. The heavy winter snows in the higher elevations remain as a slowly melting snow pack long into the summer, feeding the mountain streams flowing into the drier valleys. The interior valleys of California and Oregon, plains of Utah, Colorado, New Mexico, Arizona, parts of Spain, north Africa and a number of similar areas which have insufficient precipitation to support crops are made into fertile agricultural regions by means of irrigation water from nearby massive mountains.

At times snow may become so deep and unstable on the mountain slope as to form a terrifying avalanche or it may become consolidated to form massive layers of turquoise green-blue glacial ice to flow down the mountain as a river of ice (*see* AVALANCHE; GLACIER).

It would not be proper to end the discussion of the massive effects of snow without reference to the beauty of snow and the sport of skiing. With the rapid development in the second half of the 20th century of ski centres and ski lifts of all kinds, and the equipment demanded by the skier, the importance of snow as a valuable commodity in certain areas assumed such value that it became economically profitable to generate snow on cold nights from high-pressure air and cold water.

For an account of the history of this sport and a description of its development throughout the world *see* SKIING. *See* also SNOWSHOE.

Snow Crystals and Cloud Modification.—When the droplets in clouds cool below 0° C. (32° F.), they do not normally freeze but remain in liquid form. These subcooled droplets may cool to −30° C. or colder before they freeze, depending on whether or not they contain freezing nuclei or float near ice crystals. Whenever water subcools, it becomes unstable, so that an ice crystal alongside a liquid droplet will grow at the expense of the droplet. If nuclei suitable for ice crystal nucleation enter a cloud, they grow rapidly and either transform the entire cloud to snow or grow until large enough to overcome rising currents in the air and thus fall out of the cloud.

Since, as shown by V. J. Schaefer in 1946, it is feasible to produce local concentrations of ice nuclei in greater numbers than occur under natural conditions and having a nucleating activity better than naturally occurring particles, scientists are able to carry out experiments using clouds in the atmosphere for their subjects. Efforts are being made to take advantage of these opportunities, to prevent damaging hailstorms, to modify lightning storms, to increase snow deposits in the high mountains, to increase rainfall and to moderate or prevent the development of disastrous storms.

These local efforts may eventually lead the way to large-scale weather control. Although some evidence has been found that other methods may be useful to modify clouds, the ice nucleus and the snow particle which grows on it seem likely to remain of great importance in the science of experimental atmospheric physics.

Snow Crystal Replicas.—The form and appearance of the fragile and evanescent snow crystal may be preserved in minute detail by an extremely simple and effective process of replication. A dilute 0.5 to 3% solution of polyvinyl formal (a plastic) dissolved in ethylene dichloride (a common solvent) is required to make the replica. The solution must be cooled a few degrees below 0° C. (32° F.) before using. A small amount of the replica solution is poured onto a piece of black cardboard, sheet glass or similar material that also has been cooled below 0° C. The wet surface is then placed so that snow falls into it. When the desired number of crystals have fallen into the solution, it is placed in a cold, well-ventilated place until the solvent evaporates. This ordinarily requires 5 to 10 minutes. When the sample is dry, it may be warmed up and examined. When properly prepared, each crystal is encased in a thin plastic shell whose inner surface is an exact negative copy of the original crystal. When the crystal melts, the water passes through this thin film, leaving behind a hollow cavity which reflects and scatters light in a manner quite similar to that of the original crystal. Frost, fallen snow and related substances may be replicated in a similar manner.

See also references under "Snow" in the Index.

BIBLIOGRAPHY.—"A Lady," *Cloud Crystals, A Snowflake Album* (1864); V. J. Schaefer *et al.*, *Report of the Committee on Snow Classification of the International Association of Hydrology*, IUGG (1951); W. A. Bentley and W. J. Humphreys, *Snow Crystals* (1931); A. B. Dobrowolska, *Historja Naturalna Lodu* (1923); U. Nakaya, *Snow Crystals* (1954); H. T. Barnes, *Ice Engineering* (1928); G. Seligman, *Snow Structure and Ski Fields* (1936); H. T. Orville *et al.*, *Final Report*, Advisory Committee on Weather Control (1957); V. J. Schaefer, "Properties of Particles of Snow and the Electrical Effects They Produce in Storms," *Trans. Amer. Geophys. Un.* 28,587 (1947), "Formation of Frazil and Anchor Ice in Cold Water," *Trans. Am. Geophys. Un.* 31,885 (1950), "Experimental Meteorology," *Z. angew. Math. Phys.* 1,153;217 (1950), "A Method of Making Replicas of Snowflakes, Ice Crystals and Other Shortlived Substances," *Museum News*, Amer. Assoc. of Museums (Sept. 1941); V. J. Schaefer and D. Harker, "Surface Replicas For Use in the Electron Microscope," *J. Appl. Phys.* 13,427 (1942). (V. J. S.)

SNOWDEN, PHILIP SNOWDEN, VISCOUNT (1864–1937), British statesman, outstanding socialist propagandist and first Labour chancellor of the exchequer, was born at Ickornshaw, Cowling, Yorkshire, on July 18, 1864, the son of a weaver. He was educated at an elementary school. For a few years he was a clerk in the customs and excise department of the Inland Revenue, but in 1891 a spinal disease, which left him permanently crippled, forced him to abandon this work. When partial recovery came he devoted himself to full-time political activity for the Independent Labour Party (I.L.P.). He was perhaps the most effective of all those devoted and peripatetic political evangelists who symbolized the early activity of the I.L.P. For many years the whole of his meagre income came from lecturing and writing for socialist papers.

Unsuccessful in the 1900 general election at Blackburn, Yorkshire, he represented that constituency from 1906 to 1918, and Colne Valley, Yorkshire, from 1922 to 1931. Snowden was a formidable parliamentarian, whose character and command of fact and argument more than offset the bitterness of his manner. He initiated some of the most memorable pre-1914 debates on social and economic questions. When World War I broke out he unhesitatingly took the pacifist side, and shared with J. Ramsay MacDonald much of the resultant unpopularity. He was far from being an extremist, however. Before 1914 he had criticized the strike policy of the more militant trades unionists, and in 1921 he resigned the treasurership of the I.L.P. in protest against what he regarded as the increasing intransigence of that body.

In 1924 Snowden became chancellor of the exchequer of the first

Labour government. His first budget broke little new ground but made some well-received reductions in indirect taxation. In 1929, in the second Labour administration, he was reappointed to the treasury, and immediately scored a popular success by insisting upon British rights at the Hague conference on German reparations (1929). Deepening unemployment meant that in 1930 he faced a far more difficult budgetary situation than in 1924, but he reacted by showing himself a rigidly orthodox financier, more attached to the maintenance of free trade than to anything else. He was responsible for establishing the famous May Economy Committee of 1931, the recommendations of which, particularly those relating to the reduction of unemployment benefit, led to the breakup of the Labour government on Aug. 24, 1931.

He continued as chancellor when MacDonald formed the National government, carrying through an emergency budget, with heavy cuts in expenditure, and presiding over Britain's departure from the gold standard at the end of September 1931. He did not stand at the election in October and in November he was created Viscount Snowden of Ickornshaw and became lord privy seal. However, he resigned in September 1932, in protest against the protectionist policy adopted at the Ottawa conference. He died at Elm Lodge, Tilford, in Surrey, on May 15, 1937. Snowden was never accused of careerism; he was rigid and arrogant, but high-principled and austere. His principal books were: *The Socialist's Budget* (1907), *The Living Wage* (1912), *Socialism and Syndicalism* (1913), *Wages and Prices* (1920), *Labour and National Finance* (1920), *Labour and the New World* (1921, revised edition, 1924), and *An Autobiography*, two volumes (1934). (R. J.)

SNOWDON (Y Wyddfa, "the viewplace"), the highest peak (3,560 ft. [1,085 m.]) in England and Wales, situated in Caernarvonshire, Wales. It is the main peak in the Snowdonian mountain mass (Welsh: *Eryri*, eagle place) which, bounded by the Arfon lowlands on the northwest and the vales of Ffestiniog and Conway on the south and east, includes the peaks Glyder-fawr (3,279 ft. [999 m.]), Glyder-fach (3,262 ft. [994 m.]), Carnedd Llewelyn (3,485 ft.), and Carnedd Dafydd (3,427 ft.). The vegetation of the area is mainly rough grassland; alpine flora survives only on some high rock ledges shaded from the sun. Below 800 ft. (250 m.; the present-day limit for tree growth) are patches of deciduous forest. Snowdon is formed principally of Ordovician volcanic rocks, slates, and grits. It consists of five arêtes (sharp ridges) converging on the summit, between which lie cirques (*see* Cirque). Surrounding the peak are the Llanberis, Aberglaslyn, and Rhyd-ddu passes. A rack-and-pinion railway (opened 1896) ascends from Llanberis to the summit (4½ mi. [7.6 km.]). Several lakes occur in the glaciated valleys which radiate out from Snowdon: Lakes Padarn and Peris separated by an alluvial fan in the valley below Llanberis Pass; Lakes Gwynant and Dinas in the Nant Gwynant valley.

Snowdonia National Park covers 845 sq.mi. (2,189 sq.km.) of the mountainous tracts of Caernarvonshire, Denbighshire, and Merioneth. Noted for the beauty of its mountain, valley, and lake scenery, it includes the Snowdonian mountain mass in the north, the Cader Idris (*q.v.*) in the south, and the National Forest Park (formed in 1937) west of the Conway valley. The Forestry Commission has been responsible for extensive planting in the Gwydyr and Beddgelert forests. Conifers predominate including Sitka spruce, European larch, and Douglas fir.

Bibliography.—H. R. C. Carr and G. A. Lister (eds.), *The Mountains of Snowdonia*, 2nd ed. (1948); F. J. North *et al.*, *Snowdonia* (1949); *Snowdonia* by C. Embleton in "British Landscapes Through Maps," no. 5, ed. by K. C. Edwards (1962). (E. G. Bow.)

SNOWDROP, any plant of *Galanthus*, a small old world genus of the amaryllis family (Amaryllidaceae). All the plants of the genus have bulbs, linear basal leaves and erect flower stalks about one foot tall, bearing at the top a solitary, pendulous, bell-shaped white flower. Most especially, the term snowdrop indicates *G. nivalis*, the common snowdrop widely cultivated in the U.S. and England, often in naturalized masses in the lawn. Other distinct species are the Crimean snowdrop, *G. plicatus*, with broad leaves folded like a fan, and *G. elwesi*, the giant snowdrop, a native of the eastern Mediterranean with large (1¼ in. long) flowers.

SNOWDROP (GALANTHUS NIVALIS), ONE OF THE EARLIEST SPRING FLOWERS

All species thrive in almost any soil or position. The bulbs should be planted in autumn, 3 in. deep and about 3 in. apart, preferably by the hundreds if sheets of early white bloom—often before the winter snow disappears—are desired. (N. Tr.)

SNOW LINE, a line above which snow accumulates, temporarily (seasonal snow line) or perennially. Two types of perennial snow lines are recognized, orographic and regional. The orographic snow line passes through the lowest limit of perennial snow masses lying in protected spots. The regional snow line marks the lowest limit of an essentially continuous snow mantle, and its altitude depends primarily upon temperature and precipitation.

The regional snow line rises from near sea level in polar regions to 20,000–21,000 ft. (6,100–6,400 m.) in the horse latitudes and descends to 15,000–18,000 ft. (4,600–5,500 m.) in equatorial regions. The orographic snow line may be 2,000–3,000 ft. (600–900 m.) lower. *See* Climate and Climatology; Glacier. (R. P. Sp.)

SNOWPLOW, a device for clearing snow from streets, highways, railroad tracks, airfields, and the like. There are three general types of snowplows, known as rotary, V, and single-blade.

The rotary plow in railroad use is mounted on a carriage that is pushed by one or more locomotives. It consists of several large blades attached to a horizontal shaft powered by an engine mounted on the carriage. A housing is provided for the protection of the engine, the rotary controls, and the operators. The plow is forced into deep drifts or slides while the blades are rotating. The snow is sliced from the face of the drift and discharged to one side of the track. The rotary snowplow is generally used to remove snow in depths which cannot be handled with the V-plow.

Under some conditions flangers are used to remove snow from the inside of the top of the rail to avoid the formation of ice that might cause derailments.

The rotary plow for highway use is similar in design to that used for clearing railway tracks but is built into a self-propelled truck unit. In some cases pusher trucks may be used for additional power. An adaptation of the rotary plow is a unit consisting of two or more augers which feed the snow from the left and from the right to a centre collecting point from which a high-speed rotor discharges the snow to one side of the right-of-way. This unit is frequently mounted on the front of a standard motor grader or special heavy truck.

The V-plow consists of two concave surfaces fastened to a common centre line and sloping upward and to the rear, one to the right and one to the left. The V-plow forces the snow to either side and is used in areas of heavy snow, deep drifts or slides. For clearing railroad tracks the V-plow is mounted on a carriage pushed by one or more locomotives. In some mountainous areas the plows are mounted directly on the front of steam locomotives. The V-plow for highway clearance may be mounted on a heavy-duty truck, motor grader, or tractor.

The single-blade plow may consist of a concave surface of varying depth or a straight blade. The former moves the snow in one direction only; the straight-blade type is reversible and may be set to move the snow to the right or left. The single-blade plow mounted on a heavy-duty truck is commonly used on highways and airports. The trucks operate at relatively high speed to throw the snow well off the travel way. A wing may be attached back of and above the single-blade plow. It may be adjusted to direct snow away from the plowed way. (F. W. St.)

SNOWSHOE, a form of footgear devised by Eskimos and some Indians for traveling over deep snow. Whereas skis were

adopted in the Old World, snowshoes had more practical applications in the forested areas of North America where evidence suggests that the frame snowshoe was independently invented in several regions.

The most important parts of a snowshoe are the wooden frame, usually of birch or hickory, a toe and heel crossbar of wood or rawhide, extra strengthening bars, netting in large meshes, and foot lashing for attaching the shoe or moccasin. The size of the mesh varies, a coarser mesh being used for wet, soft snow.

Snowshoes differ in materials, size, shape, and decoration, depending on the people who make them and the conditions under which they are used. The variation in shapes is particularly great. The frames, usually 2 to 5 ft. (0.6 to 2 m.) long and 1½ ft. (0.5 m.) wide, help to distribute the body weight over a wide area to keep the wearer atop the snow. Indians like the Montagnais and Naskapi of the Labrador Peninsula, who subsist by trapping and hunting large game animals in an environment having as much as 60 to 120 in. (150 to 300 cm.) of snowfall annually, need snowshoes for a variety of kinds of winter travel. Thus their five distinct types, descriptively named from the shapes of the frames, are bear-paw, swallow tail, beaver tail, elbow, and pointed.

Walking in snowshoes is accomplished by lifting the feet slightly and sliding the overlapping inner edges of the shoes over each other, avoiding the unnatural "straddle gait" that would otherwise be necessary. Snowshoe racing, common in the U.S. and Canada, is a popular event at winter carnivals. (J. W. VanS.)

SNUFF, a powdered preparation of tobacco (*q.v.*) used by inhalation or by "dipping"; *i.e.*, rubbing on the teeth and gums. The practice of inhaling snuff became common in England around the 17th century, and throughout the 18th century it was universal.

At first each quantity was fresh grated (Fr. *râper*, whence coarser kinds were later known as "rappee"). This entailed the snuff taker's carrying with him a grater; early 18th-century graters made of ivory and other material are in existence. The art and craft of the miniature painter, the enameler, jeweler, and gold- and silversmith were bestowed upon the box. Humbler snuff takers were content with boxes of silver, brass or other metal, horn, tortoiseshell, or wood.

The mull, a silver-mounted ram's head, is a large table snuffbox.

Snuff manufacture, which requires 18 to 20 months, involves grinding the tobacco and subjecting it to repeated fermentations. Snuffs are scented with attar of roses, lavender, cloves, jasmine, etc.

SNYDERS, FRANS (1579–1657), Flemish painter of animals and still life, was born at Antwerp, where he was baptized on Nov. 11, 1579. He studied under Pieter Brueghel the Younger, and afterward under Hendrick van Balen, the first master of Van Dyck. He visited Italy in 1608. On Oct. 23, 1611, he married Margaretha de Vos, the sister of the painters Cornelis and Paul de Vos. He died on Aug. 19, 1657.

Snyders originally devoted himself to painting flowers, fruit, and subjects of still life, later turning to animal painting and executing with the greatest skill and spirit hunting pieces and combats of wild animals. His composition is rich and varied, his drawing correct and vigorous, his touch bold and thoroughly expressive of the different textures of furs and skins. Rubens frequently employed him to paint animals, fruit, and still life in his own pictures, and he assisted Jordaens similarly. In the lion and boar hunts that bear the name of Snyders the hand of Rubens sometimes appears. Snyders was appointed principal painter to the archduke Albert, governor of the Low Countries, for whom he executed some of his finest works. One of these, a "Stag Hunt," was presented to Philip III, who commissioned the artist to paint several subjects of the chase. The Prado Museum, Madrid, is rich in the works of Snyders. Others may be seen at the Museum of Fine Arts, Boston, Mass. ("Boar Hunt"); the California Palace of the Legion of Honor, San Francisco ("The Monkey and the Gander"); the Kaiser Friedrich Museum, Berlin ("Bear Hunt"); and at many other public galleries.

SOANE, SIR JOHN (1753–1837), English architect, one of the most original architects of his time in Europe, was born at Goring-on-Thames, Sept. 10, 1753, son of a bricklayer. In 1768 he entered the office of George Dance the Younger, surveyor to the City of London, whom in later life he referred to as his "revered master"; in 1772 he went to Henry Holland (*q.v.*) as an assistant; and from 1771 he attended the Royal Academy schools, winning the gold medal in 1776. In 1778 the king's traveling studentship took him to Italy. He cut short his studies abroad because of magnificent promises of employment made to him by Frederick Hervey, bishop of Derry. These were not kept, and he was an only moderately successful country house architect until in 1788 he was appointed, thanks to William Pitt, surveyor to the Bank of England. Various government appointments followed and in 1806 he succeeded Dance as professor of architecture at the Royal Academy; he was knighted in 1831. The list of his works is a long one; some of the finest, which helped establish his reputation, included his rebuilding of the Bank of England (since rebuilt again), his own house at 13 Lincoln's Inn Fields, London (left by him to the nation as a museum), and the Dulwich College art gallery, Kent. His style is characterized by a tendency to strip classical elements of design down to their structural essentials, the substitution of linear for modeled ornamentation, frequent use of shallow domes and top lighting, and an ingenious and often highly imaginative handling of interior space. Soane died on Jan. 20, 1837, in London.

Bibliography.—A. T. Bolton, *The Portrait of Sir John Soane, R.A.* (1927); John Summerson, *Sir John Soane 1753–1837* (1952); H. M. Colvin, *A Biographical Dictionary of English Architects 1660–1840* (1954). (Ms. W.)

SOAP may be defined as either a chemical compound or a mixture of chemical compounds resulting from the interaction of fatty oils and fats with alkali; *i.e.*, the salts of fatty acids. Functionally, soap is a substance possessing the characteristic "soaplike" properties of sudsing, detergency, surface tension lowering, wetting and emulsifying power, and curd and gel formation. Modern research has led to the view that these properties may be attained by inserting strongly polar (water-soluble) groups into a wide variety of long-chain molecules. Many such long-chain substances, some of them bearing no relation to the fats, have been developed, and some of them (*e.g.*, sulfated alcohols, sulfonates, etc.) are of commercial importance in the field of soaps. The functional consideration of soap brings within its scope these synthetic detergents and at the same time excludes the water-insoluble compounds of fatty acids with bases of calcium, iron, aluminum, etc. These latter are better termed "metallic soaps."

Soap appears first to have been made by boiling goat tallow and causticized wood ashes, as described by Pliny. The resulting soft potash soap was converted into hard soda soap by treating the paste repeatedly with salt. In the 13th century the industry was introduced from Italy and Germany into France, and in the 14th century into England. In America soapmaking was a household art, with few industrial developments until after about 1800.

The processes and extent of soap manufacture were revolutionized during the first half of the 19th century as a result of M. E. Chevreul's classical researches on the constitution of oils and fats and of the introduction of the Leblanc process (invented in 1791) for the manufacture of soda from brine. Large-scale manufacturing operations developed, and the industry rapidly became able to process individual kettles of soap containing from 100,000 to 1,000,000 lb. per charge.

Chemistry of Soap.—All fatty oils and fats are mixtures of glycerides; *i.e.*, compounds (esters) of the trihydric alcohol glycerol (glycerin) and some fatty acid such as palmitic acid, etc. The chemistry involved in soap manufacture may be expressed in the following equation:

$$\underset{\substack{\text{palmitin}\\ \text{(fat)}}}{C_3H_5(OCOC_{15}H_{31})_3} + \underset{\text{caustic soda}}{3NaOH} = \underset{\substack{\text{soap}\\ \text{(sodium palmitate)}}}{3NaOCOC_{15}H_{31}} + \underset{\text{glycerol}}{C_3H_5(OH)_3}$$

The reaction of the fat with the aqueous caustic solution is termed saponification. (The term may also refer to the action of water on a fat to produce fatty acid.) As seen in the equation, the saponification reaction produces both soap and glycerol. The

usual method of manufacturing soap requires a further series of operations to separate the glycerol from the soap and bring the soap to the correct physical form (or phase).

Commercial soap products are essentially soap-water systems, the properties, form, and appearance of the product depending in great degree on the proportion of soap to water. Strictly anhydrous (water-free) soap has none of the usual soaplike properties and therefore is not used.

At room temperature for most of the potash soaps and at more elevated temperatures for the soda soaps, soap-water systems exhibit a variety of physical forms, or phases. These different phases, described below, are the materials with which the soap maker works and from which his products are derived. For example, true soap solutions (called nigre phase) occur when the proportion of soap to water is under about 30% (varying from soap to soap). In potash soaps the solutions, with some modification with glycerol or alcohol, form the liquid soaps of commerce. (Special compositions of soap solutions containing electrolytes constitute the nigres of the soap kettle.) When the soap-water system contains about 40–50% water, a different phase of soap results. This phase is called middle soap, or gum soap; when derived from potash it is sometimes sold as a soft soap. Middle soap is stiff and viscous, much more so than either more dilute or more concentrated systems. At concentrations approximating 70% soap–30% water, a form called neat soap occurs. Neat soap when cooled and solidified is converted into "framed," or unmilled, bar soap. It is the starting material for the vast majority of toilet, laundry, and household soaps, since it is produced as the final product of the soap kettle or hydrolyzer; it is then further processed as occasion requires (*see* below). When neat soap is diluted with water, instead of becoming less viscous, as is the usual rule, it becomes more viscous due to formation of middle soap.

At concentrations of about 85% soap, solid (crystalline) forms of soap become increasingly predominant. Most powders, beads, and flakes are made in this low moisture range (usually below 10%) and hence are substantially crystalline in nature.

Since the great bulk of commercial soap products is derived from saponification with a soda base, it is sufficient to consider further the soda soaps alone. Dilute solutions of the order of 0.1–0.5% are used for laundering and dishwashing. Such solutions (nigre phase) are relatively thin and fluid, but more concentrated hot solutions when cooled are either opaque jellies or semisolid masses. The frothing or sudsing of soaps occurs only in the solution (nigre) state. Neat, middle, waxy, and solid forms do not foam or suds until water is added to transform them, at least partially, to nigre. Thus, while soaps are ordinarily manufactured and sold in one of the concentrated phases, their usefulness in cleansing comes about after their conversion to the nigre, or solution, state by means of water.

Neat and middle soap are anisotropic, liquid crystalline phases. Nigre is an isotropic, colloidal solution. The solid, crystalline forms of soap are called alpha, beta, delta, and omega, each of the last three contributing its distinctive properties when present in bars, powders, and pastes.

The peculiar colloidal and electrical properties of the solution, or nigre, form of soap have led to the classification of soap as a colloidal electrolyte. At very low concentrations (of the order of 0.001%) the dissolved soap is hydrolyzed to a degree by the water to form free fatty acid and small amounts of free alkali. At slightly higher concentrations, acid soap (*e.g.*, sodium hydrogen palmitate) precipitates, causing much of the cloudiness seen in dilute soap solutions and liquid soaps. The molecules of dissolved soap and soap ions in solutions of the order of 0.1% concentration begin to clump together to form "neutral colloid" and colloidal "ionic micelle," which may be hydrated. The colloidal ionic micelle (which may be thought of as an electrically charged colloidal particle) and other ions give the solution pronounced electrical conductivity in contradistinction to ordinary colloids.

The behaviour of a pure, single, anhydrous soap when heated is very complex and has led to much confusion about the melting points of soaps. The solid soap when heated is converted in turn at definite transition temperatures to subwaxy soap, waxy soap, superwaxy soap, subneat soap, neat soap, and finally to nigre.

Only the compounds prepared from fatty acids of between about 10 and 18 carbon atoms in chain length can be considered soaplike under usual conditions. Indeed, sodium stearate (18 carbon atoms) possesses practically no soaplike properties at room temperature; it is only at temperatures approaching the boiling point of water that this compound behaves like a soap.

Detergent Power.—The washing or detergent power of soap was formerly attributed to the alkalinity produced by hydrolysis of the soap. Though this explanation has been abandoned, no comprehensive substitute explanation based on adequate experimental data has yet been established in its place. The detergent action of soap is an intricate phenomenon in which wetting power, emulsification, ability to dissolve grease, etc., all play a role. A widely held and plausible hypothesis has considered that washing is dependent on an unsymmetrical structure inherent in the soap molecule, one end (polar) being water soluble and the other being insoluble in water and soluble in oil. Such molecules tend to form a concentrated film at the interface between the water and the dirt, each molecule attaching itself at one end to the dirt and at the other end to the water, and hence tending to pull the dirt into solution.

There is much evidence that the micellar (colloidal electrolyte) form of soap must be present at the dilution used for washing. Nevertheless, many authorities maintain that detergent power derives from the long-chain fatty ion, which is also present.

As a practical matter, it should be emphasized that soaps used for household laundering usually contain considerable quantities of "builders" such as sodium silicate, phosphate, etc. There may also be present a fluorescent or substantive dye which adheres to the laundered cloth to give a whiter or brighter appearance. Likewise a percentage of carboxymethyl cellulose may be included in the product to inhibit redeposition of dirt on the cloth from the wash water. These additives contribute to the final overall appearance of the cloth and to the apparent detergent power of the soap. The suds are useful in indicating that sufficient soap is present to ensure detergent activity. The contribution of the suds to the actual detergent effect, while positive, is in some instances indirect. In some types of automatic washing devices, the suds must be controlled to lower levels for most efficient washing. Because bacteria can decompose soap easily, waste water containing soap does not impose the problems of water pollution commonly encountered with synthetic "hard" detergents (*see* DETERGENTS AND WETTING AGENTS).

Soap Materials.—Almost any fat can be used in making soap, but the principal requisite for a successful commerical product is an economical and proper balance of stocks to yield a soap that will exhibit the necessary soaplike properties over a wide range of conditions. Properties of a given soap (*e.g.*, sodium stearate) are evident only under certain conditions, as noted above. This characteristic behaviour necessitates the "blending" of fats to produce a widely applicable product. Thus, tallow is much used in making soap, but tallow alone will not yield a toilet soap that will give a copious lather over a wide range of temperatures in either soft or hard water. Therefore the tallow is commonly blended with a proportion of coconut oil, although palm kernel oil may be substituted. Various blends of oils and fats are used for laundry soaps, and almost every known fatty oil at one time or another has found its way into soap manufacture.

The most generally used stocks among the animal oils are tallow and grease; of the vegetable oils, coconut, cottonseed, corn, soya, palm, palm kernel, and olive are used. Fish and whale oil were used in making soap in increased quantities after the process of hydrogenation (*q.v.*) enabled them to be hardened to proper consistency and largely deodorized. Rosin (in saponified form) is an ingredient of yellow laundry soaps. Certain alkaline materials (builders) are almost universally present in laundry soaps. The most important are sodium silicate (water glass), sodium carbonate (soda ash), and various phosphates. These materials give increased detergent action and are not to be thought of simply as fillers, such as starch and barites, which are added to

reduce manufacturing costs. For saponification of the fats, large quantities of caustic soda are required by the industry, together with lesser quantities of caustic potash for soft soap manufacture. Ammonia, triethanolamine, and various organic bases are used to make special soaps.

Manufacturing Processes.—Soap is manufactured by the boiling, hydrolyzer, or semiboiling process.

Boiling Process.—The object of the boiling process is to produce neat soap in purified condition free from glycerol, the starting material for making bars, flakes, beads, and powders. The process is conducted in a series of steps called changes, and all the changes occur in the same piece of apparatus, the kettle (U.S.) or pan (British). The soap kettle is one of the largest pieces of processing equipment used by any chemical industry.

The first step begins with the introduction of the melted fats into the kettle together with some excess of caustic (soda) solution, which is added gradually. The whole mass is boiled with open steam from perforated coils within the kettle. A gradual emulsification occurs, the caustic solution reacting with the fat (saponification) to form soap and glycerol.

To separate the glycerol from the soap, the pasty boiling mass is treated with brine or salt. The contents of the kettle "salt out," or separate into two layers, the upper a curdy mass of impure soap, the lower an aqueous salt solution containing the dissolved glycerol. The basis of glycerol removal is the solubility of glycerol and the insolubility of soap in salt solution. The slightly alkaline salt solution, termed "spent lye," is run off from the bottom of the kettle and subsequently treated for glycerol recovery.

The grainy, curdy mass of soap remaining in the kettle after the spent lye has been removed contains any (usually traces) unsaponified fat that escaped reaction during saponification and also dirt and colouring matter present in the original oils. The "strong change," calculated to remove the last traces of free fat, is carried out by adding strong caustic solution and boiling for several hours.

The final stage (pitching and settling) transforms the soap into the neat form and removes the last traces of dirt and colouring matter. The soap, after the strong change, may be given one or more saltwater washes to remove free alkali, or it may be pitched directly following the strong change. In any case the open soap is partially "smoothed" by boiling with addition of water (or open steam condensation) until the concentration is attained at which the kettle contents separate into two, or sometimes three, layers. The upper layer is neat soap (kettle soap) of an almost constant composition for a soap from a given fat (approximately 70% soap–30% water); the lower layer is a nigre, whose composition varies between about 15% and 40% soap, depending on the "closeness" of the pitch. Since colouring matter and dirt (also salt, alkali, and metal soaps) generally are soluble in nigre but relatively insoluble in neat soap and because most of the impurities are dense and tend to settle, the nigre layer removes these impurities from the neat soap. The kettle boiling process requires from 4 to 11 days from the time the saponification starts until the neat soap is ready to be drawn off. Even today the complicated stages of the boiling process are usually conducted by the soap boiler using the appearance and taste of the soap as criteria for judging its condition. The utilization of the finished neat, or kettle, soap in making the final commercial product is described below under *Soap Products*.

Hydrolyzer Process.—The boiling process requires that individual batches of fat be treated in separate units (kettles). Continuous processes have replaced the older batch process by 50–75%. Where, in the kettle, alkali is used for saponification, in the hydrolyzer process the fat is "split" into fatty acids and glycerol by means of water at high temperature and pressure in the presence of a catalyst (zinc soap). The splitting reaction

$$\underset{\text{fat}}{(RCOO)_3C_3H_5} + \underset{\text{water}}{3HOH} \rightleftarrows \underset{\text{fatty acid}}{3RCOOH} + \underset{\text{glycerol}}{C_3H_5(OH)_3}$$

is carried on countercurrently, usually in a vertical column 50 ft. or more in height. The molten fat and water are introduced continuously into opposite ends of the column, while the fatty acids and glycerol are withdrawn continuously. The water, almost insoluble in fat at low temperatures, becomes soluble to the extent of 10–25% at 400°–480° F, the reaction being greatly speeded up by the solubility effect.

In about one hour a given unit of fat passes through the column and is 99% converted to fatty acids, which, as drawn off, are continuously distilled under vacuum to effect purification. They are then neutralized with a caustic solution (NaOH or KOH) containing the proper amount of water to yield neat soap phase. As in the boiling process, this phase constitutes the base material for further processing into bars, flakes, powders, pastes, or other products.

The operations necessary to obtain final products can be combined sequentially with the hydrolyzer to give a continuous flow from the original fat to the final marketable product, the whole sequence occurring in a matter of hours as contrasted with the 4–11 days required by the boiling process. The by-product glycerol is purified and concentrated synchronously with the fatty acid production.

Considerations of time, factory space, quality, and control of product characteristics contributed to the introduction of this process into the soap industry.

Semiboiling Process.—The semiboiling process may be thought of as ending when the saponification change is completed. The fat is saponified, usually in small quantities, as completely as possible by boiling with a caustic solution. The saponified mass containing the glycerol is run into frames for cooling and solidifying or, in the case of soft (potash) soap, is packed directly into containers. This process has the advantage of allowing, by water adjustment, the direct production of neat soap, middle soap, or, in some instances, nigre (liquid) soap.

Soap Products.—Hot neat soap phase from either the boiling or the hydrolyzer process forms the starting material for many products.

Framed Soaps.—These are usually made by pumping molten neat soap into molds, or "frames," to cool. The solidified soap is then cut and stamped into cakes. Such soaps contain about 70% soap to 30% water. For laundry purposes, builders such as water glass, phosphate, etc., are "crutched," or mixed in, before solidification.

Floating Soaps.—Formerly, most of these were made by crutching air into the neat soap before cooling. Rapid chilling or "freezer" processes are now in use. The molten neat soap is either aerated and then rapidly chilled or chilled to a semisolid condition and then aerated. The two procedures give bars of somewhat different lather and other characteristics.

Milled Soaps.—Most toilet soaps of better quality are made by the milling process. Originally developed in France as a way of allowing the cold incorporation of fine perfumes that could not withstand the heat of the crutcher, the milling process is now applied to the bulk of toilet soap manufacture. It produces a glossy, waxy bar that is characteristically smooth and has more pronounced qualities of lather and solubility than unmilled soap.

The most widely applied milling process requires that the hot neat soap first be reduced to ribbon or flake form. The molten soap flows between two hollow, rotating metal rolls, one warm, one cold. The warm roll spreads the soap into a thin film that solidifies and adheres to the cold roll. The film is removed in ribbons which are dried in an oven to about 12–16% moisture. After the dried, cooled ribbons have been broken into flakes, they pass into an amalgamator, where perfume, dye, and any special ingredients are mixed into them. The amalgamated soap then passes between successive polished metal (sometimes marble) rolls, which press the perfume and any other added ingredient evenly into the soap. The sheet of soap from the final milling roll is cut into ribbons and flakes and then subjected to pressure in a "plodder," a screw device operating in a cylinder where the flakes are forced together and cemented into a continuously extruded bar or cylinder of soap. The soap coming from the plodder is cut into lengths, which are then stamped into bar or tablet form and wrapped.

Flaked or Chipped Soaps.—Soap flakes are ribbons of dried soap broken into short lengths. The development of packaged soap flakes came about mainly between 1920 and 1930 and in the United States was concomitant with the introduction of small-scale, household, power washing machines, which created a great demand for soaps that were more soluble than the old bar soaps or chips cut from bars. By 1939, 0.002-in.-thick flakes were produced, compared with early products, which were of the order of 0.1 in. thick.

Neat soap forms the basis for soap flakes. The crutched soap, containing builder if desired, is run onto the rolls (*see* above, *Milled Soaps*), spread into a film or sheet, and removed in the form of ribbons. These are then dried to proper moisture (usually 5–15%), broken into flakes, and packaged. By proper control of moisture, fat stocks, and pressure, increased transparency is sometimes imparted to the flakes.

Granulated Soaps, Beads, and Powders.—Flaked soaps took a long step in the direction of a product of increased effectiveness in use. About 1926 occurred a new development, applying spray drying methods to neat soap. Early powders and granulated products were made by grinding flakes to the desired fineness, but such products when coarse were slowly soluble and when fine lumped and "balled" in water. When hot neat soap is sprayed into a current of heated air, the atomized soap particles dry and expand into bead or bubble form. The puffed particles, of spongy texture, exhibit a very high rate of solution and do not have the dusty quality of other types of powdered soap. Denser unpuffed particles may also be made by spray drying at lower temperatures.

Medicated Soaps.—Soap itself has germicidal power against some organisms, and many of the cationic synthetic detergents are strongly germicidal. Soap is used as an enema and as an emetic. Substances such as hexachlorophene are sometimes used in soap to suppress formation of the odorous compounds of perspiration.

Textile Soaps.—In the preparation of wool fabrics soap is required in three stages: scouring of raw wool to remove wool grease, scouring the yarns and cloth after oiling, and in fulling or milling. Soap is used for degumming raw silk; for cleansing rayon, silk, and cotton before and after dyeing; in the dye bath; and in printing. For all these purposes ready solubility is required, and hence soda soaps of low titre (*e.g.*, red oil, olive, corn) and potash soaps are preferred. Synthetic detergents have entered widely into the textile field because of their stability toward acid and hardness in water. They are used in scouring both cotton and synthetic fibres.

Shaving Creams.—Shaving creams are usually mixtures of potash and soda soaps of tallow and coconut oil. The cream is prepared by semiboiling methods. Bar shaving soaps and powders are of similar composition except for water content, which is kept at a minimum. Latherless or brushless shaving creams are not strictly soap at all but modifications of vanishing or cold cream formulas. In any type of lather, water is the principal beard-softening ingredient.

Metallic Soaps.—The term metallic soap refers to the fatty soaps of the alkaline earths and heavy metals. Their chief common characteristic is insolubility in water, hence their application (copper soap, aluminum soap, etc.) as waterproofing agents. The lubricating grease industry uses large quantities of calcium, aluminum, and other soaps to produce greases of various types, since another important characteristic of the metal soaps is their ability to gel in mineral oils. The demand for extreme pressure lubricants brought about the development of transparent, or clear, greases which are usually prepared by addition of aluminum soap, alone or in combination with other soap, to oil of the desired character.

See also OILS, FATS AND WAXES; ALKALI MANUFACTURE; DETERGENTS AND WETTING AGENTS; GLYCEROL; and references under "Soap" in the Index.

BIBLIOGRAPHY.—J. Alexander, *Colloid Chemistry* (1926); G. Martin, *Modern Soap and Detergent Industry*, 3rd rev. ed. (1951); A. E. Alexander and P. Johnson, *Colloid Science*, vol. ii (1949); A. W. Adamson, *The Physical Chemistry of Surfaces* (1960). (R. H. FE.)

SOAP PLANTS, the name given to numerous herbs, shrubs, and trees that contain the poisonous glucoside saponin. Various parts of such plants form a lather in water and may be used for cleansing purposes, as the soapbark (*Quillaja saponaria*) and the soapberries (*Sapindus* species). Other examples are the soapwort, or bouncing Bet (*Saponaria officinalis*), and the cowherb (*S. vaccaria*).

In the southwestern United States various plants called *amole* by the Indians are similarly utilized, especially the small agave (*Agave heteracantha*) and the soaproot (*Chlorogalum pomeridianum*). Other soap plants of the western United States are *Zygadenus* species and the sand lily (*Leucocrinum montanum*).

SOBAT, a river of northeast Africa, enters the White Nile (Al Bahr al Abyad) south of Tawfiqiyah, Sudan. The Sobat basin includes the southwest Ethiopian Plateau and the marshy eastern part of the southern Sudan depression with indeterminate limits to south and north. The Sudan-Ethiopia border from 6° 10′ N to 8° 40′ N is marked by the Sobat's tributaries, the Jokau, Baro Pibor, Akobo, and Kaia rivers. The Sobat is about 220 mi. (354 km.) long from the confluence of the Baro and Pibor. Including its longest tributary, the Baro, the distance from source to mouth is about 460 mi. (740 km.). The Baro rises in Ethiopia at a height of 7,000 ft. (2,134 m.); it descends rapidly from the plateau, dropping 3,000 ft. (900 m.) in 45 mi. (72 km.), and passes through a narrow gorge 3,000–4,000 ft. (900–1,200 m.) deep. Below the gorge the Uaddessa joins the Baro from the north. Before Gambela is reached the river begins its course across the plains. It enters the Sudan at its confluence with the Khawr Jokau. On the north bank, 15 mi. (24 km.) downstream, lies the Khawr Machar which acts as an overflow channel into the Machar Marshes during the flood season. The Pibor drains from the south a vast area of indeterminate drainage reaching almost to the southeast extremity of the Sudan. Its main tributaries, the Gila and the Akobo, flow off the Ethiopian Plateau from the east, the Gila by a series of cascades through a break in the mountains, while the Akobo joins the Pibor at Akobo. The Sobat flows northwestward for 180 mi. (290 km.) from Nasir in a bewildering series of meanders to its confluence with the White Nile; by road Nasir is just over 100 mi. (160 km.) from the confluence. Downstream from Nasir the Sobat receives two sizable tributaries from the south: Khawr Nyanding and Khawr Ful Lus, joining the river 120 and 15 mi. (193 and 24 km.), respectively, above its mouth. Approaching the White Nile the river becomes deeper and narrower, flowing between definite banks. At its mouth it is about 400 ft. (122 m.) wide with a depth varying with the season of from 18 to 30 ft. (5–9 m.). Its discharge when in flood averages 27,192 cu.ft. (770 cu.m.) per sec. (October–November), ponding back the White Nile; the whitish sediment brought down by the Sobat accounts for this name. At low-water stage in April the White Nile flows back up the Sobat as far as Khawr Ful Lus. The Sudan government planned to construct a dam on the Baro as part of the Equatorial Nile Project (*see* NILE), with the aim of developing storage of the White Nile water nearly to its maximum.

The Sobat river system is navigable during the period June–December, via the Baro for shallow-draft steamers to Gambela, and up the Pibor to Pibor Post for small craft. Vehicles can cross the Sobat by ferry just below Khawr Ful Lus. The inhabitants of the drainage basin are Nilotic peoples including the Dinka, Nuer, Shilluk, and Anuak. The dry season pastures (Dinka, *toich*) provided by the fall in river level are a basic factor in the life of these cattle-owning people.

An Egyptian army expedition ascended the Sobat for a short distance in 1841. Nasir was founded as an anti-slavery patrol post in about 1875. Exploration above Nasir was completed by J. B. Marchand who on his return journey to France from Fashoda ascended the Baro to the foot of the Ethiopian Plateau; by the Italian V. Bottego; and by numerous British army officers (such as Capt M. S. Wellby and Brig. Gen. H. H. Austin) and administrators at the beginning of the 20th century. (H. R. J. D.)

SOBIESKI, JAN: *see* JOHN III (Jan Sobieski).

SOCAGE, a type of feudal land ownership whereby a freeman held a fee simple estate as tenant of a lord in return for render-

ing periodic agricultural or economic services. With the decline of feudalism and the rise of a money economy, "free and common socage" became the most desirable form of land tenure, and on the restoration of Charles II in 1660 almost all species of tenure were converted by statute into socage tenure. (*See* REAL PROPERTY AND CONVEYANCING, LAWS OF.) Free and common socage was also the most common tenure by which the king granted lands in the North American colonies. Since tenure has no modern significance in Anglo-American property law, socage tenure is no longer important; if its semblances are retained, it is for nonlegal reasons.

Although socage developed under Norman feudalism in England, it is Anglo-Saxon in origin. The term is derived from the Old English word soc or soke and is frequently associated with sac, in an alliterative jingle ("sac and soc"). The precise meaning of both terms is uncertain, but they refer to the district in which a lord may hold court, to the right of a tenant to make suit in such a court, and to his duty to use the court. The "sokemen," or persons holding by socage tenure, were a class of tenants midway between the military or chivalrous class and the bond tenants or villeins. Although freemen, they performed many agricultural services similar to those performed by the villeins.

If a freeman had to own land under the theory of feudal tenure, socage was the least objectionable form. Minors inheriting socage lands were not, as in military tenure, wards of the lord who held the land but came under guardianship of the nearest relative who could not inherit the lands, and this guardian, unlike the lord in a wardship, could not keep the profits during minority but had to account to the heir for them. Neither, under socage, could the lord or guardian "sell" the minor heir in marriage for his own personal profit. As in the case of military tenure, however, the death of the owner required the heir to pay the lord a relief to claim his inheritance.

At an early time the services due the lord were commuted into fixed money payments called "quitrents." These at first represented the economic value of the land, but with the long continuous fall in the value of money they became so insignificant as not to be worth collecting. Sometimes the services were purely nominal as the result of gifts to younger members of the family, to friends, or as the result of a sale for a capital sum with a nominal quitrent reserved to preserve the form of the tenure.

The 1681 Charter of Maryland granted to Lord Baltimore lands to be held in free and common socage yielding to the king as lord "two Indian arrows of those parts to be delivered every year on Tuesday in Easter week." The feudal service attached to the grant of Pennsylvania was the delivery to the king of two beaver skins annually.

In some American colonies, the king permitted the grantee in socage to establish his own manor, that is, to regrant portions of the land to tenants in return for an annual rent, although this could not be done in England after 1290. Thus some property owners in Pennsylvania rendered to the members of the Penn family a red rose at midsummer, a service that is now continued only for ceremonial reasons. In New York and some other parts of the North American colonies, those holding grants from the king sought to impose substantial quitrents on their grantees. Thus even after the Revolution the Van Rensselaer family granted lands on the upper Hudson to a settler "forever" upon a yearly rental of a certain number of bushels of wheat. So distasteful was this feudal concept that these quitrents contributed to uprisings such as Bacon's Rebellion in Virginia in 1676 and to some major antirent disturbances in New York in the mid-19th century.

(A. DM.)

SOCCER, the name given in the United States and Canada to the game called, with a few minor exceptions, football or association football in all other parts of the world. For a description of the game as played outside America, *see* FOOTBALL: *Association Football (Soccer)*.

Many different stories are current about where the first games of soccer were played in America. A marker on the Common in Boston, Mass., refers to the fact that the game was played there. There is little doubt that the first intercollegiate "football" game between Princeton and Rutgers was actually a version of soccer, and it has been fairly well established that the game was introduced into the United States by the British.

Like American football, soccer is played by two teams of 11 men each, employing a round leather- or rubber-covered inflated rubber bladder slightly larger than a volleyball and slightly smaller than a regulation basketball. It is called the booting game for good reason: only the goalkeeper is permitted to touch the ball with his hands while it is in play.

The main propulsion of the ball is given by a player's foot, although he can use any part of his body except his hands and arms to trap, deflect or project the ball. Heading the ball is considered a major movement in the game.

Should a player other than the goalkeeper handle the ball deliberately or unintentionally, a penalty is called against him and the opposing team is permitted a free kick. When the infraction occurs in the penalty area of the defending team, the attackers are allowed a penalty kick from a point 12 yd. out and directly in front of the goal.

A regulation game in soccer consists of two halves of 45 minutes' duration each. In the event that a 90-min. game cannot be played for a specific reason, the time may be shortened, but the halves still must be equal. Some college conferences and scholastic leagues play their games in equal quarters rather than halves, but no regular games last longer than 90 min. unless there is a provision for overtime play in order to determine a winner.

Play.—The maximum length of a soccer field or pitch, as it is known, is 130 yd., with the minimum set at 100 yd.; the maximum width is 100 yd. with the minimum 50 yd. The markings of the pitch include two touch lines (the length of the field) and two goal lines (the width of the field) to make a rectangle. In addition there is a midfield line across the width of the field with the centre kickoff circle (10 yd. in diameter). Before each goal is a penalty area 18 yd. long by 44 yd. wide. Inside the penalty area and also before each goal is a goal area 6 yd. in length and 20 yd. in width. The two goal posts are 8 yd. apart supporting a crossbar 8 ft. off the ground. Normally nets are required to be attached to the goal posts and crossbar and extending behind the goal.

The penalty kick area is denoted by a small (solid) circle directly in front of and 12 yd. from the goal. Quarter circles of 1-yd. radius are marked at each corner for the corner kick, which is given when the defending team last touches the ball before it goes over the defending team's goal line outside the goal. Corner flags, preferably of some bright colour, are placed in the four corners of the field.

To score a goal, the whole ball must pass over the goal line between the two goal posts and under the crossbar, either on the ground or in the air, without being carried, thrown or motivated by arm or hand of an attacking player. If a defending player accidentally kicks or deflects the ball into his own goal, it counts as a goal for the opposing side.

Three types of kicks are awarded for rule infractions including penalty kick and direct free kick from which goals can be scored directly. In order to score on an indirect free kick the ball must be played by another player before entering the goal.

When a ball goes over the touchline, it is awarded to the opponent of the team that last touched the ball before it went out of bounds. A throw-in is taken at the point where the ball went over the touchline. It is the sole instance during a game when a player other than the goalkeeper is allowed to handle the ball. The colleges have amended the throw-in rule to make it a kick-in.

The offside rule is an important rule of the game. A player is offside when there are not two defending players nearer the goal than himself at the time the ball is passed to him, with the following exceptions: from a goal kick; a corner kick; a throw-in; when he is behind the ball when the pass is given; when the ball comes from an opponent; when he is in his own half of the field.

One of the American deviations from international rules is that substitution of players is permitted in American soccer. However, the different governing bodies and sectional leagues have varying limits on substitutions. On the whole, the deviations from the international rules are rather few, which is one of the reasons why

soccer lends itself easily to international contest.

Competition and Organization.—Soccer is a relatively simple game to understand, the basic rules being neither complicated nor difficult. It is also an exciting game, with play going on continuously in the open. Compared with American football it is inexpensive to support, and there are no special physical requirements for the participant, hence wide participation is possible. While the popularity of soccer in America has not always been great, it is still expanding in schools and colleges, where it is a seasonal sport, played for about three months in the autumn.

The season of amateur and professional soccer runs for approximately ten months with a layoff during the summer. In certain areas competition is halted during the winter because of poor field conditions.

In addition to the various league and state competitions, three major national tournaments attract entries. The National Challenge cup competition for the Dewar trophy is the open tourney in which professional and amateur teams compete. The event was first held in the 1913–14 season and regularly thereafter. In the 1923–24 season, entries for the National Challenge cup became so numerous that it was decided to inaugurate the National Amateur cup competition to be restricted to amateur teams and amateur players. The National Junior Challenge cup, limited to junior teams, began in the 1934–35 season, but the final game between eastern and western sectional champions is not always played.

The United States Soccer Football association (U.S.S.F.A.), organized in 1913 and successor to other predecessor groups, is the governing body of the sport in America. This includes amateurs and professionals, who are allowed to play side by side on the same teams without any change in status. College and school groups have their own organizations and are merely associate members of the U.S.S.F.A. Leagues and state associations of amateur and professional groups are direct affiliates of the U.S.S.F.A.

Only one organized professional league is recognized, the American Soccer league, which operates chiefly along the eastern coast and includes New York, Newark and Elizabeth, N.J., Philadelphia and Baltimore. In 1960, the International Soccer league was formed to include visiting foreign teams and a New York side.

In the late 1920s and early 1930s, a professional soccer player was able to earn money enough so that he could devote full time to the game. Currently, however, even the highest paid professional cannot support his family by playing. Whereas European players devote regular days and hours each week to practice, few teams in the United States hold regular drills. A team designated to represent the United States in international competition usually has opportunity for no more than one or two practice games, which is reflected in the poor showing of United States teams in such competition. Further, few clubs, leagues or associations have the personnel to conduct programs to promote development of young players.

Soccer in the United States is an amateur rather than a professional sport. Few soccer teams have their own stadiums; most of the big international games are played in baseball parks or municipal or school stadiums. For the most part, however, regular games are played in parks or on open fields. (M. Mi.)

SOCHI, a town and resort of Krasnodarskiy (Krasnodar) Kray of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the Black Sea coast, at the foot of the western part of the main Caucasus range. Pop. (1959) 126,779. The town now extends (administratively) for about 50 mi. (80 km.) along the coast from Lazarevskoye in the north to the border of the Georgian S.S.R. in the south. The presence of hot mineral springs and the warm climate (January average 6.1° C [43° F], July average 22.8° C [73° F]) have made it a popular holiday and health resort. It has 6 tourist centres, motels, and over 50 sanatoriums. Sochi is on the main Tuapse–Sukhumi railway and has passenger vessel connections with other Soviet Black Sea ports. There are a number of food-processing industries. (R. A. F.)

SOCIAL ANTHROPOLOGY, a social science concerned with the comparative study of human societies. Sharing this aim with sociologists, many social anthropologists regard themselves as sociologists of a specialized kind. However, social anthropology has dealt primarily with what are known as simpler, primitive, preliterate peoples, while sociology has stressed study of more advanced peoples. This division has tended to disappear as social anthropologists investigate advanced Oriental and Western societies, and sociologists move out to study the development of formerly primitive societies. Social anthropologists usually study small communities such as villages or special groups in factories, camps of displaced persons, or special aspects of social life such as kinship. But both disciplines share and contribute to a common theoretical framework. Social anthropology uses above all an intensive type of inquiry that ordinarily involves living among the people to be studied, learning their language, and closely observing their day-to-day activities. Discrepancies between word and act are often of great importance to the social anthropologist. So closely may he become involved that he may engage in participant observation; that is, he cooperates with the people by engaging in their activities himself—going on a fishing expedition, taking part in a religious rite, keeping their tabus, exchanging property with them as if he were a member of the society.

It has increasingly come to be recognized that the terms "society" and "culture" refer to different aspects of the same basic human actions and relations. "Society" and "social anthropology" emphasize the human element—interpersonal and intergroup relations. "Culture" and "cultural anthropology" lay more emphasis on creative achievement, the objects and ideas that are brought into being and transmitted from one generation to another. Cultural anthropology tends to be the wider, embracing, for instance, simple technology and linguistics as well as ethnography and investigation of social structures. But the emphases shade into one another and there is a large area of common ground. Anthropological opinion is not unanimous on this point. But the essential unity in major fields of interest is shown, for example, by the occasional use of the conjoint term "cultural/social anthropology."

Modern social anthropology is characterized by theoretical studies (of social structures and processes) that are comparative, at least by having primitive societies as one pole of reference. It may be distinguished from the more descriptive field of ethnography, though much published work in social anthropology also contains descriptive material. The subject owes much to theoretical analyses by sociologists, including the older European writers E. Durkheim, M. Weber, V. Pareto, and M. Mauss (*qq.v.*), and such modern U.S. contributors as K. Davis, R. Merton, and T. Parsons. Less has been gained from psychology, though the theory of the family and kinship have been influenced by the work of Freud. Major developments of social anthropology early in the 20th century were due to W. H. R. Rivers, A. R. Radcliffe-Brown, and B. K. Malinowski (*qq.v.*) in Great Britain; A. L. Kroeber, R. Linton, A. Goldenweiser, and R. H. Lowie (*qq.v.*) in the U.S.; and R. Thurnwald in Europe. Later developments were more international; *e.g.*, the work of C. Lévi-Strauss in France. Though originally primarily a Western interest, by the 1960s social anthropology was enriched by Asian, African, and other scholars.

In the early years attention focused on structural and functional studies of family, marriage, kinship and local grouping, magic and witchcraft. In the second half of the 20th century, while kinship studies remained the core, social status and power attracted attention in the investigation of political and legal systems, and systems of social stratification. There was more concern also with systems of ritual and religious ideas. Interest shifted from African peoples, who had occupied social anthropologists for a quarter of a century, to peoples in India, Southeast Asia, Latin America, and Oceania. By the 1960s a more dynamic approach was apparent, particularly in problems of social process (as well as structure). Analyses of social change were no longer considered a separate branch of the subject, being incorporated in most studies. Attention was paid also to the significance of the time dimension in the maintenance and replacement of social units and institutions. Much greater use was made of quantitative data, as in studies of family and domestic group relations, marriage, divorce, and economic transactions.

Social Structure and Organization.—The basic data for social anthropology are usually defined as social relations arising from

the necessities of living in society. Social relations have regularities or patterns shown by repetition—as when a man plays with his children every evening after he comes home from work. These relations tend to compose a system, being linked with, and to some degree determined by, others. So a man's regular relations with his children in play are tied in with his relations with them at meals and in the house generally, and depend in large part on his relations with their mother. They may also depend on his economic relations—as when he is forced to go away from home on business—or on his legal relations—as when a court awards him custody of his children at divorce. The main elements of the interdependent system of social relations, looked at somewhat abstractly, form the social structure.

Social structures are sometimes stated (as by E. R. Leach and Lévi-Strauss) to be ideal or formal models representing the main principles of a society or social system. Less formally, social structures may be regarded as comprising the more permanent groups and other continuing features of the social system (E. E. Evans-Pritchard), the network of social relations of the system (Radcliffe-Brown, F. R. Eggan), or the crucial interdependence of role relationships in a society (S. F. Nadel). These, however, are all linked concepts, since the regularity, interdependence, and permanence of social relations depend largely on the ideal type laid down by the society, and perception of all this involves considerable abstraction of formal qualities by the observer.

Social structure is like the anatomy of the society—it provides its major shape. But the conduct of individuals varies; some follow the rules laid down and others do not. Not every man abides by the legal and moral rules of marriage—nor does every man or woman marry. Some fathers make generous provision for the education of their children and try to guide their social development. Others allow them to fend for themselves. In all fields of social action different arrangements by individuals arise from choices they make between alternatives. The total body of such actions can be termed social organization.

Linked with these concepts is the notion of social values, the system of preferences that govern action in any society. By adhering to conceptions and principles embodied in the system of social values members of the society maintain their social structure.

Structure of Primitive Societies.—The most obvious distinction of a primitive society is its size—it may consist of only a few hundred or a few thousand people, very rarely hundreds of thousands. Though demographic studies have been shown to be important for structural analysis (M. Fortes and others), size alone is not the most important factor in the structure of a primitive society.

In such a society almost all interpersonal relations are face to face—except where Western influence has introduced written communication. Moreover, for technical and social reasons, people are apt to move about much less. Hence, any one person is apt to have relationships with fewer people than in Western society and those relationships are likely to be more permanent. Almost every relationship is between people already acquainted; the impersonal day-to-day contacts of Western society—as with a sales assistant or a government official—are unlikely to occur.

This factor of proximity—expressed formally in terms of neighbourhood associations and local groupings—is extremely important in social relations. The units of household, homestead, hamlet, village, and urban ward vary greatly in different societies. They serve as a basis for economic and other cooperation; they tend to determine the character of kin grouping by virtue of the property and other rights usually associated with them. In a primitive society, the range of occupations and special social activities is much narrower than in Western society; there is far less differentiation of roles. Most of the people are primarily engaged in obtaining food with fairly similar technical equipment. This sharply limits the types of relationship possible for a given person. But the people with whom one is in contact may play several roles; relationships are many-stranded. The same person may be father, war leader, judge, economic boss, teacher, and priest.

This many-stranded quality confronts the social anthropologist with a problem of disentangling for analysis (sometimes arbitrarily) such categories as economic, political, legal, and ritual relations. The overlap leads the anthropologist to emphasize interdependence; *i.e.*, to seek the function of the relationships in any one field in maintaining the total pattern of the society.

In any society, people arrange themselves in groups—families, villages, clans, professional associations, age sets. In a primitive society, groups tend to be closed rather than open; *i.e.*, most groups recruit through birth, and membership is compulsory. In a society of this kind a man does not normally belong to groups having purely voluntary membership, though in some communities clubs and secret societies may be recruited on such a personal basis (*see* SECRET SOCIETIES, PRIMITIVE).

In a primitive society more behaviour concerns prescribed group membership than in Western society. Since there are fewer differences in specialist skill and interest, principles of age, sex, neighbourhood, and kinship are the common bases for group formation. The people are very dependent on, and rarely able to change, group membership; conformity to wishes and standards of others in the immediate environment is important to well-being, even to survival.

The rate of social change (barring outside influences, natural disaster, economic penetration or conquest) is slow, though all primitive societies have not been static, as has sometimes been assumed. But when a small primitive society comes into contact with Western civilization, the changes are bound to be rapid and may be catastrophic. New techniques alter patterns of work and cooperation; new avenues of employment open; the introduction of money greatly accelerates exchange, provides a more general measure of value, and wealth becomes more easily storable. Missions, government, and other development agencies introduce new rules and values, and affect the institutional life of the people. Traditional social controls lose force and social disintegration often results. In extreme cases (such as African shantytowns) crime may increase greatly. Yet some societies successfully raise material levels of living and increase the scale and complexity of their economic and political organization without radical alteration in morale. Some societies react with new social and ritual forms, such as the Ghost Dance of the North American Indians, with its conscious and organized revival of native culture; or cargo cults of New Guinea, ritual attempts to secure trade goods that are in the hands of Europeans (*see* NATIVISTIC MOVEMENTS). These have led to far-reaching and subtle modifications in social structure that are not simply mixtures of primitive and civilized ways; they are fresh social phenomena, new structures, demanding special study and new techniques of approach.

Sex and Age Grouping.—All societies make distinctions by sex and age. In primitive societies women are primarily responsible for young children and for most domestic tasks, while jobs that involve heavier work, greater distances from home, or violence (hunting and fighting) are male tasks. Formal control of internal group affairs and of external relations is almost entirely in the hands of men, though women may exercise great informal power. Among the Iroquois, women had the power to nominate and depose chiefs, but only men could hold the office. Queens or female chiefs, as among the Lovedu tribe in the Republic of South Africa, generally lack real power, or are symbolically equated with men. Division of labour by sex varies greatly; in one society men till the ground, in another they leave this entirely to women. Violation of local rules may cause mirth or abhorrence; the tasks of one sex are often ritually forbidden to the other. Invariably, as in Western society, sex is symbolized in differences of dress, ornament, and etiquette. It may be emphasized by ritual procedures or in myths.

The principle of seniority is equally used for social distinctions. In a few societies old people are regarded as a burden, but almost invariably old age is highly respected. Where there is no writing and where skills and cultural heritage alter little through the generations, old men are the obvious repositories of wisdom and experience. In its extreme form, the respect for the aged becomes gerontocracy (*q.v.*), rule by old men. In such societies, old men

keep for themselves special privileges, often including prior rights to marry younger women, and their position is often backed by magical knowledge and tabu. Sometimes the age principle is formalized in a system of grades. The change in status from child to adult is marked by a ceremony of initiation (usually including an ordeal). Those initiated together form a group (*see* AGE SET) with special rights and obligations. Ritual occasions, including formal transfer of insignia, may mark the movement of a group of age-mates from one status to the next. Even where no age set exists, the principle of seniority is usually of great importance.

Kinship.—A major contribution of social anthropology has been a refinement of understanding it has brought to the study of kinship (*q.v.*). Kinship is the social recognition of real or putative biological ties formed by procreation and marriage. The continuity of any society depends on the procreation and training of new members. The kinship system provides legitimacy for the procreation of children; responsible persons and social units undertake their care and training and place them in a defined position in society.

The minimum kinship unit is the elementary (nuclear, individual) family, consisting of father, mother, and children. Almost every adult is a member of two such families, called by W. L. Warner the family of orientation (that into which one is born or adopted), and the family of procreation (that in which one becomes a parent). The interlocking of these families creates a complex framework of kin relationships, consisting of consanguineal ties (by blood) and of affinal ties (by marriage). Adoption is the simulation of consanguineal ties. Most human societies use kinship in ordering economic, political, judicial, and ritual affairs.

The study of kinship in anthropology began with work in kinship terminology (*q.v.*) by L. H. Morgan. He and his contemporaries sought to classify systems of kinship terminology as stages of human evolution, equated with levels of political and technological development. Where they thought terminology was inconsistent with existing social structure, they supposed a previous state of society that would fit. Lowie, Radcliffe-Brown, and Malinowski attacked this evolutionary approach, insisting on analysis and comparison of contemporary societies to establish general principles. Speculation about origins of human institutions, that provided the main drive for the beginning of anthropology, was abandoned as unprofitable.

It is now generally recognized that kin terms cannot be understood without reference to the total pattern of kin behaviour. Nevertheless, terminologies provide important clues to the structure of a society. They reveal how people classify their kin, reflect the system of descent groups, and provide linguistic symbols for much social behaviour.

The concept of the family as a kinship unit is clearly separated from that of a domestic unit such as a household, though these units overlap in fact. Statistically the elementary family is frequently the residential norm, occupying a separate dwelling. But in many societies, as in Malaya, a young couple begin married life in one of the parental households rather than in a new home (neolocal residence) of their own. Residence with the groom's family is termed patrilocal or more correctly virilocal; with the bride's family, matrilocal or uxorilocal. Kin in such a domestic unit are called an extended family (expanded family). A household of two or more siblings, their spouses, and children may be termed a joint family—fraternal or sororal as the case may be. In some societies, such as rural Turkey, the two types of structure often merge on the death of the parental head. Modern social anthropologists increasingly realize the significance of the developmental cycle in the structure and operations of these domestic groups.

The study of descent groups is crucial in modern social anthropology. Born (or adopted) into a family, one may belong also to a wider descent group. In defining this kin group, one or both parents may be relevant. Descent reckoned through one parent only is unilineal; such a descent group is a lineage. A patrilineage consists of male descendants (through males) of a common ancestor, together with their sisters and daughters; a matrilineage consists of female descendants (through females) of an ancestress, together with their brothers and sons. A lineage is normally corporate, its members exercising rights in common and collectively subject to obligations. A lineage may comprise any number of generations, but commonly is traced through five to ten. Genealogical connections of a large lineage often are inconsistent, even obviously invented and lacking historical validity. Lineage structure may be regarded as the result of a branching process, as when two or three founders of small lineages are represented as brothers. The groups thus comprise a single larger lineage in which the smaller descent groups are segments; this segmenting process may be very extensive. Evans-Pritchard, Fortes, Leach, A. W. Southall, J. Middleton, and others have shown that many political and ritual relations in a primitive society may depend on this segmentary structure.

In many societies the lineage is exogamous; *i.e.*, a man cannot marry a woman of his own lineage, and marriages then set up relationships between lineages. The structure of patrilineal societies (*e.g.*, Nuer, Tallensi, Tikopia), where a man belongs to the group of his father, differs markedly from those with matrilineages (*e.g.*, Nayar, Trobriands, Hopi, Dobu), where a man belongs to the group of his mother's brother (*see* DESCENT, SYSTEMS OF; MATRILINY).

The prevalent pattern of residence greatly affects the operation of the unilineal principle. Among the Tallensi (*q.v.*) households under brothers or brothers' sons are usually adjacent; households in one area belong to a larger lineage segment; and so on, reflecting the lineage structure. Among the Nuer (*q.v.*) the fiction of localized lineages is found, although men often live with matrilateral or affinal kin. In matrilineal societies the problem is more complex. All the men of a lineage may live together, but their children belong to the lineages of their mothers. In the Trobriands, this difficulty is met by having sons join their mother's brothers on adolescence. If all the women of a lineage live together, no man of the lineage may be present to direct affairs—unless, as among the Hopi (*q.v.*), brother and sister normally live in the same village. Among the Yao of Nyasaland this problem is met by allowing the eldest brother of each set of sisters (sorority group) to take his own wife to live with his sisters, so that he can take charge of their affairs. Among the Dobu of the D'Entrecasteaux Islands couples alternate residence between the lineage-village of the husband and that of the wife. Among the Nayar (*q.v.*) husbands only visit their wives; the joint households contain brothers, sisters, and sisters' children by their visiting husbands. Here, as often, lineage members other than the joint family do not live adjacently, yet maintain coherence through common rights, duties, rituals, and symbols.

The word "clan," often vaguely used in older literature, usually means either a group of people claiming unilineal descent from a common ancestor but unable to give a precise genealogy, or (as G. P. Murdock suggested) a compromise kin group with residential unity. In the past the clan (*q.v.*) has been described as an exogamous unit (as it often is), but this strict definition is unnecessary. The word is now confined to unilineal descent groups although the Scottish prototype is bilateral.

In some societies both matrilineal and patrilineal descent are recognized for different social purposes. This is known as double unilineal descent. Thus among the Yakö of Nigeria (as C. D. Forde showed) the household and the farming group are based on patrilineages; matrilineal groups, although not residential, are of great importance in the ritual and political system. Under Western influences larger kin groups tend to break up as more individual rights are asserted. A. I. Richards and Fortes have shown for Bemba and Ashanti (*qq.v.*) that in matrilineal systems fathers tend to take over traditional functions of mother's brothers.

By the 1960s attention was directed to nonunilineal (ambilateral or ambilineal) systems that allow a tie through either mother or father as a basis for lineage. Among the Maori (*q.v.*), R. W. Firth showed that one's descent group may be his father's or his mother's according to his place of residence and maintenance of land rights.

Many societies have no specific descent groups. Kin groups exist, but differ for each set of full siblings. Though not forming a distinct group for any other purpose, a man's personal kin all may be held responsible for his offenses. Among the Kalinga of

Luzon, responsibility for homicide theoretically includes the offender's third cousins (in practice only second cousins).

Descent should be distinguished from inheritance of property and succession to office. Some kinds of property may pass patrilineally in a matrilineal society, or a sister's son may succeed to an office in a patrilineal society. But in societies with no clear-cut descent groups, succession and inheritance tend to be patrilineal, as most authority goes through men.

Though only one side of the family may determine descent-group membership, close ties are maintained with all who marry out of a group and their offspring; keen interest is taken in their welfare. Within any body of kin the principles of generation or kinship grade, and of solidarity of the sibling group, are important regulators of behaviour.

Marriage.—All societies distinguish between casual sexual relations, or cohabitation, and a formal union to establish a family. Several forms of union are often recognized in the same society (*see* MARRIAGE, PRIMITIVE).

Marriage generates a relatively permanent relationship between kin groups through common interest in the partners and in the offspring. Very often such a relationship between families is held inappropriate within a fairly large group, a clan, lineage, or village; such a group is then called exogamous. Exogamy (*q.v.*), by ensuring social links between potentially hostile groups, performs an extremely important integrating function. Conversely, the extreme case is that in which marriage outside a specific group is forbidden. For this the term endogamy (*q.v.*) is used and also more loosely for a tendency to marriage within a group. Endogamy is apt to characterize religious and ethnic minorities in civilized societies. In India, every small subcaste tends to be endogamous, on ritual grounds.

Strict endogamy is rare in primitive societies, but rules specifying the group from which a man must or should choose his wife are extremely common. It is often felt that an affinal tie between two groups should be preserved generation after generation. In unilineal societies this is done by a rule of cross-cousin marriage, by which a man takes his wife either from the group to which his father's sister went or from the group from which his mother came. (He may also give his sister in marriage to a man of this group.) As examined by Lévi-Strauss, Leach, and G. C. Homans with D. M. Schneider, such preferential rules support the stability of relations between groups.

In all societies the formal beginning of a new marriage is marked by ceremony; gifts are exchanged even if only food. In many societies, where groups of men hand over women, bride-price or bridewealth (*q.v.*) is paid; *i.e.*, the groom's side gives far more than the bride's. Sometimes dowry is paid with the bride. These gifts may continue as annual harvest payments, as in the matrilineal Trobriands, where a man is compensated for caring for the children of his wife's lineage.

Monogamy is common in all types of society, and polygamy is normally the exception, even where it is the ideal. But many societies permit more than one wife (polygyny) and a few permit more than one husband (polyandry). Each usually is a separate marriage, though sometimes with a minimum of ceremony. In polyandry the co-husbands are usually brothers, as in parts of Tibet, or else do not reside with their wife, as among the Nayar who practised both polygyny and polyandry.

In all societies unregulated extramarital intercourse with a wife is felt to be a wrong—at least to the husband. Reaction to adultery varies from mild claims for damages or divorce to killing both parties. But some kinsmen—a husband's unmarried younger brother or, as in the Trobriands, a brother's son—may be privileged, or special license is allowed on ritual occasions. Premarital intercourse is commonly condoned; but premarital pregnancy is serious, since it threatens to create an illegitimate member of society with no defined position.

Most societies recognize divorce, though in some it is almost unknown. Plausible attempts have been made to relate divorce rates to descent-group structure, to the structure of the household, and to amount of bridewealth, but no adequate theory has yet been found if only because the definition of marriage and of divorce in different societies makes comparison of the stability of marriage very difficult.

Rules against incest are found in all known societies; they should not be confused with exogamy. To forbid marriage within a specific group is not the same as a universal rule against sexual intercourse with close kin, particularly those in one's elementary family. Incest is widely thought to carry supernatural penalties (to be a sin); sometimes, as in Tikopia, these are almost the only penalties.

Status and Stratification.—Status is the social position an individual holds, with attendant rights and duties. The word implies stratification on a vertical scale. The spatial analogy covers several meanings. A man may be said to occupy a high position when he is able to control, by order or by influence, other people's conduct; or when he derives prestige from holding important office; or when his conduct has earned the esteem of his fellows. Relative status is a major factor determining the behaviour of people toward each other; and competition for status seems to be a prime pursuit.

One's status varies with social context. For example, the position of a man in his kin group helps determine his position in the community. Among the Hopi (as Eggan showed) the lineage, although unnamed, contains the mechanism for transmitting rights to land, houses, and ceremonial knowledge and is thus vital for personal status. Among the Tallensi (Fortes stated) a mere lad who (having lost his father) is head of a household counts as an elder; a middle-aged man still under his father's roof is formally a child. Status may be governed by occupational considerations; thus in Negro Africa blacksmiths commonly form a separate group of low status. In the Hindu caste system, sweepers are at the bottom of the scale because they handle human excrement.

Status is closely correlated with etiquette and morality, and in many societies rises with the liberal use of wealth (*see* POTLATCH). C. A. du Bois stated that manipulation of the wealth-status system (among the Alor islanders) often demands great individual effort, aggression, chicanery, and an excellent memory.

Social classes are groups arranged in a hierarchical status system; class members interact mainly within their own group and to a lesser degree with those of higher or lower status. While classes are rare in primitive societies, some kind of stratification is common. Even an age-grade system, as in many East African societies, may resemble a class structure. In some societies, clans or lineages may be ranked generally as aristocrats and commoners, or graded from a royal clan to clans that are stigmatized for lowly occupation or slave origin. True class systems obtain in the West African emirates, with urban capitals from which (as Nadel showed) a ruling class, different in dress, speech, etiquette, and morality, dominates a semipagan peasantry. Most striking is the caste system of India analyzed by S. C. Dube, M. N. Srinivas, F. G. Bailey, K. Gough, A. C. Mayer, M. Marriott, and others. In Hindu villages there are usually members of a number of small endogamous groups (subcastes) based on traditional occupations, arranged from Brahmans to Untouchables, though the exact order of the middle subcastes is not always agreed on by all parties concerned. Contact with a person of lower caste (such as eating or drinking from his hands, bodily contact) pollutes one of a higher caste and necessitates ritual purification. Escape from caste is very difficult, though Western influences have tended to loosen some rules and urban life allows latitude to many individuals (*see* CASTE [INDIAN]).

Economic Relations.—Economic behaviour is socially conditioned; it hinges on and affects the system of social relations involved in the production and exchange of goods and services (*see* ECONOMIC ANTHROPOLOGY).

In primitive society economic relations are more personalized than in industrial society; categories of employer, employed, wage earner, capitalist often do not apply. Economic services are contributed not simply to earn a living but as part of a wider system of social obligations (*see* LABOUR, PRIMITIVE).

In almost all nonindustrial societies the main economic unit is the family or household rather than the individual. All able-bodied members share, under the direction of the head of the household,

such routine tasks as getting and preparing food according to recognized patterns of division or work by age and sex. Some tasks require a wider basis of cooperation than the household: a rice-transplanting team, a canoe crew, or a hunting band. Coordination of work may then depend on the rights and duties of kinship, on central community authority, or on ritual and magic. Malinowski showed this in his analysis of the importance of the gardening magician and his rites in Trobriand village agriculture. When large working parties are assembled, their beneficiary will offer food or perhaps (as in Africa) beer. He may also make gifts, but these do not usually reflect any precise measure of the service rendered. Later, he will return the service by working for others. Technological specialists generally receive additional payment that is apt to be traditionally fixed.

Property in any society is represented by sets of rights that entitle the owners to the services of other people. In primitive society these rights vary from unconditional ownership of material objects to highly complex relationships involving shared land, ritual apparatus, or magic. The most important resource to any primitive society is land, and problems of land tenure are often complex to a Western observer (*see* LAND TENURE, PRIMITIVE). Other kinds of property are generally owned in a way more familiar to Westerners; food may be personal or household property. The house itself will belong to the head of the household, subject to the rights of other members. Tools and implements are usually personally owned. Social titles, songs, dances, magical spells, and medicines may be personal or group property; many can be sold or exchanged.

Although most essential skills are possessed by typical adults in primitive societies, goods will constantly change hands and services are continually rendered to other people (*see* GIFT EXCHANGE; TRADE, PRIMITIVE; CURRENCY, PRIMITIVE). The principle of reciprocity, or basic compensation, was strongly emphasized in the theories of Mauss, Malinowski, Firth, Thurnwald, and M. J. Herskovits; the significance of exchange of women by transfer on marriage has been demonstrated by Lévi-Strauss. A prestation (a due, paid in goods or services) is normally part of an already established relationship between two people. If the parties are of roughly equal status, a return of equal value will be made, directly or indirectly. A voluntary prestation unreturned indicates that the giver is of higher status than the receiver; the basic compensation in this case is prestige.

A primitive economic system is usually related closely to the associated political and ritual system. Productivity is believed to depend on gods and ancestral spirits, and elaborate rites are designed to secure their aid or benevolence. The system of allocating land and of distributing produce is a basic way in which chiefs and other leaders maintain status and power.

Political Relations.—Political relations concern the orderly conduct of affairs of public interest within and between communities or societies. Although they essentially involve control of people's behaviour, not even in the simplest political community are political relations based simply on force.

Many primitive societies have neither legislative or judicial machinery nor group leadership above the level of the village or camp. Yet internal and external affairs are usually conducted in an orderly fashion, and anthropologists such as Lowie have extended the use of the word "politics" beyond the field of instituted government. But primitive political systems are still tentatively classified because of their great variety. In Africa alone they range from autonomous Thonga (*q.v.*) villages, based on kinship, to large kingdoms such as Nupe and Ganda (*qq.v.*).

Fortes and Evans-Pritchard distinguished three major types of African political system: (1) small stateless societies based on kinship (*e.g.*, Bushmen); (2) larger stateless societies based on segmentary lineage systems, with uncentralized authority but with some politically significant roles (*e.g.*, Nuer); and (3) unitary states with specialized, central political authority (*e.g.*, Bemba). I. Schapera later showed that territorial as well as kinship ties are politically significant in type (1); Middleton and D. Tait noted that there are in Africa many morphologically intermediate types other than these three. In political systems of the village type there are no corporate lineages, but in central Africa, for instance, kinship and neighbourhood provide the necessary institutions for the organization of power and authority under chiefs and headmen.

Where much larger political groups exist without formal government, use is normally made of the unilineal principle of organization. Among the Nuer territorial divisions are named after the branches of the lineage system, although they do not exactly correspond to them. There is no specific authority over lineage segments. In cases of dispute no one official can order a group to accept compensation and make peace. But an elder with special powers, called by Evans-Pritchard the leopard-skin chief, usually can persuade them. Thus among this independent and quarrelsome people, group relations are ordered according to well-established moral principles.

In many parts of the world chiefs are the heads of governments. Among the Bemba (investigated by Richards) the paramount chief rules over a considerable territory. He has advisers, councilors, ritual officers, plus a whole range of courtiers and messengers, and resides in a capital considerably larger than a normal village. Members of the royal clan rule as lesser chiefs over sections of his territory, with capitals and courts similar to his. Beneath these are headmen, responsible to the chiefs. In this case, a great number of people, larger than could ever form a single kin unit, are united in one political society. The size of such a state depends primarily on the strength and efficiency of its administration, though loyalty to symbolic ties of group unity may also be important.

Relations between autonomous political groups in primitive conditions vary with size and organization of group and with ecological conditions. The normal attitude to other autonomous groups is often hostile, as in parts of inland New Guinea. Within a given political system, internal unity rests partly on opposition to external groups.

Wars between primitive communities vary greatly in aim and in ferocity, and are conducted according to specific rules. In conflicts between neighbouring groups who are kin to one another, as among the Andaman islanders, killing may be rare. Women and land, as the Maori used to say, are probably the commonest issues. Habitual raiding for cattle and slaves (once common in many parts of Africa) was without any notion of conquest. Wars of conquest are not feasible unless the conquerors have a state organization to control their subjects. Great conquerors like Shaka (or Chaka), the powerful Zulu king of the 19th century, are rare in primitive society.

Primitive military leadership and organization are often *ad hoc* and informal, though sometimes the war leader also may be the chief. Many societies have or had special war leaders with no other political tasks, as among American Indians. In eastern and southern Africa warrior regiments often existed, based on an age-grade organization. It was by reorganizing age regiments and maintaining them in royal barracks that Shaka was able to gain power and military success.

Authority in any nonliterate society must depend on the people's personal contact with the ruler or his agents. A great deal of ceremonial behaviour and story telling serves the end of transmitting norms and of ensuring that they are remembered. The society looks to the past for validation of rules and values, and explicit legislation is very rare. On the other hand, verbal transmission leaves room for continual adjustments of which the society may not be clearly aware.

Informal arbitration is found in all societies, but in many no sanction of force supports a decision, though public opinion may be a strong influence. In some American Indian societies (as Lowie showed) the function of arbitrator could be exercised only by a man who was not a war leader. In more complex societies, a political hierarchy coincides with a hierarchy of courts, with right of appeal to the paramount ruler. In Africa many societies have a constituted court and litigation is common.

At all levels of society, even where (as in Polynesia) the chief is treated with great deference, the political system provides for checks and balances. Sometimes there is provision for an assembly

of all the people, as among the Southern Bantu. Sometimes the chief, as among the Tikopia, must pass orders through subordinates who, if supported by public opinion, can tactfully evade them. Sometimes, as among the Bemba, the chief depends on the co-operation of ritual functionaries who can refuse their help. Only when a ruler can monopolize force is he able to set up a tyranny; tyranny requires a fairly complex political and military organization. Ritual sanctions behind a ruler may be strong, but they also impose on him strict rules of conduct that limit his power.

To act cohesively, people must think of themselves as a unit, and are able to do so through symbols of unity. The more powerful the emotional value of the symbol, the greater the loyalty to the group. A common and effective symbol of unity is the chief who also signifies the society's well-being and fertility. The chief is thus sacred and acquires important ritual functions. In extreme cases he may be thought of as an incarnation of a divine being. Sometimes, as among the Shilluk (*q.v.*), his own personal vigour is associated with the welfare of the society; it has been alleged that in former times when he was ailing or old he was put to death. Alternatively, the religious head of the society and the ruler may be distinct but interdependent persons.

Religious rites may also be of great importance in political relations. Among the Tallensi there were no chiefs with secular authority. Instead, ritual leaders annually conducted a great festival of autonomous clans, required during this period to settle all quarrels and refrain from hostile acts. In most primitive societies, political organization is linked with ritual and symbolism.

Ritual Relations.—Rites are formal acts not technically related to immediate purposes, but believed to have intrinsic validity. They are often symbolic in character and are usually closely related to myths and dogmas, which explain and help to validate their symbolism. Ritual behaviour ranges from the worship of God to the concocting of medicines against witchcraft or incantations to render a spear more deadly. Many authors have sought to apply the distinction current in Western society between religion and magic. J. G. Frazer suggested that religion is characterized by prayer to a personal power, magic by command to an impersonal force. Durkheim identified religion with public and obligatory rites, magic with private and optional rites. Malinowski held that magic is for immediate practical ends, religion being concerned with general well-being. But each of these criteria classifies as magic some rites that are normally called religious, and vice versa. Moreover (as Lowie, P. Radin, G. Reichard, R. Piddington, and others showed), there are too many borderline cases between prayer and command, between obligatory and optional, between practical and general purposes for these to be final criteria. The general meanings of "religion" and "magic," therefore, remain vague, though more precise classification is possible for particular societies.

All known societies include members with beliefs in a spirit world. In most primitive groups, gods and spirits are thought of as belonging to the particular society, and symbolize its unity. Dead ancestors often are important among the spirits, and belief in their powers thus reinforces lineage or other kinship systems. Myths about the spirit world are thought to explain the universe, give an account of man's place in nature, and validate moral law. They often provide a charter for existing rights and obligations between kinship groups or local groups, in such matters as ownership of lands, social status, or ritual cooperation.

Public rites may be performed simply as worship, or they may have explicit, specific aims such as the fertility and health of group members, abundance of crops, or control of the weather. The passage of an individual from one status to another—birth, initiation, marriage, death—is marked by rituals which (as the classic work of A. van Gennep and later anthropologists such as Richards showed) have important functions in social integration. Rites may also symbolize social distance, unite potential enemies, and enforce the cessation of feuds or wars. Rites such as sacrifice may expiate sins (*see* PASSAGE RITES).

Ritual and belief clearly provide individuals with important means of adjusting to their environment. Religious dogma and mythology should not be assumed to be inflexible. Skepticism, reflection, unconscious modifications in transmission, and even trial and error with ritual procedure are reported from many societies. But the system of belief provides a limiting framework within which speculation is necessarily confined. (*See* MAGIC, PRIMITIVE; RITES AND CEREMONIES; TABU.)

Social Control.—Social order consists in conformity of individual members of a society to a set of rules of conduct. People obey such rules because they wish to do so, or are conditioned to do so without conscious choice; or they are given positive or negative inducements. Social control, therefore, involves both the training of members of society and a system of rewards and punishments (sanctions) that operate to produce conformity in behaviour.

Formal education is rare outside civilized, literate societies. Children in most primitive societies learn by imitation and by participation in activity, guided by correction and praise. Initiation rites may emphasize moral rules through symbolism and explicit instruction. The solemnity of the occasion may be heightened by ordeals, including painful circumcision or scarification.

In a primitive society, formal sanctions may be carried out by community officers according to recognized procedures. But more often they are informal: unplanned individual reactions, or such group-inflicted penalties as songs of ridicule or systematic ostracism. In a small community such informal sanctions can be very strong; failure to conform, apart from material consequences, means loss of esteem and status. Even formal sanctions need not involve judicial procedure; a chief, secret society, professional body, or religious leader may act without formal weighing of evidence or statement of law.

Conduct is often controlled by the expectation that it will produce undesirable consequences apart from any human agency; *i.e.*, such an action is tabu. But belief in dire consequences may deter less than does symbolic expression of strong social disapproval. More generally, divine punishment is often said to follow moral or ritual lapses (sins); these may also be directly punished by society, especially if they are believed to endanger community well-being. But in particular sin may require confession, as R. F. Fortune showed for the Manus of the Admiralty Islands, or a sacrifice in atonement, as Evans-Pritchard and Middleton have illustrated for East African groups. The moral rules of any community are usually supported by its cosmology and by its religious myths and dogmas.

Social control is thus seen to be imbedded in the social system, a by-product of kinship, economic, political, and ritual relations. It is only in the field of law as such that social control has its own institutions.

Thurnwald and Radcliffe-Brown suggested limiting the word "law" to societies where some kind of court exists, with control of organized force to back up its decisions. By this definition many societies have no law at all. Malinowski proposed to consider as law the total body of binding custom of a society, and Lowie, K. N. Llewellyn, and E. A. Hoebel defined law in terms of action by a person against breach of a norm with communal approval or sanction. But (as Schapera showed) if such a broad definition of law be adopted to include all binding custom, there is still need to consider separately formal judicial procedures.

Primitive law involves public action to punish what are regarded as threats to society, corresponding to modern criminal law, and the settlement of private disputes, corresponding to modern civil law. The first is rare and is often confined to serious ritual offenses, especially incest, and witchcraft or sorcery. On the other hand, private disputes cover much that in Western society is criminal behaviour; *e.g.*, homicide and theft.

Specific legislation in primitive society is rare though Schapera has noted modern development of legislation by chiefs among the Tswana (*q.v.*). Where courts are found, the conduct of cases differs greatly from that in Western society. Although most cases are in the form of private disputes, the court will not hesitate to lecture both sides on their moral responsibilities—the law coincides closely with the ethical notions and public opinion of the society (M. Gluckman and P. Bohannan). Oaths and ordeals are used to establish evidence, and hearsay evidence is admissible;

professional pleaders are unusual. In punishment, fining is common but imprisonment rare; banishment or exile from the society is often the more severe penalty. (*See* LAW, PRIMITIVE.)

Values.—Social anthropologists have always paid some attention to values, usually under the head of beliefs, ideas, or institutions. Radcliffe-Brown, for example, described the social value of such phenomena as fire to the Andamanese. Later, C. Kluckhohn and co-workers made comparative studies of values among several neighbouring societies in New Mexico. R. Redfield fostered the concept of world view: the characteristic outlook of a people, and the categories they use to classify and judge all phenomena (including human actions). In a broad sense values may be described as preference qualities assigned to objects and actions in social contexts, having regard to relationships between means and ends. More narrowly, following Kluckhohn, they may be defined as conceptions of the desirable that influence selection from available modes, means, and ends of action. Some values are explicit: they can be formulated in words. Others are implicit, tacit premises implied in action. Thus the Navaho have the concept of harmony among the parts of the universe—an implicit value expressed in much behaviour but only rarely in words and expressed by singers, who are the Navaho intellectuals. Studies of individual conformity to predominant values, the relation between explicit and implicit values, and the processes of value change have received much attention from anthropologists. A problem of theoretical and practical interest (which Redfield and Nadel in particular discussed) is the degree to which anthropologists are influenced in their observations and interpretations by their own personal values.

Studies in Literate Societies.—Intensive field research and systematic analysis of social relations have not been limited to primitive or preliterate societies. Earlier anthropologists saw themselves as historians of human institutions, and made free use of materials from ancient civilizations, especially Greece and Rome. Later workers found primitive societies in transition under the rapid spread of literacy and western influences in Africa, Oceania, and other parts of the formerly primitive world.

Studies among rural communities of the folk-culture type were pioneered by Redfield in Middle America, H. M. Miner in French Canada, C. M. Arensberg in Ireland, J. F. Embree in Japan, H. Fei and F. L. K. Hsu in China, and A. Aiyappan in south India. More recent investigations of peasant communities have also been made in Europe and the Middle East. In urban communities problems of isolating the field of study have been more acute, and approach to conventional sociological methods has been closer. In recent years urban studies by anthropologists have been carried out, particularly in India and in Africa; attention has also been paid to Western industrial problems, and social and cultural factors in technical change. Social anthropologists have also developed interest in special fields such as child training and care, and the social implications of kinship ties in different societies, including those of Western countries. Social aspects of medicine and nutrition as shown in attitudes to disease and to medical care have been examined. The special problems of ethnic groups such as Negroes in Western society, and of other minority groups, in situations of differentiation and discrimination, have been studied, as have the operation and social significance of certain religious sects. Social anthropologists are thus linked with other social scientists in systematic attempts to understand the structure of complex modern societies.

See also ANTHROPOLOGY; SOCIOLOGY; and references under "Social Anthropology" in the Index.

BIBLIOGRAPHY.—References to historical works, and to important articles including work by authors cited in the text, will be found in most of the introductory and general books below.

Introductory: C. D. Forde, *Habitat, Economy and Society* (1934); R. Linton, *The Study of Man* (1936); M. J. Herskovits, *Man and His Works* (1948); E. E. Evans-Pritchard, *Social Anthropology* (1951); M. Gluckman, *Custom and Conflict in Africa* (1956); R. W. Firth, *Human Types* (1956); R. Redfield, *Peasant Society and Culture* (1956); R. Piddington, *An Introduction to Social Anthropology*, 2 vol. (1950, 1957); P. Bohannan, *Social Anthropology* (1963); R. G. Lienhardt, *Social Anthropology* (1964).

General: G. Bateson, *Naven* (1936); M. Mead (ed.), *Cooperation and Competition Among Primitive Peoples* (1937), *From the South Seas* (1939); B. Malinowski, *A Scientific Theory of Culture* (1944), *The Dynamics of Culture Change* (1945); G. and M. H. Wilson, *The Analysis of Social Change* (1945); R. H. Lowie, *Social Organization* (1948), *Primitive Society* (1949); M. Fortes (ed.), *Social Structure: Studies Presented to A. R. Radcliffe-Brown* (1949); G. C Homans, *The Human Group* (1950); M. Mauss, *Sociologie et Anthropologie* (1950); R. W. Firth, *Elements of Social Organization* (1951); S. F. Nadel, *The Foundations of Social Anthropology* (1951), *The Theory of Social Structure* (1957); A. R. Radcliffe-Brown, *Structure and Function in Primitive Society* (1952); C. Kluckhohn, H. A. Murray, and D. M. Schneider (eds.), *Personality in Nature, Society, and Culture* (1953); A. L. Kroeber (ed.), *Anthropology Today: an Encyclopedic Inventory* (1953); R. Redfield, *The Primitive World and Its Transformations* (1953); S. Tax *et al.* (ed.), *An Appraisal of Anthropology Today* (1953); R. F. Spencer (ed.), *Method and Perspective in Anthropology* (1954); G. Gjessing, *Socio-Culture* (1956); W. R. Goldschmidt, *Man's Way* (1959); G. E. Dole and R. L. Carneiro (eds.), *Essays in the Science of Culture* (1960); E. R. Leach, *Rethinking Anthropology* (1961). C. Lévi-Strauss, *La Pensée sauvage* (1962), *Structural Anthropology* (1963); E. E. Evans-Pritchard, *Essays in Social Anthropology* (1962); M. Gluckman, *Order and Rebellion in Tribal Africa* (1963); L. P. Mair, *New Nations* (1963); J. Beattie, *Other Cultures* (1964); R. W. Firth, *Essays on Social Organization and Values* (1964); C. D. Forde, *Yakö Studies* (1964).

Social Structures: F. M. Keesing, *The South Seas in the Modern World* (1941), *Cultural Anthropology* (1958); R. Redfield, *The Folk Culture of Yucatan* (1941), *A Village that Chose Progress* (1950); C. A. du Bois, *The People of Alor* (1944); M. Fortes, *The Dynamics of Clanship Among the Tallensi* (1945); E. and P. Beaglehole, *Some Modern Maoris* (1946); C. Kluckhohn and D. C. Leighton, *The Navaho* (1946); E. D. Colson and M. Gluckman (eds.), *Seven Tribes of British Central Africa* (1951); M. Wilson, *Good Company: a Study of Nyakyusa Age-Villages* (1951); K. L. Little, *The Mende of Sierra Leone* (1951); H. I. Hogbin, *Transformation Scene* (1951), *Social Change* (1958); E. H. Spicer (ed.), *Human Problems in Technological Change* (1952); S. Tax (ed.), *Heritage of Conquest* (1952); F. Barth, *Principles of Social Organization in Southern Kurdistan* (1953); D. Bidney, *Theoretical Anthropology* (1953); E. Colson, *The Makah Indians* (1953); G. Balandier, *Sociologie actuelle de l'Afrique noire* (1955), M. Marriott (ed.), *Village India* (1955); M. N. Srinivas (ed.), *India's Villages* (1955); M. Read, *The Ngoni of Nyasaland* (1956); A. W Southall, *Alur Society* (1956); E. H. Winter, *Bwamba: a Structural-Functional Analysis of a Patrilineal Society* (1956); M. Mead, *New Lives for Old: Cultural Transformation-Manus, 1928–1953* (1956). J. C. Mitchell, *The Yao Village* (1956); V. W. Turner, *Schism and Continuity in an African Society* (1957); W. Watson, *Tribal Cohesion in a Money Economy* (1959); J. J. Honigmann, *The World of Man* (1959); R. W. Firth, *Social Change in Tikopia* (1959); R. Needham, *Structure and Sentiment* (1962).

Kinship and Marriage: C. D. Forde, *Marriage and Family Among the Yakö in Southeastern Nigeria* (1941); A. Phillips (ed.), *Survey of African Marriage and Family Life* (1953); D. L. Oliver, *A Solomon Island Society: Kinship and Leadership Among the Siuai of Bougainville* (1955); F. R. Eggan (ed.), *Social Anthropology of North American Tribes* (1955); G. C. Homans and D. M. Schneider, *Marriage, Authority, and Final Causes* (1955); R. W. Firth (ed.), *Two Studies of Kinship in London* (1956), *We, the Tikopia: a Sociological Study of Kinship in Primitive Polynesia* (1957); C. S Belshaw, *The Great Village* (1957); M. Freedman, *Lineage Organization in Southeastern China* (1958); J. Djamour, *Malay Kinship and Marriage in Singapore* (1959); E. R. Leach, *Pul Eliya, a Village in Ceylon: a Study of Land Tenure and Kinship* (1961); E. Ardener, *Divorce and Fertility* (1962) H. I. Hogbin, *Kinship and Marriage in a New Guinea Village* (1963) *see* also bibliographies for KINSHIP; MARRIAGE, PRIMITIVE.

Status and Stratification: H. Kuper, *An African Aristocracy: Rank Among the Swazi* (1947); W. L. Warner *et al.*, *Social Class in America* (1949); P. M. Kaberry, *Women of the Grassfields: a Study of the Economic Position of Women in Bamenda* (1952); M. M. Tumin, *Caste in a Peasant Society* (1952); F. G. Bailey, *Caste and the Economic Frontier* (1957); L. Dumont, *Une sous-caste de l'Inde du Sud* (1957); L. Dumont and D. F. Pocock (eds.), *Contributions to Indian Sociology*, no. 1 (1957–); M. D. Sahlins, *Social Stratification in Polynesia* (1958); L. H. Palmier, *Social Status and Power in Java* (1960); M. Gluckman (ed.), *Closed Systems and Open Minds* (1964), *see* also bibliography for CASTE (INDIAN).

Economic Relations: K. H. Shih, *China Enters the Machine Age* (1944); H. T. Fei and T. Chang, *Earthbound China* (1945); I. Schapera, *Migrant Labour and Tribal Life* (1947); H. F. Codere, *Fighting with Property* (1950); M. J. Herskovits, *Economic Anthropology* (1952), A. I. Richards (ed.), *Economic Development and Tribal Change* (1954) M. Nash, *Machine Age Maya* (1958); E. and S. Ardener *et al.*, *Plantation and Village in the Cameroons* (1960); R. F. Salisbury, *From Stone to Steel* (1962); L. Pospisil, *Kapauku Papuan Economy* (1963); H C. Brookfield and P. Brown, *Struggle for Land* (1963); R. W. Firth and B. S. Yamey (eds.), *Capital, Saving and Credit in Peasant Societies* (1964); *see* also bibliography for ECONOMIC ANTHROPOLOGY.

Political Relations: E. E. Evans-Pritchard, *The Nuer* (1940), *The Sanusi of Cyrenaica* (1949); M. Fortes and E. E. Evans-Pritchard

(eds.), *African Political Systems* (1940); S. F. Nadel, *A Black Byzantium* (1942); K. A. Busia, *The Position of the Chief in the Modern Political System of Ashanti* (1951); E. R. Leach, *Political Systems of Highland Burma* (1954); J. A. Barnes, *Politics in a Changing Society* (1954); L. A. Fallers, *Bantu Bureaucracy* (1956); J. Middleton and D. Tait (eds.), *Tribes Without Rulers* (1958); J. M. Gullick, *Indigenous Political Systems of Western Malaya* (1958); A. I. Richards (ed.), *East African Chiefs* (1960).

Ritual Relations: B. G. M. Sundkler, *Bantu Prophets in South Africa* (1948); R. M. Berndt, *Djanggawul* (1952); M. N. Srinivas, *Religion and Society Among the Coorgs of South India* (1952); C. D. Forde (ed.), *African Worlds* (1954); S. F. Nadel, *Nupe Religion* (1954); A. J. A. Elliott, *Chinese Spirit-Medium Cults in Singapore* (1955); A. I. Richards, *Chisungu: a Girls' Initiation Ceremony Among the Bemba of Northern Rhodesia* (1956); E. E. Evans-Pritchard, *Nuer Religion* (1956); P. Worsley, *The Trumpet Shall Sound: a Study of Cargo Cults in Melanesia* (1957); M. Wilson, *Rituals of Kinship Among the Nyakyusa* (1957), *Communal Rituals of the Nyakyusa* (1959); V. Lanternari, *Movimenti religiosi di libertà e di salvezza dei popoli oppressi* (1960); J. Middleton, *Lugbara Religion* (1960); C. Geertz, *The Religion of Java* (1960); K. O. L. Burridge, *Mambu* (1960); G. Lienhardt, *Divinity and Experience: the Religion of the Dinka* (1961); M. Gluckman (ed.), *Essays on the Ritual of Social Relations* (1963); C. Lévi-Strauss, *Totemism* (1963); J. Middleton and E. H. Winter (eds.), *Witchcraft and Sorcery in East Africa* (1963); *see* also bibliographies for MAGIC, PRIMITIVE; RELIGION, PRIMITIVE.

Social Control, Values: O. F. Raum, *Chaga Childhood* (1940); M. Reay, *The Kuma: Freedom and Conformity in the New Guinea Highlands* (1959); P. Bohannan (ed.), *African Homicide and Suicide* (1960); *see* also bibliographies for LAW, PRIMITIVE; MORALITY, PRIMITIVE.

Modern Developed Societies: C. M. Arensberg and S. T. Kimball, *Family and Community in Ireland* (1940); W. L. Warner *et al.*, "Yankee City Series" (1941–); J. West, *Plainville, U.S.A.* (1945); J. F. Embree, *A Japanese Village: Suye mura* (1946); M. J. and F. S. Herskovits, *Trinidad Village* (1947); R. Landes, *City of Women* (1947); K. L. Little, *Negroes in Britain* (1948); W. Goldschmidt, *As You Sow* (1947); E. Hellmann, *Rooiyard: a Sociological Survey of an Urban Native Slum Yard* (1948); H. Powdermaker, *Hollywood: the Dream Factory* (1950); A. D. Rees, *Life in a Welsh Countryside* (1950); S. Patterson, *Colour and Culture in South Africa* (1953); M. Mead (ed.), *Cultural Patterns and Technical Change* (1953); J. A. Pitt-Rivers, *The People of the Sierra* (1954); E. Z. Vogt, *Modern Homesteaders* (1955); M. Freedman (ed.), *A Minority in Britain: Social Studies of the Anglo-Jewish Community* (1955); H. B. Hawthorn (ed.), *The Doukhobors of British Columbia* (1955); M. Young and P. Willmott, *Family and Kinship in East London* (1957); E. Bott, *Family and Social Network* (1957); S. Collins, *Coloured Minorities in Britain* (1957); M. Banton, *West African City: a Study of Tribal Life in Freetown* (1957); E. Clarke, *My Mother Who Fathered Me: a Study of the Family in Three Selected Communities in Jamaica* (1957).

(R. F.)

SOCIAL CONTRACT, in political philosophy, a term applied to certain theories of political obligation. Although similar ideas can be traced back as far as the Greek Sophists, social contract theories had their greatest currency in the 17th and 18th centuries, and are associated with such names as Thomas Hobbes, John Locke, and Jean Jacques Rousseau (*qq.v.*). What distinguished these theories from other doctrines of the period was their attempt to justify political authority on grounds of individual self-interest and rational consent. The method used was to demonstrate the value and purposes of organized government by comparing the advantages of civil society with the disadvantages of the state of nature, a hypothetical condition characterized by a complete absence of governmental authority. The purpose of this comparison was to show why and under what conditions government is useful, and ought therefore to be accepted by all reasonable men as a voluntary obligation. These conclusions were then reduced to the form of a social contract, from which it was supposed that all the essential rights and duties of citizens could be logically deduced.

Although social contract theorists were agreed as to method, the conclusions they drew from it were widely different. According to Hobbes (*Leviathan*, 1651), the state of nature was one in which there were no enforceable criteria of right and wrong. Each person took for himself all that he could; man's life was "solitary, poor, nasty, brutish and short." The state of nature was therefore a state of war, which could be ended only if men agreed to give their liberty into the hands of a sovereign, who was thenceforward absolute, on the sole condition that the lives of the citizens were safeguarded by his power. Locke (the second of *Two Treatises on Civil Government*, 1690) differed from Hobbes insofar as he described the state of nature as one in which the rights of life and property were generally recognized under natural law, the inconveniences of the situation arising from insecurity in the enforcement of those rights. He therefore argued that the obligation to obey civil government under the social contract was conditional upon the protection not only of the person but also of private property. Rousseau (*Du contrat social*, 1762) held that in the state of nature man was unwarlike and even timid, and that his rational powers were too little developed to enable him to have any conception of rights and duties. Laws resulted from the combination of men who agreed for mutual protection to surrender individual freedom of action, and thereby acquired for the first time a sense of moral and civic obligation. In order to retain its essentially moral character, government must therefore rest on the consent of the governed, the *volonté générale* ("general will"). Other social contract theorists provided yet other interpretations of the common theme.

These theories are historically important for the part they played in the development of modern ideas of government, especially the idea of government by popular consent. Toward the end of the 18th century, however, with the rise of modern historicism, they became increasingly unacceptable, not so much because of their content as because of the abstractly rationalistic terms in which they were formulated. The more perceptive social contract theorists, including Hobbes, had always recognized that their concepts of the social contract and the state of nature were unhistorical, and could be justified only as hypotheses useful for the clarification of timeless political problems. David Hume, in particular, is absolutely explicit on this point. But as people grew increasingly interested in history, and tried to understand political problems in terms of the historical origins and growth of specific institutions, the earlier types of abstract analysis seemed ever less relevant. Thus the term social contract fell gradually into disuse, and ceased to play a significant part in the vocabulary of modern politics. (F. M. W.)

SOCIAL INSECTS are insects that live in colonies in which the adults perform different tasks associated with differences in structure and function—for example, the queen lays eggs and the workers tend the nest and procure food. Social organization has evolved independently in several groups of insects in two insect orders, the Hymenoptera (ants, bees, wasps) and the Isoptera (termites). It is found in varying degrees in some species of bees and wasps, but is fully expressed in all species of ants and termites. Maurice Maeterlinck in his lyrical *The Life of the Bee* (1901) attributed the unity and communal cooperation within the colony to an elusive "spirit of the hive [that] disposes pitilessly of the wealth and the happiness, the liberty and life, of all this winged people; and yet with discretion, as though governed itself by some great duty." This spirit of the hive is now known to be a complex system of communication by means of chemicals called pheromones, or social hormones, which influence behaviour and growth of individuals and appear to be the integrating factors in all insect societies.

INSECT SOCIETY: A SUPRAORGANISM

An insect society may be compared to an individual organism, the individual members of the society being comparable to organs or cells of the organism. Such a comparison does not say anything new about insect societies but it suggests interesting approaches. An individual organism is an open system with adaptation to, and interaction with, its environment; it maintains a relatively steady-state by self-regulation of internal conditions; it has a capacity for reproduction; it exhibits functional specialization of its parts; and it shows change with development. The insect society is a supraorganism—a collection of organisms exhibiting all of the organismic properties mentioned above. The protoplasmic contact between the parts, characteristic of the organism, is replaced in the insect society by behavioural and biochemical interaction. Communication through behavioural specializations and trophallaxis—mutual exchange of food, exudates, and pheromones—is characteristic of insect societies. Analogous similarities between levels, such as nonreproductive somatic cells in the organism and

nonreproductive castes in the insect society, result from natural selection of whole units favouring division of labour, integration, and improved self-regulation (homeostasis). All the detected trends of evolution in insect societies and in organismic evolution are toward improved homeostasis.

The societies of man are analogous to insect societies in such attributes as division of labour among adults of the same sex, communication, and integration of individuals into systems (social institutions, tribes, nations). Human societies also show an evolutionary trend toward increased homeostasis to provide better conditions for existence. In these senses, human societies are also supraorganisms. They differ from insect societies, however, in having a social organization based upon learned symbolic communication. There are no genes for a particular language, a particular style of architecture, or a particular type of agriculture, as there are among social insects. The evolution of insect societies is slow, whereas the evolution of cultural traits in man is rapid. Insect societies are rigidly organized and are characterized by stereotyped and automatic responses of members; human societies exhibit much greater versatility and functional adaptability.

The success of insect societies presumably results from a division of labour among different individuals of a colony. Except in very primitive societies, the individuals performing different functions differ also in structure. Thus, the egg-laying queen is quite different from a worker or (in termites) from a soldier. These various forms are called castes. The existence of castes (a condition termed social polymorphism, or polyphasy) and their evolution and determination are subjects of continuing study.

ORIGIN

Except perhaps for some groups of bees, true social insects arose from insects that live in family groups consisting of a mother and her offspring but in which the mother dies or leaves the group after the eggs hatch and before the offspring reach maturity. Such a way of life, called subsocial, is in contrast to that of the majority of insects, in which the mother lays her eggs and leaves them immediately or dies before they hatch. In subsocial forms, the young may be both protected and fed by the mother or sometimes by both parents.

The social wasps arose from solitary wasps of the subfamily Eumeninae, family Vespidae. Solitary Eumeninae construct a cell, provision it with food, *e.g.*, paralyzed caterpillars, lay an egg in the cell, and seal it. The food is adequate to provide for the entire growth of the single larval wasp; such a cell is said to be mass provisioned. There is no contact between the larva and its mother, although the mother provided a protecting cell and all of the larval food. Wasps more immediately ancestral to social forms show various degrees of continual, or progressive, provisioning of cells; cells are left open during larval growth, and food is added as the larvae require it. Such wasps are strictly subsocial, since there is direct contact between mother and growing offspring. True social wasps all provision cells progressively.

Social bees arose, not once but repeatedly, from solitary forms nearly all of which use mass provisioning. Since most social bees also have mass provisioning, it is obvious that progressive feeding is not a prerequisite for establishment of social behaviour. Ants are related to solitary vespoid wasps, but there is no living family from which they might have evolved. Some bethylid wasps, which belong to the general group from which ants also possibly arose, have subsocial behaviour in which larvae are attended to by their mothers. These larvae are external parasites on very much larger host insects paralyzed by the repeated stings of the adult female bethylid. The larvae are therefore not in individual cells; larvae of ants also are not in separate cells but are reared grouped together in masses.

Termites arose from roaches. Subsocial behaviour of the sort that possibly was found among the ancestors of termites exists today in wood-eating roaches (*Cryptocercus*) that live in small family groups and harbour wood-digesting flagellate protozoa similar to those that serve the same function in primitive termites.

Subsocial family groups are known in a number of insect orders, but except for termites, true societies occur only in Hymenoptera. Perhaps the most important single feature increasing the potentiality for social development in Hymenoptera is the possession of sensory and central nervous equipment good enough to permit the insect to find food and then successfully return, repeatedly, to its nest. Solitary bees and wasps, which have such abilities well developed, constitute a great pool of organisms from which social forms could evolve repeatedly.

SOCIAL WASPS

The social wasps include more than one thousand species of the family Vespidae. Some of the larger and more aggressive species, called yellow jackets and hornets, are respected for their very effective stings. Although adult social wasps are attracted to flowers, decaying fruits, and the like for their own sustenance, they collect animal food (largely chewed up insects) to feed their larvae, along with liquids from the same sources as part of the adult food. One genus of the New World tropics (*Brachygastra*), however, stores honey in papery cells. Social wasps construct paperlike nests composed of chewed dry plant material, usually wood, mixed with saliva. These nests usually consist of one or more layers of cells covered with a papery envelope. The cells, hexagonal in cross section, are usually vertically arranged with the openings downward. Depending on the kind of wasp, the nests may be found in cavities in the soil, in tree trunks, or hanging from leaves, branches, or eaves of buildings. Eggs, larvae, or pupae are found in the cells, one per cell. Eggs and small larvae are held in the cells by an adhesive material; the older stages retain their position by means of their snug fit. The mature larva spins a cocoon, lining and capping the cell, before it pupates.

When they are fed, larvae emit from their mouths secretions that are much in demand by the adult wasps. This exchange of substances—the larva receiving food brought to it by the adult and the adult receiving the secretion of the larva—and other sorts of mutual stimulation are termed trophallaxis. It is very widespread among social insects and is among the integrative mechanisms that make the colony a successful operating unit.

Workers rarely or never develop into egg layers. In the large subfamily Polybiinae, especially common in the American tropics, the colonies often contain many queens, as well as even larger numbers of workers. In this group queens and workers are scarcely distinguishable externally, although in ovarian development, physiology, and behaviour they are quite different. New colonies are founded by the swarming of several queens with many workers, a type of colony establishment not found in other social wasps.

The most familiar social wasps in the temperate zones are species of *Polistes* (subfamily Polistinae) and of the genera *Vespa, Vespula,* and *Dolichovespula* (subfamily Vespinae). These wasps and hornets have colonies that last a single season. There is ordinarily only a single queen in a colony except in autumn when new young queens and males are produced. In autumn the wasps perish except for the recently mated females, each of which may establish a new colony in spring.

The *Polistes* queen builds a few cells attached to a single stalk hanging downward in a sheltered place, often under the eaves of a house or shed. Her first progeny consists entirely of females (workers). After their appearance, the queen no longer leaves the nest, for food is brought to her by the workers. Workers add new cells laterally to the enlarging comb, which ultimately may attain a size of 500 cells or more. The comb of *Polistes* is naked, without a paper envelope. Nests of the other genera listed above are established in a similar manner, although those of *Vespula* are usually in underground cavities. The combs are surrounded in these genera by an envelope consisting of several layers of paper with a hole at one point through which wasps can come and go. As the colony grows, additional combs are constructed in tiers beneath the original comb. The paper envelope must be enlarged as new combs are added; this is done by removing inner layers of the envelope and adding new layers on the outside. Nests of Vespinae may become very large, up to two feet in diameter in the temperate zones and even larger in tropical Asia. Both the queen and the workers are equipped with stings with which they can defend themselves and the colony.

Socially parasitic wasps that have lost the worker caste have evolved in both Polistinae and Vespinae.

SOCIAL BEES

The bees (superfamily Apoidea) are wasplike insects that differ from wasps primarily in that they feed their larvae on pollen and honey instead of on insects or other food of animal origin. Of the probably 20,000 species of bees, 1,000 or more are primitively social and probably less than 500 are as highly social as bumblebees and honeybees.

Primitively social bees are those in which workers and queens are scarcely, or not at all, distinguishable externally (although usually they differ in average size). Of course there are striking behavioural and physiological differences between castes. The most familiar primitively social bees are among the sweat bees (Halictidae). All intergradations from solitary to social exist in this family. The nests consist of burrows, usually in the ground, leading to cells. Each cell is a cavity, lined with waxlike material, in which sufficient food (pollen mixed with nectar) is stored to provide for the growth of a larva. After an egg is laid on top of the food mass, the cell is sealed. In many species the female dies before emergence of her progeny, but in some species several females, probably sisters, live together and make cells in a single nest. In such cases the rare unfertilized female may be relatively short lived but do a great deal of work during her life, thus being workerlike. In certain species about half of the females in the colony are workerlike in this sense. Finally, in a large number of species, there exist behaviourally distinct castes that usually differ in average size. Colonies consisting of a queen and two or three workers are common, although there are species that have colonies of up to a few hundred individuals. In temperate climates the colonies are started each spring by overwintered, fertilized queens working alone, as in temperate zone social wasps. The first offspring produced are mostly workers, which enlarge the nests and take over foraging activities. In some species new queens and males are produced throughout the summer so that queens are replaced from time to time; in other species overwintered queens survive through the summer, and new queens and males appear only in late summer.

Another group of primitively social bees consists of genera related to *Allodape*, all found in the tropics or southern continents of the Old World. They nest in burrows in stems or wood, the larvae being reared together in the burrow and fed progressively. Queen and worker castes, differing from one another about as much as those of halictids, have been recognized in one Australian genus, *Exoneura*.

Bumblebees (genus *Bombus*) are the most primitive of the thoroughly social bees. About 300 species are found in the north temperate and Arctic regions, and a few extend into the tropics. As with temperate social wasps and halictid bees, the fertilized female, or queen, overwinters, usually in the ground, and seeks a nest site in the spring. Most species use empty mouse nests or animal burrows in the soil, but a few construct nests elsewhere from feathers or grass stems. After the nest site is selected, the queen gathers pollen and nectar, mixes them into a paste, and builds a circular wall of wax around the lump of paste on the floor of the nest cavity. Wax is produced from both dorsal and ventral abdominal glands. She lays several eggs in a clump on the lump of paste and closes over the wax cell. Soon after the first batch of eggs is laid, the queen builds a large wax honeypot at the entrance of the nest and in this stores honey for her own use. The eggs hatch in a few days, and the larvae soon devour the paste, after which they are fed progressively by the queen, who forages for food. The wax cell is enlarged as the larvae grow; after ten days or so the larvae spin separate cocoons of silk for pupation. The resulting adults are all workers, usually less than half the size of the queen. They relieve her of foraging activities, enlarge the nest, build new cells, and store more honey and pollen in newly constructed pots of wax. For a considerable period of time, only worker progeny are produced, and the colony grows until it contains 50 to 600 bees. In late summer, with the large population of workers bringing in abundant food, males and new queens are produced. After the queens are fertilized, all individuals except the mated young queens die. Even in the tropics, bumblebee colonies are short lived and are established by lone queens just as in the temperate zone, not by swarms as with the bees discussed below.

Socially parasitic bumblebees of the genus *Psithyrus* evolved from *Bombus*; *Psithyrus* lacks the worker caste and pollen-collecting apparatus.

Several hundred species of tropical bees without stings are called the stingless honeybees. There are two principal genera, *Melipona* and *Trigona*. Although stingless, these bees are often well able to defend their large colonies for they swarm over the intruder, often biting with their jaws and, in one subgenus of *Trigona*, working a highly irritating substance into the skin with their jaws. The queens are specialized in that they have lost the pollen-gathering and pollen-carrying apparatus and therefore are unable to establish colonies by themselves but must have help from workers. Each colony ordinarily contains only a single queen. In *Melipona*, queens come from cells mixed with, and indistinguishable from, those producing workers; moreover, queens are produced in large numbers, and all but a few are killed off by workers. In *Trigona*, as in the true honeybee, queen cells are much larger than the others, and only a few queens are produced each year. Nests are found in hollow trees, more rarely in cavities in the soil or in walls of houses and the like. A few species make their nests only in cavities in nests of ants or termites. Nest construction is with a mixture of wax and resin. Brood cells are usually in horizontal combs, the cells opening upward, but they may be in irregular clusters.

Pollen and honey are stored in pots, quite different from brood cells, and varying in size among different species, from the size of a pea to an inch or more in diameter. The adult bee population varies, among species, from a few hundred to more than 50,000. After each brood cell is provisioned, the queen lays an egg in it and then it is closed by a worker; mass provisioning is practised even in these highly social bees.

To establish a new colony, workers locate a suitable site within their flight range and carry construction materials to it from the mother nest, build part of the new nest, and even take honey and pollen to it. After a few weeks, a considerable number of workers are living there, and ultimately a newly emerged queen also flies there to become queen of the new colony. Food may still be provided from the parent colony for several additional weeks. Colonies of the stingless honeybees are perennial.

There are no true social parasites among the stingless honeybees, but the genus *Lestrimelitta*, derived from *Trigona*, consist of robbers. Workers of *Lestrimelitta* do not visit flowers but steal food from colonies of *Trigona* and carry it back to the *Lestrimelitta* colony.

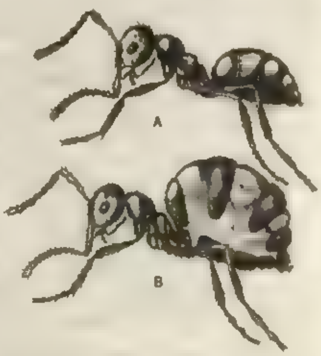
True honeybees constitute the genus *Apis*, which contains the common domestic honeybee and a few other species found in the tropical Orient. The common honeybee (*Apis mellifera*) is a native of Europe, western Asia, and Africa; it has been widely introduced throughout the rest of the world. The natural location for its nests is hollow trees, although cavities in cliffs and even the ground are sometimes used. Man-made hives merely take advantage of behaviour patterns of the insect, which have probably not been changed by domestication. These bees construct hanging vertical combs of wax (for details *see* BEE and BEEKEEPING). Feeding of larvae is progressive, the cells being left open until the larvae are nearly ready to pupate. Much of the activity of the young bees in the nest consists of feeding the larvae; there may be many thousands of larvae in the colony at one time, and in the last day of larval life, for example, each receives about 2,000 feedings.

As with stingless honeybees, the queen is unable to forage alone, and colony establishment is by means of swarms. Shortly before the appearance of adult new queens in early summer, the old queen and thousands of workers leave the nest in a swarm. They fly to a tree or other support where they settle while scouts locate suitable nesting sites; in due course the swarm flies to such a site and establishes itself there.

ANTS

The ants (family Formicidae) constitute the largest group of social insects. Varying degrees from relatively simple to highly specialized social behaviour exist among the more than 5,000 species. Nearly 600 species and subspecies occur in the United States. Many of them excavate galleries in the soil or in fallen logs, but many species are mound builders on the surface of the ground, or are arboreal, living in natural or prepared cavities of plants or constructed nests. The nests are always loosely organized and lack the cells and similar precise construction of those of bees and wasps. Eggs are gathered by the workers in masses, and larvae are carried to favourable temperature and moisture conditions in the nests and may be sorted by stage of development by worker nurses.

Ant workers, unlike those of bees and wasps, are always wingless. In primitive ants (for example, most Ponerinae) workers are of one size (monomorphic) and structurally similar to the queen, although lacking simple eyes, wings, and well-developed ovaries. (Occasionally workers may lay eggs, however.) More specialized ants in most subfamilies have workers of differing sizes (polymorphic), and in many genera nutritive and protective functions may be somewhat divided among different-sized workers. Commonly the largest workers have relatively larger heads and jaws and are called soldiers. Soldier ants are not a clearly distinct caste like soldier termites (*see* below); they are merely the largest of the ant workers. Large workers may be behaviourally specialized for protection in that they are aggressive and attack disturbing objects. In some genera, however, the large-headed individuals are nonaggressive and apparently serve an important function in crushing seeds, which would not otherwise be available for food.



FROM W. M. WHEELER, "ANTS"; COLUMBIA UNIVERSITY PRESS

FIG. 1.—ANT (PRENOLEPIS IMPARIS)

(A) An ordinary worker; (B) a worker greatly distended with syrup for transport to nest

Most male and queen ants are winged. They are produced at certain seasons; in most species their emergence from parental nests is synchronized among numerous colonies so that crossbreeding is possible. Males mate with females during the mating flight or on the ground after the flight. After mating, the queen breaks, or chews off, her wings and seeks a place to establish her nest. The site depends on her species, but in the commonest temperate zone forms it is a small chamber excavated in the ground. The queen's large thoracic wing muscles, no longer useful after the flight, are absorbed and presumably assist in nourishing the queen and her first larvae, which she feeds progressively by salivary secretions. The new queen takes no food except her own eggs; she does not leave her closed chamber. This behaviour is called cloistral colony establishment. In the primitive subfamilies Myrmiciinae and Ponerinae, cloistral nest establishment does not occur; the queen leaves her nest chamber at intervals to seek food. As soon as her first young, all of which are workers, are produced, they undertake foraging. The colony may grow to contain many thousands of workers before any reproductives appear.

In a few groups like the army ants (Dorylinae), the queen is wingless, and new colonies are formed by division of the old colony rather than by nuptial flights. In some ants (for example, *Formica ulkei*), new queens are accepted back into the old nests after the nuptial flight, and new colonies are formed only later, by division of the populations. Some species of *Formica* are temporary social parasites, that is, queens start their colonies by being social parasites in the nests of some other species of the same genus. They are accepted in nests of the temporary hosts and lay their eggs; ultimately the host queen is killed, and the young of the parasitic species are reared by workers of the host species; the host workers finally die of old age and are replaced by the workers of the new queen. Once this process is complete, no evidence remains of the parasitic nature of such species. A few ants are permanent social parasites, having no workers and no independent colonies but living always in colonies of the hosts.

The food of ants is much more diverse than that of social wasps or bees. Primitive ants resemble their wasp ancestors in preying upon living insects. A few specialize, but most are general predators. The army and driver ants are important raiders, mainly on insects of the earth surface. Most of these primitive predatory ants possess the sting of their wasp ancestors. Many more specialized ants are scavengers, feeding alike on dead or living insects and larger animals, decaying fruit, etc. Some ants supplement their diet with the sweet excreta of aphids; shelters are sometimes constructed for these "ant cows."

"Honey ants," like those of the genus *Myrmecocystus*, exhibit a remarkable specialization for conserving sweet liquids. Honeydew, or nectar, is stored in the crops of some workers until they become so distended with sweet liquid that they become immobilized. Such workers, called repletes, can only hang from the roof of the nest and regurgitate honey as required by other workers. So far as known, repletes are morphologically identical to other workers in the colony. If the repletes are removed from a colony, other individuals soon become distended to replace them.

Harvesting ants collect, husk, and store seeds in cavities in their nests. The fungus-growing ants feed only on special fungi, different for each species of ant. They cultivate the fungi in chambers of their nests, usually underground. The primitive fungus ants use a substratum of insect droppings as a culture medium, the more specialized species cut pieces of leaves and petals and transport them to their underground chambers, where they are chewed into a moist spongelike mulch on which the fungus grows. In these ants there is much worker polymorphism. The largest workers function as soldiers, the middle-sized workers do most of the harvesting of leaf fragments, and the smallest workers, which never leave the nests, tend the fungus gardens. Before her nuptial flight, the queen fungus ant takes a pellet of fungus into a pouch below her mouth. After mating and establishing herself in a chamber in the soil, she places the pellet on the floor of the chamber manuring the fungus with her droppings. Her eggs are laid in the developing fungus garden, and the larvae are fed on fungus or on some of her other eggs. The first small workers appear in about 40 days and soon start foraging for leaves to enlarge the fungus garden. When larger workers appear, the smallest no longer leave the nest. The larger species of fungus ants (*Atta*) make some of the largest nests, consisting of many chambers up to 18 in. in diameter down to depths of 16 or 20 ft. and with over a thousand openings spread over a wide area from which the ants issue to defoliate trees.

In the slave-making ants of the genus *Formica* (subgenus *Raptiformica*) the colony is founded by a queen invading, as a temporary social parasite, a nest of the slave species (another *Formica*) and adopting some of the larvae and pupae. The workers that later emerge from such adopted brood feed and attend the slave-making queen and care for her brood like that of their own species. The queen of the slave species ultimately is killed or dies. The progeny of the slave-making queen emerge as workers and raid slave species in the vicinity to replenish the dwindling population of slaves. These raids may have evolved from predatory hunting by ancestral species. The loosely organized group of slave makers finds a nest of the slave species, invades it, and captures grown larvae and pupae in addition to an occasional young adult worker. These are carried off to the home nest and when mature perform the foraging, feeding, and nest-building activities of the colony. Such slave makers are also capable of living and performing their own work without slaves. The proportion of slaves varies among species and also among colonies of the same species.

Amazon ants of the genus *Polyergus* are obligatory slave makers. New colonies are established as with *Raptiformica*. The slave raids are made by workers of *Polyergus* that seem to have no other function. One or two hundred slave makers emerge from their nests and travel in an excited manner directly to the nest of a neighbouring *Formica*. There the slave makers crowd around the opening of the nest and enter as rapidly as possible. Any *Formica*

worker offering resistance is immediately killed by the sickle-shaped mandibles of the Amazons. After 10 or 15 minutes a line of Amazons returns to their nest carrying their captives. These ultimately develop into mature slaves and function in the colony as if it were their own. In artificial nests one slave can keep ten Amazons well fed, but without slaves the slave makers starve even in the presence of abundant food. Slave making among ants appears possible because adults emerging from enslaved pupae exhibit at least some of the distinctive behaviour patterns of the host species. Artificially, for example, adults of a normally timid species reared in a colony of an aggressive species are more aggressive than is usual for their kind.

Ants may become highly adapted to certain plants with natural cavities in their leaf petioles, thorns, twigs, trunks, or roots. In some cases evolutionary adaptation may be largely in the ants rather than in the plants. For example, ants of the genus *Oecophylla*, in tropical Africa and the Orient, construct their arboreal nests of leaves attached to one another by means of silk spun from the salivary glands of the ant larvae. Some workers hold the leaf edges together, while others, holding larvae in their jaws, weave the silk-secreting larvae back and forth from one leaf to the other, thus fastening leaves to form a protective nest. On the other hand, some plants appear to have evolved along with the ants associated with them. For example, the hollow thorns of the bull-thorn *Acacia* are always inhabited by ants of the genus *Pseudomyrmex*, some species of which do not occur except on these plants. These species feed only on the nectar secreted by glands on the leaves of the tree and by proteinaceous Beltian bodies produced by the young leaves. The tree thus provides protection and food for the ants, and the aggressive ants effectively protect the *Acacia* from leaf-feeding insects and browsing animals; moreover, the ants bite off vines and other plants that grow into the foliage of the *Acacia*. The association is so firm that the survival of this species of *Acacia* is thought to be dependent on the presence of these ants. Many similar cases of interdependence of ants and ant plants exist, especially in the tropics.

Inquiline insects or "guests" occur in the nests of many social insects. Numerous groups of insects and a few other animals are predators or scavengers in ant nests, but the more specialized myrmecophiles have trophallactic relations with the ants, often producing secretions that attract the host ants. Tufts of hairs, called trichomes, that disperse liquids and scents from glands at their bases have evolved convergently in myrmecophiles, for example, in several different families of beetles.

TERMITES

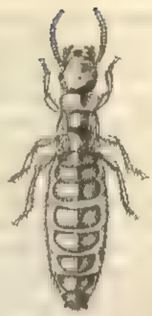
Termites, sometimes misleadingly called "white ants," show a remarkable social convergence to ants though, as mentioned earlier, they were derived from roaches. They develop by incomplete metamorphosis through nymphal stages instead of by complete metamorphosis through larvae and pupae, as in Hymenoptera. The primitive *Mastotermes* of Australia, unlike other termites, has a large lobe on the hind wing and lays eggs in groups, in both respects resembling roaches. Termites feed principally on dead plant material: wood, dead grass, dead leaves, dry droppings of grazing animals, etc. One group in the Old World tropics cultivates and eats fungi, as do the fungus ants of the American tropics. Termites of the north temperate zone are mostly wood feeders.

There are over 2,000 known species of termites, mostly tropical. The social organization differs from that of Hymenoptera in that each caste consists of both sexes. Instead of having males and females, with the latter segregated into queens and workers, as in Hymenoptera, the castes are reproductives (kings and queens), soldiers, and, in specialized termites, workers.

FROM C. L. MARLATT, "ENTOMOLOGICAL BULLETIN", U.S. DEPT. OF AGRICULTURE

FIG. 2.—RETICULITERMES FLAVIPES, FULLY WINGED MALE

The winged primary reproductives make a dispersal flight, usually synchronized among the colonies of a neighbourhood and occurring at a time of high atmospheric humidity. They leave the parent colony, fly through the air, and descend to the ground. Then they break off their wings along breakage sutures near the bases. Usually it is only after this event that the male is attracted to the female by odour. Once the pair meet, they walk in tandem across the ground as she seeks a crevice or place to excavate into dead wood or into the earth. The first progeny of a pair of primary reproductives among the lower families of termites mostly remain as working nymphs for a long time, although a few soon become soldiers. The colony usually attains a population of thousands before primary reproductives are produced. Among the specialized termites, the first progeny mostly become workers, although a few, perhaps 10%, become soldiers. Again, the colony must attain a considerable size before the first primary reproductives are produced. Some termite colonies become the most populous of any social insects, having several million individuals.



FROM N. BANKS AND T. E. SNYDER, "REVISION OF THE NEARCTIC TERMITES"; SMITHSONIAN INSTITUTION

FIG. 3.—RETICULITERMES VIRGINICUS, BRACHYPTEROUS QUEEN, WITH SHORT WING PADS

Among specialized termites, the queen develops enormous ovaries, and the abdomen becomes greatly distended (up to four inches long); after several years such a queen may lay 8,000 eggs per day continuously for years. Queens of some primitive termites grow little from their size at the dispersal flight. The king lives with the queen throughout life, and mating occurs periodically. Upon the death of the primary king or queen, one or more nymphs may develop into unwinged secondary reproductives and substitute for the lost primary one.

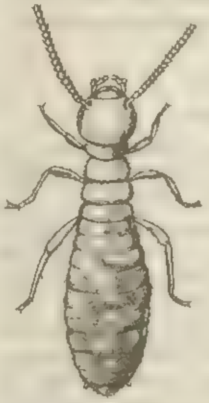


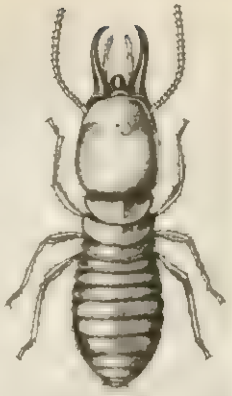
FIG. 4.—RETICULITERMES FLAVIPES WORKER

All termites possess soldiers for defense except two genera in which this caste has been secondarily lost. Soldiers have no wings, no eyes, and much reduced reproductive organs. Adult soldiers serve only for defense; being incapable of feeding themselves, they must be fed by nymphs or workers. Soldiers of most termites have enlarged heads and jaws, but some Termitidae have soldiers with reduced jaws; instead, the front of the head is prolonged into a sort of squirt gun for forcibly ejecting a sticky and irritating liquid from a large gland in the head. This defense is highly effective against marauding ants. Rarely two or three different types of soldiers occur in the same colony. An extreme case is found in two tropical American genera. Here the largest soldiers have large defensive mandibles, but the small soldiers have reduced mandibles and a large gland in the head that secretes a substance that runs down the upper lip, or labrum, and volatilizes as a repellent gas.

In termite colonies there are large numbers of immature forms, or nymphs, which do much of the work of the colony, enlarging galleries, feeding reproductives, carrying eggs away from the queen, and feeding younger nymphs that are not yet able to eat by themselves. Growth of termites is rather slow, especially among the primitive families, so that a large population of such working nymphs exists. Only in the specialized families Rhinotermitidae and Termitidae does a true worker caste exist. Workers are merely individuals whose development stops while they still have the form and behaviour of the working nymphs. Being adults, workers have lost the potentiality possessed by nymphs of continuing development and becoming soldiers or reproductives.

Most primitive termites are wood eaters, living in galleries chewed out of wood. Such termites build partitions from their excrement but do not construct nests. The more specialized Rhinotermitidae usually use wood in contact with the ground and make tunnels into the ground, for which reason many are spoken of as subterranean termites. Often the queen is in galleries in the soil, and the burrows may extend for some distance from this cen-

tral location or nest to the wood. Sometimes covered passageways extend, for example, across the surface of the foundation of a house and make available wood that is not in direct contact with the earth. The most advanced family of termites (Termitidae) contains most of the known species, nearly all of them tropical. Most of them make distinctive nests to which workers and working nymphs bring food. Some are excavated in the soil, whereas others are mounds rising up to 30 ft. high. Others make globular or oval nests of bits of wood and excrement cemented together with saliva and attached to trunks or branches of trees. The most elaborate nests are those of the subterranean genus *Apicotermes*, noteworthy for the geometrically arranged ventilation pores in the walls. The nests result from group behaviour, and symmetry, replication, homology, recapitulation, and vestigial features can be recognized in their structure, just as they are noted in the structure of the termites themselves.



FROM N. BANKS AND T. E. SNYDER, "REVISION OF THE NEARCTIC TERMITES"; SMITHSONIAN INSTITUTION

FIG. 5.—PRORHINOTERMES SIMPLEX, SOLDIER

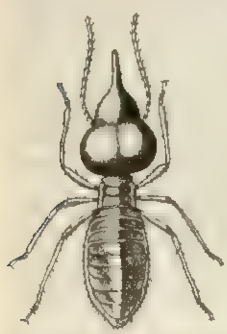
No social parasites are known among termites, although there are a few Termitidae that nest only in mounds of other species of the same family. The termites, especially the Termitidae, harbour in their colonies a host of termitophilous beetles, flies, bugs, caterpillars, etc., with remarkable exudatory organs adapted apparently to trophallactic exchange.

LIFE OF THE COLONY

Sex and Caste Determination

In nearly all Hymenoptera, unfertilized eggs develop into males and fertilized eggs become females. There are exceptions to this generalization, however: homozygous fertilized eggs in certain parasitic Hymenoptera produce males; unfertilized eggs of certain ants (*Oecophylla* species) produce females, as do eggs laid by unfertilized workers of the common honeybee occasionally and by the South African race of honeybees regularly.

In most social Hymenoptera the queen lays both fertilized and unfertilized eggs, thus producing offspring of both sexes. Eggs occasionally laid by workers, which rarely mate, are usually unfertilized. A high percentage of the males of some of the halictid bees may result from worker-laid eggs; the workers in this group have ovaries often sufficiently developed to produce eggs. In many hymenopteran colonies the loss of the queen permits ovarian development among workers so that many worker-laid eggs and ultimately many males may be produced before the colony finally dies out for lack of a queen.



FROM N. BANKS AND T. E. SNYDER, "REVISION OF THE NEARCTIC TERMITES"; SMITHSONIAN INSTITUTION

FIG. 6.—NASUTE SOLDIER OF TENUIROSTRITERMES TENUIROSTRIS

After several matings the queen honeybee has a supply of sperm adequate for her entire life; these are stored in the sperm sac, or spermatheca. She can release a few sperm onto an egg just before it is laid by relaxing a muscle that closes the spermathecal duct. Little is known about the factors that cause queens to fertilize some eggs and not others, but in the honeybee the structure of the cell in which the queen is laying is involved. If she lays in a small (worker) cell where her abdomen fits snugly, she is stimulated to release sperm cells and lay a female-producing egg. The same is true if she lays in an extraordinarily large (queen) cell. But when she lays in a cell similar to the worker cell but slightly larger so that her abdomen does not fit snugly, she is not stimulated to release sperm, and a male egg results. In a sense, therefore, the workers are ultimately responsible for determining the sex of eggs laid by the queen, since they construct the cells.

Why is it that some female eggs develop into queens and others into workers? The environment (food or other aspects) plays a most important role in determining the caste. Caste determination can occur at almost any stage in development from egg to adult and may occur at different stages among individuals of a single species. In some forms having weak caste differences, determination occurs in the adult stage. This is true in some halictid bees (*Lasioglossum marginatum*), as well as in the polistine wasp *Polistes gallicus*. In the latter, if several overwintered potential queens start a common nest in the spring, a socially dominant one becomes the queen, while the others, although mated, have the short life, small ovaries, and behavioural characteristics of workers. If the dominant individual is removed, however, one of the others replaces her as queen. In the same species, workers of the ordinary sort produced in the summer are determined in the larval stage instead of in the adult.

In the honeybee the castes are determined in the young larval stage. Larvae in queen cells are fed exclusively bee "milk" or royal jelly, a secretion of the pharyngeal glands of young adult workers, while larvae in worker cells are fed such a substance for about 2½ days, after which they receive bee "bread," a food consisting largely of honey and pollen. Female larvae can be shifted from queen to worker development or vice versa providing that transfer to the new type of cell occurs before the larva is 2½ or 3 days old. It is evidently the size and shape of the cell that stimulates nursing workers to feed the larva in a cell in a particular manner that results in one caste or the other.

In spring, queen cells are constructed in small numbers in nearly every colony. However, at other times of year they are constructed only if the queen dies or is removed. The result is production of replacement queens after the original queen is lost; such queen cells can be constructed around eggs or young larvae already present in the colony but not yet determined as to caste. Mechanisms determining construction of queen cells (other than those normally constructed in the spring) have been studied in detail. The presence of a laying queen inhibits the construction of other queen cells and also inhibits ovarian enlargement among workers. Queen cell construction and ovarian development are also inhibited in queenless colonies if workers are continuously introduced from the presence of a queen or if bodies of dead queens or extracts of queens are introduced. Mere sight, sound, or odour of queens have little effect. The inhibitory influence is "queen substance," a pheromone identified as 9-oxodec-2-enoic acid. This substance is stored in the mandibular glands of the queen and is released through the ducts of these glands opening at the bases of the mandibles. It is obvious that in a populous colony, *e.g.*, 50,000 bees, relatively few individuals can contact the queen each day; yet queen cell construction begins only a few hours after the loss of a queen. This must be a reaction to the decrease in the total concentration of queen substance among the aggregate of workers; the substance is transferred by mutual food exchange and licking among the workers and thus influences all members of the colony continuously. There is evidence that queen substance is not the only pheromone involved in inhibition of queen cell construction, but other pheromones have less noticeable effects.

Among some ants, larvae are undetermined as to caste late into the last larval stage, while in other species caste is fixed in the first larval stage, and a strong caste bias may exist even in eggs. Such environmental factors as kind and quantity of food and the developmental stage that the larva has reached when it goes into winter quiescence are involved in caste determination. In species where caste bias appears in eggs, important factors are the temperature of the queen and the kind of food that she is receiving at the time the eggs are produced.

Only in the stingless honeybee, *Melipona*, is there an indication that queen and worker eggs differ from one another genetically. Kerr (1950) has shown that under certain conditions queens are produced at the rates approaching 25% (for one species) and 12.5% (for other species) of the total female production. Most young queens are killed off by workers; they cannot be useful in such enormous numbers. These rates of queen production suggest a genetic agency.

In termites the control of social polymorphism is due to a very complicated mechanism involving several pheromones. Probably

all individuals have the initial tendency to develop as reproductives. For some reason that is not known, large numbers achieve this caste in established colonies at certain seasons when winged reproductives are produced and leave the colony. Otherwise, pheromones, different for the two sexes, secreted by the king and queen, licked from their body surfaces, and transmitted among the members of the colony by interchange of food and mutual licking, inhibit nymphs from developing toward the reproductive condition. On the death of a queen, however, her inhibitory pheromone is no longer produced, and one or more female nymphs of the appropriate age and sensitivity to the reduction in concentration of the pheromone develop as secondary queens. The great majority of nymphs are inhibited from developing as reproductives and tend to become soldiers. Soldiers also secrete inhibitory pheromones, in such quantities that, if roughly 10% of the adults in a colony are soldiers, inhibition of development toward soldierhood occurs. If the percentage of soldiers drops below 10%, then some nymphs are able to become soldiers. Nymphs that are prevented by inhibitory pheromones from becoming reproductives or soldiers probably tend to remain for long periods in the nymphal stage in the more primitive families of termites or mature as adult workers in the more specialized families. Detailed studies based on *Kalotermes* indicate that the above summary is simplified and that there are pheromones that promote as well as inhibit development toward certain castes. In a termite colony having an excess of reproductives or soldiers, the excess will be killed and eaten by workers and nymphs until the normal relative abundance of castes is reestablished (another group of pheromones stimulates the eating of excess reproductives or soldiers).

In the highest termites (some Termitidae) castes are in part genetically controlled. This is suggestive of *Melipona* among bees. In both bees and termites, evolution has seemingly been toward the taking over by genetic mechanisms of caste control that was originally environmentally regulated.

Foraging and Communication

There are obvious advantages to any social insect that forages far from its nest to have the ability to communicate to its colonymates about a food source it has found. There is no evidence of such communication in primitively social bees, in bumblebees, or in ants that make use of widely scattered small food sources like grass and weed seeds. In some other social insects, the communication related to foraging is merely social facilitation; *i.e.*, a successful forager stimulates other workers of appropriate age and physiological condition to go out foraging, but no information is conveyed as to the location of the food sources. Such stimulation occurs in certain kinds of ants (*Formica*), in social wasps, and in stingless honeybees (some *Trigona*). In many ants, however, trail-marking pheromones integrate the foraging procedure. Such ants generally exploit food sources too large to be carried back to the nest by a single ant. Probably only a small percentage of the workers, presumably the oldest and most experienced, go out from the nest as individual foragers or "scouts." On finding food, a scout takes a sample back to the nest, using her knowledge of landmarks, odours, the position of the sun, and the like, to take a more or less direct course back. On the way back she marks her trail by dragging or lowering her abdomen and releasing a pheromone. This substance stimulates other ants of the same species to follow the marked trail and come upon the food source. If the food source is rich, the trail becomes better marked and therefore increasingly used. By using the gland substance it is possible to make artificial trails that ants will follow. Chemically marked trails permit the colony to utilize relatively young and inexperienced workers that cannot find their way as lone foragers. The trail pheromones of some ants are extremely volatile so that only relatively short trails are possible and foraging occurs close to the nest. For other species, the trail pheromones are more lasting, and trails can sometimes be followed by ants after days or weeks of disuse and even after rain.

Some stingless bees (*Melipona* and some *Trigona*) use a similar device. After a forager has returned several times from a newly found food source to the nest, she alights at frequent intervals leaving odour spots (mandibular gland secretion) on the ground or on vegetation. Once back in the nest, she stimulates other potential foragers by a high-pitched buzzing sound and then leads a group of them out of the nest and along the trail of odour spots to the food. In contrast to the ants, both a leader and an odour trail are involved, possibly because the pheromonal trail cannot be continuous in the case of a flying insect.

Some termites are also capable of leaving odour trails using a gland on the underside of the abdomen.

The remarkable system of communication in honeybees was discovered by Karl von Frisch. A honeybee that has found a rich food source marks the food with an odour emitted from a scent gland on the apex of its abdomen. After returning to the hive, the bee may impart the odour of the food to other bees, but the main items of importance—distance and direction to the food—are conveyed by "dances" on the vertical combs in the hive. For northern European honeybees, if the food is 50 m. or less from the hive, the scout performs a round dance making circles first in one direction and then the other, about an inch in diameter, running on the surface of the comb. If the distance is between 50 and 100 m., the dance includes a short straight run between turns, and the abdomen is wagged from side to side during the straight run. For distances greater than 100 m., the straight run between two half circles is considerably longer. The speed of the dance, as well as sounds associated with it, varies with distance. For distances under 100 m., direction is scarcely indicated, but over 100 m., the direction of the straight wagging run indicates the direction of the food in relation to the position of the sun. If the food is toward the sun, the straight run is vertical, up the comb. A downward run indicates a direction away from the sun. A deviation of, for example, 10° to the right of the vertical indicates the direction of 10° to the right of the direction toward the sun. Any angle in the compass can be indicated by this system. Of course, the dance is modified as the sun moves. If the sun is straight up (noon during certain days in the tropics), communications stop from a few minutes before to a few minutes after noon. During a period of light overcast, bees orient by perceiving cloud-penetrating ultraviolet radiation from the sun. If any blue sky is visible, they get their cues from the polarization of light from the sky. Similar communication is used by swarming bees to transmit information about suitable nesting sites to members of the swarm. Minor heritable differences exist among communication systems of various geographical races of honeybees. Bees in colonies made by mixing subspecies misunderstand the dances of individuals of races other than their own.

Martin Lindauer examined in detail the communication by which workers learn for themselves the current necessities of the hive. They do this by extensive patrolling and inspection. Apparently their alertness to the needs of the hive makes every bee available for community work at any moment, without the necessity for a direct summons from other bees. Pheromonal stimulants to aggression occur in ants and bees. In the honeybee such a pheromone is liberated when a sting is inflicted, stimulating other bees to sting also.

Bibliography.—*Social and Subsocial Insects:* W. M. Wheeler, *Social Insects, Their Origin and Evolution* (1928), *Ants* (1910); C. D. and M. H. Michener, *American Social Insects* (1951); O. W. Richards, *Social Insects* (1953); J. S. Kennedy (ed.), *Insect Polymorphism* (1961); E. O. Wilson, "Chemical Communication in the Social Insects," *Science*, vol. 149, no. 3688, pp. 1064–71 (Sept. 3, 1965); H. F. Schwarz, "Stingless Bees (Meliponidae) of the Western Hemisphere," *Bull. Amer. Mus. Nat. Hist.*, vol. xc (1948); K. von Frisch, *Bees: Their Vision, Chemical Senses, and Language* (1950), *Dancing Bees* (1955); C. R. Ribbands, *The Behaviour and Social Life of Honeybees* (1953); J. B. Free and C. G. Butler, *Bumblebees* (1959); M. Lindauer, *Communication Among Social Bees* (1961); A. H. Forel, *Social World of the Ants Compared with That of Man* (1929); W. S. Creighton, "The Ants of North America," *Bull. Mus. Comp. Zool. Harv.*, vol. civ (1950); N. Banks and T. E. Snyder, "Revision of the Nearctic Termites," *Bull. U.S. Nat. Mus.*, no. 108 (1920); A. E. Emerson, "Regenerative Behavior and Social Homeostasis of Termites," *Ecol.*, vol. xxxvii (1956); S. H. Skaife, *Dwellers in Darkness* (1955); R. Schmidt, "The Evolution of Nest-Building Behavior in Apicotermes," *Evolution*, vol. ix (1955).

Insect and Human Society: A. E. Emerson, "Structure et Physiologie des Sociétés Animals," *Colloq. Int. Cent. Nat. Rech. Sci.*, no. 34 (1952),

"Dynamic Homeostasis: a Unifying Principle in Organic, Social, and Ethical Evolution," *Sci. Mon.*, vol. lxxviii (1954); T. C. Schneirla, "Problems in the Biopsychology of Social Organization," *J. Abnorm. & Soc. Psychol.*, vol. xli (1946). (C. D. M.)

SOCIALISM, a broad term that generally denotes a system of public ownership and management of the means of production and distribution of goods as contrasted to capitalism (*q.v.*), that emphasizes private ownership and management. There are many varieties of socialism and the term has had a long and a complicated history. The presentation of the subject in this article follows historical lines and is based on the following outline:

- I. Early Socialists and Their Predecessors
  - 1. Origins
  - 2. The Saint-Simon and Fourier Doctrines
  - 3. Owenism
  - 4. First Anticapitalist Writers
  - 5. Christian Socialism
  - 6. French Socialism
- II. Karl Marx and His Influence
  - 1. The Communist Manifesto
  - 2. The First International
  - 3. Social Democracy
  - 4. Fabianism
  - 5. Socialism in the U.S.
- III. Later National Developments
  - 1. France
  - 2. Germany
  - 3. The Second International
  - 4. Syndicalism
  - 5. Guild Socialism
  - 6. The War Issue
- IV. Russian Socialism
  - 1. Socialist Revolutionaries
  - 2. Social Democrats
  - 3. Bolsheviks and Mensheviks
  - 4. Communism
- V. Modern Trends
  - 1. Western Europe
  - 2. North America
  - 3. Australia, New Zealand, and South Africa
  - 4. The Far East
  - 5. Contemporary Socialism
  - 6. Interpretation of Socialism

I. EARLY SOCIALISTS AND THEIR PREDECESSORS

1. Origins.—The words socialist and socialism came into use in Great Britain and France soon after 1825, and were first applied to the doctrines of certain writers who were seeking a complete transformation of the economic and moral basis of society by the substitution of social for individual control and of social for individualistic forces in the organization of life and work. Socialist seems to have been used first in Great Britain (*Co-operative Magazine,* 1826) to describe the followers of Robert Owen, and the word *socialiste* in France (*Globe,* 1832), with reference to the followers of Claude Henri de Rouvroy, comte de Saint-Simon. In Great Britain the followers of Owen officially adopted the name Socialists in 1841. The word socialism as the antithesis to individualism was popularized by P. Leroux and J. Reynaud in their *Encyclopédie nouvelle* and in their other writings, and by 1840 had come to be used freely in Europe to describe the schools of Saint-Simon, Charles Fourier, Owen and others who attacked the existing system of commercial competition and put forward proposals for a new way of life based on collective control. Later, these early schools of socialism were categorized by Karl Marx and Friedrich Engels as "utopian socialism," in contrast with the "scientific socialism," based on the materialist conception of history, of which they were the pioneers.

Later writers, delving back into the past, found anticipations of socialist doctrine in many earlier theorists, and applied the name over a wide field—for example, to Plato among the ancients, to a number of medieval writers, to Sir Thomas More (*Utopia,* 1516), T. Campanella (*City of the Sun,* 1623) and other utopia makers, and to such 17th-century social critics as the curé Jean Meslier in France and the Digger Gerrard Winstanley (*The Law of Freedom,* 1652), Peter Chamberlen (*The Poor Man's Advocate,* 1649) and the Quaker John Bellers in England. Among 18th-century writers the name is applied to Gabriel Bonnot de Mably (*Entretiens de Phocion sur le rapport de la morale avec la politique,* 1763) and Morelly (*Code de la nature,* 1755), the foremost critics of economic and social institutions among the philosophers who prepared the way for the French Revolution.

Among those who took part in the Revolution the name socialist is given to Jacques Pierre Brissot (1754–93), the Girondin leader, and sometimes to Jacques René Hébert, and above all to François Noel (Gracchus) Babeuf (1760–97), leader of the Société des Égaux, whose unsuccessful attempt to overthrow the Directory government in 1796 is commonly regarded as the starting point of the modern socialist movement. Babeuf's revolutionary group set forth the first practical program, and the *Manifeste des égaux* was the first socialist pronouncement. Babeuf and his followers regarded socialization of land and industry as necessary to complete the revolution begun in 1789. They proclaimed the equal natural right of all men to the enjoyment of the goods provided by nature, the universal obligation to labour, the universal right to education, and the necessity of abolishing both riches and poverty in the interests of human happiness.

Babeuf sought to establish his system by revolution carried through under a revolutionary dictatorship. His ideas were popularized later by his collaborator, Philippe Michel Buonarroti (1761–1837), whose *Conspiration de Babeuf* (1828), translated by James Bronterre O'Brien, influenced the English Chartists. His views on landed property had something in common with those of the British 18th-century land reformers (Robert Wallace, 1697–1771; William Ogilvie, 1736–1819; Thomas Spence, 1750–1814; and Thomas Paine, 1737–1809), but he went far beyond them in the amplitude of his social designs. Spence is nearest to him, and the Society of Spencean Philanthropists, formed by Spence's followers in 1816, can be claimed as the first organized socialist body in Great Britain. Spence stood for the collective ownership of the land by local communes, which would take over the government and provide by federation for the simple needs of administration under a communal system. Paine, on the other hand, was not a socialist, but a social reformer who, apart from his doctrines about land, anticipated many modern social service projects in part ii of his *Rights of Man.*

Nor can William Godwin (1756–1836) be rightly described as a socialist. In his *Political Justice* (1793) he declared for the abolition of all government, and advocated a free society in which property would go to those who could make the best social use of it. Godwin believed in the possibility of a completely rational social system without government, based on right education in moral and social principles. He inspired much later socialist as well as anarchist thinking, and had a considerable influence on Owen's ideas as well as on those of the British radicals of the 1790s, including Wordsworth, Southey and Coleridge in their early days. Shelley's *Prometheus Unbound* was largely based on the ideas of Godwin, whose daughter he married.

2. The Saint-Simon and Fourier Doctrines.—Socialist thought took a quite different turn with the comte de Saint-Simon (1760–1825), a French aristocrat whose outlook was formed not only by the French Revolution but also by a keen awareness of the impending revolution in industrial technique and in the role of science in human affairs. Saint-Simon held that the new social forces that had been unloosed by political revolution and scientific advance called imperatively for planned organization and control in the general interest, and was the first to see clearly the dominant importance of economic organization in the affairs of modern society and to grasp the concept of economic evolution as the key factor in social adjustment. In his most important works (*Introduction aux travaux scientifiques du XIX^e siècle,* 1807; *Nouvelle Encyclopédie,* 1810; *De la Réorganisation de la société européenne* [with the historian Augustin Thierry], 1814; *L'Industrie,* 1816–18; *Nouveau La Politique,* 1820; *Catéchisme des industriels,* 1823–24; *Nouveau Christianisme,* 1825) he put forward, not a system, but a series of originative ideas which were worked up by his disciples into a system. He stood, not for a struggle between employers and workers, but for a community among the productive classes against the parasitic nonproducers, on a basis of social ownership of the means of production and their administration by men of the requisite scientific and business knowledge. He believed in unequal rewards,

corresponding to real differences in quality of service, and in the conferring of large powers on a directing authority constituted on a basis of merit. The state, he held, was under an obligation to provide work for all, and all were under an obligation to labour for society according to their powers. Saint-Simon vigorously denounced the exploitation to which the labourer was subject under the existing system of property rights, and anticipated Marx in holding that the property relations sustained by any social order conferred upon it its essential character. He believed that human society tended in the march of history toward a system of universal association, which would be the guarantee of peace; and he regarded the Roman Catholic Church as having been a step toward this association. Holding its doctrines outmoded, he put forward in his *New Christianity* (1825) a nontheological religion for the new era, based on science and expressing his faith in a "science of sciences"—a field of thought in which he was the predecessor and inspirer of Auguste Comte. Saint-Simon's great contribution to socialist thought was his insistence on the duty of the state to plan and organize the use of the means of production so as to keep continually abreast of scientific discovery, and his insistence on the master function of the industrial experts and organizers, as against the politicians and the mere "men of business," in the society of the future.

Charles Fourier (1772–1837) was a younger contemporary of Saint-Simon. His first book, *Théorie des quatre mouvements*, was published in 1808 and was followed by others in which he repeated and expanded the same ideas (*L'Association domestique agricole*, 1822, revised as *Théorie de l'Unité universelle*, 1841; *Le Nouveau Monde industriel et sociétaire*, 1829; *La Fausse industrie morcelée*, 1835–36). Fourier is best known as the advocate of the *phalanstère*, a form of co-operative settlement which he believed to be the essential unit of social organization. It was to consist of 1,500 or 1,600 persons—a number which Fourier held to be the minimum for securing a right mingling of the many different tastes and temperaments. It was to be at once a working unit and a community of persons living together in a single great building; but the members were not to be equals in income or status, and each was to be rewarded for his own share in the common labour. Moreover, Fourier accepted property rights, providing for the division of the total product five-twelfths to labour, four-twelfths to capital and three-twelfths to special talent. His proposed division was in fact more favourable to labour than this makes it appear; for he advocated a very high rate of interest on small investments of capital, falling very sharply as the amount belonging to any one holder increased. In effect, he contemplated that most people would be capitalists as well as workers.

Fourier's system rested on the belief that most forms of labour could be made highly attractive if they were rightly organized, and that no one need, or should, be made to work at anything except of his own free will. He held that everyone should have many different jobs, never spending more than an hour or two consecutively at the same work, and that all work should be organized by groups or series of workers, between whom there would be natural emulation to do well. He stressed the predominance of intensive agricultural and horticultural work both over other forms of agriculture and over manufacturing labour, which he believed should be cut down to a minimum by making things to last and by the elimination of unnecessary products. The "dirty work" he said should be done by children, to whom it could be made attractive by the spirit of emulation and service. He advocated education through doing things rather than through learning about them, and gave children a large part in the work of his utopia. He stood for the equality of men and women, regarding the advancement of women as the crucial test of social progress.

Fourier's theories were based throughout on a psychological approach. He believed that civilization erred in seeking to repress the passions, which could by right social organization, under his system, be allowed to work beneficently, as God had meant. He claimed that his system would work without any need to change human nature, which he regarded as immutable. Fourier's methods of expounding his doctrines were often fantastic, but men found much sound sense in his writings. He was entirely opposed to state action for the establishment of his system, holding that its basis must be wholly voluntary, and believing this to be possible because he thought he had discovered the form of association that really fitted in with men's natural desires. Though usually regarded as a socialist, he is in reality rather the ancestor, jointly with Owen, of voluntary producers' co-operation.

Both Saint-Simon and Fourier were the inspirers of influential schools of thought. The disciples of Saint-Simon created a formal school, headed after his death by Saint-Amand Bazard (1791–1832) and Barthélemy Prosper Enfantin (1796–1864), who separated when Enfantin developed the religious sides of his master's doctrine on extravagant lines, including a form of community living which involved "free love." This aberration over, Enfantin, like others influenced by Saint-Simon (notably the famous engineer, Ferdinand de Lesseps), became a leading industrialist, while Comte went on to build his positive philosophy on Saint-Simonian foundations. Of Fourier's disciples, the most important was Victor Prosper Considérant (*La Destinée sociale*, 1834–38; *Manifeste de l'école sociétaire*, 1841), and in England Hugh Doherty (*The Morning Star*, 1840). In the United States Fourier greatly influenced Albert Brisbane (1809–90), C. A. Dana, Horace Greeley, Margaret Fuller, Hawthorne and Emerson; and the famous Brook Farm community (1841–47), described in Hawthorne's *Blithedale Romance*, was largely Fourierist in inspiration.

3. Owenism.—Robert Owen (1771–1858), generally regarded as the founder of British socialism, had much in common with Fourier, but approached the social problem from an essentially different angle. A highly successful master cotton spinner, at the head of the great factory at New Lanark which he made a model for good labour conditions, Owen was a powerful critic of the manufacturing system introduced by the Industrial Revolution and of the exploitation of labour which it involved. He denounced competition as leading necessarily to this exploitation, and advocated in its place a co-operative system based on villages of co-operation in which men and women would work and live together on a basis of communal enjoyment of the fruits of labour in both field and factory. Fourier advocated differences of individual payment and the allowance of interest on capital. Owen was prepared to allow interest until the owners of capital voluntarily gave it up, as he believed they would; but he favoured fully communal living in his villages and elsewhere payment to producers on a basis of "labour-time" instead of money—that is, of the time necessary for the various kinds of work, which Marx later called "socially necessary labour time."

The basis of Owen's doctrine was that the social character of men (*i.e.*, their social behaviour) is made for them and not by them: it is the product of environment and education and can be changed by improving these things. He was led to denounce all established religions because they taught that men were responsible for their evil doings, instead of attributing them to bad environment, and he set out to preach social rather than moral reformation. These views were first expounded in Owen's *New View of Society* (1813–14) and more fully in his *Report to the County of Lanark* (1821). They came into prominence after the end of the Napoleonic wars, when Owen appealed to the government both for factory legislation to protect the workers and for state action to cure unemployment by setting up villages of co-operation. Unlike Fourier, Owen did not exclude state action: he appealed first to the state and then, when it failed to respond, next to private philanthropists and subsequently to the working classes directly.

Owen's views attracted a large working-class following. From about 1830 to 1834 he was in effect the leader of a vast trade union movement aiming at the emancipation of labour by setting up co-operative productive societies, labour exchanges (for the exchange of the products of different trades on a basis of labour time), and great national associations, such as the Guild of Builders, designed to take over whole industries. On the collapse of this movement in 1834 Owenism remained the creed of a substantial body of followers, who turned back to the attempt to establish villages of co-operation. Some of these had been tried already—New Harmony, in the U.S., in 1825; Orbiston, in Scotland, in 1826;

and Ralahine, in Ireland, in 1831. A last Owenite community was started at Queenwood, in Hampshire, in 1839, on a basis that incorporated some elements of Fourier's doctrines. It failed in 1845; and Owenism thereafter gradually faded away, giving place on the one hand to secularism and other forms of "rational religion" and on the other to the great successful movement of consumers' co-operation, instituted in 1844 by the Rochdale Society of Equitable Pioneers, who were ardent Owenites. (*See* CO-OPERATIVES.)

Meanwhile, in France, Fourier had a successor in utopia building in Étienne Cabet (1788–1856), author of *Voyage en Icarie* (1840) and of an unsuccessful attempt to found a community in Texas and later in Illinois. Cabet's work was much influenced by Sir Thomas More's *Utopia*. It presented the picture of a perfectly symmetrical state, owning all the main industries and dividing the product on a basis of equality. Cabet's rulers were to be chosen by popular election, but were to have large powers, analogous to those of Saint-Simon's industrial leaders. Cabet borrowed some of the features of Fourier's *phalanstères*, but departed from Fourier in laying great stress on the state and on uniformity of dress, status and behaviour among the citizens. Like Saint-Simon, he put much emphasis on the role of science and, like Fourier, on education, which, with Owen, he regarded as the means of training children in the right social ideas. Pending the establishment of Icaria, he was a strong advocate of minimum wage laws and of progressive taxation of the rich. (*See* UTOPIAN COMMUNITIES.)

4. First Anticapitalist Writers.—Cabet was the last of the group of utopia makers who represent socialism in its first phase of evolution. Saint-Simon, though contemporary with this group, does not belong to it; and Owen is notable not only as a utopia maker, but even more as a leader of great working-class movements and an acute critic of the doctrines of orthodox political economy. In this last field he was the first of a group of writers who took the accepted principles of classical economics and based on them anticapitalist or socialist deductions. Owen's *Report to the County of Lanark* in many respects anticipates Marx; and Owen's leading follower in the realm of theory, William Thompson (1785–1833) (*Principles of the Distribution of Wealth Most Conducive to Human Happiness*, 1824), develops further the doctrine that, as labour is the sole true source of value, the labourers by hand or brain should receive in common the entire product, to be distributed on a basis of the greatest practicable social equality. Thomas Hodgskin (*Labour Defended Against the Claims of Capital*, 1825; *Popular Political Economy*, 1827; *The Natural and Artificial Right of Property Contrasted*, 1832) similarly turned the doctrines of the Ricardians against capitalism, but, following Godwin, arrived at anarchist rather than socialist conclusions about the new social order. Apart from Godwin, the sequence of British anticapitalist writers began with Charles Hall (*The Effects of Civilisation*, 1805) and continued with the Spencean Thomas Evans' *Christian Polity* (1816), Piercy Ravenstone's *Doubts on the Correctness of Some Opinions Generally Entertained on the Subject of Political Economy* (1821) and John Gray's *Lecture on Human Happiness* (1825). The Christian Owenite John Minter Morgan contributed *On the Practicability of Mr. Owen's Plan* (1819), *The Revolt of the Bees* (1826) and *Hampden in the Nineteenth Century* (1834); and William Thompson followed up his magnum opus with *An Appeal of One-Half the Human Race* [women] (1825), *Labour Rewarded* (1827) and *Practical Directions for the Establishment of Communities* (1830). T. R. Edmonds' *Practical, Moral and Political Economy* (1828), John Gray's *The Social System* (1831), J. F. Bray's *Labour's Wrongs and Labour's Remedies* (1838–39) and Mary Hennell's *Outline of the Social Systems and Communities Founded on the Principle of Co-operation* (1844) are the other leading works of the British Owenite and anticapitalist schools. Most of these writers mingled socialist deductions from Ricardian economics with ingredients drawn from Owen's co-operative doctrines. Bray wrote in England, but was born in America.

After the collapse of the Owenite trade-union and co-operative movements in 1834, the emphasis in Great Britain shifted to Chartism, with its demand for parliamentary reform based on manhood suffrage and annual parliaments. *The People's Charter*, published in 1838, crystallized this program, behind which was the driving force of acute economic distress arising out of the new Poor law of 1834 and the prolonged industrial depression that began in the 1830s and continued into the "hungry '40s." Chartism, resting on economic distress, had no clearly defined economic program: it was a hunger revolt, without much theoretical basis in its early stages. Under the leadership of Feargus O'Connor (1794–1855) it developed in the 1840s a "back-to-the-land" program based on individual peasant holding in collectively acquired land settlements (O'Connorville, Charterville), as against the Owenite advocacy of collective tillage in villages of co-operation. The Chartist land scheme collapsed in 1848, and thereafter Chartism broke up into a number of sects, with which it will be more convenient to deal later.

5. Christian Socialism.—There had been in England in 1848 and the following years a movement for what was called "Christian socialism" (*q.v.*) but was in fact co-operativism. This was mainly due to John Malcolm Ludlow (1821–1911) who had been much influenced by Philippe Buchez, Louis Blanc and the development of co-operative workshops in France, and succeeded in enlisting John Frederick Denison Maurice (1805–72), Charles Kingsley (1819–75), Edward Vansittart Neale (1810–92), Thomas Hughes (1822–96) and others in support of a movement to set up similar workshops in England, on a definitely Christian basis. The English Christian socialists were moved by horror at the appalling conditions prevailing in the workshops and factories and at the un-Christian spirit of the spreading industrial system, but were skeptical of the Chartist remedy of political reform. Kingsley announced that in his view the "French cry, Organization of Labour" was worth a dozen of the *People's Charter*. Immediately, the Christian socialists set out to establish small working associations which they financed, in the hope of sowing the seeds of a new social order. These soon failed; but they were led to give great help to the growing consumers' co-operative movement, especially in securing for it a legal status under the Industrial and Provident Societies act of 1852. Later, the Christian socialists made many further attempts to foster producers' co-operation and also to secure the acceptance of profit sharing and copartnership in capitalist industry. They also transferred their activities to education, founding the Working Men's college in London (1854); but by the later '50s the Christian socialist movement, as such, had practically ceased to exist.

Christian socialism developed also on the continent, in somewhat different forms, largely connected with the Roman Catholic Church. In France, its forerunner was Bishop Claude Fauchet (1744–93), the Girondin, who preached a form of Christian communism at the time of the Revolution. Hugues Félicité Robert de Lamennais (1782–1854), co-founder of *L'Avenir* (1830–32), can also be regarded as a forerunner (*Paroles d'un croyant*, 1834). Buchez, as we have seen, influenced Ludlow, and there has always been in France a Catholic social movement with some affinity to Christian socialism. The movement, however, developed much more strongly in Germany, partly under the influence of the Protestant Victor Aimé Huber (1800–69), founder of the Association of Christian Order and Liberty and leading advocate of co-operative production, and partly under that of the Roman Catholic bishop Baron Wilhelm Emmanuel von Ketteler (1811–77), who also advocated co-operative production as a means of curing the ills of capitalism and promoting reconciliation of classes. The movement spread to Austria and Belgium, but, like English Christian socialism, it belongs rather to the realm of social reform than to that of socialism in the usual sense of the word.

Also on the borderline of socialism is the German economist, politician and landowner Karl Rodbertus (1805–75) (*Social Letters*, 1850–51; *Illumination of the Social Question*, 1875), who expounded the subsistence theory of wages and the labour theory of value, and proposed state action to increase the labourer's share in the product by preventing the owners of capital from appropriating the fruits of rising productivity. Rodbertus regarded society as in process of evolution from the wage system to one in which the means of production would be publicly owned, and private

property would survive only in respect of income, which would correspond to the value of each man's labour. He also threw much light on the theory of economic crises, and considerably influenced the later formulation of Marx's ideas.

A much more central figure in the development of German socialism is Ferdinand Lassalle (1825–64), founder of the Universal German Working Men's association (*Allgemeiner Deutscher Arbeiterverein*) (1863), the first organized German Socialist party, in opposition to the Liberal party headed by Franz Schulze-Delitzsch. Lassalle, who was a great orator, caused this body to spread with extraordinary rapidity, but he was killed in a duel the year after its definite formation. It languished in his successors' hands and was finally fused in 1874–75 with the rival Marxist Social Democratic party formed in 1866 under the leadership of Wilhelm Liebknecht (1826–1900) and August Bebel (1840–1913). Lassalle laid great stress on the "iron law of wages," *i.e.*, the theory that wages tended under capitalism always to subsistence level, and attacked proposals for credit associations, consumers' co-operative societies and similar bodies as useless to help the workers in face of this inexorable law. Instead, he urged that "the working class must become itself a monster employer," not in little producers' co-operative societies, but by taking over large-scale industry "in the massed and concentrated form of the factory, with its enormous advantage in productivity." As the workers could not do this by voluntary action, it was necessary to invoke the power of the state, which "exists only to expedite and assure the march of culture." The state should place the necessary credits at the disposal of the associated workers, who should proceed to take the conduct of industry into their own hands, and should receive, in addition to their wages, dividends based on the profits of the several collective undertakings. The state, says Lassalle, "is the consolidated people"; but it can be made an instrument for the proposed change only by universal and direct suffrage, which must therefore be won by means of universal association and agitation. Under universal suffrage, the state will stand ready to execute the people's will (*Open Letter to the German Working Men's Association*, 1863). Lassalle was thus an advocate of socialization, to be achieved by means of political agitation and action, through a democratization and democratic conquest of the machinery of the state. But he was also, pending this democratization, ready to appeal to the existing state to extend help to the workingmen's associations. He discussed his plans with Bismarck, who undoubtedly learned something from Lassalle both in constituting the German Reich and later in introducing the social measures sometimes called "Bismarckian state socialism."

6. French Socialism.—While Chartism was running its course in Great Britain, socialism was taking on new forms in France after the revolution of 1830. The revolt of the Lyons weavers in 1831 is often regarded as the starting point of the French proletarian movement; and under the monarchy of Louis Philippe began the insurrectionary career of Louis Auguste Blanqui (1805–81), who passed more than 40 years in prison for various conspiracies. Blanqui, whose essential writings were collected in 1885 as *La Critique sociale*, was the successor of Babeuf in the advocacy of the seizure of power by a revolutionary minority, which would at once proceed to institute a socialist system. He organized conspiracies in 1836 and again in 1839. Released in 1848, he promptly attempted a new conspiracy, and was in prison until 1859. In 1861 he was again imprisoned for four years. In 1870 he headed the extreme left in Paris, but was imprisoned by Louis Thiers before the Commune, in which his followers played an important part. Blanqui's theories of revolutionary dictatorship contributed much to the French syndicalist doctrine of *la minorité consciente*, but were opposed by the Marxists, with their rival theory of mass dictatorship of the proletariat as a whole.

Blanqui was always the leader of small groups of ardent socialist revolutionaries. Under Louis Philippe, the main body of French working-class and socialist opinion followed, not him, but either Louis Blanc (1811–82) or Pierre Joseph Proudhon (1809–65). Blanc can be claimed in many respects as the forerunner of modern social democratic doctrines. His most important work, *L'Organisation du travail* (1839), demanded that the state should give full recognition to the "right to work," and should provide capital for "national workshops" to be managed collectively by the workers under public ownership. In the revolution of 1848 Blanc became a member of the provisional government, but was unable to get his ideas carried into effect, the so-called "national workshops" started by his opponents being mere relief works which were meant to fail so as to discredit his proposals. Blanc went into exile when Louis Bonaparte came to power, and spent many years in England, writing a series of important books on British social and political conditions. His other main socialist writings are *Le Socialisme: droit au travail* (1849); *Catéchisme des socialistes* (1849) and *Histoire de dix ans, 1830–40* (1841–44). Blanc was the originator of the socialist formula "From each according to his abilities: to each according to his needs." He stands for a socialism resting on public ownership, combined with workers' control in industry, and for a democratic parliamentary system as the instrument of socialist advance.

The idea of co-operative workshops, without that of state control, was also actively urged in the 1830s by the follower of Saint-Simon, Philippe Buchez (1796–1865), who became the foremost advocate in France of Christian socialism and largely inspired the English Christian socialists. His views are embedded in his periodical, *L'Européen*, and in his *Essai d'un traité complet de philosophie* (1839) and *Traité de politique et de science sociale* (1866).

Much more important than Buchez was that singular genius, P. J. Proudhon (1809–65), whose work belongs essentially to the history of anarchism rather than of socialism. From *Qu'est-ce que la propriété?* (1840), with the startling answer "*La Propriété, c'est le vol*" ("Property Is Theft"), to *De la capacité politique des classes ouvrières* (1865), Proudhon poured out a spate of books in which he mingled devastating criticism of the institutions of capitalism with emphasis on the inherent capacity for self-organization of the working class. He became the leading advocate of *mutuellisme*, whether it took the form of co-operative societies or labour banks or trade societies, or any form of working-class organization in which might be descried the germ of the classless, nongovernmental society of the future. Proudhon called himself a socialist as an opponent of capitalism and individualism and private property; but his positive doctrine always laid the main stress on personal freedom and voluntary association in a society freed from the tyranny of monopolist property rights. He favoured personal property and inheritance under a moralized system of property rights designed to exclude exploitation, and laid great emphasis on the abolition of interest and the provision of free credit to the producers. His best-known books, in addition to those already cited, are *Système des contradictions économiques ou, Philosophie de la misère* ("The Philosophy of Poverty") (1846), which provoked a famous reply from Marx: *Misère de la philosophie* ("The Poverty of Philosophy") (1847); *Confessions d'un révolutionnaire* (1849); *Idée générale de la révolution au XIX^e siècle* (1851); *La Révolution sociale* (1852).

Proudhon was addicted to complicated logical arguments, based largely on the idea of inherent "contradictions" in capitalist society. He used, or, as Marx held, misused, the Hegelian dialectical method extensively. He was not a system maker but a critical and suggestive writer of great force. Marx classed him as a *petit bourgeois*, because he had little grasp of historical evolution or of the power of great industry, and because his mutualist doctrines were taken up by the declining class of small-scale producers and were inapplicable to large-scale capitalist conditions. Proudhon's ideas had great influence on the development of French syndicalism and he ranks, with Mikhail Bakunin, as the principal philosopher of 19th-century anarchism.

II. KARL MARX AND HIS INFLUENCE

Up to the 1840s socialism developed almost exclusively as a French and British movement. Industrial evolution was much more advanced in Great Britain than elsewhere; and France had been, from the 18th century, the principal seedbed of revolutionary political ideas. In Germany, which was politically divided and economically much more backward, socialism began as a movement when the more radical followers of Georg Hegel came into

contact with advanced French social ideas in the course of the '30s and '40s. Johann Fichte (1762–1814) has sometimes been claimed as an ancestor of German socialism, because he exalted, as a nationalist and a social reformer, the claims of the state; but there is no substance in this view, unless all statist doctrines (including modern national socialism) are to be regarded as socialist.

German socialism really begins with the materialist philosophy of Ludwig Feuerbach (1804–72), who much influenced Marx, and with the "left" Hegelians, such as Bruno Bauer (1809–82), Moses Hess (1812–75) and Karl Grün (1817–87). Beginning as a realist or materialist critique of religion and of the established German philosophy (*see* Karl Marx and Friedrich Engels, *German Ideology*), this "left" Hegelianism, in which Marx was brought up, broadened out under French influence into Marxist "scientific socialism," taking the materialist conception of history, first clearly formulated by Karl Marx (1818–83) and Friedrich Engels (1820–95) in the *Communist Manifesto* of 1848, as its essential foundation. This famous manifesto, the starting point of the modern socialist and communist movements, was issued in the "year of revolutions," 1848, by the Communist league, which Marx had created on the basis of the German societies of exiles that had been preaching various forms of utopian socialism from the late '30s. Wilhelm Weitling (1808–71) (*Guarantees of Harmony and Freedom*, 1842) was the principal leader of this pre-Marxist German socialist movement, but was pushed out of the way by Marx and Engels as a moralist-utopian as soon as they had formulated their essential ideas. The League of the Just (1836–47) gave place to the Communist league, which at its congress in Dec. 1847 authorized Marx and Engels to draft the *Communist Manifesto*.

1. The Communist Manifesto.—This most celebrated of all socialist documents is the first clear proclamation of the revolutionary role of the proletariat, or working class. It sets out from an interpretation of all human history as a sequence of vast class struggles for power. Behind these struggles, and as their cause, lies the development of the "powers of production," or, in other words, of man's command over the forces of nature. Every stage in this development demands a corresponding organization of social forces for its exploitation—a particular arrangement of property rights and class relationships, sustained by an appropriate political and ideological system. This political and ideological system, however, though it is used to enforce obedience to the class conditions required, is not their cause but their result. The driving force lies in the powers of production themselves; and, as these develop, a disharmony arises between them and the social and political structure based on what they were at an earlier stage. This disharmony expresses itself in class struggles and leads to social revolution, which, when it comes, rapidly destroys the obsolete social structure and replaces it by a structure in harmony with the changed needs of the time. Such social revolutions involve and include revolutions in ideas as well as in institutions: they are the major turning points of human history.

Marx and Engels were far from supposing, as some of their critics have taken them to suppose, that all historical events could be explained in terms of this formula. It applies only to the great revolutions of history, and not to day-to-day events outside moments of historic crisis. Nor did Marx and Engels ever suggest that men are moved only by economic or self-interested motives. That is not the point, which is that, whatever may be the motives that move men, they are impelled by historical necessity to adapt their social structures and ideas to the requirements of the development of the powers of production. "Man," Marx writes, "always makes his own history," but he goes on to say that man makes it within the limiting conditions set by the problems and material realities of his own time and place.

The *Communist Manifesto* throws out, on the basis of this conception of history, a resounding challenge to the governing class, which is warned that the proletarian revolution is on the march. It includes also a searching criticism of other schools of socialists who fail to take their stand upon the principle of the class war. The work of the utopian socialists is passed under review; the followers of Proudhon are castigated; adverse comment is passed on Christian socialism and on the so-called "professorial socialism" which was becoming fashionable among a school of German economists. The communists, on the strength of their understanding of the historical role of the proletariat, put themselves forward as the natural leaders of the coming revolution, and sweep aside the *petit bourgeois* radicals whom they regard as in process of being rapidly superseded by the development of economic forces. This is not meant as an adequate summary of the *Communist Manifesto:* there is no space to deal with it adequately. All that can be attempted is to select the points most important for the history of socialism.

The defeat, in Germany and all over Europe, of the revolution of 1848 naturally involved a setback for European socialism. The Communist league disappeared; the leaders of socialism were again scattered and mostly in exile. But from this time on Marxism, though not yet fully developed as a system, found adherents in most of the leading countries. In England it influenced the later stages of Chartism, which became more consciously socialist as it ceased to be a mass movement. Of the Chartist leaders the three who have some claim to a place in socialist history are George Julian Harney (1817–97), Ernest Charles Jones (1819–69) and James Bronterre O'Brien (1805–64). Harney, who first published the *Communist Manifesto* in England, was a revolutionary who looked back to 1789 and regarded the proletarian revolution as the necessary completion of the French. He was influenced by Marx, but presently Marx discarded him as too unstable, and transferred his hopes to Ernest Jones, who became the principal English exponent of Marxism in his speeches and journals (*Notes for the People*, 1851–52; *The People's Paper*, 1852–58). O'Brien, also much influenced by the French Revolution and especially by Babeuf, never became a Marxist. Originally an Owenite, he developed into a state socialist, looking forward to nationalization of the land, public utilities and industries subject to monopoly. His principal work is his unfinished "Rise, Progress and Phases of Human Slavery" (in *Reynolds' Political Instructor* 1849, republished 1885). By 1860 Chartism in these later forms had flickered out, and socialism hardly existed in England except in the form of sympathy for utopian socialism in the writings of John Stuart Mill and some other leading economists.

2. The First International.—Meanwhile Marx, in London, had published in 1859 his *Critique of Political Economy*—virtually an advance sketch of *Capital*, and containing a plain exposition of the essentials of the materialist conception of history. The first volume of *Capital* appeared in 1867—the only volume published during Marx's lifetime. But before this Marx had succeeded in bringing into existence in London the International Working Men's association, or "first international" (1864) (*see* INTERNATIONAL, THE), consisting at the outset mainly of continental exiles and British trade-union leaders who sympathized with foreign revolutionary aspirations, however little they were disposed to foment revolution in their own country. The project of an international association had come in the first instance largely from the followers of the liberal nationalist leader Giuseppe Mazzini; but Marx succeeded in pushing them aside and in securing a body ready to follow his own lead. He drafted the address and statutes, and soon branches were founded in a number of European countries including Germany, France, Belgium, Italy, Switzerland and Spain. The new international played a considerable part in the labour troubles of the later '60s in France and elsewhere; and the outbreak of the Franco-Prussian War in 1870 found it active all over western Europe. In France and Belgium it was at first largely dominated by the followers of Proudhon and the Belgian collectivist socialist Jean Guillaume César Alexandre Hippolyte Colins (1783–1859) in Germany by the Marxists, in Italy and Spain and parts of Switzerland by the anarchists, of whom the Russian Mikhail Bakunin (1814–76) became the acknowledged leader. The first international held congresses in London (1865), Geneva (1866) Lausanne (1867), Brussels (1868), Basel (1869) and, after an interval caused by the war, at The Hague (1872). Its statutes and addresses were widely drawn, to attract adherents from a number of different schools; and its general council was composed largely of British trade-union leaders, who left the running to Marx and the group of exiles he had gathered round him. In 1867 there was

a struggle between the Marxists and the followers of Proudhon, who were driven out; and thereafter until 1871 Marx was practically in control.

The testing time for the first international came with the proclamation of the Paris Commune in 1871, after the defeat of France by Prussia and the fall of Napoleon III. Marx had advised his French followers against the rising, but when it took place he rallied at once to its defense. In Paris, members of the first international took part in the Commune side by side with the followers of Blanqui and of Proudhon. The Paris Commune was the first occasion on which the working class actually achieved political power, albeit only for a few weeks; and the rising was of immense symbolic importance for socialists everywhere. Marx, in the manifesto upon it (published as *The Civil War in France*, 1871) which he wrote in the name of the international, seized on its significance for the strategy of socialism as a whole. The proletariat, he held, could not assume power merely by taking over the existing apparatus of the bourgeois state. It needed to break the old state in pieces and to construct in its place a new proletarian state of its own. He praised the Communards for what they had done in this way, while criticizing them for having proceeded too slowly and too mildly to meet the need. At a time when the entire capitalist world was denouncing the Commune for its revolutionary excesses, Marx stressed rather the hesitancy and the clemency with which the proletariat had employed its new power.

There was acute controversy in socialist and labour circles about the Paris Commune. Marx's defense was too strong meat for many of the English trade-union leaders who had hitherto lent their names to the International Working Men's association; and in all countries the more timid or reformist adherents of socialism were scared off. The defeat of the Commune by Thiers and the reactionary French national assembly which wiped it out in blood was fatal to the international, already torn asunder by the dispute between the Marxists and the anarchist, or antipolitical, followers of Bakunin. At the Hague conference of 1872 Marx, in order to get rid of what had become an embarrassing child and to prevent its appropriation by the followers of Bakunin, obtained a vote removing the seat of the international to the United States, where, after lingering nominally for a few years, it was ingloriously wound up. The followers of Bakunin maintained for a time their rival international in Europe; but it struck no lasting roots except in Spain.

The question posed by Marx in connection with the Paris Commune—should the proletariat aim at conquering the bourgeois state in order to use it as the instrument of proletarian power, or should it set out rather to destroy the bourgeois state and replace it by a new, proletarian state of its own?—continued to divide the socialist movement. This very question came to a head when it was proposed, a few years after the Commune, that the German followers of Marx and of Lassalle should combine to form a single Social Democratic party. Lassalle, as we have seen, had stood for the conquest and use of the state, which he had regarded as needing only to be based on universal, direct suffrage in order to become an instrument of social progress. Marx, on the other hand, held that the bourgeois state was incapable of such conversion, and indeed that every state was essentially a class institution, made in the image of a particular ruling class—so that each class, on coming to power, would need to break the state of its predecessor and make a new state of its own as the instrument of its "dictatorship." This is the theory of the dictatorship of the proletariat, which the Bolsheviks claimed to apply in the Soviet Union after 1917.

3. Social Democracy.—Liebknecht and Bebel, Marx's leading supporters in Germany, were not prepared in 1875 to jeopardize union with the Lassalleans by insisting on this principle. The Gotha program, on the basis of which the German socialist forces were united in that year, accepted by implication Lassalle's view of the state. Marx, to whom the draft was sent, protested violently in a memorandum to the German Social Democrats (subsequently published as *Critique of the Gotha Programme*); but the leaders in Germany actually suppressed his criticism, and pushed the amalgamation through. The German Social Democratic party thus came to birth with a contradictory set of objectives. Most of the language of its program was that of Marxism, but embedded in it was the objective of winning power by taking over the control of the bourgeois state, rather than by overthrowing it, together with an advocacy of co-operative production on the lines proposed by Lassalle. In effect, the German Social Democratic party became a parliamentary party, aiming at the assumption of political power by constitutional electoral means. Its leadership, however, was in the hands of men who regarded themselves as devout Marxists but who refused to accept the dictatorship of the proletariat as part of the essential Marxian doctrine.

This was of vital importance for subsequent socialist history, for over most of western Europe German social democracy became the model on which national Social Democratic parties were organized. In France Jules Guesde (1845–1922), after several years of agitation, was joined by Marx's son-in-law Paul Lafargue (1842–1911) in founding the Marxist Parti Ouvrier Français in 1880. In Belgium the followers of César de Paepe (1842–90) formed a similar party, but closely linked up with the trade-unions and co-operative societies, in 1885. In Spain Pablo Iglesias (1850–1925) created a Marxist party, in rivalry to the anarchists, in 1879. In Scandinavia, the Danish party dates from 1879, the Norwegian from 1887, the Swedish from 1889 and the Finnish (the first to secure a parliamentary majority) from 1899. The Italian Socialist party dates from 1892, the Polish Socialist party from the same year and the Dutch Social Democratic party from 1894. The Russian Socialist party was founded, as the Emancipation of Labour group, by G. V. Plekhanov (1856–1918) and P. B. Axelrod (1850–1928) in 1883. Thus Marxian socialism spread over Europe during the final quarter of the 19th century. In some countries, such as Russia and Poland (mainly under Russian rule), the political conditions made parliamentary activities impossible, and the parties maintained a decisively revolutionary and underground character. In most countries, however, they became national parliamentary parties, compelled by the exigencies of electioneering to put forward programs of social reform to be pressed for in advance of the full conquest of political power and the establishment of a socialist system.

In Great Britain the attempt to establish a Marxist party was made by Henry Mayers Hyndman (1842–1921), who was chiefly responsible for founding the Democratic federation in 1881. This body had at first only a semisocialist program, as it was attempting to attract into its ranks the radical workingmen's clubs which had followed Charles Dilke and Joseph Chamberlain in their radical phase. But in 1884 the federation became the Social Democratic federation and adopted a completely Marxist program. The same year it split, William Morris (1834–96), the socialist poet and artist, and others, with the support of Engels, breaking away to form the Socialist league. Neither of these bodies attracted a large following; but their influence was felt indirectly in the great outburst of strikes among the less skilled workers about 1889 and in the formation of new trade-unions, with a largely socialist outlook, among the dockers (London dock strike 1889), gasworkers and other classes of previously unorganized workers.

At the same time there arose among the miners, under the leadership of James Keir Hardie (1856–1915), a movement for the legal minimum wage and the legal eight-hour day, backed by the newly formed Miners' Federation of Great Britain (1888). The socialists, notably John Burns (1858–1943) and Tom Mann (1856–1941), placed themselves at the head of the less skilled workers, adding the demand for the "right to work" to those for the minimum wage and the eight-hour day. Hitherto British trade-unionism had been for the most part attached to the Liberal party under W. E. Gladstone; but from the late '80s there arose a widespread movement for independent labour representation, and the local bodies formed on this basis joined forces in 1893 in the Independent Labour party (I.L.P.) under the leadership of Keir Hardie, with an evolutionary socialist program. Powerfully reinforced by the *Clarion*, founded in 1891 by Robert Blatchford (1851–1943), whose *Merrie England* (1894) sold more than 1,000,000 copies, the I.L.P. set out from the first to induce the trade-unions to enter politics on the socialist side. In 1900 it was successful in bringing into existence the Labour Representation

committee as a federation of trade-unions and socialist bodies, but without a fully socialist program.

4. Fabianism.—We must now retrace our steps to 1884. The Social Democratic federation, after the split, fell almost completely under Hyndman's control and pursued a narrow and doctrinaire form of Marxism which alienated the trade-unions interested in immediate reforms and drove one working-class leader after another out of its ranks. Morris' Socialist league, which had seceded partly as a protest against Hyndman's dictatorial methods and partly because it objected to the running of Socialist candidates aided by Tory gold in order to split the Progressive vote, turned wholly antipolitical and was captured by anarchists, who drove William Morris out of it and deprived him of his paper, the *Commonweal,* in which originally appeared his socialist romance, *News From Nowhere* (1891).

In the same year as the Socialist league (1884), the Fabian society was founded, of which Sidney Webb (1859–1947) and George Bernard Shaw (1856–1950) speedily became the outstanding leaders. The Fabians were a small group of intellectual socialists, standing for an evolutionary conception of socialism in sharp contrast with Marxism and endeavouring, by progressive reforms and the nationalization and municipalization of industries, to turn the existing state into a "welfare state." The name they adopted was derived from that of the Roman general known as Fabius the Delayer because of his deliberate, long-range strategy. They declared themselves the inheritors of the Benthamite tradition of promoting "the greatest happiness of the greatest number," and maintained that this principle now pointed to extensive state intervention as plainly as it had once seemed to point to *laissez faire.* In *Fabian Essays* (1889) and in a host of well-informed Fabian tracts the society laid down its essential doctrine and the practical applications of it; and the Fabians also proclaimed a policy of "permeation," by which they meant a readiness to get their ideas taken up by any party or person that would listen to them. Their work considerably influenced the policy of Keir Hardie's Independent Labour party, and their hold on the working-class movement was strengthened by the careful and constructive studies of trade-unionism and other working-class organizations carried out by Sidney Webb in collaboration with his wife Beatrice Webb (1858–1943). The Fabians can be regarded as the group which first clearly worked out the philosophy and practical implications of that form of "gradualist" socialism which became in practice the policy of the Labour and Socialist parties in most of the countries with parliamentary constitutions, whatever theoretical doctrines they continued to proclaim as the basis of their faith. (*See* further FABIAN SOCIETY.)

5. Socialism in the U.S.—In the United States, socialism followed a different course. After the decline of Owenism and Fourierism, which had at one time a considerable hold, socialism practically died out. The single tax doctrines of Henry George (1839–97) obtained a large following in the 1870s and helped to prepare the way for socialism, which was reintroduced mainly as the doctrine of immigrants from various European countries, each bringing with him his particular brand. Largely under German influence, the Socialist Labor party was formed in 1877 and ran its first presidential candidate in 1892. It continued to grow until 1898; but then came a split, Daniel De Leon (1852–1914), its leader, insisting on its duty to form a socialist trade-union movement in opposition to the American Federation of Labor, which was strongly antisocialist. De Leon was a Marxist of the extreme left, following Marx's view of the state and the necessity for the dictatorship of the proletariat, and his attempts at industrial unionism were the forerunners of the later attempt to create the Industrial Workers of the World (1905) (*q.v.*), with the difference that De Leon aimed at industrial unionism closely linked with a revolutionary Socialist Labor party, whereas the Chicago I.W.W., led by W. D. ("Big Bill") Haywood, was antipolitical and syndicalist in outlook. In 1901 the seceders from De Leon's S.L.P. joined forces with other groups to form the American Socialist party, under the leadership of Eugene V. Debs (1855–1926), and the new party soon greatly outnumbered the S.L.P., polling nearly 1,000,000 votes for Debs in the presidential election of 1912.

III. LATER NATIONAL DEVELOPMENTS

1. France.—In the 1880s and 1890s the developing Socialist parties in western Europe were being compelled to define their political attitudes. In France, the Parti Ouvrier had split in 1882 into two sections, Guesdists, or Marxists, and Possibilists, the followers of Paul Brousse (1844–1912), with a more moderate program. The Possibilists split again in 1890, when a left wing broke away under Jean Allemane (1843–1935). The followers of Blanqui held together in a separate Comité Révolutionnaire Central; and yet another party, the Independent Socialists, grew out of a group originally created in 1885 by Benoît Malon (1841–93), the author of an extensive *History of Socialism* (1881–84) and the apostle of *Le Socialisme intégral*—the title of his best known book (1891–92). The Independent Socialists attracted a group of highly influential leaders, including among them Jean Jaurès (1859–1914) and Alexandre Millerand (1859–1943). All these groups wooed trade-union support; but the French trade-unions, under the influence of Fernand Pelloutier (1867–1901), held aloof from politics and went over gradually to the doctrines of revolutionary syndicalism and direct action, which were definitely proclaimed, with the general strike, as its policy at the Lyons congress of 1901. This tendency was already well established when, in the middle 1890s, all France was set in turmoil by the Dreyfus affair. This famous affair caused a big reorientation of political forces, including the establishment in 1899 of the ministry of "republican defense" of Pierre Marie René Ernest Waldeck-Rousseau, in which the socialist Millerand was offered and accepted office. The ensuing dispute led to the regrouping of French socialism in two rival parties, the Parti Socialiste de France (Guesdists and Blanquists) and the Parti Socialiste Français (Broussists, Allemanists and Independents), the latter supporting the participation of socialists in a progressive capitalist government. The split lasted until 1905, when the rival parties joined forces in the Parti Socialiste Unifié, under Jaurès, which pursued a stormy course up to the outbreak of World War I in 1914. The assassination of Jaurès, the outstanding leader of French socialism, on the eve of war in that year was a severe blow to French socialism and opened the door wide to fresh dissensions.

2. Germany.—The rift in French socialism occurred on the practical question of socialist participation in a capitalist government. In Germany the dispute, with the same fundamental basis, took a different form. In 1891 the German social democrats, in their new Erfurt program, eliminated the Lassallean clauses of the Gotha program of 1875. In 1899 Eduard Bernstein (1850–1932) published *Die Voraussetzungen des Sozialismus*, translated under the title *Evolutionary Socialism*, in which he declared the need for a fundamental revision of Marxist doctrine. Bernstein had lived long abroad, first as an exile in Switzerland and from 1888 as correspondent in London of the Social Democratic *Vorwärts;* and he had been greatly influenced by Fabianism and by the British Independent Labour party. Bernstein denied the imminent collapse of capitalism, the progressive crushing out of the middle classes, the concentration of capital in fewer and fewer hands, the "increasing misery" of the working class and the necessarily reactionary character of the state. He affirmed the importance of working for immediate social reforms rather than for revolution, and asserted that the gradual movement toward socialism was everything, and the ultimate goal of socialism in effect nothing. He attacked the orthodox formulation of the materialist conception of history and emphasized the importance of noneconomic and moral factors in the making of social change; and he regarded the Marxian theory of value as a merely abstract concept remote from actual conditions. Finally, he denied that the industrial proletariat could hope, by itself, to conquer political power and, relying on parliamentarism as against dictatorship, pleaded for a moderate socialism which could be regarded as the heir of liberalism and for an appeal to men and women of all classes rather than to the workers alone.

The demands of "revisionism" led to a tremendous controversy, in which Karl Kautsky (1854–1938) appeared as the principal champion of Marxist orthodoxy. The revisionists were defeated at the congress of the German Social Democratic party, and the

letter of Marxism was faithfully preserved. But in practical politics the German Social Democrats to an ever increasing extent accepted Bernstein's precepts, though the very limited elements of democracy in the German constitution greatly limited their practical influence and effectively prevented them, up to 1918, from being confronted with any such problem of participation in the government as had disturbed the French Socialists in 1899.

3. The Second International.—The rise of national Socialist parties in the leading countries was followed by attempts to create a new socialist international. Periodical international socialist congresses were held from 1889; and in 1900 the International Socialist bureau (the second international) was set up to act as a permanent link between the national parties. Unlike the first international, however, it was not a centralized body claiming to dictate policy to its sections, but a loose federation of independent parties, each claiming to follow its own line in accordance with national conditions. It was much occupied with the questions of antimilitarism and the prevention of war, but refused to endorse the policy of the general strike against war sponsored mainly by the French, and succeeded only in passing ambiguous resolutions which furnished little practical help in the crisis of 1914.

4. Syndicalism.—The years immediately before World War I were a period of great labour unrest in many countries. They were marked by a reaction against parliamentary social democracy, by a rapid growth of trade unionism in various forms and by the development of new doctrines of direct action and syndicalism radiating mainly from France. The French syndicalists proclaimed the preponderance of economic over political power, and contended that parliamentarism necessarily involved compromise and a weakening of the will to revolution. The workers, they maintained, must fight their battle on their own terrain—that of industry—by strike action, which could be used to secure political as well as industrial concessions. They must act alone, without invoking the aid of politicians, and must take as their aim the transference of industry into their own hands. The French syndicalists, influenced by Proudhon and Pelloutier, looked forward to the reorganization of France into self-administrating communes under the control of the workers' *syndicats* and *unions de syndicats*, loosely federated nationally and internationally on a basis of pure "workers' control." They were strongly antimilitarist and antipatriotic, favouring the general strike both as a measure against war and as a means of accomplishing the social revolution. In the hands of certain theorists, notably Georges Sorel (*Réflexions sur la violence*, 1908) and Hubert Lagardelle (*Le Socialisme ouvrier*, 1911), the general strike was elevated to the rank of a "social myth," powerful in inspiring the workers to resolute action, even if it never happened in the form supposed.

This syndicalist doctrine owed much to anarchism and had much in common with the anarchist-communism of the great Russian revolutionary Prince Peter Kropotkin (1842–1921), author of *Fields, Factories and Workshops* (1898), *Mutual Aid, a Factor of Evolution* (1902) and other books of wide international influence. It was much like the doctrine maintained in Spain by the Iberian Anarchist federation (F.A.I.) and the Spanish National Confederation of Labour (C.N.T.). Transplanted to the U.S., where capitalism was much more advanced and centralized, it assumed a different form, the attempt to form a single, all-inclusive big union, designed to overthrow capitalism by revolutionary direct action, and to take the great industries into the workers' hands through the Industrial Workers of the World. In Great Britain it appeared in an industrial syndicalist movement led by Tom Mann (*The Industrial Syndicalist*, 1912) and in a modified form in guild socialism, an attempt to reconcile state socialism with the syndicalist doctrine of "workers' control." (*See* SYNDICALISM.)

5. Guild Socialism.—First preached by A. R. Orage (1873–1934) and S. G. Hobson (1868–1940) in the *New Age* about 1911, guild socialism (*q.v.*) spread rapidly during World War I, under the auspices of the National Guilds league. The guild socialists stood for state ownership of industry, combined with "workers' control" through delegation by the state to national guilds including all the workers by hand and brain and organized internally on democratic lines. About the state itself they differed, some believing that it would remain more or less in its existing form, and others that it would be transformed into a federal body representing guilds, consumers' organizations, local government bodies and other social structures (G. D. H. Cole, *Guild Socialism*, 1920; S. G. Hobson, *National Guilds and the State*, 1920). Guild socialism was much stimulated during World War I by the rise of the left-wing shop stewards' movement, demanding "workers' control" in the war industries. After the war, the building workers, led by Hobson and Malcolm Sparkes, founded building guilds which built houses for the state; but after the slump of 1921 the state withdrew financial help, and the movement collapsed. The shop stewards' movement also disintegrated when the war ended; and the guild socialist movement fell to pieces, leaving behind it an effect in the incorporation of some element of "workers' control" into the programs of trade unionism and of the Socialist and Labour parties.

6. The War Issue.—Over most of the world, World War I sharply divided the socialist forces. In Great Britain, France and Germany the majorities in the Social Democratic and Labour parties supported their respective countries, whereas in Italy and Russia the majorities were against the war. In Great Britain there was no formal split, but the Independent Labour party, while remaining part of the federally organized Labour party, opposed the war. In Germany, in 1915, the Social Democratic party virtually split, a minority breaking away to become the "Independents," under Kautsky and Hugo Haase, and a group further to the left, the *Spartakusbund* (Spartacus union), appearing a little later under Karl Liebknecht (1871–1919) and Rosa Luxemburg (1870–1919), one of the principal latter-day theorists of Marxism (*The Accumulation of Capital*, 1913). In France, *majoritaires* (for the war) and *minoritaires* (antiwar) waged incessant conflict without positively forming rival parties. In Italy Benito Mussolini broke away from the Socialist party to lead a militant prowar group, which formed the nucleus of the Fascist party. In Poland Jozef Pilsudski, previously a Socialist, sided with Germany in the hope of securing Polish liberation from Russia, while Rosa Luxemburg's followers, who had formed a Polish Social Democratic party in 1893, sided with the Bolsheviks in opposing the war.

IV. RUSSIAN SOCIALISM

The outbreak of the Russian Revolution in 1917, followed by the Bolshevik revolution later in the year, exercised a deep influence on the socialist movement throughout the world. Russian socialism, at the time of the revolution, was divided into four main parties—Socialist Revolutionaries, Left Socialist Revolutionaries, Bolsheviks, and Mensheviks.

1. Socialist Revolutionaries.—Conditions in Russia were so different from those in western Europe that Russian socialism had followed a largely independent course. The earlier Russian exponents of socialism, Aleksandr Herzen (1812–70), Peter Lavrov (1823–1900) and Nikolai Chernyshevski (1828–89), sought to apply western socialist ideas to Russian conditions, whereas Bakunin applied ideas that were indigenously Russian to the conditions of the west. Herzen believed that the survival of the commune as the basis of Russian village life made Russia, despite its industrial backwardness, an appropriate country in which to institute the socialist revolution, which he conceived as above all else a peasant revolution. Chernyshevski developed these ideas, regarding the overthrow of tsarism and the landlord system and the institution of political democracy as the first steps toward agrarian communism. Lavrov (*Letters on History*, 1868–69), the philosopher who inspired the Narodnik (People's) movement among the Russian intellectuals, preached a gospel of self-devotion by the educated classes to the cause of the people, and then looked forward to the achievement of socialism by the leavening of the people with socialist ideas, which could be built on the foundations of the communal tradition of Russian village life. Out of the Narodnik movement sprang the Narodnaya Volya, the idealist movement of organized terrorism which was responsible in 1881 for the assassination of Alexander II. This in turn was the precursor of the Socialist Revolutionary party, which came into formal existence

in 1901 and, representing many different tendencies, was by far the largest of the Socialist parties at the time of the revolution of 1917. From it split off, under M. Spiridonova, the Left Socialist Revolutionary party, which joined the Bolsheviks in the second revolution, and for a short time occupied a place in the soviet government.

2. Social Democrats.—As against the Socialist Revolutionaries, whose main appeal was to the peasants, the Russian Social Democrats had their strength among the urban workers. Russian factories were relatively few, even in 1917, but the few were mostly very large and used mass-production methods. The urban proletariat was small, but highly concentrated, and lived under conditions of intense exploitation. We have seen how in 1883 Plekhanov and Axelrod took the lead in setting up in Geneva the Emancipation of Labour group, based partly on local societies established inside Russia in the '70s. This group spread the knowledge of Marxism in Russia. In 1895 Vladimir Ilyich Ulyanov (Lenin; 1870–1924) united the groups in St. Petersburg (Leningrad) into an Emancipation of Labour society, which put itself at the head of the industrial struggles then in progress. Out of this and other local bodies arose the Russian Social Democratic party, which was formed at Minsk in 1898 but did not take definite form until after the foundation of *Iskra* ("The Spark") (1900), a social democratic journal published abroad in which Lenin collaborated with Plekhanov and Axelrod.

3. Bolsheviks and Mensheviks.—The second (and first real) conference of the Social Democratic party was held in Brussels and London in 1903. It adopted a revolutionary program largely drafted by Lenin, including an assertion of the rights of national self-determination, a series of demands for land reform in the interests of the peasants, and a pronouncement in favour of the dictatorship of the proletariat. A dispute arose over the conditions of admission to the party, Lenin demanding a closely knit, disciplined party, whereas the majority favoured open membership for anyone who accepted the declared principles. The leader of the majority against Lenin was Julius Ossipovich Zederbaum, known as Martov (1873–1923), subsequently the leader of the Mensheviks. But on the other issues Lenin's group won, and they were thereafter known as the Bolsheviks (majority) whereas Martov's followers were the Mensheviks (minority). In 1904 the split became definite, the Bolsheviks and Mensheviks holding separate congresses and organizing two rival parties. The basic difference was that the Mensheviks wished to create a Social Democratic party on the model of the national parties in western Europe, and to work in alliance with the bourgeois left for a democratic revolution, holding that the time for socialism in Russia would come when the country had passed through a regime of capitalist democracy, whereas the Bolsheviks wanted a centralized and disciplined party that would aim directly at power and would seek to rouse the peasant masses to rally to the side of the industrial proletariat. The Bolsheviks were prepared to work with the bourgeois left for democratic revolution, but set out to capture at once the leadership of the revolution and refused to make concessions to the *bourgeoisie* in order to enlist its help.

This is not the place to tell the story of the Russian Revolution —of the disintegration of the large but amorphous Socialist Revolutionary party, the conquest by the Bolsheviks of a majority in the revolutionary soviets of workers and peasants, and the crushing out of the Mensheviks on the Bolshevik assumption of power. (*See* RUSSIAN REVOLUTION.) What is relevant here is that the Bolshevik revolution was based on Lenin's interpretation of Marxism, putting in the forefront the dictatorship of the proletariat, the complete destruction of the old state machine and the creation of a new proletarian state as the instrument of the dictatorship (Lenin, *The State and Revolution*, 1917; *The Proletarian Revolution and Kautsky the Renegade*, Eng. trans., 1920). This view of Marxism was combated not only by the Mensheviks in Russia but also by most of the leading social democratic theorists in western Europe, notably Karl Kautsky (*The Dictatorship of the Proletariat*, 1918; Eng. trans., 1920), and by Émile Vandervelde (1866–1938), the Belgian president of the Labour and Socialist international, which was established in 1920 and reorganized in 1923 as the successor of the pre-1914 second international and continued its traditions practically unchanged, as well as by the British leaders.

The Bolsheviks sent out to the workers of the world a new *Communist Manifesto* (1919) and set out to organize a new third international—the way for which had been prepared by the antiwar international Socialist conferences held at Zimmerwald (1915) and Kienthal (1916), and by the attempt to organize a world conference at Stockholm in 1917—with the design of stirring up revolution throughout the world. At this stage it was widely believed that the Russian Revolution could not succeed in establishing itself unless it were speedily followed by socialist or communist revolutions in the advanced capitalist countries. Leon Trotsky (1879–1940), at that time Lenin's closest collaborator and the creator of the Red army, was especially associated with this view, which was part of his doctrine of "permanent revolution." Trotsky was also an advocate of rapid Russian industrialization as a means of consolidating the power of the revolution and an opponent of concessions to the peasants, which he held would delay this process.

4. Communism.—As developed theoretically by Lenin and later by Joseph Stalin (1879–1953) and others, communism was no new doctrine. It was Marxism, applied practically without change to the conditions of contemporary Russia. It could be so applied, because Russia in 1917 was still much like the countries of western Europe at the time when Marx and Engels formulated their essential doctrines. Western social democracy, while it retained its nominal allegiance to Marxism, had in fact so modified Marx's doctrines, especially about the state and dictatorship, as in effect to deny them. It had been working in practice for the peaceful democratic conquest of state power, and for an evolutionary socialism that would set out to use the existing state machine and to transform it gradually. The Russian socialists could not possibly have attempted so to use the tsarist state machine, as was plainly shown in the events that followed the abortive Russian revolution of 1905. They could only seek to overthrow it and to replace it either by a state modeled on western parliamentarism (the Menshevik policy) or by a state based on proletarian dictatorship (the Bolshevik policy). The western social democrats, wedded to parliamentarism, supported the Mensheviks and denounced the Bolsheviks as betrayers of democracy, whereas the Bolsheviks denounced them as counterrevolutionaries and traitors to the socialist cause.

The world revolution hoped for by the Bolsheviks did not come, but neither did the crushing out of the Russian Revolution by the capitalist countries. Through many vicissitudes, the Soviet power was able gradually to consolidate its strength. Trotsky was driven into exile; Stalin became the unquestioned leader and set out to, as he conceived of it, build up socialism within a single country under the successive five-year plans and the statute of nationalities, which conferred wide autonomy on the many peoples of the Soviet Union. In 1930 the policy of agricultural collectivization was taken seriously in hand, a form of village co-operative (the *kolkhoz*, or "collective farm") replacing individual peasant cultivation. Industrialization was vigorously pushed, and in World War II the Soviet Union plainly demonstrated its might and power of survival. In 1935 Sidney and Beatrice Webb, the theorists of Fabianism, surprised many by publishing their monumental work (*Soviet Communism: a New Civilisation*) largely in support of the policies and achievements of Russian Communism. This did not mean that they regarded the Soviet form of revolution as applicable to western Europe. They did, however, sharply differ from the social democratic theorists in regarding it as having justified itself fully under Russian conditions.

V. MODERN TRENDS

1. Western Europe.—In western Europe the end of World War I brought with it a great increase in the strength of the Social Democratic parties. In Great Britain, the Labour party had been up to 1918 only a small minority group. Profiting by the disintegration of liberalism, it reorganized itself in 1918 under Arthur Henderson (1863–1935) as a nationwide party standing for evolu-

tionary socialism (*see* its manifesto, *Labour and the New Social Order*, 1918, written by Sidney Webb). By 1924, under James Ramsay MacDonald (1866–1937), it formed the first (minority) Labour government, which lasted only a year. In 1929 it returned to office, again in a minority, but the second Labour government collapsed in the crisis of 1931, MacDonald seceding with others to form a "national" coalition government. During the 1930s Labour gradually regained strength. It remained a moderate, evolutionary socialist party, and was a partner in the war coalition formed in 1940. In 1945, in the election that followed hard upon the end of the European phase of World War II, the Labour party for the first time gained a considerable clear majority over all other parties, winning not only the industrial areas, but also most of the suburban residential districts near London and other big cities. Labour thereupon formed its third government, with a definite mandate for the policy of evolutionary socialism on which it had fought the election. This policy was set out in its manifesto, *Let Us Face the Future* (1945). In 1951 the Conservatives were returned and held control until 1964 when Labour won a very narrow parliamentary majority.

In Germany, the fall of the Hohenzollerns and the proclamation of the Weimar republic raised the Social Democrats to the position of the first party in the state. But they made no attempt to establish socialism and were soon involved in suppressing left Socialist and Communist groups. The German workers became disastrously divided into rival Social Democratic and Communist parties; and severe economic distress reinforced feelings of national humiliation in stimulating the rise of Adolf Hitler's Nazi, or National Socialist, party, which ultimately seized power in 1933. "National Socialism" (*q.v.*) was not a form of socialism, but its bitterest enemy. The so-called "Socialist" element in it consisted in its exaltation of the claims of the national state, regarded as a metaphysical being beyond good and evil, above all individual or sectional claims; but in practice it worked as the ally of large-scale capitalism and militarism and remorselessly crushed out all free working-class organization. It was the national state, organized to the limit for imperialist war and conquest—the very antithesis of the socialism described in this article. After 1939, however, Nazism won the allegiance of a few notable socialist renegades, such as the French neosocialist Marcel Déat.

Fascism (*q.v.*), of which Nazism was the German variety, made its first appearance in Italy, under the leadership of Mussolini. Italian socialism, which had opposed the war, adopted a semi-revolutionary line after 1918. The Italian workers in 1920 occupied the factories but made no attempt to turn their industrial struggle into a political revolution. Their defeat prepared the way for the march on Rome, by which Fascism seized power in 1922. Italian Fascism, like German Nazism, glorified the state, to which it professed to give a corporative form, based on corporations representing industries and professions. (*See* CORPORATE STATE.) In fact, behind this facade, it meant dictatorial control by the Fascist party, working in alliance with large-scale capitalism and suppressing all independent working-class or political organizations.

In France, after 1918, the communists were a majority in the Socialist party, which they converted into a Communist party. A large section broke away and reconstituted the Socialist party (known as S.F.I.O.—Section Française de l'Internationale Ouvrière), of which Léon Blum became the leader. In France, as in Germany, socialism was greatly weakened by the conflict between communists and socialists, and the position was made worse by the development of two rival trade-union movements. At length, in 1935–36, the trade-union split was healed, and both communists and socialists allied themselves with the radicals in the popular front. A wave of strikes followed, and many concessions were gained. But the front broke down, and French politics were in disorder at the outbreak of World War II. Following the collapse of 1940, most of the French socialists joined the resistance movements associated with Gen. Charles de Gaulle, and the communists, while maintaining their separateness, took part in the Gaullist and postwar coalition cabinets from 1944 to 1947. In the latter year they were excluded from the government.

In Spain, where the working-class forces were traditionally divided between socialists and syndicalists (with a strong anarchist element, chiefly in Catalonia), a popular front, extending to the radicals, achieved power in 1931, but was overthrown in the ensuing civil war by the Fascist counterrevolution under Gen. Francisco Franco, supported by Germany and Italy, while the western parliamentary countries stood aloof. The U.S.S.R. gave some help, but on conditions that split the socialist forces. In Austria, where the Social Democrats, headed by Otto Bauer (1881–1938), a leading theorist of the movement, were the strongest party in Vienna but had little hold on the Catholic countryside, a Catholic semi-Fascist dictatorship was set up by Engelbert Dollfuss in March 1933, and in Feb. 1934 a rising of the Viennese workers was crushed and the Socialist movement was disrupted or driven underground. In 1938 the Nazis seized Austria and incorporated it in "greater Germany," applying the entire apparatus of suppression.

The social democratic movement enjoyed its only successes between the wars in Scandinavia, especially in Sweden, where, under the leadership of Karl Hjalmar Branting (1860–1925) and later of Per Albin Hansson (1885–1946), it carried through a moderate policy of liberal semisocialism with considerable economic success. Denmark had a similar experience; and in Norway, where socialism was on the whole farther to the left, much progress was also made. The Scandinavian parties made no attempt to introduce socialism, but adopted advanced policies of social reform and economic planning without disturbing the institutions of capitalist property. Their fortunate economic position made this possible to a degree not possible elsewhere. In eastern Europe, under the tutelage of the U.S.S.R., Socialist parties by mid-century had disappeared or had been absorbed by the Communist parties.

2. North America.—In the U.S., socialism remained backward. (*See* SOCIALIST PARTY [U.S.].) The U.S. Communist party took away a good deal of the former support of the Socialist party, but failed to gain a really large following. The American Federation of Labor remained strongly antisocialist and even more anticommunist, refusing all association with the trade-unions of the U.S.S.R. The newer Congress of Industrial Organizations, which arose in 1935 as the result of a breakaway from the A.F. of L., had a more leftist though hardly less anticommunist policy, but did not result in the growth of any nationwide Labour party, although Labour parties arose in a few states, notably New York. The U.S. produced no great socialist thinker, though Upton Sinclair became world famous for his socialist novels (*The Jungle*, 1906; *The Brass Check*, 1919; *Oil*, 1927) and there was much acute social criticism, notably in the work of Thorstein Veblen (1857–1929) (*Theory of the Leisure Class*, 1899; *Theory of Business Enterprise*, 1904). In Canada, socialism made advances after the establishment in 1932 of the Co-operative Commonwealth federation, the Canadian Labour and Socialist party, which appealed to farmers as well as industrial workers.

3. Australia, New Zealand, and South Africa.—In Australia the Labor party, based mainly on the trade-unions, was a powerful political force. It formed the government in 1941 and remained in power for eight years. In New Zealand the Labour party, having won the elections of 1935, remained in power until 1949. It returned to power in 1957 but only for three years. In South Africa, mainly because of the divisions between African and white workers, socialism made no headway.

4. The Far East.—In the far east, Chinese socialism began with Sun Yat-sen (1866–1925), the leader of the revolution of 1912, whose famous *Testament* is an application of socialist principles to Chinese conditions. The Chinese National party, the Kuomintang, was at the outset partly under socialist inspiration; but following the rise of a Communist movement in China, under Russian influence, there were violent dissensions leading to civil war (1929) until the Chinese communists came to power in 1949. The new regime, although closely allied to the Soviet Union and entirely dominated by the Communist party, was adaptable enough to use the services of noncommunists, including small capitalists as well as socialists, in many specialist services and branches of production. Great projects of industrial development were embarked upon with Soviet help, and rapid progress was made with

various forms of rural collectivization and co-operation. Many in the west hoped that Chinese communism would before long deviate sharply from the Russian model.

In Japan a Socialist party was formed in 1901 but was speedily suppressed, and there was renewed persecution during the Russo-Japanese War of 1904–05, which the socialists opposed. After 1918 there was for a time a renewal of socialist activity; but no stable socialist movement was established, and by the 1930s the movement had again been wiped out by militarist persecution, or driven wholly underground. Japanese socialism became politically important after 1945, establishing itself as the leading opposition force in the postwar parliament.

Meanwhile, in India, independence put into power a Congress party government, headed by Jawaharlal Nehru, himself a socialist; and this government's second five-year plan, issued in 1956, included large provisions for public investment as well as for welfare development. The Indian National Congress became formally committed to a sort of democratic socialism as a long-term objective, but continued to be largely under nonsocialist leadership and to claim to stand for national unity. The Indian Socialist party operated at first inside the Congress, but presently seceded from it and thereafter experienced considerably fluctuating fortunes and underwent a split in 1955, when a section headed by R. M. Lohia broke away from it. In 1964 the two groups merged to form a new United Socialist Party with a parliamentary representation of 21.

In the Arab countries substantial Communist and Socialist parties made their appearance; but nationalism remained the predominant issue, and the socialists were active mainly in the left wing of the various nationalist movements, though in Syria and Lebanon, as well as in Egypt, social questions—especially land reform—were forcing their way to the front.

In Indonesia the communists became a powerful opposition group; the socialists after playing a considerable part in the early years of independence subsequently lost ground. Burma established itself under broadly socialist rule. In 1964 Burma became a one-party state with the Socialist Program Party founded by Gen. Ne Win as the only authorized party.

The Asian Socialist parties, after refusing to join the postwar Socialist international, established an Asian Socialist international of their own, which entered into fraternal relations with it.

5. Contemporary Socialism.—The wartime alliance between the Soviet Union and the west temporarily eased animosity between European socialists and communists but did nothing to overcome their fundamental differences. Later, such factors as the communist seizure of power in Czechoslovakia, the Korean War, and Stalin's growing despotism caused further deterioration of this relationship. But the death of Stalin led to a substantial modification of Soviet totalitarianism by 1956, in which year the 20th congress of the Communist party of the Soviet Union was marked by a sharp denunciation of Stalin's methods and by a recognition of the possibility of advance toward socialism by nonrevolutionary methods in certain countries.

Immediately great debates opened both inside the Communist parties of all countries and among left-wing democratic socialists, who foresaw some prospect of improving international working-class relations and of further approaches toward world *détente*. As against this the western socialist parties and the Socialist international uniting them maintained their anticommunist vehemence. The suppression of the Hungarian revolution in Nov. 1956 suggested that the Soviet government was potentially as ruthless as ever and by hardening anticommunist feeling in the west thoroughly confused the over-all situation.

The issue of parliamentary democracy continued to occupy socialist thinkers. In Yugoslavia and, subsequently, Poland one-party rule appeared capable of substantial liberalization; nevertheless democratic socialists continued to stress the virtues of free elections and rival political parties. Doubtless these were indispensable wherever their traditions were established; it was not clear whether they could be insisted upon in countries lacking parliamentary experience and where nonrevolutionary advance was blocked by feudal and absolutist rule. Nor was parliamentary progress toward socialism possible in countries such as France and Italy where the working-class movement was divided into factions unable to co-operate in any way.

In September 1964 socialists met in Brussels, Belg., to celebrate the 100th anniversary of the founding of the first International. The secretary of the organization reported that it had affiliated with it 44 Labour and Socialist parties with a combined membership of nearly 12,000,000 persons. Parties affiliated with the International were in control of the governments of Great Britain, Denmark, Norway, Sweden, Israel, and the Malagasy Republic.

6. Interpretation of Socialism.—The foregoing historical account has been designed to describe the socialist movement by denotation, rather than to attempt any precise definition of socialism, which cannot indeed be accurately defined. Socialism is both a movement and a theory, and takes different forms under different historical and local conditions. It is fundamentally a movement aiming at a classless society based on the socialization of property in the essential instruments of production and appealing primarily to the working class as the exploited class whose historic mission it is to bring the class system to an end. This description (it is not meant as a definition) admits of many different views of the methods of social and economic organization appropriate to a socialist society and of the steps to be taken for its establishment.

Socialism can be either statist or libertarian, Marxist or "liberal," revolutionary or gradualist, cosmopolitan or only internationalist in a sense that lays stress on the importance of the national unit.

Marx drew a sharp contrast between "utopian socialism," basing itself on moral principles conceived as absolute, and "scientific socialism," resting on the materialist conception of history, with its corollary, the class struggle. As Marxism developed, the first great controversy was between the Marxist advocates of disciplined and centralized proletarian mass action and the anarchist (or later syndicalist) upholders of the theory of local mass spontaneity and federalist "free communism," sponsored by the followers of Bakunin and, in a less revolutionary sense, by those of Proudhon (the *mutuellistes*).

Still later, after the rise of social democracy, the controversy shifted to the question of whether socialists should set out to capture and use the existing state (Fabianism, evolutionary social democracy) or should seek to destroy the existing state and build up a new one by means of proletarian dictatorship (bolshevism, communism).

In these controversies, means and ends were mixed up together. It was partly a question of revolutionary versus evolutionary methods, partly a question of parliamentary democracy versus dictatorship, partly a question of libertarianism versus centralized state authority. Naturally, doctrines and emphasis varied from time to time and from country to country. Marxism, in its emphasis on the scientific basis of the materialist conception of history, tended at the extreme to throw over altogether the moral basis and to regard morality simply as a reflection of the forces of economic evolution. It is, however, difficult to see why anyone should trouble to work for socialism merely because he believes that it is in the line of historical evolution, or unless he also believes it to be good and that what is good ought to be pursued. Moral indignation is not the least of the weapons in Marx's own powerful armoury.

In essence, what Marxism opposes is the view that socialist judgments of value can be reached a priori, without taking account of the conditions of time and place. Man makes his own history; but the choices open to him are limited by the circumstances in which he has to act, and among these the condition of the "powers of production" holds a key place.

Socialism is essentially a movement for the promotion of the well-being and happiness of individual men and women (the "greatest happiness of the greatest number"), and not of any metaphysical entity such as the state, which should exist only to promote the ends of individuals. That these individuals are grouped in classes, and that social action in its major manifestations takes a class form in no wise invalidates this generalization, which is com-

mon to all schools of socialists (but not to "national socialists" or to some forms of so-called "Christian socialism"). Anarchist communism, social democracy and Leninist (though not Stalinist) communism all aim in their several ways at the establishment of a classless society and at the liberation of the human spirit, as well as at the abolition of physical want. Communist dictatorship, however perverted in practice, is intended only as a transitional measure, leading up to the "withering away" of the state. The ideal of world-wide human brotherhood and liberty is common to all schools of socialists, by whatever route they attempt to advance toward it. This holds good even for those socialists who reject "idealism" as a bourgeois doctrine and insist on speaking in terms of class rights and historical necessity rather than of humanistic ideals.

See also CHARTISM; CHRISTIAN SOCIALISM; COMMUNISM; FABIAN SOCIETY; GUILD SOCIALISM; INTERNATIONAL, THE; MARXISM; biographies of prominent socialist figures; and references under "Socialism" in the Index.

BIBLIOGRAPHY.—The bibliography of socialism is immense, and only a very few titles can be given here. K. Marx, *Capital*, 3 vol. (first published, in German, 1867–94, Chicago, 1915; Moscow, 1954–); F. Engels, *Anti-Dühring* (first published, in German, 1877; Moscow, 1954); G. Bernard Shaw (ed.), *Fabian Essays* (first published, 1889; with postscript, 1948); F. S. Nitti, *Catholic Socialism* (1908); L. Levine, *Labour Movement in France* (1912); F. Mehring, *Die deutsche Sozialdemokratie* (1879), *Karl Marx* (1936); M. Beer, *History of British Socialism* (1940); H. W. Laidler, *Social-Economic Movements* (1944); L. Trotsky, *History of the Russian Revolution* (1934); A. Rosenberg, *History of Bolshevism* (1934); E. Burns (comp.), *Handbook of Marxism* (1935); S. and B. Webb, *Soviet Communism: a New Civilisation*, 3rd ed. (1945); N. Lenin, *Selected Works*, 2 vol. (1953); E. Dolléans, *Histoire du mouvement ouvrier, 1830–1952*, 3 vol. (1936–53); G. D. H. Cole, *History of Socialist Thought*: vol. i, *The Forerunners, 1789–1850* (1953), vol. ii, *Marxism and Anarchism, 1850–90* (1954), vol. iii, *Second International, 1889–1914* (1956), vol. iv, *Communism and Social Democracy, 1914–1931* (1958), vol. v, *Socialism and Fascism, 1931–1939* (1960), *The Meaning of Marxism*, rev. ed. (1948), *The Simple Case for Socialism* (1935); Carl Landauer and others, *European Socialism: A History of Ideas and Movements*, 2 vol. (1959); Émile Durkheim, *Socialism and Saint-Simon* (1958); J. A. Schumpeter, *Capitalism, Socialism, and Democracy* (3rd ed. 1950); Norman Thomas, *Socialism Re-examined* (1963); Ludwig von Mises, *Socialism: An Economic and Social Analysis* (rev. ed. 1959). (G. D. H. C.; X.)

SOCIALIST PARTY (U.S.). On March 29, 1900, representatives of the Social Democratic Party of America, of which Eugene V. Debs (*q.v.*) and Victor L. Berger were leading members, met in New York city with representatives of the moderate branch of the Socialist Labor party led by Morris Hillquit, Max Hayes and others, and agreed to nominate Socialist candidates for president and vice-president in the 1900 campaign. Their nominees were Debs, former president of the American Railway union, for president, and Job Harriman, California attorney, for vice-president. The two groups waged a joint campaign that resulted in a vote of 96,878 for the Socialist ticket.

Following the campaign the two groups issued a call for a "unity convention" to be held July 29, 1901, in Indianapolis, Ind., with the view toward officially launching a united party. At this convention the older groups dissolved and formed the Socialist party to serve as "the party of the working class and those in sympathy with it" dedicated to the bringing about of "a system of collective ownership by the entire people of the means of production and distribution." The immediate demands in the party's platform included planks for woman suffrage, old-age pensions, unemployment insurance, health and accident insurance, the increase of wages, the reduction in working hours, direct legislation, and public ownership of public utilities.

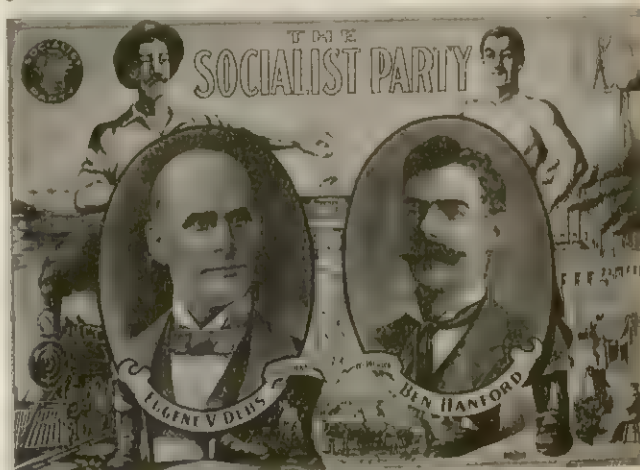
During the next 11 years, from 1901 to 1912, the party grew steadily in numbers and influence. Its dues-paying membership increased from 16,000 to 118,000; its presidential vote, with Debs as candidate, from 96,000 to nearly 900,000; its representation in public office from zero to more than 1,000, including 56 mayors, over 300 aldermen, numerous state legislators and one congressman, Victor L. Berger from Milwaukee.

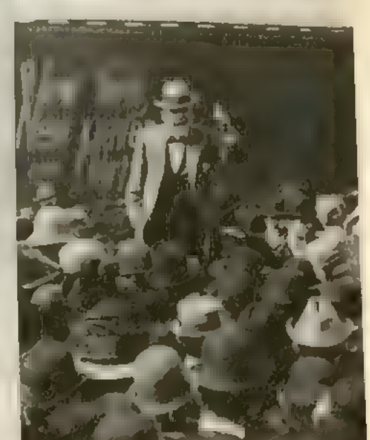
The year 1912 likewise found the party engaged in a vigorous internal struggle with the followers of William D. Haywood, leader of the Industrial Workers of the World (I.W.W.), who emphasized direct industrial action, including sabotage and the general strike, as a means to a co-operative social order, based on syndicalistic principles. The majority of the party adhered to their belief in parliamentary action and amended the constitution to exclude from membership advocates of sabotage and violence. In 1913 Haywood was expelled from the party's executive committee. The controversy led to the withdrawal of numerous "direct actionists" from the party.

In the following year, 1914, the European war broke out. Some Socialists felt that the United States should immediately enter the war on the side of the Allied powers. The majority took the position, as did millions of other Americans, that the United States should remain neutral and later offer itself as mediator or arbitrator, with a view toward bringing about a just and lasting peace. In 1916 the party nominated Allan L. Benson, writer and antimilitarist, as its presidential candidate and rolled up a vote of more than 585,000.

When the U.S. entered the war in April 1917, the Socialist party at its St. Louis convention declared "its unalterable opposition to the war just declared." During the next year many of the party's journals and meetings were suppressed by the government, and Debs was arrested for delivering an antiwar speech at Canton, O. and sentenced to a federal penitentiary for ten years. While Socialist activity came to a standstill in many parts of the country, it took on renewed vigour in other sections. In New York city, Morris Hillquit, Socialist candidate for mayor on a platform that called for an early and democratic peace, received a record vote of 146,000. He helped to bring into office ten assemblymen and seven aldermen.

On Nov. 7, 1917, the Russian Bolsheviks staged their successful revolution. Many members of the Socialist party were of the opinion that the United States was ripe for a similar revolution and urged the adoption by the Socialist party of Bolshevik tactics and the joining of the Communist International. To aid their cause, they swelled the party membership rolls with recent immigrants from countries of eastern Europe who were better ac-





(ABOVE) THE GRANGER COLLECTION; (RIGHT) BROWN BROTHERS

(ABOVE) POSTER USED IN THE PRESIDENTIAL CAMPAIGN OF 1904; (RIGHT) EUGENE V. DEBS EXHORTING RAILROAD WORKERS TO OPPOSE U.S. PARTICIPATION IN WORLD WAR I, 1917

quainted with the institutions of their homelands than with those of the United States.

The struggle over socialist *v.* communist tactics culminated in the Chicago convention of the party in late Aug. 1919. The left wing elements in the party were defeated and withdrew to form the Communist and Communist Labor parties. Though the party's membership during the struggle was greatly reduced (numbering 26,766 after the convention), the party's vote in 1920 with Debs, still in prison, again as its nominee, exceeded 900,000.

During the early 1920s labour was faced with widespread unemployment, with a concerted attack by employers on the trade union movement, and with attempts of the government to suppress civil rights. At the same time government corruption reached new heights. As a result of these and other conditions, many trade unions turned to independent political action as one method of fighting reaction and an unstable economy. In 1922 a number of leaders of organized labour, particularly in the railway trades, led by William H. Johnston, president of the machinists, formed a Conference for Progressive Political Action. Hillquit was invited to represent the Socialist viewpoint on the conference's governing board. The conference called with others a convention to meet in Cleveland on July 4, 1924, to nominate a Progressive presidential candidate. It nominated Sen. Robert M. La Follette of Wisconsin.

In his acceptance speech, La Follette expressed his belief that a new party would be born following the election. Socialists, seeing in this campaign the beginning of a Progressive or Farmer-Labor party, supported the La Follette ticket, which also received the endorsement of the executive council of the American Federation of Labor. Although the Progressive ticket received nearly 5,000,000 votes, no new party materialized and the Socialists returned to their traditional policy of supporting Socialist candidates, while continuing to work for a national farmer-labour political alignment.

Debs, five times Socialist presidential candidate, died in 1926, two years after the La Follette campaign. In the 1928 elections the party turned to Norman Thomas, writer, lecturer, former Presbyterian minister and co-executive director of the League for Industrial Democracy, as its new standard bearer. In that year, when many were predicting that the "new capitalism" was guaranteeing to the country permanent prosperity, Thomas received but 267,200 votes. Four years later, however, in the midst of the depression, his vote increased to 884,781.

During the next four years various factions developed within the party. They differed in opinion as to "the road to power" and the possibility of limited cooperation with the Communists in the fight against world fascism and reaction. This controversy led in 1936 to the withdrawal of several hundred party members and their organization of a Social Democratic federation. The 1936 convention decisively repudiated any political united front with the Communist party. During the next few years antagonism to the Communist party was progressively sharpened by the American Communists' support of Stalin's policies.

During this four-year period (1933–1936), the Franklin D. Roosevelt administration, in its drive to surmount the depression, embodied in its "New Deal" legislation many immediate demands of the Socialist party platform in the fields of social security, housing, public power and

CULVER PICTURES, INC.

NORMAN THOMAS SPEAKING AT A SIDEWALK RALLY IN NEW YORK CITY, ON AUG. 23, 1940, THE 13TH ANNIVERSARY OF THE EXECUTION OF SACCO AND VANZETTI

labour-management relations. The enactment of section 7-a in the National Industrial Recovery act gave an impetus to labour in its drive to organize many of the giant industries, while labour leaders came into more cordial relationship with the president of the United States than ever before in the nation's history. As a result of these developments, many trade unions felt politically obligated to give active support to the Democratic administration, and made it increasingly difficult for their members to remain active in the Socialist party and other minority political groups. In the 1936 campaign, therefore, labour's support fell off and the vote for Thomas decreased to 187,720.

During the 1940s and 1950s the party continued to nominate presidential candidates—Thomas in 1940, 1944 and 1948, and Darlington Hoopes, Pennsylvania lawyer and former Socialist assemblyman, in 1952 and 1956. After the outbreak of World War II, and before Pearl Harbor, Socialists opposed America's entrance into the war. When war was declared by the U.S. the majority of Socialists gave it their critical support, while urging the maintenance of civil liberties, democratic controls over war and postwar collectivism and the winning of the earliest peace that promised permanency, followed by the establishment of an inclusive international organization.

In its presidential campaigns the party emphasized increasingly that its economic goal was not the public ownership of all industry, but an economy aimed at equality of opportunity in which public, co-operative and private ownership existed side by side. The 1956 platform urged "the ownership by genuine cooperatives and publicly owned and democratically managed corporations," of public utilities, banks, insurance companies and certain basic industries. Its immediate demands included the expansion of public educational, health, recreational, housing and social security services; the conservation of natural resources; the strengthening of labour legislation; the end of racial discrimination; economic and cultural aid to underdeveloped countries; and international agreements for disarmament with strict inspection and controls.

In 1957 the Social Democratic federation reunited with the Socialist party, forming the Socialist party-Social Democratic federation (S.P.-S.D.F.). Frank Zeidler, mayor of Milwaukee, was elected chairman of the executive committee. A few months after the merger the party was readmitted as a full member into the Socialist International. In 1958, emphasizing its increasing role as a political educational, rather than an electoral organization, the party voted to permit its individual members to enter the primaries of other parties and to support labour and avowedly progressive candidates of other political parties. In 1960 and again in 1964 the majority of the party decided against the nomination of a presidential candidate, but campaigned for the party platform, emphasizing the necessity of an honest, meaningful political realignment.

While the party's electoral successes have been less than its founders had hoped, since the beginning of the century the party has served, its advocates hold, as a potent educational force in social and political life. (H. W. L.; N. M. TH.)

BIBLIOGRAPHY.—D. D. Egbert and Stow Persons (eds.), *Socialism and American Life*, 2 vol. (1952); Norman Thomas, *A Socialist's Faith* (1951); Harry W. Laidler, *American Socialism* (1937), *American Socialism: a Brief History*, pamphlet (1952).

SOCIALIZED MEDICINE: *see* MEDICAL CARE, GOVERNMENT.

SOCIAL PSYCHOLOGY: *see* PSYCHOLOGY, SOCIAL.

SOCIAL SCIENCES. The social sciences are concerned with the orderly investigation of the behaviour of man in society with the aim of cumulating a body of relevant theory. Theories concerning group behaviour, whether economic, political, or social are legion. They are implicit in men's actions, transmitted in customary beliefs, and elaborated in literature. The distinctive quality of social science is that it attempts to formulate such theories unambiguously and in a way which permits the testing of theory against fact.

Method.—Much of the difficulty in science, contrary to popular misconception, is not to produce theory but to test it. For this reason social scientists tend to be as much concerned with method

as with result in their inquiries. Indeed what John Maynard Keynes said of economics is likewise true of all social science: "It is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions." Moreover, though the social sciences may be defined as the study of man in society, it must not be concluded that this is held to be the only method available for the understanding of social phenomena. On the contrary, much insight into the nature of communities, associations, classes, nations, and races comes from the poet, the novelist, the traveler, and the moralist. The social sciences can claim a special place in the vast body of writing on social life only insofar as they aim to produce generalizations which are verifiable because they are arrived at by scientific procedures.

The lack of maturity of the social sciences compared with the physical and biological sciences is often attributed to the use of quantification in the latter and its relative absence from the former. Certainly science is refined by measurement, and the conversion into numbers of such concepts as social distance, moral density, or political allegiance sets peculiarly intractable problems. Nevertheless it must be recognized that the use of quantitative methods distinguishes the natural from the social sciences only in degree; such methods are used with conspicuous success in economics and demography, and statistical techniques form part of the indispensable equipment of all modern social scientists.

More generally it is characteristic of the modern development of the social sciences that they are modeled on the aims and methods of the natural sciences. What these aims and methods are remains, to some extent, a matter for debate and throughout the 20th century, especially in America, there has been much criticism of misguided imitation of the natural sciences in the quantitative pursuit of the trivial, the substitution of scientism for science, and susceptibility to "testomania and quantophrenia," as Pitirim A. Sorokin said in his *Fads and Foibles in Modern Sociology and Related Sciences* (1956). Nevertheless it is increasingly recognized that there is a unity of method in all generalizing or theoretical sciences which is sometimes called the hypothetico-deductive method and which is admirably expressed in the writings of Karl Popper, for example, in *Poverty of Historicism* (1957), *Logic of Scientific Discovery* (1959), and *The Open Society and Its Enemies* (1945).

"What is important to realise is that in science we are always concerned with explanations, predictions, and tests and that the method of testing hypotheses is always the same. . . From the hypothesis to be tested—for example a universal law—together with some other statements which for this purpose are not considered as problematical—for example some initial conditions—we deduce some prognosis. We then confront the prognosis whenever possible, with the results of experimental or other observations. Agreement with them is taken as corroboration of the hypothesis, though not as final proof; clear disagreement is considered as refutation or falsification." (Karl Popper, *Poverty of Historicism*, p. 132, Routledge and Kegan Paul, London, 1957.)

Three Examples.—This brief statement of the fundamentals of theory and method in all science, natural or social, may be illustrated by reference to three examples, one from economics, one from sociology, and one from political science. The first comes from Lord Keynes's *General Theory of Employment, Interest and Money* (1936). The pre-Keynesian or neoclassical theory of the level of employment and wages held (1) that the demand for labour is determined by the marginal product of labour and (2) that the supply of labour is determined by the marginal disutility of work. These two schedules when put together would determine the wage of labour and, in consequence, would assure full employment, apart from "frictional" and voluntary unemployment. From this theory it could be deduced that the mass unemployment in the United States in 1932, for example, must have been due basically to a refusal by the unemployed to accept wages corresponding to their marginal productivity. This deduction or prognosis, Keynes argued, was not supported by the facts. He therefore set to work to show that the classical postulate—that the general level of real (as distinct from monetary) wages is determined by the character of the wage bargain—was an illicit assumption which limited the validity of classical theory to the special case of full employment. He then tried to generalize the theory by putting forward a new hypothesis concerning the forces which determine the general level of wages, income, and employment, emphasizing, in particular, the level of effective aggregate demand. In the new theory some initial conditions—for example the skill and quantity of available labour and the existing technology—were taken for granted. The level of employment (and the national income) was then explained in terms of variations in three variables—the propensity to consume, the marginal efficiency of capital, and the rate of interest—which determine the rate of investment and saving. In contrast to "classical" theory, the rate of interest was explained not in terms of savings and investments but as a function of a new concept of liquidity preference together with the quantity of money. In this way it was possible to deduce the existence of underemployment equilibrium in an economy.

A second example may be taken from the work of the German sociologist Max Weber. His theory that the development of Western bourgeois capitalism was necessarily preceded by Calvinist and other Puritan religious movements as creators of indispensable psychological preconditions offers an illustration of theoretical advance by methods akin to those used in the natural sciences. Weber started from the then well-established Marxist theory of the genesis of capitalism through an emerging *bourgeoisie* seizing favourable opportunities for capital accumulation, the creation of a propertyless proletariat, and the exploitation of a new technology permitting factory production. He then extended the theory by showing that these "material" conditions, though necessary, were not sufficient. They existed elsewhere, whereas Calvinism or its equivalent social counterpart of capitalism did not. As a result Weber contributed to the cumulation of theory in the social sciences. From the point of view of method Weber's work is one of the best examples of the use of history as a kind of natural laboratory for the testing of a hypothesis. He first established, in his famous essay on "The Protestant Ethic and the Spirit of Capitalism" (1904), that there is a high degree of congruence between the ethic of Protestantism and the organization of economic activity in capitalist societies. He then turned to an immense comparative study of other religions and other civilizations to seek for the negative instance of indigenous capitalism without the ascetic spirit of Protestantism. Of course the vagaries of history could not provide more than a rough approximation to the rigorous controls of an experiment in chemistry, and the Weber thesis remains controversial; but the plan of the work is an impressive demonstration of Popper's doctrine of the unity of method in science.

The third example, taken from political science, begins with German economist and sociologist Robert Michels' "Iron Law of Oligarchy" expounded in his *Political Parties* (1911), which held that political parties and other membership organizations inevitably tended toward oligarchy, authoritarianism, and dictatorial leadership. This example again illustrates the paradoxical nature of science as advancing by looking backwards critically at the existing state of theory and specifying the limits within which its hypotheses are valid. Thus a later American study (Seymour M. Lipset, Martin Trow, and James Coleman, *Union Democracy* [1956]) produced a set of specific hypotheses concerning the factors which will support or undermine member participation and democratic control in private organizations and thereby showed the limited conditions under which Michels' law applies and replaced a somewhat vague generalization by a clearer set of related propositions. The later study also illustrates a method increasingly followed in social science in that it used the economy of investigating the deviant case in order to delimit application of a hypothetical generalization—in this case the International Typographical Union, as the one outstanding exception to the general rule among private organizations of government by one-party oligarchy. The results constituted another example of cumulative research, the purpose of the study being, as the authors remarked, "not to 'refute' Michels or other previous workers in this area, but rather to refine and build on their insights and findings, paying them the respect

of using them more often than we quote them."

Historical Origins.—All science can be traced back to antiquity. In the modern world the disparity of achievement between the natural and the social sciences is sometimes put down to the youth of the latter or, on a more sophisticated view, to the relatively small number of hours spent on it. But it is doubtful whether either explanation is borne out by the facts. What is certainly true, however, is that the sciences of man and of nature have been the objects of oscillations in the attentions of scientifically minded men. In the contemporary world the natural sciences are plainly dominant. From roughly the beginning of the 17th century the advance of first the physical and later the biological sciences has not only outstripped that of the social sciences but has been the basis of a continuous revolution in human life. There have been previous dramatic advances in the material and cultural conditions of mankind—for example the Neolithic revolutions of some seven or eight thousand years ago which gave men control over their food supply through the development of the techniques of cultivation—but none of these stages in social evolution are comparable, as judged for instance by their effect on population size, with the rise of modern industrial society based on the application of natural science. Nevertheless the relative importance accorded the social sciences is not inevitable and indeed it is arguable that, at least at one stage in the development of Greek science, the works of Plato and Aristotle had put politics ahead of physics.

The crucial period in the growth of the modern disparity between the sciences was the 150 years from the beginning of the 17th century. Scientific interests had penetrated but never dominated the humanism of the Renaissance. Niccolò Machiavelli's *Prince* (1513) stands out as a monument to the release of men's minds from medieval preoccupations and is still read by the modern political scientist. But, as natural scientists, Leonardo da Vinci, Tycho Brahe, Kepler, and Copernicus in the 15th and 16th centuries have to be seen as precursors of the great astronomers and physicists of the 17th century, above all of Galileo and Newton. Even in the 20th century the social sciences are often said to await a Galileo, and only mathematical economics is thought of as having passed through its Newtonian revolution.

Again it must be emphasized that the remarkable concentration of human intelligence on the sciences of nature in the century and a half after 1600 was not absolute. In the same period the Italian Giambattista Vico had published his *Principles of a New Science Dealing With the Nature of Nations, Through Which are Shown Also New Principles of the Natural Law of Peoples* (1725) in which he set out to study human history by the methods which Francis Bacon had proposed for the study of the world of nature. In Vico's *Scienza Nuova* we find, as Edmund Wilson has put it (*To the Finland Station*, Doubleday, p. 2, 1940), "the modern sociological and anthropological mind waking amidst the dusts of a provincial school of jurisprudence of the end of the 17th century and speaking through the antiquated machinery of a half-scholastic treatise." But Vico's vision was not shared by his contemporaries, and his work was left to be rediscovered by Jules Michelet a century later. Similarly in England in the 17th century the scientific pursuits of the Royal Society had as a minor theme the study of human life by similar principles of empirical science. Sir William Petty's population studies, John Graunt's analysis of mortality rates in the London plague, and Edmund Halley's construction of the first life table mark the beginning of an English tradition of "political arithmetic" which, adding moral and administrative concern for the consequences of urban industrialism to its scientific interests, has flourished down to the present day.

However, the beginnings of modern social science may be traced directly to the 18th-century Enlightenment with the first flowering of rationalism in France and the stern pursuit of a secular moral philosophy in Scotland. The revival of interest in society at this time also illustrates the general truth that social science is the child of social change. The rise of capitalist society, with its attendant social upheaval of urban growth, industrialism, and mobility of men and ideas, imparted a powerful impetus to social inquiry. In France, through the work of the physiocrats, economics was launched as an empirical science, and in books like Voltaire's *Age of Louis XIV* (1751) the notion of progress was involved in the study of history and society in a way which broke sharply with medieval thought and to a degree which was unshaken until the rise of functionalism in anthropology at the beginning of the 20th century. The second half of the 18th century also saw notable advances in Scotland by a remarkable group of moral philosophers connected with the Scottish universities, including Adam Smith, John Millar, Francis Hutchison, and Adam Ferguson—all influenced by a growing mood of skepticism and empiricism in philosophy, of which perhaps the greatest exponent was David Hume. Adam Smith's *An Inquiry Into the Nature and Causes of the Wealth of Nations* (1776) remains, with Karl Marx's *Das Kapital* (1867) and Alfred Marshall's *Principles of Economics* (1890), to this day among the great comprehensive treatises in economics, but the Scottish moralists also contributed in their lesser-known works some of the foundations of modern sociology, psychology, and anthropology. Thus, for example, John Millar's *Observations Concerning the Distinction of Ranks in Society* (1771) is probably the first scientific analysis of social stratification, and Adam Ferguson's *An Essay on the History of Civil Society* (1767) is an admirable early example of the comparative study of social institutions.

Specialization and Synthesis.—In the 19th century the location of initiative in the social sciences shifted from Scotland to Germany and other continental European countries. During the course of the century what would now be thought of as the main divisions of social science—economics, anthropology, sociology, political science, and social psychology—began to emerge and, toward the end of the century, to receive academic recognition. The 19th century was also characterized by an opposite tendency toward the construction of a synthesis of social science, especially at the hands of Auguste Comte, Karl Marx, and Herbert Spencer.

In economics the first three-quarters of the century was devoted to the building of the synthesis of theory known as "classical economics"—a term invented by Marx to refer to Ricardo and his predecessors but which was later extended to include also the followers of that tradition, notably John Stuart Mill and Alfred Marshall. This phase came to an end with the clarification of the doctrine of marginalism (*see* ECONOMICS: *The Marginal Utility Analysis*). In the last quarter of the century attention was transferred to methodological problems—especially to the *Methodenstreit* controversy in Austria and Germany between Karl Menger and Gustav von Schmoller, the latter asserting that economics could be no more than a historical science in the sense that its laws could only apply to the unique circumstances of particular historical periods; the former insisting on the status of economics as a theoretical science searching for universal laws governing the behaviour of supply, demand, price, and the level of employment. The victory of Menger's point of view has since proved to be an emancipating force not only for economics and economic history in Germany but for the social sciences everywhere. On the other hand the models of rationality (*e.g.* "economic man") which permitted conspicuous advances in 19th-century economics left problems for the 20th-century development of theory in the social sciences, particularly that of finding a place for the nonrational determinants of human behaviour. For the most part these tasks fell to sociology and psychology, but economic theory itself was also extended to include problems of choice and expectations through applications of the mathematical theory of games.

The 19th century also saw considerable advances in those branches of the social sciences which were oriented more toward biology and history. The social scientists of this period inherited the 18th-century achievements of Linnaeus and Bouffon in the classification of living organisms and the Lamarckian hypothesis that the hierarchy of species was due to a natural process of evolution. In seeking to classify and date types of culture and to formulate generalizations concerning the social evolution of man, scholars like Lewis Morgan (*Ancient Society* [1877]) in the United States, Herbert Spencer (*Principles of Sociology* [1876–96] and *Descriptive Sociology* [1873–1934]), Sir Henry Maine (*Ancient Law* [1861]), and Sir Edward Tylor (*Primitive Culture*

[1871]) in England, Johann Bachofen (*Das Mutterrecht* [1861]) in Switzerland, and Friedrich Ratzel (*Völkerkunde* [1885]) in Germany laid the foundations for modern comparative anthropology and sociology. Spencer gave modern sociology most of its terminology (the word sociology is attributed to Auguste Comte), advanced the theory of supraorganic evolution before Darwin's *Origin of Species* appeared in 1859, and attempted a comprehensive classification of societies according to their social structures. Morgan too produced a developmental theory of cultural evolution through three "ethnical periods"—savagery, barbarism, and civilization. Maine followed the same line of inquiry in respect to legal and political institutions and Bachofen did similar work on kinship systems, attempting in particular to establish that matriarchy was an earlier evolutionary form than patriarchy.

These advances in the social sciences, though the basis for specialization in ethnology, physical anthropology, comparative sociology, etc., also linked the social sciences to geology, paleontology, and biology. In other words they linked history to natural history to yield a scientific account of the vast sweep of human fortune since the first appearance of man. There is no more exciting reading in all the literature of the social sciences than the modern summary of this branch of knowledge which is to be found in the writings of V. Gordon Childe (*Man Makes Himself* [1936], *Social Evolution* [1951]). These integrative developments were further stimulated by Darwin who had himself been influenced in his thinking about natural selection by the previous work of social scientists and particularly by Thomas Malthus' *Essay on the Principle of Population* (1798). Thus sociology and social anthropology became recognized and, under the banner of the great synthesizing idea of evolutionism, respected as worthy to march in intellectual company with economics.

There were also other movements toward a synthesis of the social sciences in the 19th century, including earlier the positivism of Auguste Comte and later the elaborate conceptualizations of German sociology. But the most powerful of these, and itself much influenced by classical economics and evolutionary theory, was Marxism. If biological and cultural evolution explained the development of nature and the primitive history of man, Marx seemed to offer the laws of motion governing not only the past but also the future of civilized societies.

Thus, by the beginning of the 20th century the social sciences gave promise of following quickly on the heels of the biological sciences into scientific maturity and academic acceptance. True, the basis for existing generalizations was being challenged and especially the theories of progress and unilinear evolution. But there was widespread optimism that the emerging specialties, or specialisms, would provide evidence for a more soundly based general sociology, the role of which in the 20th century would be to correct the bias of each particular discipline and relate it to the general body of theory. Meanwhile the universities were opening their doors to the scientific newcomers. Political economy had gained recognition in all the major countries by 1830. Edward Tylor was appointed as an anthropologist at Oxford in 1884. The University of Chicago recognized sociology from its foundation in 1890, and all of the social sciences were to be studied at the London School of Economics and Political Science founded in 1895. In Germany in 1897, Max Weber went to teach at Heidelberg and, in France, Émile Durkheim held a chair of education at the Sorbonne from 1902.

20th-Century Developments.—If the outcome in the 20th century has fallen short of Victorian hopes, and certainly the continued advance in the power of natural science with its unending implications for social, economic, and political change has lent urgency to the challenge faced by the social sciences, the advance of the sciences of man nonetheless has been considerable. The grand theoretical structures of 19th-century evolutionism, Marxism, and Comtean sociology have not survived. But four modest plants of 19th century origin were established in sturdy growth.

First may be cited a tendency for high standards of empirical method to spread throughout the specialisms in alliance with modest theory (sometimes called middle-range). The revolt against overambitious theorizing may be seen, for example, in the anti-evolutionary movement championed by the Polish-born British anthropologist Bronislaw Malinowski and by such men as Franz Boas in the United States. The subsequent trend in anthropology toward the detailed study of particular societies and communities within complex societies, using meticulous methods of observation and recording, has produced numerous examples of empiricism at its best. Social surveys in the 20th century have amassed great quantities of carefully collected data on the social structure of urban industrial societies while political science has built up a store of precise knowledge on voting behaviour, political parties, pressure groups, and the social composition of political elites, against which may be tested new theories of political behaviour.

Perhaps the greatest benefit of this period of concentration on more modest and detailed work was that the old and still valid problems of economic growth and social development could be approached anew after World War II. The urge toward improved standards of life through the spread of industrialism over the "underdeveloped" areas of the world stimulated renewed search for theories of political, social, and economic development adequate to encompass the accumulated body of fact. The problem of economic growth loomed large in postwar economic writing as may be seen from such examples as W. Arthur Lewis' *Theory of Economic Growth* (1955), W. W. Rostow's *The Stages of Economic Growth* (1960), or Simon Kuznets' *Six Lectures on Economic Growth* (1959).

This substantive concern with industrialization and social and political development has had integrating effects in the social sciences in that no one specialism is adequate to tackle the problems involved. The consequent tendency to a breakdown of specialist boundaries has involved the second trend which has been characteristic of the 20th century, namely the spread of a sociological approach into the various specialisms; *i.e.*, the recognition of the interdependence of the social, political, and economic forces which determine behaviour. Historical studies have also been affected by this development. Thus Richard Hofstadter's *The Age of Reform* (1955) is an example of a sociological orientation in history, and Neil Smelser's *Social Change in the Industrial Revolution* (1959) is a direct application of sociological theory to the Lancashire cotton industry between 1770 and 1840. Similarly, books like David Apter's *The Gold Coast in Transition* (1955) illustrate the use of sociological theory in the analysis of political change. It is from the interpenetration of the established disciplines that a new synthesis of theory is looked for and, indeed, at least one 20th-century attempt to build a unifying conceptual framework for sociology, anthropology, and social psychology has appeared in the writings of Talcott Parsons (*The Structure of Social Action* [1937; 2nd ed. 1958], *The Social System* [1951]).

The third noteworthy development has been the rise of several branches of psychology having relevance for the analysis of social behaviour. Among the more important of these are psychoanalysis, behaviourism, Gestalt theory, and the study of individual differences. Above all the influence of Freud on the social thought of the 20th century is comparable only with that of Marx and Darwin in the 19th century. Quite apart from his voluminous clinical writings Freud made direct excursions into anthropology (*Totem and Taboo* [1918]) and social psychology (*Group Psychology and the Analysis of the Ego* [1921], *Civilization and Its Discontents* [1930]). But no branch of the social sciences has escaped the influence of psychoanalytical ideas. They were introduced into political science by Harold Lasswell in his *Psychopathology and Politics* (1930) in a way which gave fresh impetus to the neglected insistence by earlier writers like Graham Wallas (*Human Nature in Politics* [1908]), Arthur F. Bentley (*The Process of Government* [1908]), and Charles Merriam (*New Aspects of Politics* [1925]) on the relevance of psychological and sociological factors in the study of political man. Psychoanalytical ideas were carried further into social anthropology by Abram Kardiner and into sociology by Talcott Parsons.

Movements in psychology other than those inspired by Freud and his earlier associates have made contributions both to theory and more especially to rigorous methods of experiment, testing, and measurement. Examples include the contribution of Clark

Hull to learning theory (*A Behavior System* [1952]), Wolfgang Köhler's work on perception (*Gestalt Psychology: an Introduction to New Concepts in Modern Psychology* [1947]). Kurt Lewin's influential *Field Theory in Social Science* (1951), and Donald Hebb's integration of Gestalt theory with neurophysiology in his *Organization of Behavior* (1949). Out of these movements there has also developed a recognized specialism of social psychology concerned with such topics as the scientific study of attitudes, public opinion, the social determinants of perception and motivation, and the interaction processes of small groups; and each of these topics boasts a considerable literature.

The fourth line of advance in 20th-century social science has been the steady improvement and extension of quantitative methods. At the end of the 19th century mathematical and statistical techniques for social analysis were in their infancy. Simple quantification had always been part and parcel of demography and the political arithmetic tradition. But, at least until after World War I, mathematical economics was not taken seriously in the English-speaking world, being largely confined to France, Italy, and Switzerland and such men as Léon Walras, Augustin Cournot, and Vilfredo Pareto. Thus the first systematic exposition of economic theory in mathematical terms, Walras' *Éléments d'économie politique pure* (1874–77), was not available in English until 1954. However, during the course of the 20th century mathematical treatments have tended to displace literary discussion, and statistical techniques have generated a new discipline of econometrics and transformed research in economics, sociology, and political science. Sampling procedures based on probability theory were introduced between World Wars I and II to make social surveys, public opinion polls, and electoral studies both economical and reliable. Psychological research has developed refined methods of factor and multivariate analysis, tests of ability and aptitude, and assessments of the significance of differences in the distribution of attributes between groups. Advances in computer technology have also revolutionized the possibilities of complex large-scale analysis of social phenomena.

See also ANTHROPOLOGY; ECONOMICS; PSYCHOLOGY, SOCIAL; SOCIOLOGY; SCIENCE, HISTORY OF; SCIENTIFIC METHOD.

BIBLIOGRAPHY.—An excellent *Reader's Guide to the Social Sciences* (1959) is available, edited by Berthold F. Hoselitz. *Methodology and Research:* K. Popper, *Poverty of Historicism* (1957); Max Weber, *Methodology of the Social Sciences* (1949); Paul Lazarsfeld (ed.), *Mathematical Thinking in the Social Sciences* (1954); Paul Lazarsfeld and M. Rosenberg (eds.), *The Language of Social Research* (1955); L. Festinger and D. Katz (eds.), *Research Methods in the Behavioral Sciences* (1953). *Anthropology:* E. Evans-Pritchard, *Social Anthropology* (1951); A. L. Kroeber *et al.* (eds.), *Anthropology Today* (1953), *Selections,* ed. by S. Tax (1962). *Economics:* Paul Samuelson, *Economics*, 5th ed. (1961). *Political Science:* H. Eulau *et al.* (eds.), *Political Behaviour: a Reader in Theory and Research* (1956). *Social Psychology:* E. E. Maccoby, T. M. Newcomb, and E. L. Hartley, *Readings in Social Psychology*, 3rd ed. (1958). *Sociology:* Kingsley Davis, *Human Society* (1949); George Homans, *Social Behavior: Its Elementary Forms* (1961). *Theoretical History:* Patrick Gardiner (ed.), *Theories of History* (1959). (A. H. HA.)

SOCIAL SECURITY, a term that originated in the United States and came into common use throughout the world in the first half of the 20th century to denote programs established by law to provide for the economic security and social welfare of the individual and his family. Social security systems vary from country to country but in general they provide for benefits to the aged, widows and young children, the disabled, the unemployed, and for some form of medical care. The three principal means of providing such benefits are by social insurance, public assistance, or public service. Social insurance, as the name implies, is based on insurance principles, although other elements such as the number of dependents may also be taken into consideration. It is usually financed through contributions made to a special fund by employees and employers (and in some cases, government); the benefits depend upon the individual's earnings or contributions and are described in the statute establishing the system. Public assistance programs provide cash payments and services to needy individuals (such as widows with dependent children) to help them meet their basic needs. Under public service programs, cash payments or services are provided directly by the government to every member of the community who qualifies, regardless of his economic need or his contribution to an insurance fund.

Origin and Growth.—In most countries, until late in the 19th century, the relief of poverty was left to private charity, to churches, to workingmen's provident associations, or to the government acting under so-called poor laws (*see* POOR LAW). In 1883–84 Germany, under the leadership of Bismarck, became the first European country to introduce a modern form of social security—a contributory health insurance program (at that time called "sickness insurance"), a national compulsory accident insurance program and old-age pensions. Germany's example was soon followed by Austria and Hungary and by the early years of the 20th century public opinion in most European countries had come to favour some means of providing old-age pensions for the needy and of dealing with such contingencies as sickness and unemployment among workers. Great Britain in 1908 established a system to provide noncontributory pensions to needy persons who had reached the age of 70, and in 1911 it became the first country to adopt a national system of unemployment insurance (*q.v.*). The compulsory health insurance program adopted in Great Britain in 1911 was greatly extended in 1920. A system of contributory old-age, widows', and orphans' pensions began in Great Britain in 1925 and unemployment insurance was put on a new basis by the Unemployment Assistance Act of 1934, which set up at the same time a national system of unemployment assistance. (For the first U.S. Social Security Act, see *United States*, below.)

In no country does social insurance cover all needs and all contingencies; public assistance is therefore provided to aid those persons in need who are not entitled to some form of insurance benefit (or to supplement an insurance benefit). Whereas social insurance implies that benefit payments are made to contributors according to a formula stated in the insurance law, public assistance payments are based on a minimum level of subsistence. Public assistance as the sole or main form of social security is now used in only a few countries and is gradually being replaced by social insurance.

Social insurance programs are usually compulsory in the sense that specified categories of employers and employees are required by law to participate by paying contributions. But in some countries the system has voluntary or quasi-voluntary features.

After World War II a great deal of attention was given to social security in most European countries and in other parts of the world as well. Each year after 1945 saw the introduction of social security measures into some countries for the first time; by the mid-1960s they were in operation in more than 100 countries.

UNITED STATES

The development of social security legislation in the United States was slow, and the recognition of social welfare as a governmental responsibility was late in developing. Because an agricultural economy set the pattern of life in the 19th and early 20th centuries (except in larger towns and cities) there was little unemployment in the modern sense of that term, and family groups assumed responsibility for their members who were in need. The 19th century was characterized by a generally accepted belief that personal initiative and industry were all that was necessary for the achievement of material success. This attitude tended to support the traditional belief that indigence was synonymous with incompetence; there was a definite stigma attached to any governmental aid given the financially unfortunate. The traditional concept that public welfare was a function of local rather than state or federal government also tended to retard improvements in welfare programs. Only gradually, as welfare needs increased, did the states come to recognize that they would have to assume at least some of the burden of caring for the poor. In the first three decades of the 20th century many states made arrangements for granting public assistance to certain categories of the population, but this type of assistance was extremely limited in its coverage. Workmen's compensation (*q.v.*), which had its beginning in 1908 was the first type of social insurance to be developed in the United States, and, with the single exception of an unemployment insurance bill passed by Wisconsin in 1932, it was the only form of

social insurance in effect in the United States in 1935 when a broad national program was recommended.

Establishment of a Federal Program.—The national system of social security established by the Social Security Act of 1935 was part of Pres. Franklin D. Roosevelt's "New Deal." It provided for a system of federal old-age benefits for workers; a federal-state system of unemployment insurance; and federal financial aid to the states to help them provide public assistance to dependent children, the needy aged, and the needy blind. These programs were administered at the federal level by a bipartisan Social Security Board of three members. The other programs included in the act—provision for maternal and child health services, services for crippled children, child welfare services, vocational rehabilitation, and assistance to the states for public health—were administered by other federal agencies.

In 1939 the Federal Security Agency was created and six organizations were placed under its jurisdiction—the Social Security Board, the Public Health Service, the Office of Education, U.S. Employment Service, the Civilian Conservation Corps, and the National Youth Administration. In 1940 the Food and Drug Administration was transferred to this agency. During the next dozen years there were numerous administrative changes, followed by the creation in 1953 of the Department of Health, Education and Welfare, headed by a secretary who is a member of the president's cabinet. The Federal Security Agency was abolished and all its functions were transferred to the new department.

Old-Age, Survivors and Disability Insurance.—The old-age insurance program established by the original Social Security Act was designed to provide a basic retirement income at age 65 for workers insured under the system. Before 1935 various retirement plans had been established for public employees and private pension plans had been set up by some business organizations. But by 1930 the number of employees covered by such private plans was only about 3,500,000. The Social Security Act of 1935 provided old-age insurance for millions of additional workers, and later amendments brought in millions more. The original act covered only employees in industry and commerce—about six out of ten workers in the labour force—for old-age benefits. Thirty years later the law provided almost every wage earner, salary earner, and self-employed person with social insurance protection against three basic risks—old-age, long-term disability, and death of a breadwinner.

Coverage.—Chiefly because of the administrative difficulties involved, certain types of employment were not at first covered by the Social Security Act. Among the types of work not covered were agricultural labour; domestic service; casual labour not in the course of the employer's trade or business; service as an officer or member of the crew of a merchant ship; employment with any federal, state, or local government or its instrumentality; and service with certain religious, charitable, humane, and educational institutions not operated for profit. Railroad employment was excepted because it was covered by a separate system established by the Railroad Retirement Act. All employment after age 65 was also excepted.

Under the 1939 amendments some of the types of employment previously excepted were brought under the system. These included service on U.S. vessels and employment in national banks, building and loan associations, and certain similar organizations. The exclusion of employment after the age of 65 was removed as of Jan. 1, 1939, thus enabling many workers previously ineligible because of age to qualify for benefits and enabling others to increase the amount of their benefits. At the same time, the amendments added certain exceptions. The term "agricultural labour" was redefined to exempt some services that were closely related to farming. Certain part-time or intermittent employment providing only nominal wages, services of newsboys under 18 years of age, family employment, and employment for foreign governments were also excluded by the 1939 amendments.

Under the 1950 amendments, coverage was broadened to include the self-employed (except farmers and those in certain professions) and regularly employed agricultural and domestic workers. Voluntary coverage was permitted for employees of nonprofit institutions and of state and local governments not already under a retirement system. Certain civilian employees of the federal government not under a retirement system were also included. Amendments in 1954 extended coverage to about 10,000,000 additional workers, most of whom were self-employed farm operators. Coverage was given to additional members of state and local government retirement systems, and to larger numbers of agricultural and domestic workers by liberalizing the definition of the term "regular employment." In 1956, coverage was extended to several additional groups of the self-employed. Larger numbers of farm owners and operators were brought within the coverage of the act, and, except for doctors of medicine, most self-employed professional people were included. Coverage was also extended to almost 3,000,000 members of the uniformed military services on active duty. In 1965 self-employed physicians were covered and the only large group outside the system were the civilian employees of the federal government who were covered by other federal systems.

Contributions (Taxes).—In the beginning, old-age and survivors insurance was financed solely by a payroll tax levied on both employees and employers. The tax rate was 1% per year on the first $3,000 of the wages of each employee, matched by the same amount paid by the employer. An increase of ½% was to be made every three years until in 1949 a total contribution of 6% would be payable (3% by employers and 3% by employees). The 1939 amendments, however, eliminated the ½% increase for the years 1940–42, and, because of subsequent acts of Congress, the tax did not actually increase to 1½% until Jan. 1, 1950.

Under the 1950 amendments the 1½% rate of tax was retained for 3 more years, and the tax for the newly covered self-employed persons was set at 2¼% of net earnings. At the same time a new maximum earnings base of $3,600 was established. In 1954 the combined employer-employee tax rate was 4% (2% each) while each self-employed person contributed 3%. The same rates were retained for 1955–56 but the maximum taxable earnings base was increased to $4,200. In 1957 the schedule of rates was increased by ¼% to cover the cost of disability insurance benefits, with the combined employee-employer rate a total of 4½% of the earnings base while the self-employed began paying 3⅜%. Scheduled increases in 1959, 1960, 1962, and 1963 brought the contribution rate for 1965 up to 7¼% for employees and employers (3⅝% each) and 5⅜% for the self-employed. The tax was then paid on the first $4,800 of earnings. The 1965 amendments increased benefits, adjusted the contribution schedule, and increased the maximum earnings base to $6,600 beginning in 1966. The 1966 rate was established at 7.7% for employers and employees (3.85% each) and 5.8% for the self-employed. The schedule provided increases in steps until the tax would reach 9.7% (4.85% each) and 7% respectively in 1973. The law provided for additional contributions for hospital insurance beginning at $\frac{7}{10}$ of 1% in 1966 ($\frac{35}{100}$% each) and reaching by progressive steps 1.6% in 1987 ($\frac{8}{10}$% each).

Financing.—Under the act of 1935 an old-age reserve account was established in the United States treasury wherein receipts from the payroll taxes were to be deposited, with future benefit payments to be made from this account without drawing on the general revenues of the government. The act specified that reserve funds in the account were to be invested in government obligations or government-guaranteed obligations bearing 3% interest. Although the amendments adopted in 1939 temporarily increased the costs, Congress, as noted above, held the contribution rate at 1% until 1950. In spite of this, the trust fund grew more rapidly than had been expected, partly because of high wartime employment and earnings, and partly because of increased employment opportunities for older workers during the war years. Between 1950 and 1956 there were three changes in the benefit formula; one nearly doubled the benefits while the other two provided for more moderate increases. Although individual benefits increased substantially and contributions remained at 1½% until 1954, when they were raised to 2%, the total assets in the fund continued to increase. In 1957 a separate disability insurance trust fund was established to provide payments to persons who became totally disabled. Increased costs of the program were

accompanied by increases in the contribution rates every year or two.

As a result, the program was fully supported by the social security taxes paid by workers and their employers without any general subsidy. For a time during the 1940s the law had included authorization to appropriate from general revenues any additional money that might be needed to finance the benefits, but additional money was not needed. In 1950 Congress made clear its intention that social security taxes, plus the interest earned on accumulated funds, should be sufficient to meet all of the costs of the program without any subsidy from general revenues.

Benefits.—The 1935 act provided for the payment of monthly benefits, beginning in 1942, to insured workers who were 65 years of age or over when they retired. Benefits were to be based on the employee's total wages during the qualifying period (exclusive of amounts received in excess of $3,000 per year from any one employer). The formula used to compute benefits was weighted somewhat in favour of the lower-paid worker. The maximum monthly benefit was set at $85, the minimum at $10. A lump-sum payment equal to 3½% of the employee's total wages was provided for those reaching age 65 without qualifying for monthly benefits; and a death benefit of the same amount was provided subject to deduction of any benefits the worker might have received during his lifetime.

The 1939 amendments made material changes in these benefits by advancing the start of monthly benefit payments to 1940 instead of 1942, increasing benefit rates, and bringing greater security to the family group by providing benefits to certain dependents and survivors of insured workers. Because of this latter extension of the system, the title was changed to "old-age and survivors insurance," abbreviated as OASI. Instead of computing the benefit rate on the basis of total wages received, a new formula based on average monthly earnings was adopted.

From 1939 through 1950 the average monthly wage was determined by dividing a worker's total wages, exclusive of annual amounts in excess of $3,000, by the number of months he could have worked under the system; *i.e.*, the number of months after 1936 (or his 22nd birthday, if later) and the date of retirement or death. In 1950 the benefit formula was amended to make it possible, beginning in 1951, to include in the computation earnings up to $3,600, and to use only those earnings credited after 1950, if this would yield a higher benefit rate. One reason for this "new start" was that the amendments of that year brought in many new groups who had no earnings to their credit for the earlier period. In 1955 the maximum earnings taken into account in computing benefits was increased to $4,200 a year, in 1959 to $4,800, and in 1966 to $6,600.

In addition to increasing the maximum earnings on which benefits were based, the formula for determining benefits was several times revised to provide larger payments. Between 1940 and 1950 benefits were related not only to wages but also to years under the system; for each year of coverage an increment of 1% was added. A worker with an average monthly wage of $200 in the 1940–1950 period would have been entitled to a benefit of $41.50 per month if his association with the system had been limited to three years. However, if he had been in covered employment for the entire period his monthly benefit would have been $45.20. Beginning in 1951, the benefits rates were increased, but the 1% increment for each year of coverage no longer applied. A person with average earnings of $250 a month, provided he met other conditions of entitlement, would have received $72.50 per month and if his earnings were $300 per month (the maximum then creditable) his monthly benefit would have been $80. The 1958 act provided $95 a month for the worker with average annual earnings of $3,000 (equivalent to the former $250 per month) and $127 a month for the worker with the maximum creditable average annual earnings of $4,800. The 1965 amendments increased benefits 7% retroactive to Jan. 1, 1965. Minimum benefits were increased from $40 to $44 a month. An increase in the contribution and benefit base to $6,600 a year made possible a maximum monthly benefit of $168. Average old-age benefits were $23 a month in 1940, $44 in 1950, and $84 in 1965.

As originally passed, the Social Security Act provided benefits only for insured persons. As noted above, the 1939 amendments made supplementary benefits available for certain family members of a retured worker and certain survivors of a deceased insured person. Beginning in 1940, benefits equal to one-half the worker's benefit were paid to a wife 65 years of age and older and also to each unmarried dependent child under 16 years of age (or under 18 if regularly attending school). In case of the death of an insured person, three-quarters of the insured's earned benefit would be paid to his widow if she was 65 years of age or over, or to his widow, regardless of age, if she had dependent children. Benefits equal to one-half the worker's benefit would be paid to each of these dependent children and, if there were no other family members eligible for benefits, to dependent parents. With the 1950 amendments, the wife of a retired worker, regardless of her age, was entitled to benefits if she had in her care a child of his who was eligible for monthly benefits. For the first time benefits were payable to a dependent husband (age 65 or over) of a retired or deceased woman worker. Benefits for the first surviving child were raised to 75% of the insured person's benefit. The 1960 act extended this percentage rate to all children, while legislation in 1961 increased the aged widow's or widower's rate to 82½% of the basic rate, which also applied to a dependent parent (75% for each if there were two such parents). Total family benefits were subject to statutory maximums, with the 1958 act setting up the maximum monthly figure of $254 or 80% of the average wage whichever is smaller, but in no case less than 1½ times the worker's basic benefit. The 1965 amendments modified the method of computing the family maximum and raised the monthly maximum to $368 for a family.

Under the 1956 amendments women became eligible to draw benefits at age 62, reduced on an actuarial basis for each month before age 65 if they claimed benefits as wives of retired workers—or without a reduction if they were widows. The 1961 amendments extended a similar provision to men. The amount of reduction was adjusted for persons who retired at any time between age 62 and 65. The 1965 amendments permitted a widow at age 60 to elect a benefit actuarially reduced for each month she is under age 62.

To make the social security system one that would more nearly provide lifetime social insurance protection for insured workers their dependents, and their survivors, the beginnings of a disability insurance program were included in the 1954 amendments. This provided for a "freezing" of the worker's old-age and survivors account at the beginning of the year prior to his becoming totally disabled. This was somewhat analogous to the "waiver of premiums" feature of commercial insurance programs. The period of disability was to be disregarded in determining the amount of eventual benefits. The 1956 amendments provided for payments to disabled individuals who had reached age 50, and to dependent sons and daughters of retired and deceased workers, if the child's disability began before age 18. The 1960 law removed the age qualification for entitlement to disability insurance benefits. The benefits payable to a disabled worker and his dependents were made the same as if the worker had attained age 65 at the time he became disabled. In 1965 further changes were made in the disability insurance benefit including liberalization of the definition of disability and inclusion of reimbursement from social security funds for the rehabilitation of disabled persons.

In addition to the periodic benefits, a lump-sum payment was made available on the death of the insured worker. From 1940 through 1950 these payments were made (if no monthly survivor benefits were payable) to relatives or to the person who paid the funeral expenses. The amount of the payment so made was then six times the worker's monthly benefit rate. After 1950, an amount equal to three times the worker's monthly benefit rate was made to the spouse, if one survived, or to the person paying the funeral expenses. Under the 1939 act the maximum lump-sum death benefit was $360 (assuming 50 years' coverage). The 1950 amendments set a maximum of $240, and the 1954 act increased this to $255.

Eligibility for Benefits.—To qualify for benefits, a person must

have had earnings in covered work over a specified length of time, which may vary with the individual's date of birth, sex, and the type of benefit that is being applied for. Since 1939 the act has provided for two kinds of insured status—"fully insured" and "currently insured." The 1939 act required that a worker was not fully insured until he had received at least $50 in covered employment in each of 40 calendar quarters or in each of enough calendar quarters to equal half the calendar quarters elapsing between Dec. 31, 1936 (or his attainment of age 21, if later) and the quarter in which he died or attained the age of 65, but never any fewer than six calendar quarters. The method of determining eligibility on the basis of calendar-quarter credits was modified somewhat by later amendments. This was necessary because new groups were brought under the act and because self-employment income was not credited until after 1950. After 1950, most self-employment net earnings of at least $400 in a taxable year yielded four quarters of coverage. Under the 1961 act persons were considered fully insured if they had quarter-of-coverage credits equal to at least the number of years after 1950 (or after age 21, if later) and up to age 65 (62 for a woman), or up to the date of death, if this occurs earlier.

If a worker dies before acquiring "fully insured" status, certain survivors benefits may be paid if he is "currently insured," *i.e.*, if he had credit for at least a year and a half of covered work within the three-year period prior to his death. For a dependent husband or widower to be eligible for benefits from the work of a woman worker, she must have been both fully and currently insured at the time of her death, attainment of age 62, date of disability, or the time of filing of her application for old-age benefits, if after age 62. Generally speaking, an applicant for disability insurance benefits, in addition to having "fully insured" status as of the date of disability, must also have credit for covered work in 5 out of the 10 years ending when he became disabled.

Once benefits have started, certain events can cause them to be suspended for one or more months or be terminated altogether. Under the 1935 act, retirement was required for receipt of old-age benefits—receipt of any earnings in covered employment was sufficient cause to suspend benefits for the month or months in which they were earned. The 1939 act modified the retirement requirement and allowed benefits to continue if earnings in covered employment were less than $15 a month. In 1950 the act was amended to provide that benefits would be paid to persons age 75 and over, irrespective of retirement, and for those under age 75 benefits could be paid if covered earnings did not exceed $50 a month, and this figure was increased to $75 by the 1952 amendments. The 1954 amendments allowed all benefits to be paid during a year when earnings did not exceed $1,200 (including all earnings whether in covered employment or not). Beginning in 1966 a beneficiary could earn up to $1,500 a year in any employment without loss of benefits. For each $2 of the next $1,200 of earnings, $1 of benefits was withheld, and for each $1 of earnings over $2,700, $1 was withheld. For beneficiaries age 72 or over there was no retirement requirement. Some other events that may terminate benefits are remarriage on the part of a widow, divorce of a wife, adoption by a person outside the family, marriage of a child receiving benefits, or recovery of health or demonstration of ability to work on the part of a disability beneficiary.

Unemployment Insurance.—The provisions of the 1935 Social Security Act relating to unemployment insurance were designed to stimulate adoption of state systems that would insure employees against total loss of income during periods of unavoidable unemployment. Because of constitutional law problems, this approach was considered preferable to a wholly federal system of unemployment insurance. Unemployment insurance bills had been presented in a number of state legislatures before Congress considered the Social Security bill in 1935. The burden that such legislation would place on state industries in competition with industries in states without such legislation had deterred action in all states except Wisconsin. By levying a federal tax on employers and then allowing credit against this tax for contributions made to a state unemployment fund, the federal act helped to remove the interstate competitive factor, and by June 30, 1937, all states had enacted unemployment insurance laws. Under the terms of the federal act, wide latitude was allowed the states in adopting unemployment insurance systems, but the state laws nevertheless had many basic similarities.

Financing.—The unemployment insurance provisions of the Social Security Act levied a federal payroll tax on employers who had eight or more employees, certain types of employment excepted. The rate of this tax was set at 1% of wages up to $3,000 per year in 1936, 2% in 1937, 3% from 1938 to 1961 and $3\frac{1}{10}$% in 1961 and thereafter. Federal legislation in 1954 required employers of four or more employees to pay this tax beginning in January 1956. Against this federal tax, employers were allowed credit for the amounts they contributed to unemployment insurance funds under a state law approved by the Social Security Board. The maximum credit allowed was 90% of the 2.7% federal tax, thus providing funds for the federal government to meet necessary administrative costs. The federal law allowed an employer credit for any reductions in his state rate of contribution resulting from favourable experience in regard to payment of unemployment benefits. As a result of the experience-rating provisions, the average employer contribution to state unemployment insurance funds was 2% of taxable payroll. The employer's contribution, like the federal tax, was ordinarily based on the first $3,000 of earnings. By the 1960s, however, several states had raised the taxable wage base to $3,600 or more. Employee contributions were required by only a few states.

All money collected by the states for unemployment insurance is deposited in the U.S. treasury in an unemployment trust fund to be drawn upon by the states to pay benefits. Until 1954 the unemployment tax paid to the federal government became part of the revenue of the federal treasury, and funds necessary for administration were appropriated annually. Unemployment tax collections consistently exceeded these appropriations, and thus parts of the tax funds were not used for the employment security program. A 1954 amendment earmarked these taxes collected in excess of administrative expenses, and the funds became available for noninterest-bearing loans to states whose funds fell below a specified level. This amendment substantially increased a continuing ability to pay benefits in states with relatively high levels of unemployment.

Coverage.—The 1935 federal law excepted from the employer's tax wages paid for the following services: agricultural labour; domestic service; shipping on navigable waters of the United States; employment for the immediate family; federal, state, and local government service; and employment by certain religious, charitable, and educational organizations not organized or operated for profit. Railroad employment was exempted from coverage when the federal Railroad Unemployment Insurance Act was enacted in 1938.

The amendments of 1939 extended coverage of the federal unemployment insurance tax to wages paid for employment by national banks, building and loan associations, and similar organizations, and in 1946 maritime employment was added. In 1954 a federal employment insurance law provided compensation to federal employees for income lost because of unemployment. Most state laws closely followed the occupational exclusions of the federal unemployment tax act. By the mid-1960s workers covered by unemployment insurance represented about 60% of the labour force and 80% of all civilians earning wages or salaries.

Benefits.—From the beginning, all state laws provided that weekly benefits be related to the weekly wage of the individual worker. In 1939 average benefits in the country as a whole were about 40% of the average wage of covered workers, and although benefits increased substantially thereafter, the increases lagged behind rising wages. Thus, by the mid-1960s the average benefit had dropped to less than half the weekly wage.

Both the amount of the benefit and the length of time for which it could be paid have differed from state to state and from year to year. By 1964 average weekly benefits for the United States as a whole were $36. Maximum weekly benefits varied widely from state to state. Unemployment insurance in the United States was designed to cover only relatively short periods of unemploy-

ment. In the early years of this legislation most laws provided benefits for a maximum of 16 weeks. By the mid-1960s practically all states set the maximum at 26 weeks and about a fifth of all states had a maximum exceeding 26 weeks. In virtually all states a waiting period of one week was required before benefits became payable. In 1958 and again in 1961 the federal government provided for a temporary extension of unemployment compensation for workers who had exhausted their regular state benefits. All the state laws denied benefits to claimants who were unavailable for work or were unable to work; were involved in a labour dispute; voluntarily left their jobs or refused work without good cause; or were discharged for misconduct. When a worker applied for unemployment benefits he registered at a state employment office, reported to that office each week as evidence that he was able and willing to work, and was obliged to accept any "suitable" employment (as defined by the state) offered him.

Public Assistance.—The act of 1935 made provision for granting federal funds to the states to aid them in giving financial assistance to three groups of needy persons—the aged, the blind, and dependent children. In contrast to social insurance benefits, which are available to all persons who qualify as to age, disability, or inability to find a job and who meet the requirements as to length of time in covered employment, public assistance is available only after proof of need—a so-called means test. The states adopt and administer their own public-assistance plans, but each state plan must be submitted to the federal government for approval. Standards for approval of state plans specified by the Social Security Act of 1935 were designed to assure efficient administration and equitable distribution of assistance. Each state, however, was free to determine the amount of the assistance grant and its own eligibility conditions so long as such provisions were not more restrictive than those required for approval by the federal act. The state laws thus differed considerably, some being distinctly more liberal than others.

The 1935 act fixed the amount of federal funds available to the states for old-age assistance and aid to the blind at one-half the state's total assistance payments, up to a monthly federal-state maximum of $30 to each recipient. For aid to dependent children, the act provided federal grants equal to one-third of the amount paid by the state. As a basis for computing the federal share of aid, the act set a maximum of $18 a month for the first child and $12 for each additional child in the same home. A dependent child was defined as one who is under age 18 (until 1956 children between the ages of 16 and 18 had to be regularly attending school), is without parental support because of the death, disability, or continued absence of a parent (or because of unemployment of a parent after the 1961 amendments), and who is living in his own home or with certain near relatives (or receiving foster care under conditions specified in the 1961 and 1962 amendments). The federal government also made grants to the states that covered one-half the cost of administration. In 1962, in order to encourage the states to expand their social services for the rehabilitation of needy persons, the federal share of providing social services under the state plans was increased from one-half to three-quarters of the cost. After 1935 the federal grants to states for public assistance were increased several times, and in 1950 these grants also became available for aid to the needy who are permanently and totally disabled. In 1960 grants were made available for medical assistance for aged individuals needing help only in meeting their medical expenses. The 1965 amendments added two health insurance programs for the aged (*see* below) and also extensively revised the medical assistance program and broadened it to enable the federal government to help finance state payments for medical services for practically all needy persons. By the mid-1960s medical assistance had become a substantial and still growing part of the public assistance program. As with cash payments, the medical assistance provisions of state programs varied widely with respect to groups covered, income and other eligibility conditions, and the scope of medical services covered. As a result of legislation during the early 1960s, federal funds were made available to states and localities for work relief and training programs for needy persons on public assistance.

Health and Welfare Services.—The original Social Security Act, in addition to its provisions for social insurance and public assistance, included federal grants to help states develop maternal and child health services, services for crippled children, and child welfare services. At the federal level, the state programs were administered by the Children's Bureau of the Welfare Administration. The sums granted by the federal government for these services were increased several times over the years and on various occasions, funds were earmarked for specific problems such as maternity and infant care, for the health of school and preschool children, and to provide day care for children of working mothers. The act of 1935 also provided funds for distribution to the states to assist them in extending public health services. The approval of the surgeon general of the United States was required for all state programs, and the federal Public Health Service assisted the states in improving their work in this area. As a result of separate legislation grants were made for general public health, to aid in the control of tuberculosis, cancer, venereal disease, mental illness, heart disease, for the health of migrant workers and those chronically ill and aged, and for construction of hospitals and community mental health centres. In 1963 and 1965 grants were also made under the Social Security Act to combat mental retardation.

Federal grants to states for vocational rehabilitation were made as early as the 1920s. The Social Security Act of 1935 continued this program and increased the grants. Separate legislation in 1943, 1954, and 1965 expanded the program. The purpose of such legislation was to assist the states in providing disabled persons with services that would enable them to engage in remunerative employment. Any person who had a physical or mental disability and could reasonably be expected to become employable was eligible to participate in the program.

Workmen's Compensation.—It was not until about 1900 that the tremendous losses due to industrial accidents in the United States gained recognition and the movement for workmen's compensation got under way. Before enactment of workmen's compensation laws, an injured worker who sued his employer for compensation had to prove that the employer had been negligent. The employer was also entitled to various specific defenses under the common law, and the court procedure was slow, costly, and uncertain. Workmen's compensation legislation introduced a new legal and economic principle. the concept of individual fault was abandoned, and the cost of industrial injuries was considered to be part of the cost of production. (*See* WORKMEN'S COMPENSATION.)

Social Insurance for Special Groups.—Retirement and unemployment insurance are provided by special legislation for railroad workers. The Railroad Retirement Act of 1935, with later amendments, included the same types of benefits provided under federal old-age and survivors insurance, but higher contributions were required and the benefits, in most cases, were more liberal for railroad workers. A federal railroad unemployment insurance law was passed in 1938, and after 1946 this legislation also included temporary disability benefits. These benefits were financed initially by a tax on employers fixed at 3% of the first $300 of monthly wages. After 1948, however, because of high employment and a large reserve, employers were permitted to pay a tax much lower than 3%, the actual tax for some years being only ½% of wages. In 1964 the rate was 4% on the first $400 of monthly wages. This act was similar in many respects to the general state unemployment insurance acts. The average benefits, however, were higher (about $50 a week in the early 1960s in contrast with an average benefit paid under the state laws of about $35). Another difference between this federal act and the state laws was that all railroad employers paid the same tax rate, whereas the state acts allowed the individual employer's contribution to vary on the basis of the amount of unemployment of his workers.

There are many different types of retirement systems for employees of state and local governments. After 1951 it was possible for employees with no retirement systems of their own to be brought under the federal old-age and survivors insurance law on

a voluntary basis. After 1954 employees covered by a state or local retirement system could be brought under federal old-age and survivors insurance.

A federal civil service retirement system was established in 1920 and was improved and liberalized many times. The vast majority of the employees of the federal government were covered by this system. In 1950 the federal old-age and survivors insurance law was amended to cover a considerable group of federal workers (mainly temporary employees) who were not covered by the civil service retirement system. The civil service retirement system is financed by contributions of 6½% of the employee's total wage or salary. The government made no specific contribution but assumed responsibility for any deficiencies in the fund. The benefits were based on the employee's salary and length of service. They were in most cases much larger than those paid under federal old-age and survivors insurance, although civil service benefits might be lower for employees with short periods of service.

Health Insurance (Medicare).—The Social Security amendments of 1965 established two coordinated health insurance programs for the aged (65 or older) and were popularly known as "Medicare." The basic plan, which went into effect in July 1966, covered hospital and related services and was financed by increasing the social security taxes paid by employers, employees, and the self-employed. The hospital insurance taxes began at 7/10 of one percent (half paid by the employer and half by the employee) and were scheduled to reach 1.6% in 1987. The cost of benefits for those not covered by social security was met out of general revenues. The voluntary supplementary plan paid for physicians' services and was financed by a $3 monthly contribution, beginning in July 1966, by each individual 65 years of age or older and an equal contribution by the federal government from general revenues.

The basic plan provided for payment of most of the aged individual's hospital costs for up to 90 days of care for each spell of illness; the patient was required to pay the first $40 of his hospital expenses and $10 a day after the first 60 days. The plan also provided (effective in 1967) for up to 100 days of extended-care services in a nursing home (after 3 or more days of hospitalization) with the beneficiary paying $5 for each day beyond the 20th. In addition, the plan provided up to 100 health services (such as a visit by a nurse) to a patient at home following discharge from a hospital or nursing home, and outpatient hospital diagnostic services with the patient paying the first $20 and 20% of the cost of services from the same hospital during a 20-day period. Inpatient psychiatric services were covered, but with a lifetime limit of 190 days of care.

The voluntary supplementary plan was open to persons 65 or older who chose to enroll in it. It provided for paying doctor bills and related medical expenses and was administered through private insurance firms. Under this plan the patient paid $50 of his annual medical expenses and the government paid 80% of the patient's bill (above the $50) for the following services: physicians' and surgeons' services, whether furnished in a hospital, clinic, office, home, or elsewhere; home health services (with no requirement of prior hospitalization) for up to 100 visits during each calendar year; X-rays; laboratory tests; surgical dressings; wheelchairs; artificial legs, etc. Outside-the-hospital treatment of mental illness was limited during any calendar year to $250 or 50% of the expenses, whichever is smaller.

(A. J. A.; E. H. Hn.; W. L. M.; Wr. J. C.)

SOCIAL INSURANCE IN GREAT BRITAIN AND THE COMMONWEALTH

A comprehensive new system of social insurance in Great Britain came into operation on July 5, 1948, as a result of the report of Sir William (later Lord) Beveridge. It greatly expanded and integrated the various social insurance programs that had been adopted during the preceding 40 years. The statutes under which the various parts of the scheme were initiated were the National Insurance Act, 1946, the National Health Service Act, 1946, and the National Assistance Act, 1948. The cost was to be shared by insured persons, employers, and the state. (See *Contributions*, below.) In general, every person in Great Britain over school-leaving age and under pensionable age became insured as a member of one of the following classes: (1) employed persons; (2) self-employed persons; and (3) nonemployed persons. Married women might choose whether or not to contribute under special conditions applicable to them.

Benefits.—The rates of benefit were generally similar regardless of the basis for need (unemployment, sickness, widowhood, or old age). Under the National Insurance Act, 1964, for an adult (over the age of 18), other than a married woman, the weekly rate was £4, with an allowance for one child of *22s. 6d.*, 14*s.* 6*d.* for each additional child and 50*s.* for an adult dependent. The benefits for children were in addition to the family allowance payable for other than the first child. A married woman over 18 received 55*s.* unless she was supporting an invalid husband or not residing with and unable to obtain any financial assistance from her husband, when she was paid the full £4 rate. Boys and girls under 18 without dependents received 45*s.* 6*d.*

Unemployment Benefit.—To qualify for unemployment benefit a person had to be available for employment and to have paid at least 26 contributions. The standard rate of benefit was payable during a benefit year if in the previous contribution year 50 contributions had been paid or credited. Unemployment benefit was payable for 180 days, plus additional days calculated on the claimant's recent record of contributions and benefit. After that he would not be entitled to further benefit until he had been back at work for 13 weeks and had paid 13 contributions.

Sickness Benefit.—To qualify for sickness benefit a person must have paid 26 contributions since entry into insurance. The standard rate of benefit was payable during a benefit year if in the previous contribution year 50 contributions had been paid or credited. If less than 50 contributions had been paid or credited, benefit was payable at a reduced rate.

Maternity Benefits.—These consisted of maternity grants and maternity allowances. A maternity grant of £22 was payable for each baby born if either the mother or her husband had paid 26 contributions before the date of confinement and had at least 26 contributions paid or credited in the last complete contribution year before the benefit year in which the confinement took place. A maternity allowance of £4 weekly was payable to an employed or self-employed woman who gave up paid work for the period for which the allowance was granted. This was normally a period of 18 weeks beginning with the 11th week before the expected week of confinement.

Widows' Benefits.—There were three kinds of widows' benefits, each being dependent upon the husband's contributions only: (1) widow's allowance; (2) widowed mother's allowance; (3) widow's pension. The widow's allowance of £5 12*s.* 6*d.* was payable for 13 weeks together with allowances for children (in addition to family allowance). The allowance was not payable if the widow was over 60 and her husband was entitled to a retirement pension, as she would then be entitled to a pension with her husband.

If the widow had a family which included a child under school-leaving age she would receive a widowed mother's allowance of £4 per week when she had finished drawing her widow's allowance, with an extra allowance for each child. In certain circumstances a widow could obtain a widow's pension at the rate of £4 a week when her other allowance ceased.

Retirement Pension.—The expression "old-age pension," for which provision was made in previous statutes, was replaced by "retirement pension." This was paid at a rate of £4 a week to an insured man aged 65 or over or to an insured woman aged 60 or over who had retired from regular employment and satisfied certain prescribed conditions.

Where a wife received a retirement pension on her husband's insurance the rate was 50*s.* An allowance of 22*s.* 6*d.* was payable to a retirement pensioner for his eldest dependent child and 14*s.* 6*d.* each for younger children.

"Retirement" as understood under the act did not necessarily mean that the person had ceased to do any paid work. He could be treated as retired although he was engaged in a gainful occupa-

tion provided that such occupation was only occasional or inconsiderable in extent or otherwise not inconsistent with retirement. The amount a retirement pensioner could earn without reduction of his pension was limited. At the age of 70 (65 for women) the full rate of retirement pension then earned would be paid though the pensioner might continue to work. There were certain advantages if a person remained at work after pensionable age. He could still obtain unemployment and sickness benefit within certain limits up to the age of 70 (65 for women) and could also earn a higher rate of pension. For every 12 contributions paid as an employed or self-employed person during the five years after reaching pension age, the retirement pension would be increased by 1*s.* a week. The rate of the wife's retirement pension would also be increased.

Death Grant.—A death grant to help meet funeral expenses was payable on the death of certain insured persons, or on the death of the wife, widow, husband, or child of their family, at a rate of £25 for an adult and smaller sums for persons under 18 years of age. Death grants were not payable for men born before July 5, 1883, or for women born before July 5, 1888, and they were payable only at a reduced rate of £12 10*s.* for men born between July 5, 1888, and July 5, 1898. The reasons for these disqualifications was that death grants had not been provided by the state insurance programs prior to July 5, 1948.

Contributions.—The weekly contribution paid by an adult employed man under the National Insurance Act, 1964 (including, provided that he was earning at least £9 a week, an additional contribution, graduated according to his earnings, entitling him to a higher flat-rate retirement pension), was 13*s.* 8*d.*; that of an adult employed woman was 11*s.* 5*d.* This contribution was reduced to 9*s.* 5*d.* for a man and 7*s.* 8*d.* for a woman earning £5 a week or less. Boys and girls under 18 years of age contributed 9*s.* and 7*s.* 6*d.* respectively. The corresponding employers' contributions were: 12*s.* 11*d.* for each man and 11*s.* 2*d.* for each woman, increased to 17*s.* 2*d.* and 14*s.* 11*d.* respectively when the employee's remuneration was £5 a week or less. Employers paid 9*s.* for boys and 7*s.* 6*d.* for girls. For self-employed persons the rates were: men, 18*s.* 8*d.*; women, 15*s.* 4*d.*; boys, 10*s.* 6*d*; girls, 9*s.* For nonemployed persons the rates were 14*s.* 11*d.* for men between 18 and 65 years of age and 11*s.* 7*d.* for women between 18 and 60 years of age; and for boys and girls under 18, 8*s.* 5*d.* and 6*s.* 10*d.*, respectively.

National Assistance.—The National Assistance Act, 1948, provided a unified national system of financial assistance according to need in place of various services provided previously by the state or local authorities. It was placed under the administration of the National Assistance Board, with offices throughout the country. It was the duty of the board to assist persons who were without resources or whose resources (including benefits receivable under the National Insurance Acts) were insufficient to meet their requirements. The question of whether a person was in need of assistance and the nature and extent of any assistance to be given were decided by a local officer of the board in conformity with general regulations made by the minister of national insurance, but provision was made for consulting a local advisory committee in special cases. The Board's officers themselves, however, used considerable discretion in applying the regulations. There was special provision for blind persons and persons who suffered a loss of income in order to undergo treatment for pulmonary tuberculosis. The cost of national assistance was met out of national taxation.

Noncontributory Old-Age Pensions.—Noncontributory old-age pensions, as distinct from contributory pensions, were payable under the Old Age Pensions Act, 1936, subject to a means test. They were a direct charge on the national exchequer and did not come out of the insurance fund. They applied to a diminishing number of old people who had not become entitled to retirement pensions by contributing under the national insurance acts. Unlike retirement pensions, which were payable at 65, a noncontributory pension was not payable until the age of 70.

Workmen's Compensation.—Provision was made under the National Insurance (Industrial Injuries) Act, 1946, for the compulsory insurance of all persons engaged in insurable employment against personal injury caused by accidents, or the contraction of certain industrial diseases arising out of or in the course of such employment. The provisions of the act therefore replaced the former provision for the payment of workmen's compensation under the various workmen's compensation acts, the first of which was passed in 1897. (*See* also WORKMEN'S COMPENSATION.)

Health Services.—Various health services without direct payment were made available to the whole population under the National Health Service Act, 1946. These included: (1) hospital and specialist services; (2) certain domiciliary services provided by local health authorities; and (3) general medical, dental, pharmaceutical, and ophthalmic services.

Everyone became entitled to free medical services and had the free choice of a medical practitioner from whom treatment might be obtained, provided the medical practitioner had agreed to serve under the program. The doctor was paid a fee for each person on his list, together with certain additional allowances. Medicine prescribed by the doctor could be obtained free from any pharmacist taking part in the service. Dental treatment, for which a fixed charge was imposed later, was also available from dentists participating in the service. Their remuneration was paid according to scales of fees for various types of work. General responsibility for the National Health Service was vested in the minister of health, who acted through regional hospital boards and hospital management committees; the domiciliary service was administered by the county and county borough councils. (*See* MEDICAL CARE, GOVERNMENT.)

National Insurance Administration and Finance.—The minister of national insurance, later designated the minister of pensions and national insurance, was made responsible for the administration of the program. His department worked through regional and local offices in close association with the local employment offices of the Ministry of Labour and National Health Service. A national insurance advisory committee was established to assist the minister in discharging his functions under the act with the special task of considering any draft regulations before their submission to Parliament. There were also local advisory committees throughout the country which assisted local officers of the ministry. Contributions were paid by the purchase of special stamps at any post office; one of these stamps was to be affixed to the insured person's contribution card for each week of employment but special arrangements could be made for stamping in bulk. When the national insurance plan came into operation the invested assets of the former health, unemployment, and pensions plans, amounting to about £900,000,000, were taken over. Of this, £100,000,000 was credited to the national insurance fund (under the control of the minister) to meet the payment of insurance benefits and the cost of administration. Into this fund was to be paid the money received from the sale of insurance stamps. The remaining £800,000,000 of the old funds was used to constitute a reserve fund. The interest on the investment of this fund was to be paid into the main fund. In addition to bearing the cost of benefits under the national insurance plan, contributions had to be made from the fund toward the cost of the National Health Service.

British Commonwealth Systems.—At mid-20th century schemes of social security were operating in most Commonwealth countries to a varying degree. In Australia there was no contributory insurance program, but social security services were provided by the Commonwealth Department of Social Services and the cost was met by the exchequer. Maternity allowances and child endowment were paid without a means test. For unemployment and sickness benefit there was a means test on income (although not on property); and for widows' pensions, old-age pensions and invalid pensions (except for blind persons) there was a means test on both property and income. In New Zealand the general social security benefits were also subject to an income test, but legislation provided for the gradual elimination of the means test for old-age pensions by the substitution of a superannuation benefit available to persons aged 65 and over who satisfied certain residential qualifications. A person in

receipt of any other benefit under the Social Security Act, 1938, except family benefit, was not entitled to superannuation benefit. The rates for this benefit (consolidated by the Superannuation Act, 1964) were £4 16*s.* weekly for a husband, the same for his wife (provided she had reached the age of 65), and £5 6*s.* for a single person.

In Canada the first social security plan, established in 1940, provided only for unemployment insurance. In 1952 a comprehensive federal scheme was introduced covering old age, disability, and workmen's compensation. Public assistance was provided by the provinces. Old-age pensions were financed through a 3% sales tax, a 3% tax on net corporation income, and subject to a maximum limit, a 4% tax on individual incomes.

See also SOCIAL WELFARE and references under "Social Security" in the Index. (JN. M.)

BIBLIOGRAPHY.—*United States:* U.S. Department of Health, Education, and Welfare, Social Security Administration, *Social Security Bulletin* (monthly) gives regular statistics; the Department's *Basic Readings in Social Security* (1960) is a bibliography. William Haber and Wilbur J. Cohen (eds.), *Social Security: Programs, Problems, and Policies* (1960) is a compilation of readings. Two textbook treatments are Eveline M. Burns, *Social Security and Public Policy* (1956); John G. Turnbull *et al.*, *Economic and Social Security: Public and Private Measures Against Economic Insecurity*, 2nd ed. (1962). Wayne Vasey, *Government and Social Welfare* (1958) discusses roles of federal, state, and local governments in administering welfare services. *See* also articles in *Social Work Year Book* (triennially).

Great Britain and the Commonwealth: Statutes, regulations, and orders relating to national insurance published by HMSO; explanatory leaflets issued by the Ministry of Pensions and National Insurance. Eleanor Rathbone, *The Case for Family Allowances* (1940); *Britain: an Official Handbook* (annual); *International Labour Review* (monthly); *International Social Security Association Bulletin* (monthly); U.S. Department of Health, Education, and Welfare, *Social Security Programs Throughout the World* (1961); International Labour Office, *Social Aspects of European Economic Co-operation* (1956), *Industry and Labour* (1961); Commonwealth Department of Social Services, *Commonwealth Social Services* (1961); official yearbooks of individual countries.

SOCIAL SERVICE is organized philanthropic activity to promote human welfare. Its objective is to help those who, because of personal factors or forces within their cultural, social, and economic environment, are prevented from realizing their fullest potential. As a professional discipline and social institution, social service has emerged out of centuries of man's humanitarian efforts to help his fellow man. For the history of these efforts in the form of personal and religious charity *see* PHILANTHROPY. For the recognition by the Elizabethan Poor Law of 1601 of public as well as moral and religious responsibility to aid those in need *see* POOR LAW. For the development of modern programs *see* SOCIAL WELFARE and CHILD WELFARE. *See* also POVERTY; SOCIAL SECURITY; WELFARE STATE.

Following a brief survey of the historical background of social service, this article deals with its development as a professional discipline and the emergence of the profession of social work.

HISTORICAL BACKGROUND

The desire of people to help one another in times of need and stress is deeply ingrained in the human race. The helping hand of neighbour to neighbour, the sporadic impulse to help as expressed in the dropping of a coin in an outstretched hand, the charitable acts which characterize certain seasons of the year all reflect man's humanity to man. Throughout the ages and in all parts of the world, these humanitarian impulses have marked the beginnings of social service.

Personal Charity.—Mutual aid and individual charity can be considered early social service. The giving of alms was accepted as a religious duty or as a self-imposed social obligation of the privileged toward the underprivileged. In some countries, particularly those influenced by Islamic traditions of mutual aid and neighbourly love, personal charity continues as a primary motivating force and predominant charity method. Recognition of deficiencies in almsgiving came early to the countries of the Western world. As early as the 16th century, Juan Luis Vives in Belgium and, a little later, St. Vincent de Paul in France sought to relieve destitution through orderly methods of investigation of the needs and problems of the poor. The Elizabethan Poor Law of 1601 in England established the principle of public responsibility for those in need, a principle which was adopted by the American colonists who carried the English poor law to the new land.

Organized Charity.—Changing economic and social conditions in 19th-century England and America brought new problems of massive poverty and, also, new approaches toward solutions of the problems of destitution. These attempts flowed from compassion and paternalism, strongly tinged with a desire to protect society from the risks and consequences of pauperism. The Rev. Thomas Chalmers in Glasgow saw in public poor relief a permanent obstacle to the development of thrift, industry, and temperance in the poorer classes. In 1819, he put forward in St. John's parish his proposals to solve the problem of poverty. His program, which combined direct personal service to the poor with careful investigation of need and resources, demonstrated to the satisfaction of many the power and superiority of private charity. Chalmers' work did not long survive his departure from Glasgow, but the ideas he espoused found fertile soil in the charity organization movement, fostered by Edward Denison, Octavia Hill, and other leading exponents of private charity.

The London Charity Organisation Society, founded in 1869, became a potent force under the direction of Charles Stewart Loch who served as secretary to the council of the society from 1875 to 1914. It brought to the forefront a new conception of social service as an organized activity designed to lead individuals and families to self-help and economic independence through thrift, industry, and strength of character. In the 11th edition of *Encyclopædia Britannica* (1910–11), Loch stated the principles and outlined the methods that guided this movement. The principles and methods of modern social service have expanded and changed in many ways, but the roots lie in the charity which Loch described as "a science based on social principles and observation." He pointed out that charity should be organized and given a definite social purpose to which members of the community must be associated if the purpose is to be fulfilled. He stated, further, that this "depends on the recognition of common principles, the adoption of a common method, self-discipline and training, and co-operation." In such ideas and in the use of volunteer friendly visitors, the charity organization movement laid the foundations for the later development of a body of knowledge and a methodology for the profession of social service and for the emergence of the professionally educated social worker.

The charity organization movement found welcome acceptance in the United States. By 1900 there were 150 societies in the U.S. operating under charity organization principles. The charity organization movement contributed significantly to the development of social service for individuals and families and to the aims and methods of community organization.

Other movements in that century also gave birth to programs and ideas that did much to shape social service as it is known today. In the 1830s Samuel Gridley Howe began his work with blind, deaf-mute, and mentally defective children. His successful demonstration of their educability laid the groundwork for increasingly effective training and social service for the handicapped. In the same period the inspired crusade of Dorothea Lynde Dix on behalf of the insane demonstrated that mental illness is a disease capable of treatment and cure. The child-saving movements fostered at mid-19th century by William Pryor Letchworth and Charles Loring Brace to rescue children from the almshouses and the slums gave rise to child-serving agencies and child welfare programs. A strong reform dimension was added to philanthropy at the end of the century by the social settlement movement. Initiated in London by Samuel Barnett at Toynbee Hall, the movement was brought to its fullest flower by Jane Addams at Hull House in Chicago. From it social service derived the beginnings of group work, a community approach, and social action.

Mitigating Misery.—The 19th-century prototypes of the social service worker centred their efforts on mitigating human misery, created and often cruelly neglected by a rapidly changing

social and economic order. The underlying causes of the distress that charity workers and poor law officials were attempting to alleviate received too little attention. In fact, officials' and charity workers' perception of their task reflected content with the society in which they lived. They saw little amiss in holding the destitute responsible for their destitution. Jane Addams (*q.v.*) summed up the prevailing philosophy of the times in these words:

> Formerly when it was believed that poverty was synonymous with vice and laziness, and that the prosperous man was the righteous man, charity was administered harshly with a good conscience; for the charitable agent really blamed the individual for his poverty, and the very fact of his own superior prosperity gave him a certain consciousness of superior morality. ("The Subtle Problems of Charity," *The Atlantic Monthly*, 83:163–178, 1899.)

The palliative and protective aspects of social service have by no means disappeared, even in affluent societies. The services that are rendered and the views about relationships that should exist between the giver and the beneficiary have undergone great change. In some countries, notably in the less-developed regions of the world, social service may still focus on economic dependency and the conditions associated with it. The problems in those countries may be similar to many of the problems of an earlier period in the Western world. To Western experience, the countries of Asia, Africa, and Latin America are adding their own evolving experience in raising the levels of living of predominantly rural populations. Indigenous methods, such as community development, can provide new solutions to the age-old problem of poverty, enlarging social service experience throughout the world.

Prevention and Social Reform.—A marked shift in emphasis in the conception of social service became apparent at the turn of the 20th century. Passive adjustment to the status quo had begun to arouse vague feelings of discomfort in the charity workers and clearly did not satisfy the Hull House reformers who pioneered for social justice in the late 19th century and the opening decades of the 20th. (*See* also HULL HOUSE; SOCIAL SETTLEMENTS.)

Crusading reformers were not new in social service. In other countries and in earlier periods there were men like Robert Owen and Lord Shaftesbury who battled for factory and child-labour legislation. Crusaders like Elizabeth Fry and John Howard devoted their lives to the reform of the prison system. Beatrice and Sidney Webb, in London, shared a vision of a better life; Jacob Riis, in New York, worked to wipe out the slums; other settlement houses were hard at work in New York, Boston, Pittsburgh, Buffalo, and Cleveland. But in that small group in Chicago there was a combined force and intensity that had never before in social service been mustered so successfully in the interests of the common man.

Prevention was the byword. Through slum clearance, decent wages and working conditions, abolition of child labour, juvenile courts for child offenders, mothers' pensions, and similar means they sought to remove the causes of economic distress and social breakdown. Their leadership associated social service with the emerging liberalism of the times. Private charity was not always sympathetic to their goals, particularly those that involved public support and administration of social service programs, but the tide was running in favour of social reform.

In the United States the first White House Conference on the Care of Dependent Children, called by Pres. Theodore Roosevelt in 1909, paved the way for the protection and care of dependent children in their own homes through publicly supported and administered mothers' pensions; the first laws were passed in Illinois and Missouri in 1911. With that step and with the earlier enactment of the first workmen's compensation (*q.v.*) law, a new and better future was charted for the public social services. The road was open to the later development of the social insurance, public assistance, and public child-welfare measures embodied in the Social Security Act of 1935. (*See* further SOCIAL SECURITY.)

Methods of Work.—The paternalistic and moralistic charitable work of the 19th century would be repugnant to the modern social worker. The early work contributed concepts that are still of value, however: the concept that each person and his needs represented a unique situation; helping people to help themselves; the value of personal relationship expressed in Octavia Hill's phrase "not alms, but a friend"; and friendly visiting, early confirmed as a key method of helping through personal influence. The importance of knowledge of the person in his social situation was demonstrated in procedures for investigation and in painstaking assemblage of information. When the experience of that period was formulated by Mary Richmond in *Social Diagnosis* (1917), it was clear that social service had acquired the professional method of social casework (*see* below).

Organized Agencies.—Charitable and philanthropic work was represented, in the early decades of the 20th century, by many agencies, programs, and services under a variety of auspices, governmental and nongovernmental, sectarian and nonsectarian. They included child welfare, family welfare, public welfare, juvenile courts, probation and parole service, medical social service, school social service, mental health, programs for the handicapped and other special groups such as the aged, settlement houses, and leisure-time services.

Community Organization.—An early aim of the charity organization movement had been the coordination of charitable efforts to eliminate duplication and overlapping. The larger coordinating effort was later assumed by councils and federations of charities and philanthropies that came into being early in the 20th century. Social planning and federated fund raising were the features of these agencies. Their experience in organizing communities to create and support social welfare programs gave to social service the professional method of community organization.

The reform movement of the early 20th century highlighted social action. From their day-to-day exposure to the consequences of a faulty social order, the settlement house workers and, increasingly, the charity workers assumed responsibility for assessing causes and proposing remedies. Social research and social action followed naturally the shift in attention from deficiencies in the poor to insufficiencies in their environment.

The Workers.—Charity workers were a mixed lot but distinguishable by the end of the 19th century. Clergymen of many faiths, volunteer friendly visitors, wealthy and influential citizens on agency boards and state boards of charities and corrections, overseers of the poor, and paid charity agents were associated in a common effort. But as yet there was no bond of common education or united purpose to bind them together.

Inspired by the teaching and personal influence of Francis G. Peabody, young graduates of Harvard University made a career of charitable work; so, also, did men past middle age; and many women who started as volunteers remained as paid workers. This heterogeneous group founded the National Conference of Charities and Corrections, now the National Conference on Social Welfare. The paid workers of Boston and Philadelphia early met in "Monday" clubs for the purposes of professional association.

The expanding charity organization societies early faced the problem of training volunteer visitors and paid agents. A number of British and American pioneers in philanthropic work were able economists and social philosophers who left a considerable volume of reports and recorded experience. Their commitment to "scientific charity" produced theories and principles that were available to the neophyte. Mary Richmond was one of the first to proclaim that good intentions and common sense were not enough for charity work. She saw the limitations of apprenticeship programs that had developed in both Britain and the United States. With her successful proposal in 1897 that a training school of applied philanthropy be established, charity work took a giant step in the direction of social service as a profession.

EMERGENCE OF SOCIAL SERVICE AS A PROFESSION

Mary Richmond's call for special preparation for charity work met with immediate response. The New York Charity Organization Society launched a summer school of philanthropic work in 1898. A more significant step was taken in 1899 when an Institute for Social Work Training was established in Amsterdam and offered a two-year course combining study of general sociological knowledge, socio-economic problems, and legislation with supervised practical training in various fields of social work. By 1903

the training activities of the London Charity Organization Society had led to the development of a two-year program of theory and practice in a formal School of Sociology. In 1904 the New York Charity Organization Society offered a one-year program in the New York School of Philanthropy, the first school of social work in the United States (now the Columbia University School of Social Work). Boston followed New York in 1904, Chicago and Berlin in 1908. The movement spread rapidly within Europe, throughout the United States and by 1920 it had reached Latin America, where the first school was established in Santiago, Chile.

Charitable and philanthropic work had become generally known as social work or social service; its personnel were identified and designated as social workers. The road to professional status was open but rugged. In 1915 Abraham Flexner disturbed the growing confidence of social service in its professional status with his conclusion that it met some but not all of the criteria of a profession. He noted particularly that it did not have its own "technique communicable by an educational process."

A significant milestone was reached with the appearance in 1917 of *Social Diagnosis*, Mary Richmond's outstanding contribution to social service as a professional discipline. The systematic method of social study and diagnosis which she outlined gave to social service distinctive processes and techniques thereafter considered communicable in formal education programs.

With the entry of the United States into World War I, social service was faced with assisting servicemen's families in need and, later, servicemen under psychiatric care. At the same time psychiatry made a new and tremendous impact upon social service. Virginia Robinson's *Changing Psychology in Social Case Work* (1930) cited the major contributions of psychoanalytic psychology to social service: (1) the concept of determinism in psychic life; (2) emphasis on the need (as opposed to the intellectual) basis of human behaviour; and (3) insight into the effect of family relationships on individual development. To these should be added a revolutionary new concept of the meaning and professional significance of the relationship between the social worker and his client as the medium through which change in the person and his social situation could be effected.

Mary Richmond's formulations represented a synthesis of the knowledge of her time and provided primarily a sociological basis for the practice of social casework. The family in its social environment lay at the centre of her concern. Psychoanalysis supplied a psychological basis that, for a number of years, centred attention on the individual, almost to the exclusion of all else. In the depression of the 1930s, social service was brought back to its primary concern with man interacting in his social situation. It had by that time emerged as a professional discipline with its own communicable methods of practice.

Awareness of professional identity led to the establishment of national professional associations. The schools of social work in the United States and Canada launched in 1919 a central standard-setting and service organization, now the Council on Social Work Education. As early as 1918 American practitioners came together in various associations and, in 1921, they established the American Association of Social Workers, the major predecessor of what is now the National Association of Social Workers.

Knowledge, Methods, and Scope.—United Nations surveys of training for social work indicate a large measure of international agreement on the basic body of knowledge and the skills required by social workers. The humanities and the social and behavioural sciences contribute to the social worker's general understanding of man and his environment. From medicine (particularly psychiatry), psychology, law, and education, knowledge is borrowed and reshaped to the purposes of social work. Increasingly, however, the social work profession is developing its own distinctive body of knowledge.

As already noted, casework early emerged as a professional method in social work practice. Mary Richmond's first formulation (1915) stated simply that "Casework is the art of doing different things for and with different people by co-operating with them to achieve at one and the same time their own and society's betterment." The many definitions that followed have clarified and refined the meaning of casework as a complex helping and problem-solving process, but all contain certain essential ingredients: (1) the social worker's commitment to the worth of the individual and his belief in man's capacity for growth and change; (2) the artistic use of knowledge in a professional relationship in which the individual is helped to cope more effectively with his problems; and (3) effective use, in the problem-solving endeavour, of all relevant resources within the social environment and within the person seeking help. Social workers also may lead groups to help in individual and group achievement. Community organization as a professional social work method shares with casework a common ancestry in the charity organization movement. At one time seen largely as a process of adjustment between community resources and community welfare needs, community organization has begun to emerge as a means of helping the community to engage in a systematic process of planned change toward community improvement.

The West, particularly the United States and Canada, pioneered in the delineation of the professional social work methods of casework, group work, and community organization. All countries which recognize social service as a professional discipline have used one or another or all three of the methods in translating social goals into direct service to people. Dame Eileen Younghusband of the United Kingdom noted in the third United Nations survey of social work training:

> The conclusion would seem to be that in casework, group work and community organization and certain aspects of community development, as these have evolved in different cultural settings, some fundamentals of working with people have emerged which hold true under all circumstances. (United Nations, *Training for Social Work: Third International Survey* (1959, p. 95.)

Early concern with broad social reform and later emphasis upon individual treatment are more and more becoming fused into a unified effort to induce social as well as individual change, to prevent social problems as well as to alleviate their end results, to affect the destiny of large population groups as well as to help individuals and families and small groups. This transitional state is also reflected in new attempts to synthesize sociological and psychological knowledge and to undergird professional practice with social work's own developing theory.

The emerging new conception of social service as a force and instrument in the promotion of planned social change enlarges the scope of professional social work activity, which, traditionally, has been associated with such fields of practice as child and family welfare, medical and psychiatric social work, school social work, corrections, group services, community welfare planning, industrial social work, and various forms of social or public assistance. Social workers have provided protective, palliative, and rehabilitative (and, to a lesser extent, preventive) services for the disadvantaged, the disabled, the displaced, the unemployed, the mentally, emotionally, and physically ill, the criminal and the socially irresponsible, the deprived young and the neglected old, and all the ugly results of social disorganization and social injustice. To these significant age-old tasks, the social work profession, in the view of practitioners, should now add newer responsibilities of joining with other disciplines to attack mass poverty at its source, to seek out and prevent the problems that blight the lives of people in affluent as well as struggling new societies, and even, it has been said, to reshape social structures in directions thought to provide new and better opportunities for productive and satisfying living. A United Nations survey of programs of social development disclosed that within the first ten years after World War II responsibility for social service as an instrument of social development had been proclaimed in new constitutions or introduced, by amendments, into the former constitutions of 45 countries.

Social Service Personnel and International Organization.—Nations old and new share the goal of staffing their social services with professional personnel, but in no country are there enough qualified workers to meet the need. The United Nations early recognized that steps must be taken through technical assistance and other means to promote social work education if social welfare

policy is to be translated into effective social services. International fellowships, consultant services, regional seminars, and special studies and training projects have been introduced.

Approximately 300 schools of social work were in operation in over 40 countries in 1950 when the United Nations made its first international survey of training for social work. Fifteen years later, more than 400 schools were preparing social workers in more than 50 countries and the totals were increasing every year. Most of the growth occurred in Asia and the Middle East.

The pattern in the United States and Canada places social work education as a graduate professional discipline within universities, which grant a master's degree upon completion of a two-year program. An increasing number of graduate schools of social work in the United States also offer programs leading to the doctoral degree. In other parts of the world, schools of social work may be organized within or in relation to universities (usually at the level of undergraduate study) or as independent educational institutions under government, religious, or secular auspices. The university pattern of social work education appears to be finding increasing acceptance in Asia and some parts of Latin America and Africa, where schools of social work are relatively new.

The International Conference of Social Work brings together in a biennial world forum persons concerned with social welfare problems. The International Association of Schools of Social Work sponsors a biennial congress and a range of activities designed to promote high standards of professional education. The International Federation of Social Workers brings together the practitioners affiliated with national associations of social workers. Each of these organizations has been granted consultative status by the Economic and Social Council of the United Nations. Together, they cooperate in many areas, including the publication of an official journal, *International Social Work*.

Social service is influenced in the countries of the world by social, political, economic, and cultural forces that vary greatly from country to country. It is also influenced by an awareness that peoples everywhere share common needs and aspirations.

BIBLIOGRAPHY.—Edith Abbott, *Social Welfare and Professional Education* (1942); Jane Addams, *Twenty Years at Hull-House: with Autobiographical Notes* (1910); Frank J. Bruno, *Trends in Social Work, 1874–1956: a History Based on the Proceedings of the National Conference on Social Work* (1957); Clarke A. Chambers, *Seedtime of Reform: American Social Service and Social Action* (1963); Nathan E. Cohen (ed.), *Social Work and Social Problems* (1964); Grace L. Coyle, *Group Experience and Democratic Values* (1948); Edward T. Devine, *When Social Work Was Young* (1939); Alfred J. Kahn (ed.), *Issues in American Social Work* (1959); Porter R. Lee, *Social Work as Cause and Function* (1937); Helen Harris Perlman, *Social Casework, a Problem-Solving Process* (1957); Helen U. Phillips, *Essentials of Social Group Work Skill* (1957); Norman A. Polansky (ed.), *Social Work Research* (1960); Kenneth L. M. Pray, *Social Work in a Revolutionary Age* (1949); Ralph E. and Muriel W. Pumphrey, (eds.), *The Heritage of American Social Work: Readings in Its Philosophical and Institutional Development* (1961); G. A. Rennison, *Man on His Own—Social Work and Industrial Society* (1962); Mary E. Richmond, *Social Diagnosis* (1917); Virginia Robinson, *A Changing Psychology in Social Case Work* (1930); Murray G. Ross, *Community Organization: Theory and Principles* (1955); United Nations, Bureau of Social Affairs, *Social Progress Through Community Development* (1955), Department of Social Affairs, *Training for Social Work: an International Survey* (1950), Department of Economic and Social Affairs, *Training for Social Work: Third International Survey* (1959), Educational, Scientific and Cultural Organization, *The Contribution of Social Sciences in Social Work Training* (1961); Eileen L. Younghusband, *Social Work and Social Change* (1964); National Association of Social Workers, *Encyclopedia of Social Work* (1965); David Owen, *English Philanthropy 1660–1960* (1964); Marie-Hélène de Bousquet, *Le Service Social* (1965).

Periodicals: Informations Sociales, Revue Mensuelle de l'action sociale, L'Union Nationale des Caisses d'Allocations Familiales, Paris; *International Social Service Review,* United Nations, Department of Economic and Social Affairs, New York; *International Social Work* (quarterly), International Conference of Social Work, Bombay, India; *Social Casework* (October–July), Family Service Association of America, New York; *Social Service Quarterly,* National Council of Social Service, London; *Social Service Review* (quarterly), University of Chicago Press, Chicago; *Social Work* (quarterly), National Association of Social Workers, New York; *Journal of Education for Social Work,* Council on Social Work Education, New York (biannual); *Social Work,* a quarterly review of family casework, Council of the Family Welfare Association, London; *Social Worker* (quarterly), Canadian Association of Social Workers, Ottawa (English and French). (K. A. K.)

SOCIAL SETTLEMENTS are neighbourhood social welfare agencies, also called settlement houses, community centres, or neighbourhood houses, centres, associations, or guilds. In Great Britain there are both residential settlements, where members of the staff live as well as work, and community centres founded by city and borough councils. Settlements, under a variety of names, are to be found in most countries of western Europe and in Southeast Asia and Japan.

History.—The settlement movement began with the founding of Toynbee Hall (*q.v.*) in London in 1884. Samuel A. Barnett (*q.v.*), then vicar of St. Jude's Parish, invited a number of university students to join him and his wife in "settling" in a deprived area. The movement spread to the United States when Stanton Coit, an early visitor to Toynbee Hall, established Neighborhood Guild, now University Settlement, on the Lower East Side of New York City in 1886. In 1889 Jane Addams, another Toynbee Hall visitor, and Ellen Gates Starr bought a residence in the Near West Side of Chicago, and started Hull House (*q.v.*). In the same year Jane E. Robbins and Jean Fine (Mrs. Charles B. Spahr) opened the College Settlement in New York City. Two years later Robert A. Woods established Andover House, later called South End House, in Boston. The movement spread to other countries mainly through the influence of visitors from those countries to settlements in England or the United States.

Purposes.—The purpose of a settlement is the development and improvement of a neighbourhood or a cluster of neighbourhoods. It differs from other social agencies in that it is concerned about neighbourhood life as a whole. In a sense the neighbourhood is its client. Settlements seek to develop harmonious relationships among community groups of different cultural, economic, religious, and social characteristics. While they may be sponsored and supported by a religious organization, their purpose is to serve all kinds of people who live in the neighbourhood.

Settlements try to help people act together to improve their living conditions and environment. Under the leadership of pioneer workers such as Lillian D. Wald (*q.v.*) at the Henry Street Settlement in New York City, Jane Addams, and the many persons who served as residents and volunteer leaders (Joseph Moss and Sophonisba Breckinridge, to mention only two from Hull House, for example), social settlements helped win legislation providing for juvenile courts, mothers' pensions, workmen's compensation, and the regulation of child labour; they promoted playgrounds and recreation centres, maternal, infant, and child health clinics and services, and adult education programs, and took part in housing and other reforms.

Services.—The services of settlements, while they vary, are all committed to bringing together people who live in the same geographical area for the purpose of following common interests and of improving neighbourhood life generally. The settlement's program is determined by the needs and interests of the neighbourhood, changing in accordance with changing conditions. One settlement serving two or more neighbourhoods that differ in composition or condition may offer different services to each. In general, services fall under three main categories: work with individuals and families, with groups, and with, or in behalf of, the neighbourhood as a whole.

Work with Individuals and Families.—Informal counseling and home visiting are done by almost every member of a settlement's staff. Where specialized skill is needed, settlements have employed caseworkers, psychologists, psychiatrists, home economists and vocational counselors. In neighbourhoods lacking adequate health services, medical or dental clinics have been provided, and settlements have campaigned for the extension of public health services. Nursery schools and day care for children have been provided for families needing them because of employment of mothers or other reasons.

Work with Groups.—Group work is intended to enrich the social life and experience of the individual members and to provide opportunities for the group as a whole to contribute to neighbourhood improvement. Settlements serve both membership groups sponsored by the settlement and autonomous neighbourhood organizations. Membership groups—including friendship

clubs, classes, athletic teams, and interest or hobby groups—enroll people of all ages, interests, and abilities. The settlement provides a place to meet and a group leader, teacher, or coach, and, for younger groups program materials and equipment. Members join for amusement, and often for personal and social development. The settlement's worker seeks to develop feelings of neighbourliness, mutual respect, and social responsibility. Such group work requires most of the settlement's space and staff time.

Independent neighbourhood organizations and groups often request certain services of the settlement, sometimes the use of a room for meetings, sometimes the guidance of a staff member. Settlements have helped to develop parent-teacher associations in neighbourhood schools; have helped local churches to develop social and recreational youth programs; and have provided guidance in the development of youth, recreational, or welfare programs for labour unions and veterans, civic, and fraternal organizations.

Work with the Neighbourhood.—At Toynbee Hall and at certain settlements in other countries, some professional staff members as well as university students and other volunteer workers live in the settlement's building so that they may become in fact as well as in feeling a part of the neighbourhood. In all settlements, the staff workers are expected to become familiar with neighbourhood people, conditions, and problems and to identify themselves with attempts to improve conditions, working with the people, with governmental officials, and with citizens of the larger community. They encourage and help local individuals and groups to organize and to take action to secure enforcement of housing, safety, and sanitation codes. They make available to the neighbourhood people reliable information about proposed public improvements, such as highways, slum clearance, and urban redevelopment, and encourage them to express their concerns to governmental agencies and officials.

Neighbourhood action may be directed toward legislation as well as to administrative decisions. Settlements seek to acquaint social welfare and other planning bodies with facts about living conditions and social needs in their neighbourhoods, and they participate, through their representatives on community welfare councils or similar groups, in social planning for cities and metropolitan areas. Social settlements are thus agents of social reform.

Administration.—The responsibility for operating a settlement is vested in a board of directors composed of citizens, not all of whom live in the neighbourhood. The board establishes policy and is responsible for securing and administering funds. The administrative and supervisory staff is usually composed of persons with professional education in social work. Workers trained in informal education, nursing, early childhood education, physical education, and recreation also are employed extensively.

In the United States most settlements secure financial support from Community Chests or United Funds. Support may also be provided by endowments, in both the United States and Great Britain; by government agencies, especially in Europe; by local, national, or international religious organizations; by industrial corporations and private individuals; and by charitable trusts and foundations.

Buildings vary greatly. Many settlements have a modest central building but use rooms in buildings owned by housing authorities, public schools, and churches for some of their activities. Clubs and other groups may meet in private homes of members. Leadership, rather than meeting space, is considered the primary contribution of the settlement. However in a congested and deteriorating city neighbourhood, the settlement house often furnishes a hospitable common ground where diverse and transitory groups gain a feeling of security they cannot find elsewhere in the neighbourhood.

The Settlement and the City.—Historically the social settlement was a method of solving problems peculiar to urban living, particularly for the poor. Crowded conditions, preoccupation with one's own urgent problems, and the lack of concern of one's neighbours made difficult the maintenance of neighbourly feeling and responsibility characteristic of rural and village life. In the United States, before the adoption of the Quota Laws of 1921 and 1924, waves of immigrants successively crowding into transitional, depressed, or slum areas faced additional problems of coping with strange customs and a strange language. The settlement attempted to restore under urban conditions some sense of belonging and of caring which seemed necessary in a humane society.

The difficulty of finding financial support in smaller, poorly organized (for purposes of social welfare) communities had tended to confine settlements to metropolitan areas. However, because of the concern of religious organizations particularly, some small agricultural and newer industrial communities have been helped to establish and maintain a community centre, similar to a settlement. Dispersion of industry outside metropolitan areas has brought into being new towns which need this general-purpose type of community service. Communities of migratory agricultural workers, particularly where children are involved, have claimed the attention of religious organizations in the United States, with the result that a settlement type of work has been established.

City, National, and International Organizations.—In the larger cities of the United States and in some cities of Europe the settlements have formed voluntary city federations. In most instances, these have not employed any staff but have provided opportunities for employed workers especially, and for some board members, to discuss and sometimes act together on common problems. The first such federation was formed in Chicago before the turn of the 20th century and the second in New York City shortly thereafter. In a few smaller cities all of the settlements have combined to form one agency which operates several centres in different neighbourhoods.

The founders of early settlements in the United States met periodically, especially with reference to conditions which seemed to require national co-operation and action of the federal government. In 1911 they established the National Federation of Settlements, with Jane Addams as its first president. Robert Woods of South End House, Boston, was the first (unpaid) executive secretary. From 1929 the federation employed a full-time executive secretary to facilitate the co-operation of the settlements on matters of national concern and to help improve standards of neighbourhood work throughout the country. The British Association of Residential Settlements was established for the same purposes. Several European countries also have national organizations of settlements, most of them founded after World War II.

The first International Conference of Settlement Workers, held in London in 1922, led to the organization in 1926 of the International Federation of Settlements. Except for a hiatus immediately before and during World War II, this federation has met every four years. After World War II its headquarters were at the French Association of Social Centres in Paris till 1956, when they were moved to the Dutch Federation in Utrecht. The International Federation is represented by observers at the United Nations.

Bibliography.—Jane Addams, *Forty Years at Hull-House* (1935); J. A. R. Pimlott, *Toynbee Hall; Fifty Years of Social Progress, 1884–1934* (1935); Mary K. Simkhovitch, *Neighborhood: My Story of Greenwich House* (1938); Graham Taylor, *Pioneering on Social Frontiers* (1930), *Chicago Commons Through Forty Years* (1936); Lillian D. Wald, *Windows on Henry Street* (1934); Robert A. Woods, *The Neighborhood in Nation-Building* (1923); Robert A. Woods and Albert J. Kennedy, *Settlement Horizon: a National Estimate* (1922). *See* also Lorene M. Pacey (ed.), *Readings in the Development of Settlement Work* (1950); Fern M. Colborn, *Buildings of Tomorrow: Guide for Planning Settlements and Community Buildings* (1955); National Federation of Settlements and Neighborhood Centers, *Neighborhood Centers Today: Action Programs for a Rapidly Changing World* (1960), *Neighborhood Goals in a Rapidly Changing World* (1958), *Standards for Neighborhood Centers* (1960). (J. McD.)

SOCIAL WELFARE is considered, in this article, as including the attempts made by governments and voluntary organizations to help families and individuals by maintaining incomes at an acceptable level, by providing medical care and public health services, by furthering adequate housing and community development, by providing services to facilitate social adjustment, and by furnishing facilities for recreation. In addition, social welfare includes legislation and facilities designed to protect those who

might be subject to exploitation, and to care for those groups considered to be the responsibility of the community. Educational services, usually considered as social welfare provisions, are covered in other articles (*see* EDUCATION [ARTICLES ON]).

All nations have developed social welfare services as they have moved from agrarian to industrial economies. In preindustrial societies the family group cares for its own young, its own aged, its own sick and disabled; the food and shelter available are distributed among all members; social maladjustment is dealt with by the family. Such a system of mutual responsibility becomes impossible under modern industrial organization, which demands mobility and the congregation of the population in crowded urban centres. It has become necessary, therefore, for society to provide substitutes in the form of social services, subsistence, medical care, and protection. In all countries these have been provided by both private associations and the state.

HISTORY

The early history of social welfare in most western nations is the history of the Poor Law (*q.v.*) and of the private charities and foundations which supplemented it. (*See* also CHILD WELFARE.) Until the latter half of the 19th century, therefore, most social welfare services carried only negative implications. They were intended to assuage abject poverty and suffering; to prevent epidemics of deadly disease; to protect the population against begging and thieving; to appease the lower classes and so prevent revolution and maintain order. Earlier improvement in social welfare services was deterred by two enduring concepts: first, that government should avoid to the greatest extent possible any interference with the economy or social relationships; second, that all was well with the economic and social organization and that poverty or social maladjustment could be attributed to personal inadequacies or moral infirmities.

These two concepts were emphasized during the 18th-century Enlightenment and became most effective as the middle classes obtained political power. The demand for limited government, a reaction to the excesses of monarchies and mercantilism, was, of course, at the heart of the struggle for modern constitutional government. Social welfare services, except those provided by private benevolence, could not be developed as long as the function of government was thought to be limited to the protection of private property and the conduct of foreign affairs. Less theoretical concern about the propriety of governmental responsibility for social welfare services was expressed during the latter part of the 19th and early 20th century, particularly in the United States. There the public services were usually manned by political appointees with no civil-service tradition or security of tenure. Graft and other political chicanery were to be expected.

The second concept had its origins in Judaeo-Christian philosophy and was articulated in secular terms during the 18th century. Recognizing first the supreme importance of the individual, the philosophers of the Enlightenment saw the greatest fulfillment of the individual and therefore the greatest welfare of society in the operation of the free market. Since each individual, according to natural immutable laws, was bound to operate rationally in his own self-interest, any interference with the freedom of an individual to choose, for example, a path which leads to poverty instead of wealth would be in violation of natural law and inimical to the general welfare.

These ideas, which found ready acceptance among the middle classes, were buttressed by the theories of such men as Joseph Townsend, English clergyman, geologist, and author of *A Dissertation on the Poor Laws* (1786), who considered the spectre of poverty and suffering necessary to entice men to work at menial occupations; T. R. Malthus, who considered any aid given to the poor as encouraging them to increase their numbers and so to diminish the subsistence available to the rest of the population; and David Ricardo, who saw in the "wage fund" the impossibility of improving the condition of the labouring classes. Also delaying improvement in social welfare was Protestantism in general and Puritanism in particular, with its emphasis on hard work and frugality and its equating of worldly success with divine election. The acme of this trend in ideas was achieved by the Social Darwinists during the last half of the 19th century. Herbert Spencer, for example, believed that men who advocated social welfare measures were "blind to the fact that under the natural order of things society is constantly excreting its unhealthy, imbecile, slow, vacillating, faithless members." Suffering by some must be endured so that the welfare of the many might eventually be achieved.

Poor Law Reform (1834).—The Poor Law reform of 1834 in England was the most specific expression in legislation of such ideas concerning the nature of poverty and the proper role of the state in its relief. All relief to the able-bodied in their own homes was forbidden, and all who wished to receive aid must move into the workhouse. This law established officially the principle of "less eligibility" which was to plague the development of social welfare in the United States and England for the next hundred years. According to this principle—since poverty was due to individual fault, since any encouragement to laziness or thriftlessness was a threat not only to the proper operation of the economic system but to the very survival of the nation—all efforts should aim at making the application for and receipt of poor relief as odious and degrading as possible. In the United States during the early 19th century no such national statement could have been expected because of the limited powers and responsibilities of the federal government at that time. The Yates Report of 1824 in New York State and the Quincy Report of 1821 in Massachusetts, however, both recommended that all the poor be cared for in public almshouses, the latter stating "that of all the modes of providing for the poor, the most wasteful, the most expensive, and most injurious to their morals and destructive to their industrious habits is that of supply in their own families."

Voluntary Welfare.—The ideas of the Enlightenment, which were so congenial to the middle classes during the 19th century, also had the effect of encouraging voluntary social welfare. While the discipline of *laissez-faire* economics and devotion to natural law impelled the commercial classes to permit suffering and allow poverty to go unrelieved by government, the humanitarian aspects of the Enlightenment emphasized the importance of good works and the benefits of charity to the giver. Benjamin Franklin, for example, was an insistent opponent of all forms of public relief yet was a well-known philanthropist whose time and money had been given to the creation of a large number of welfare organizations. This humanitarianism also caused men of influence to be repelled at the condition of the working classes resulting from the free operation of economic forces. The paradoxical quality of their attitude toward social welfare is illustrated by the fact that in England the first substantial Factory Act, providing for the appointment of factory inspectors and requiring that employed children between the ages of 9 and 13 attend school for two hours each weekday, was passed during the year prior to the restrictive Poor Law reform of 1834.

Lavish expenditures for private charity were possible during the early period of industrialization because of the accumulations of great wealth in the hands of a few individuals. Many hospitals, schools, universities, and charitable foundations were endowed in the era of unrestricted capitalism in England, the United States, and western Europe. The unrestrained giving typical of the 19th century, however, was disturbing to the conscience of the middle-class liberal as it too could pauperize the recipients, discourage thrift, sacrifice, and hard work, and so interfere with the free operation of natural economic law. Their answer was "scientific philanthropy" which aimed at separating the "deserving" from the "undeserving" poor and assisting the former to self-support. In England and the United States this rationalization of benevolence was carried on by the Charity Organization societies.

The humanitarian impulses of liberalism also gradually led to a less rigid attitude toward the supposed evils of public poor relief. While the leaders of private charity were concentrating on saving the "deserving" poor from the evils of the workhouse or almshouse, the public authorities were attempting to make a similar distinction. Special provisions were made first for the mentally ill, then for dependent children. Relief in their own homes was permitted the aged and infirm. By the early 20th century in En-

gland and later in other countries special provisions were instituted for the able-bodied who were unemployed through no personal fault (*see* UNEMPLOYMENT INSURANCE). By categorizing the poor on the basis of the supposed causes of their destitution, it was possible to relieve suffering without moral damage to the recipients of relief and without interfering with the free operation of the economic system.

Social Insurance.—Even more acceptable to the middle classes than categorization of the poor was social insurance. Here was a term and a concept borrowed from the commercial world. Social insurance would place the relief of poverty on a "businesslike" basis. It would encourage thrift and provision for the future and would in no way pauperize the recipients of its benefits. Social insurance in most countries, therefore, developed during the early 20th century and has gradually been expanded to include most common risks and to cover most of the population. Poor relief (*i.e.*, general assistance for the needy who do not qualify for one of the categorical programs for aid to dependent children or to the aged, blind, or disabled), however, has never recovered from the slander of the 19th-century liberals and remains in most places a "residual category."

Credit for the development of modern social welfare services should not be given to the humanitarianism of the middle classes exclusively. The rise to political influence of the urban working class was perhaps more important. It was not possible to convince the labourers that an economic system which brought hardship and suffering to so many of their numbers was what nature intended or was divinely sanctioned. They had experienced or witnessed poverty which could not be attributed to moral infirmities but was the result of economic conditions beyond their control. The evil of government intervention was much less serious to them than the evils of free competition bridled only by monopoly. The impact of the rise of organized labour to a position of political significance can be seen most clearly in the case of England. There the acceleration in the enactment of labour and social welfare legislation coincided with the entrance of the trade unionists into Parliament beginning in 1874.

On the continent of Europe the influence of socialism was probably more effective than either the labour movement or the humanitarianism of the liberals. In the United States the influence of organized labour was not crucial to the furtherance of social welfare measures until the 1930s. The earlier advances in the U.S. can be attributed to agitation by middle-class reformers and intellectuals.

In a sense it can be said that modern social welfare programs are the result of willingness on the part of both the middle classes and the urban workers to abandon their apocalyptic hopes and to attain their ends within the framework of constitutional democracy. The middle classes were the first to move from their original position. As the progress of industry led to larger accumulations of capital, men with small or middle-sized financial resources found it harder and harder to survive in competition with wealthy rivals. In practice it became apparent that the free market was an obstacle to all but a few. Although free competition continued to be considered the basis of society, genuinely competitive conditions called for positive action by the state, for tariff protection and antimonopolistic and other forms of remedial legislation.

By the 20th century most socialists also realized that they would be able to achieve their aims only by reforming the existing system through established institutions, and they too turned to positive state action. The working-class movement had been able to accomplish much improvement in working and living conditions by strikes and other nonpolitical means, but these gains could be consolidated only through legislation. Organized labour had also been able to protect its members to some extent from the risk of sickness, injury, old age, and unemployment by collective bargaining and by mutual benefit societies, but in order to provide wide coverage and adequate benefits it was necessary to turn to compulsory insurance which could be enforced by the taxing power of the state and subsidized from general revenues.

In the United States the movement from the poor laws to modern social welfare provision followed much the same pattern as in England and on the continent but at a much slower pace, probably because the nation was dominated almost exclusively by the middle classes during the 19th and early 20th centuries. Not only were there no groups comparable to the nobility and peasantry of Europe but, in addition, a politically conscious and influential labour movement did not appear until relatively late. Early reforms, therefore, were essentially the product of middle-class activity and were concerned more with the regulation and control of business and with attempts to improve legislative and administrative processes than with the provision of social welfare services. Only with the economic depression of the 1930s and the rise to political influence of the labour movement did public social welfare arrangements begin to equal those of Europe.

The change in attitude toward the poor and other unfortunates and toward the responsibility of the community which occurred in western nations generally between the early 19th century and World War I cannot, of course, be attributed solely to the disillusionment of the middle classes or to the extensions of suffrage which brought the influence of the working classes to bear directly upon public policy. Other forces subtly altered the outlook of the public toward the social ills that beset nations that were being industrialized rapidly. For example, there was the fear of revolution threatened by socialists, communists, and anarchists. Concern for the poor and for the plight of the working classes was encouraged by the papal encyclical *Rerum novarum* (1891) and by the Christian Socialists; in the United States the Social Gospel movement encouraged Protestants to seek a solution to the social problems of the day (*see* LEO: *Leo XIII;* CHRISTIAN SOCIALISM). Improved means of communication and the spread of literacy brought to increasing numbers of people the facts of poverty and suffering and the remedies proposed by militant reformers. Recurrent business depressions began to dispel the idea that all was well with the economic and social system and that poverty could in many cases be attributed to individual fault.

Whatever the causes, the results of this change in attitude can be illustrated by the extent to which social insurance measures had been adopted by western industrialized nations prior to 1914. Laws providing compensation for industrial accidents were in force throughout practically all of Europe, in many of the British dominions and colonies, and in about half of the United States. There were compulsory sickness insurance programs in Germany, Austria, Hungary, Great Britain, and Norway and voluntary insurance was subsidized by government funds in most of the remaining European countries. Compulsory old-age insurance was operating in France and Germany. The first compulsory unemployment insurance system had been instituted in Great Britain, and subsidized voluntary unemployment insurance coverage was available in many large European cities and nationally in Norway and Denmark.

In most countries the progress of social insurance continued after World War I. By 1930 practically the entire industrialized world had workmen's compensation in some form or another. Practically all of Europe had some form of social health insurance. Of the 27 countries in Europe, 23 had either old-age insurance or old-age pensions or both. By 1935, 19 European countries had nationwide unemployment insurance, either compulsory or voluntary. Also in 1935 the United States adopted the Social Security Act, which provided a nationwide program of contributory old-age insurance, unemployment insurance, grants to the states for public assistance to the aged, the blind, and dependent children, and for various health and child welfare services. (*See* further HEALTH INSURANCE; MEDICAL CARE, GOVERNMENT; SOCIAL SECURITY; WORKMEN'S COMPENSATION.)

Other Developments.—Social welfare developments during the early 20th century were not, of course, limited to those in social insurance and public assistance. A few examples concerning the welfare of children will illustrate the range and nature of other improvements. As children were removed from the workhouses and almshouses there was resort first to free foster homes and private and public specialized institutions. Later, emphasis was placed on boarding homes supervised by professionally staffed social agencies. The first juvenile court was established in Chi-

cago in 1899 and the idea spread rapidly throughout the world (*see* CHILDREN'S COURT). Great Britain and Canada adopted juvenile court laws in 1908, France and Belgium in 1912, Hungary in 1913, Austria and Argentina in 1919, and Germany and Brazil in 1923. Except for the Scandinavian countries, which adhered to the theory that children should not be tried in courts at all and gave responsibility for delinquent children to child welfare councils, the juvenile court system was gradually adopted throughout most of the world. Treatment of delinquent children was greatly improved by the development of probation services, and all nations gained from the experiences of the Borstal institutions for young offenders established in England in 1908 (*see* BORSTAL SYSTEM). The importance of family life to child development was emphasized by the first mother's pension act passed by the state of Illinois in 1911. (*See* also CHILD WELFARE.)

Perhaps the greatest advances in social welfare during this period were in the area of health and medical care. Medical research developed means to immunize whole populations against major epidemic diseases. Improved medical care and knowledge both decreased infant mortality and increased longevity. Advances in public health removed important obstacles to further urbanization and industrialization. Improved facilities for sewage disposal, water purification, and food preservation diminished the dangers to health inherent in city living. The assumption of responsibility by the state for town and country planning, for standards of housing, and for the provision of recreational facilities all contributed to the health and welfare of industrial communities. (*See* also PUBLIC HEALTH.)

A final example of the change in attitude toward social and economic conditions and individual maladjustment is the growth of the new profession of social work, particularly in the United States and Great Britain. From the "scientific philanthropy" of the Charity Organization societies of the 19th century the knowledge and methods of social work developed to a point where university degrees were awarded for their mastery. By 1930 professional social workers were being called upon in increasing numbers to staff the public social services and to provide the leadership necessary to ensure that services provided would be of the greatest possible assistance to individuals for whom they were designed. (*See* further SOCIAL SERVICE.)

After the 1930s.—The worldwide economic depression of the 1930s and World War II added new dimensions to the concept of social welfare and furthered its acceptance by all classes in the major industrial democratic nations. The trend emerging during the 19th century had been from palliative and punitive measures to the provision of services and financial benefits which would compensate for the failings of the economic and social systems. Emphasis during the postwar period was placed upon the responsibility of the community through voluntary and governmental action to correct such deficiencies.

The Atlantic Charter of 1941 proclaimed the intention of democratic nations to abolish poverty and establish freedom from want. It was realized by all countries that another prolonged depression might lead to the destruction of democracy and submersion of the private enterprise system. In addition, the industrial output and progress during the war that made possible a condition of full employment of all resources served as a challenge to all nations to maintain full employment after hostilities ceased. If full employment could be maintained, freedom from want could be realized, and each nation could establish a "national minimum" standard of living. Increases in productivity accompanying full employment would give to each person a basic income that would supply his essential needs no matter what vicissitudes he experienced.

The ideas of full employment and of a national minimum both received detailed exposition in the Beveridge Report published by the government of Great Britain in 1942. This report gained worldwide attention and was the inspiration for many social welfare developments during the postwar years. Among other matters the report emphasized the importance of family or children's allowances, as it had been determined that about a quarter of all want in Great Britain was due to failure to relate income to the size of family. The influence of the Beveridge Report on western nations is illustrated by the rapid increase in the number of family allowance plans following the publication of the report. Ireland and Canada established universal family allowance plans in 1944, making payments to all children of certain ages regardless of the income of their parents. New Zealand made its plan universal in effect during 1946, and Norway and Sweden instituted such programs in 1946 and 1947 respectively.

The Beveridge Report was able to assume that full employment could be maintained not only because of the wartime record but also because of the "new economics" developed by John Maynard Keynes (*q.v.*) and his disciples throughout the world. Their theories rejected the classical economist's fear of government interference and instead emphasized the positive role which government must play in maintaining full employment and expanding production.

These economists believed that there are many ways in which an economy can be helped by government toward full employment, but all of them must be aimed at encouraging consumption and investment and discouraging excess saving. Because of the emphasis on consumption, social welfare expenditures have come to be considered as instruments of fiscal policy as well as means for the relief of destitution and the improvement of social conditions. Their effectiveness from a fiscal point of view depends upon the extent to which they serve to redistribute income and maintain or increase consumption during periods of threatened economic recession.

An extension of this approach has been the proposal of a "negative income tax," or guaranteed minimum annual income for every individual, as advocated, for example, by Milton Friedman in his *Capitalism and Freedom* (1962), and by Robert Theobald in his *Free Men and Free Markets* (1963). The guaranteed income would assure families of a minimum standard of living (and a minimum level of consumption) regardless of whether wage earners can find jobs. Critics claim such a program would rob people of incentive to work, or find work; proponents say that that objection can be met by regulating the size of income grants. In at least one respect the proposal represents a major departure from established welfare programs in that since it would put money directly in the hands of persons with incomes below the poverty level, it would in many cases replace programs based on categories of need.

It would be impossible to summarize all the social welfare developments since the 1930s; their complexity in any one country defies analysis. Progress has been made by all nations, due to a great extent to increases in productivity and thus in national income. In order to establish freedom from want a nation must be producing sufficient goods and services to provide a basic minimum for all its inhabitants. For underdeveloped nations, therefore, the greatest contribution to overall social welfare has been an increase in investment to increase productivity.

Some appreciation of the range in the proportion of total income devoted to social welfare services can be obtained from studies conducted by the International Labour Office. One of these, published in 1964, attempts a comparison of the total expenditures by various nations during 1960 for compulsory social insurance, certain voluntary insurance schemes, family allowances, special schemes for public employees, public health services, public assistance, and benefits granted to war victims. The percentage of gross national product being expended for these various services was for the United Kingdom 11.1%, for Canada 8.9%, and for Australia 7.9%. In western Europe the percentage for the Federal Republic of Germany was 16.1%, for France 13.9%, for Italy 12.7%, for Sweden 12.4%, and for Denmark 11.1%. The percentage for the United States was 6.3%. To account for these differences it would be necessary to study in detail the social, economic, and political situations in each nation. Decisions regarding the appropriate level of social welfare expenditure are essentially political. Crucial factors affecting such decisions are the total productivity available, the nation's concept of its national minimum, and the relative influence of competing interest groups.

A survey by the U.S. Department of Health, Education, and Welfare showed that in the 1960s more than 100 countries had

some form of social security system in operation. The great majority of more than 75 statutory old-age pension, invalidity, and survivors' benefits programs reported were full insurance measures, but a few still provided pensions subject to a means test. Approximately 30 countries, including all Western European countries except Portugal, had some form of unemployment insurance. In 1964 the Council of Europe prepared a code to promote the development and standardization of social security measures in order to facilitate labour mobility among the member states.

See also WELFARE STATE.

MAJOR TRENDS OF GOVERNMENTAL RESPONSIBILITY

Some idea of the progress made by wealthier nations can be obtained by examining the history and status of social welfare provisions in Great Britain and in the United States as representing the two major trends of governmental responsibility. In Great Britain, the trend has been toward the guarantee of a national minimum social security, with children's allowances, comprehensive health insurance, etc., available to all. In the United States, although there has been extensive development of various forms of social insurance under both public and private auspices, "means test" programs for special categories are usually relied upon to protect the population from dire poverty. For example, the insurance based Medicare provides limited physician and hospital services to eligible aged, while the means test based Medicaid can provide the aged with comprehensive medical care. Under the "War on Poverty" the role of government has been seen as providing opportunities for all to participate fully in the economy (*see* further *The "War on Poverty"* below).

GREAT BRITAIN

Social welfare in Great Britain during the 19th century was characterized on the one hand by the brutal and repressive philosophy of the Poor Law reform of 1834 and on the other by a new reform movement by which the principle of governmental intervention in the interest of the well-being of the people was firmly established. As in most other western nations this progress toward modern social welfare was accomplished first by members of the middle and upper classes who were disturbed by the suffering occasioned by the free operation of the economy. They were joined during the latter part of the century by organized labour and socialists.

Child Labour Regulation.—The movement began with a campaign against child labour. In 1802 Sir Robert Peel secured passage of a bill restricting the employment of pauper apprentices to 12 hours a day between 6 A.M. and 9 P.M. In 1819 Peel was joined by Robert Owen in bringing about the passage of a law forbidding the employment of all children under 9 years and restricting the work of children under 16 years to 12 hours between 5 A.M. and 9 P.M. This act had limited usefulness as it applied only to the cotton industry and no effective means of enforcement were specified. Britain's coalfields also were the scene of important struggles concerning the regulation of working conditions, especially of women and children, and the development of trade unions. The Factory Act of 1833 both restricted child labour and provided for inspection. This protection granted children was the first of a long succession of laws designed to safeguard labour by providing better living and working conditions, by promoting health and sanitation, by developing housing, and by encouraging education. By 1847, for example, the working day for women and all persons under 18 years was limited to ten hours. (*See* also CHILD LABOUR; CHILDREN, LAWS CONCERNING; LABOUR LAW.)

Chartism, Christian Socialism, Fabianism.—Labour's first attempt to gain influence in politics occurred during the decade which ended in 1848 when there were demonstrations and rioting aimed at achieving a people's charter which involved, among other matters, manhood suffrage (*see* CHARTISM). With the collapse of Chartism the workers turned from politics to the improvement of their economic status. In addition to bargaining for higher wages and better working conditions, they developed consumers' cooperatives, stemming from the teachings of Robert Owen. The beginnings of social insurance also were apparent as trade unionism emphasized mutual benefits through insurance against sickness, unemployment, and old age. The improved position of labour resulting from these activities won respect; in 1867 suffrage was granted to urban workers, and in 1874 the first trade unionist entered Parliament.

The defeat of Chartism also brought the socialists to influence; of these the most important groups from the point of view of social welfare were the Christian Socialists and the Fabians. The Christian Socialist movement, started in 1848 by J. M. Ludlow, Frederick Denison Maurice, and Charles Kingsley, emphasized cooperative associations among the workers and education for the working classes. It did much to foster a sympathetic relationship between the church and the socialists and labour leaders. The Fabian Society (*q.v.*) was founded in 1883, and among the nucleus which determined its policies were George Bernard Shaw, Sidney Webb, Sydney Olivier, and Graham Wallas. The Fabians were concerned with immediate and practical reforms which they helped accomplish by issuing scores of publications and by having individual members enter politics and public administration. The far-ranging interests of the society included women's suffrage, an eight-hour day, a minimum wage, housing, public education, and poor relief. (*See* also SOCIALISM.)

Social Investigation.—During the last quarter of the 19th century two other groups had at least indirect effect on England's progress in the field of social welfare. The first was the Charity Organisation Society, founded in 1869 in London by Edward Denison and Octavia Hill and guided for 39 years by its general secretary, Sir Charles Loch. Although the society espoused the principles of the reform of 1834 and opposed all forms of public relief, it did recognize that there were some deserving poor to whom assistance might be given if the purpose of the aid was to preserve the family, strengthen character, and prevent poverty by helping people to be self-supporting. In addition, in the process of social investigation to separate the deserving from the undeserving, much knowledge was gained concerning the living conditions of the poor and their habits. The formal and informal reports of the visitors did much to convince the well-to-do that all poverty was not due to individual irresponsibility and that general social reform was needed.

The second group of social investigators included parliamentary committees, royal commissions, and individuals such as Edwin Chadwick, whose report on sanitary conditions to 1842 set forth facts which showed how environmental conditions contributed to ill health and mortality (*see* CHADWICK, SIR EDWIN). The most influential investigator was Charles Booth (*q.v.*), who undertook in 1886 his study reported in *Life and Labour of the People in London* (1903). He studied people by trades and by the districts in which they lived; he surveyed the conditions under which they worked, the size and quality of their homes. Also examined were hours of labour, rates of pay, and the extent of unemployment and irregular employment. It was a study of all the people, and he found that one-third of the population was on or about the line of poverty or below it. These findings were confirmed for a small city by Benjamin Seebohm Rowntree's study of York (1900), where it was discovered that 28% of the total population was in poverty.

The struggle between uncompromising individualism and the concept of government as a social force continued throughout the 19th century and into the 20th when an attempt to resolve it was made by the Royal Commission on the Poor Laws and Relief of Distress appointed in 1905. The important protagonists were represented on one side by representatives of the Poor Law Division and the Charity Organisation Society, who constituted a majority, and on the other by representatives of the Fabian Society and Socialist-Labour interests. As might have been expected, although there was some fundamental agreement among the commissioners, no single set of conclusions and recommendations was acceptable to all; majority and minority reports were issued. Although the commission's responsibility was to concentrate on the Poor Law, its deliberations and two reports had implications for all of social welfare.

The two reports were in agreement in their rejection of the spirit of the Poor Law reform of 1834. Indicative of this was their insistence that the title public assistance be substituted for that of poor law or poor relief and that the emphasis in administration should be changed from deterrence to help, prevention, and social provision. Both also agreed that there should be a national system of employment exchanges, unemployment and invalidity insurance, and regularization of employment. The crucial difference between the two reports was in their ideas as to the future role of poor relief. The majority report saw the existing functions of the Poor Law carried out by public assistance committees appointed by the county borough councils with treatment and preferred assistance being provided by voluntary aid committees in each community. The minority report, led by Beatrice Webb, wished to do away with the Poor Law entirely and assemble its activities under existing agencies. These old Poor Law functions would be assigned to divisions in the county councils responsible for education, health, pensions, and mental disease, with relief to the able-bodied to be administered by a Ministry of Labour to be created.

Although the recommendations of the commission were not enacted into law immediately as was the case in 1834, they did signify a new attitude toward poverty and social welfare and were soon followed by a rapid extension of social insurance and public assistance outside the Poor Law and finally by the abolition of the Poor Law itself.

Social Legislation, 1906–34.—The general election of January 1906 also provided evidence of a new attitude and was a revelation of the new forces that had been released by the 19th-century enfranchisement of the workers. The Liberal Party received an enormous majority and labour itself had substantial representation. The leadership in Parliament saw government as a means of social change and proceeded rapidly to enact national social welfare measures. The first invasion of the Poor Law came with the Provision of Meals Act of 1906 whereby local education authorities were empowered to supply meals to children unable to take full advantage of education by reason of lack of food. This was followed during 1907 by the Education (Administrative Provisions) Act, which provided for medical inspection and attention to the health and physical condition of schoolchildren. In 1908 came the Old Age Pensions Act, which provided a flat weekly benefit to all those over 70 years whose income was below a certain amount and who, until 1919, were considered to be "deserving." Also in 1908 the Coal Mines Regulation Act established the principle of the eight-hour day. In 1909 the Labour Exchanges Act created a system of labour exchanges designed to attack unemployment by increasing the mobility of labour, and the Trade Boards Act made a beginning at setting minimum wages.

The culmination of this period of social legislation came with the National Insurance Act of 1911, which provided for compulsory health and unemployment insurance for certain industrial workers. Leadership in sponsoring this program was provided by David Lloyd George. The intention of the government to adopt the principle of compulsory insurance was announced by Winston Churchill, and the details were worked out by Sir Hubert Smith in collaboration with William Henry (later Lord) Beveridge. Both the health and unemployment plans were to be financed by contributions from employers, employees, and the state, setting a precedent to be followed by future social insurance programs in Great Britain. Participation in the insurance was limited to manual labourers and to persons in other occupations receiving less than a prescribed amount of wages. Health insurance was administered through approved nonprofit groups organized by friendly societies (*see* FRATERNAL ORGANIZATION) or labour unions or as adjuncts of commercial insurance companies. Unemployment insurance was administered through a national system of employment exchanges. In 1925 the principle of insurance was extended to cover the eventualities of old age and death by the Widows', Orphans' and Old Age Contributory Pensions Act of that year.

The social legislation beginning in 1906 made it possible for many more persons to avoid applying for poor relief when their earnings were interrupted but did nothing to alter the structure of the Poor Law itself. In 1918, however, some of the harshness was reduced by the removal of the disenfranchisement provisions. In 1919 responsibility for national supervision was placed in the newly created Ministry of Health, and in 1929 the local boards of guardians were abolished and their functions turned over to the county councils. By 1930 the tenets of 1834 had been discarded to the extent that aid was regularly being given to the able-bodied unemployed in his own home. This change in philosophy of Poor Law administration can be attributed, in part at least to the failure of unemployment insurance to meet the need occasioned by the widespread unemployment which followed World War I and continued into the 1930s. Those forced to apply for relief as a result of the decrease in the number of jobs available could hardly be held personally responsible for their plight.

By 1931 the unemployment insurance plan had ceased to have any actuarial basis and the fund had long been insolvent. This situation had been caused by successive extensions of the period of benefits and the payment of "uncovenanted" benefits which assumed that contributions would follow when employment was obtained. The National Economy Act of that year put the scheme back on an actuarial basis by increasing contributions and decreasing benefits, and a system of transitional benefits from national funds was instituted for those no longer eligible. Need for these payments was determined by county public assistance committees who also administered poor relief. Although it was required that these allowances be in cash and that no work test be imposed, there was great dissatisfaction with the plan because of the great variation in amounts paid from county to county from national funds. A centralized national administration of public assistance to the unemployed was finally established by the Unemployment Act of 1934 which created the Unemployment Assistance Board. Allowances set on a national basis were administered by national civil servants whose decisions were subject to review by local appeal tribunals. War-related responsibilities as well as supplementation of pensions and insurance benefits were assumed by the agency, which in 1940 became the Assistance Board.

Post-World War II Legislation.—By the beginning of World War II in 1939 Britain had a fairly comprehensive public social welfare system. There were public medical services, a national health and unemployment insurance system, widows' and orphans' pensions, old-age pensions, public assistance, and provision for the control and subsidization of housing and the supervision of town planning. These were supplemented by an extensive network of voluntary services. Wartime experience, however, emphasized that there were serious gaps and weaknesses in the existing system. The problems of evacuation, for example, showed that there were considerable inequalities between some of the medical services provided in the towns and in the country.

The aspirations of the people of Britain for a better and more economically secure life after the war were given official expression in 1942 in the Beveridge Report on *Social Insurance and Allied Services*. It recommended the extension of social insurance so that a true national minimum and "cradle to the grave" social security would be available to all no matter what exigencies might occur This interdepartmental report signed by Beveridge insisted, however, that this plan would be successful only if there was a system of children's allowances, if there were comprehensive health and rehabilitation services for all, and if mass unemployment could be avoided.

With the end of World War II, Britain's public social welfare services were reorganized and enlarged to form such a comprehensive system. A series of acts forms the framework within which this system continues to develop. The Family Allowances Act, 1945, the National Insurance Act, 1946, and the National Insurance (Industrial Injuries) Act, 1946, were all based on the Beveridge Report. The National Health Service Act, 1946, established the machinery for operating the new health services, while the New Towns Act, 1946, and the Town and Country Planning Act, 1947, created the conditions necessary for rebuilding the nation in a rational and orderly way. The National Assistance Act of 1948 removed the last traces of the old Poor Law by providing that anyone in need should be assisted out of central government

funds, while the Children Act of 1948 provided for a more nearly equal chance in life for the child deprived of normal parental care.

With the Poor Law specifically repealed in 1948, the only qualification for public assistance is need according to standards provided in regulations which are subject to parliamentary approval and are revised from time to time according to cost of living trends. The amount of grant payable is determined by comparing any resources already available to the applicant with the figure at which his needs are assessed according to regulations. The figures are subject to addition for rent according to the amount paid and may be altered when particular circumstances so require. The provision of welfare services under the act, as distinct from financial aid, is the responsibility of the local authorities.

Child welfare services are also the responsibility of the local authorities and their cost is shared by the localities and the exchequer. By the Act of 1948 it is the duty of the local authorities to receive into their care any child under the age of 17 who appears to them to have no parent or guardian, or who has been abandoned, or whose parents are unable to provide for him temporarily or permanently. They must also accept children committed to them by the juvenile court or under the Matrimonial Proceedings Act in exceptional cases resulting from divorce. Children in care are, if possible, boarded out with foster parents or may be placed in a children's home managed by a local authority or a voluntary organization. The children's officer of each locality is assisted by a staff of social workers who undertake inquiries, give help and advice in planning for the care and upbringing of individual children, and supervise them in foster homes.

The development of comprehensive public social services in Great Britain has not deprived the voluntary agencies of a useful role in social welfare. They complement the statutory services and can often undertake work outside the scope of national and local agencies. They are aided by public funds, and in many branches of social service they act as agents of the public authorities. The larger societies concerned with children are Dr. Barnardo's Homes, the Church of England Children's Society, and the Catholic Child Welfare Council, all members of the National Council of Associated Children's Homes. Valuable community service is still rendered by the 50 or so social settlements found in the poorer districts of the cities. In all large and many small towns there are Citizens' Advice bureaus. The Family Welfare Association, successor to the Charity Organisation Society, receives in its area offices in London applications for help from over 4,000 families and individuals a year. About 180 other welfare agencies help families in difficulty or distress, marriage guidance is given in over 100 centres, and the Family Planning Association maintains about 200 clinics. Welfare services for the sick continue to be provided by the British Red Cross Society, the Order of the Hospital of St. John, St. Andrew's Ambulance Association, and many others.

The United States

The history of social welfare in the United States cannot be traced in as orderly a fashion as that of Great Britain. In the United States social welfare functions have traditionally been considered as being exclusively within the sphere of the state governments with financial and administrative responsibility devolving to counties, townships, and municipalities; until the 1930s there were only isolated instances of federal concern. The status of social welfare at any particular time, therefore, must be described in general terms, since each state developed its provisions at a pace which reflected its needs, the status of its industrialization, and the attitudes of politically dominant groups toward social welfare and particularly toward poor relief. Any consistency results from the fact that as each state enacted social welfare programs it benefited from the experience of those which preceded it. In addition, it is difficult to assess the influence of interest groups and changing economic and social philosophies. In both state and federal governments dissatisfactions tend to be met by changes in policy by one or both major political parties rather than to be expressed in a separate political movement. The influence of organized labour on social welfare, for example, is difficult to appraise. Prior to the 1930s it was the avowed policy of the American Federation of Labor to concentrate on collective bargaining and not to sponsor remedial social legislation. With the formation of the Committee for Industrial Organization (later called Congress of Industrial Organizations) and a change in leadership of the A.F. of L. the labour movement began to assume a positive political role and exerted a powerful influence on public policy, but always within the framework of the two-party system.

Early History.—The early colonists had brought with them from England the methods of the Poor Law of the 17th century with its emphasis on local responsibility and administration. Because of the necessity for individual initiative and hard work in developing and expanding a new country, harsh attitudes toward those unable to support themselves were common and the settlement laws designed to protect localities from being burdened by paupers from some other place were stringently enforced (*see* Poor Law: *United States*). As in England at the time, persons who were dependent because of physical and mental disabilities were treated no differently than those unemployed because of some other factor. By the beginning of the 19th century the general attitude toward the poor and toward public relief became even more punitive as the ideas of the Enlightenment had their effect and as legislation in the United States reflected experiences in England leading to the reform of 1834. Increasingly the almshouse was considered the most suitable and economical method of caring for those in need. Many states made county or township almshouse care mandatory.

The reforming and humanitarian zeal typical of the middle classes of this period was concentrated on the abolition of slavery and the struggle for free public education. Some attention, however, was paid to the problems of poverty. In 1841 Robert M. Hartley founded the New York Association for Improving the Condition of the Poor, to save the "deserving" poor from the almshouse and to emphasize moral aid rather than material relief. In 1853 Charles Loring Brace founded the New York Children's Aid Society, a pioneer agency in removing children from the almshouses and orphanages and placing them in foster homes. During this period beginnings were also made for separate provision for special classes of dependents under state government auspices. A school for the "deaf and dumb" was established in Connecticut in 1816 and another in Kentucky in 1822, both of which received some federal aid. In 1841 Dorothea Lynde Dix (*q.v.*) began her crusade for the improvement of the treatment of the insane. She carried her cause personally to every state east of the Rocky Mountains, and in almost every instance her memorial to the state legislature was followed by the erection of a state hospital or the improvement of an existing one. Her memorial to the Congress finally resulted in passage of a bill which would have granted public lands to the states for the care of the insane. The bill was vetoed by President Pierce on the basis that it would establish the principle of federal responsibility for "all the poor in all the states." This veto was used as a precedent for denying aid to the states for social welfare until the 1930s.

Social Reforms to World War I.—The period between the Civil War and the opening of the 20th century was one of rapid industrialization characterized by ruthless and predatory business practices. Although evidence of social deterioration, such as the growth of slums and the poverty of immigrant and native workers, did not go entirely unnoticed, all could be explained by the theories of *laissez faire* and Social Darwinism. Some attempts at improving conditions, however, were made by three groups. First there were the state boards of charity and corrections. By 1867, 16 of the 36 states had established such boards following Massachusetts' example of 1863. Their responsibilities were initially limited to inspection and supervision but later included the direct administration of prisons, reformatories, mental hospitals, and other state welfare institutions. The second group comprised the Charity Organization societies which were found in 92 cities by 1877. Their philosophy and methods were similar to those of the London Charity Organisation Society, with emphasis on individual treatment of deserving applicants for aid and opposition to all forms of public relief. Their "scientific philanthropy" developed over the years into modern social work and its principal method, case-

work. Third were the social settlements (*q.v.*) the first of which was established in New York City in 1886. Perhaps the most important from the point of view of social welfare reform was Hull House founded in Chicago by Jane Addams in 1889, since the group assembled there were outspoken in their belief that government has a responsibility for the well-being of all the people.

While Great Britain was embarking on its great period of social legislation beginning in 1906, the United States was in the midst of the Populist revolt and the Progressive Party movement, which began in the 1890s and continued until World War I. Primary emphasis during this period was on reform of state and local government and the control of business, but progress in social welfare was also made. After overcoming constitutional difficulties in 1908, for example, reformers began the process of achieving state laws on industrial accidents, occupational diseases, factory inspection and safety measures, regulation of child labour, and special provisions to protect women in industry. Attempts to achieve a national child labour law were twice defeated by adverse Supreme Court decisions; not until 1938 were they successful. In 1909 Pres. Theodore Roosevelt called the first White House Conference on Children, which emphasized the importance of the family and the home and one result of which was the establishment in 1912 of the U.S. Children's Bureau. Among the most important of the new social welfare programs enacted during this period of economic, political, and social reform were workmen's compensation, mothers' aid, and pensions for the blind. The first established the principle of social insurance in the United States and the others achieved the first breakup of the poor laws.

In 1908 a federal Compensation Act covering civil employees of the federal government provided a stimulus to the states to enact workmen's compensation laws. The first law to be held constitutional by the state courts was passed in 1911, and by 1920 workmen's compensation laws were in effect in 43 states, Alaska, and Hawaii. The White House Conference had encouraged concern for the welfare of children left orphaned or abandoned or taken from poverty-stricken parents and so laid the groundwork for the development of mothers' aid legislation. The first statewide mothers' pension law was enacted in Illinois in 1911, and 17 states had followed its example by 1913. The needs of the blind also gained early recognition, the first special pensions being provided by Ohio in 1898 and by Illinois in 1903.

Other important social welfare developments during this period included the expansion of public parks and playgrounds and the establishment of voluntary recreational agencies; the growth of state and local public health departments and the improvement of public health services; the strengthening of child welfare services and the spread of the juvenile court idea; and the rise of philanthropic foundations devoted to social welfare such as those of Carnegie, Rockefeller, and Russell Sage (*see* FOUNDATIONS, PHILANTHROPIC). Finally there was the development of Community Chests and Councils of America to coordinate the financing and activities of the many social welfare agencies by then found in all population centres.

The 1920s.—The social reform movement which halted during World War I did not resume following the Armistice until the economic depression of the 1930s. During this period of relative complacency, however, some advances were made. In 1920 the federal government offered grants-in-aid for the vocational rehabilitation of disabled persons. The following year funds were provided the states to help support maternal and child health services. All but three states were receiving grants under this program when it was discontinued in 1927. Further breakup of the Poor Law occurred when Montana and Nevada passed old-age assistance laws in 1923. By the end of 1929, 11 states had such provisions.

The 1930s.—The depression of the 1930s found the social welfare provisions of the United States wholly inadequate to cope with the needs of the millions of unemployed men and women. Except for the few categories already mentioned, poor relief was almost entirely locally administered and locally financed. The rapid increase in relief loads during 1930 and 1931 placed an impossible burden on local, and particularly municipal, finances. The first shift in responsibility was to the states. New York was the first to establish a temporary relief administration and was followed by seven states prior to May 1932. In that month Congress passed the Federal Emergency Relief and Construction Act marking the beginning of federal responsibility, which offered emergency loans to the states to help meet the costs of relief. Only six states did not take advantage of the offer.

Franklin D. Roosevelt's inauguration as president in March 1933 marked the beginning of a new era. On the basis of the countrywide support he received in 1932 and 1936 he was able to insist upon the enactment of a wide assortment of social welfare legislation. Important in this regard was the backing he received from the now politically articulate labour movement.

The first measure enacted to relieve the plight of the unemployed established the Civilian Conservation Corps (March 1933 to July 1942), which provided work-camp employment to over 3,000,000 young, unmarried men during the ten-year period. In May 1933 the Federal Emergency Relief Administration was established and given authority to make grants to the states for both work relief and direct relief. The administration was created for two years and all powers were to be exercised by the administrator, Harry L. Hopkins. This program lasted until the autumn of 1935 and expended a total of $3,000,000,000. The FERA exerted a lasting influence on the administration of public assistance, particularly through its requirement that public funds must be administered by public agencies and its encouragement of relief payments in cash rather than in kind. Between November 1933 and March 1934 the FERA was temporarily discontinued and replaced by the Federal Civil Works Administration, which was a gigantic federal works agency offering to pay wages to 4,000,000 unemployed if state and local governments would put them to work. Other social welfare programs enacted by Congress prior to 1935 were the Public Works Administration (1933–42) which aimed at stimulating the economy with large-scale construction projects, the U.S. Employment Service (1933), and the Federal Surplus Commodities Corporation (1933—later called Surplus Marketing Administration), which purchased farm surpluses and distributed them to public assistance agencies.

When the FERA was discontinued it was replaced by the Works Progress (later Work Projects) Administration (1935–42), designed to give work to all needy persons who could work and to return responsibility for the "unemployables" to the states. The National Youth Administration (1935–42), first a part of the WPA and later an independent agency, provided part-time employment for needy school and university students and others, helped establish job training and counseling programs, and encouraged the development of constructive leisure-time activities for youth.

In June 1934 President Roosevelt set up a special cabinet committee on economic security, to make recommendations concerning a comprehensive program relating to old-age security and unemployment, sickness, and health insurance. A report of the committee was transmitted to Congress on Jan. 17, 1935. The resulting Social Security Act, which became law on Aug. 14, 1935, marked a new stage in the acceptance by the federal government of responsibility for the welfare of all citizens. It established a national contributory old-age retirement annuity system for workers in industry and commerce and laid the basis for a nationwide system of insurance to protect persons against the risks of short-term unemployment. In addition it provided for federal grants-in-aid to the states for old-age assistance, aid to the blind and to dependent children, and for maternal and child health services, child welfare services, services for crippled children, and vocational rehabilitation.

Since 1935 most important social welfare developments have been extensions or improvements of the Social Security Act. Three later federal acts, however, must be mentioned. The Fair Labor Standards Act of 1938 prescribes national standards for wages, hours, working conditions, and child labour applicable to firms producing goods destined for interstate commerce. In 1946 Congress passed the Employment Act, which declares that it is the continuing policy and responsibility of the federal government to promote maximum employment, production, and purchasing power. The National Mental Health Act (1946), designed to provide a method

for financing research and training programs and to assist the states in establishing community mental health centres, was followed in 1963 by the Mental Retardation and Community Mental Health Centers Act that made available funds to the states for the construction of community facilities. Two other federal measures not related to the Social Security Act but which have been important social welfare developments were the Manpower Development and Training Act of 1962 and the Economic Opportunity Act of 1964. Both of these are described below.

Social Insurance.—By 1966 the limited old-age benefits of the Social Security Act of 1935 had become the Old Age, Survivors, Disability and Health Insurance Program. The health insurance or Medicare feature was added in 1965. It provided for hospital, nursing home, and home nursing care for the elderly, plus a voluntary medical insurance plan to help pay the cost of physician's services including surgery whether rendered in a hospital, clinic, doctor's office, or private home. (*See* further *Medical Care* below.) Coverage of the old age, survivors, and disability insurance programs had been successively extended until in the 1960s nine out of ten workers were included. Since 1950 Congress has regularly revised the benefit schedule to reflect changes in the cost of living.

Unemployment insurance, adopted by all states shortly after the passage of the Social Security Act, underwent little change during the ensuing years and in some aspects regressed. Coverage remained as limited as in 1935 except that in 1954 it was extended from employers of eight or more persons to employers of four or more. No standards as to amount and duration of benefits had been imposed by the federal government, and benefits stated as a percentage of average wages had generally decreased since the program was introduced.

Workmen's compensation had remained since its beginnings in the early 1900s a state-controlled program providing benefits and medical care to persons injured, and benefits to survivors of those deceased, in the course of employment. Because of the complete lack of centralized supervision and standard setting, there was a confusing array of coverage provisions, benefit levels, and administrative arrangements.

Public Assistance.—Despite the extension of social insurance, about 4 persons out of every 100 in the population were dependent upon some form of public assistance during the 1960s. They were in need and either ineligible for benefits provided by other public programs or, if eligible, received insufficient amounts. This assistance was provided by the states and localities with the help of the federal government under the three categories of the Social Security Act of 1935, a fourth category, aid to the permanently and totally disabled, established in 1950, and Medicaid, introduced in 1966. Those in need but not eligible for any of the federally aided programs depended upon general assistance or poor relief, which in most states was still locally administered and financed.

All 50 states had programs of old-age assistance which provided aid to persons over age 65 who were in need, aid to dependent children which assisted needy families with children who had lost support or care because of a parent's death, incapacity, or absence from the home, and aid to the blind. Aid to families with dependent children was also provided by 18 states to families where loss of support was due to the unemployment of a parent. All but one state had legislation providing aid to the permanently and totally disabled. Financial participation by the federal government in the case of each of these programs was computed as a percentage of the average grant paid by each state up to a specified maximum. Since 1958 this percentage has varied to some extent in accordance with differences in per capita income among the states.

Federally aided public assistance is superior to poor relief primarily because of the requirements state programs must fulfill in order to be eligible for grants-in-aid. These requirements are aimed at protecting the needy person's right to public assistance if he is eligible according to standards set by the state. For example, assistance must be available in all subdivisions of the state, all persons must have an opportunity to apply for assistance and have their applications acted upon within a reasonable time, and an opportunity for a fair hearing before the state agency must be provided anyone whose claim is not acted upon within a reasonable time. The federal government also requires that assistance be given in the form of money or medical care. This provision has been interpreted as preventing any restriction on the recipient's use of the money payment. He may spend it as do others in the community and continue to live in his own home without interruption of family life. In 1956 Congress declared that a major purpose of public assistance administration was to promote the well-being of the nation by encouraging the states to place greater emphasis on helping to strengthen family life and helping needy families and individuals attain the maximum economic and personal independence of which they are capable. The 1962 amendments to the Social Security Act affirmed this intent by offering additional funds to the states that provide public assistance recipients with prescribed social services.

The federal government in its requirements sets no standards as to what constitutes a national minimum for health and decency. Each state, therefore, is responsible for determining who are needy people and how much they will receive under a given program. States differ both in their definition of need and the degree to which they decide to meet need as determined. One state may provide for some medical needs in its old-age assistance program, but not for its aid to dependent children program, while another state may make no provision whatsoever for medical care. In 1966 average payments per recipient for old-age assistance, for example, ranged from a low of $40 per month in Mississippi to more than $123 in Wisconsin. States not only differ in the amounts of assistance they are able or willing to pay but also in their definitions of what resources an applicant may retain and still be eligible. Policies vary greatly, therefore, with respect to the responsibility of relatives, the value of life insurance, cash reserves, and the ownership of real property.

The federal government also permits the states to require a certain period of residence in the state before an individual can be eligible for aid. Such provisions, reminiscent of the settlement provisions of the old Poor Law, occurred in the legislation in almost all of the states. For old-age assistance, however, less than half the states required as long a period as the Social Security Act permits; *i.e.*, five of the last nine years. The common residence requirement for aid to dependent children was one year.

Those needy individuals and families who cannot qualify for one of the federally aided categories of public assistance must resort to general assistance, which is administered in most states on a local basis in much the same manner as the Poor Law of the 19th century. Such aid is usually given in cash but frequently in the form of orders for groceries, rent, and other items. In most states it is provided only to unemployable persons, and nonresidents are aided only in emergencies with arrangements being made for their return to their state of residence as quickly as possible.

The "War on Poverty."—During the early 1960's public concern was directed to the large number of families living in poverty despite the economic prosperity of the nation as a whole. Resulting social welfare policy emphasized the need of persons in poverty for education, training, work experience, and social services rather than direct measures of income redistribution. The 1962 Public Welfare Amendments to the Social Security Act provided the states with funds to finance such services for public assistance recipients. In the same year Congress passed the Manpower Development and Training Act by which funds were made available for the costs of tuition and subsistence for persons requiring vocational training or retraining. The Economic Opportunity Act (1964) had two principal aims. One was to relieve the plight of young people unable for various reasons to qualify for jobs available in an expanding and technologically oriented economy. For this purpose the Jobs Corps provided education and training in residential centres, and the Neighborhood Youth Corps offered work experience to youths within their own communities. Most far-reaching, perhaps, was the Community Action Program which provided funds to local groups attempting to relieve some of the causes of poverty and deprivation. Although each locality determined the nature of the program to be supported, "Operation

Headstart" providing preschool youngsters from disadvantaged homes cultural and educational experiences to prepare them for school attracted the greatest measure of public interest and support. Most innovative in its approach to poverty was the demand that the poor themselves have a voice in determining the policies of community action programs. (*See* also POVERTY.)

Medical Care.—The growth of publicly supported general medical care has been very slow in the United States. Until 1965, protection against the costs of illness, although widespread, was available only through private contractors. The movement for compulsory state-supported health insurance which began during the early years of the 20th century did not achieve any success until 1965. In that year Medicare was established under Social Security which benefited persons aged 65 and over, and met part of the costs of hospital and home nursing care. A voluntary medical insurance plan, with the federal government matching the premium of the insured, was also instituted to help pay the costs of physician's services. More important in terms of possible future impact were the provisions for Medicaid also enacted in 1965. That program, administered by state public assistance and health agencies, permits greatly expanded hospital and physician's services to those considered unable to afford such care from their regular income, whether or not they are eligible for other forms of public assistance. The federal government contributes generously to all expenditures made by the states if certain minimum requirements are fulfilled. The potential of this program to provide free comprehensive medical care is illustrated by the state of New York where a family of four persons with an income of less than $6,000 per year is eligible for completely free comprehensive hospital and physician services. Most protection against the costs of illness is provided by voluntary plans contracted for by individuals or by employers for their workers. (*See* also HEALTH INSURANCE; MEDICAL CARE, GOVERNMENT; SOCIAL SECURITY.)

Voluntary Programs.—Private social welfare agencies continue to fulfill an important function despite the expansion of governmental programs. Their services are so numerous and so varied as to make summarization impossible. In a typical urban community will be found both sectarian and nonsectarian family service agencies, child welfare agencies, mental health clinics, home nursing services, medical out-patient clinics, and a variety of recreation and group work agencies. Certain national agencies such as the American Red Cross and the National Traveler's Aid Association also have branches in many localities. The financial support for these agencies is for the most part obtained by private donations although public subsidy of private child welfare agencies and mental health clinics is common. Fund raising for local agencies and local branches of some national agencies is usually accomplished through a federated appeal. In most communities the coordination of services is the responsibility of a council of social agencies or health and welfare council to which most public and private agencies belong.

INTERNATIONAL SOCIAL WELFARE

International social welfare programs have increased and expanded with the growing realization that the development of human resources is the real key to a better way of life for the peoples of all countries and that relieving poverty, ill health, and misery is an important part of the task of maintaining peace. The focus of such activity is in the United Nations, where the Department of Economic and Social Affairs of the UN Secretariat administers a worldwide program of research, technical assistance, community development, and information. The principal organ of the United Nations in the social field is the Economic and Social Council. Policy recommendations for this group are developed by the UN Social Commission and carried out by the Bureau of Social Affairs and Advisory Welfare Services.

Other social welfare activities are carried on by specialized agencies which have been established by intergovernmental agreement and are coordinated by the UN Economic and Social Council. The World Health Organization (WHO) helps combat disease and epidemics throughout the world. The United Nations Educational, Scientific and Cultural Organization (UNESCO) assists governments to eradicate illiteracy. The Food and Agriculture Organization of the United Nations (FAO) seeks, among other things, to overcome chronic food shortages. The UN High Commissioner for Refugees supervises the application of the international convention relating to the status of refugees. The oldest of the specialized agencies is the International Labour Organization (ILO) which gives special attention to problems of income security and labour legislation.

See also SOCIAL SECURITY; SOCIAL SERVICE; WORKMEN'S COMPENSATION; and references under "Social Welfare" in the Index.

BIBLIOGRAPHY.—Emerson F. Andrews, *Philanthropic Giving* (1950); Sir William Beveridge, *Social Insurance and Allied Services* (1942); British Information Service, *Social Services in Britain* (1963); Eveline M. Burns, *Social Security and Public Policy* (1956); Karl de Schweinitz, *England's Road to Social Security* (1943); France, Ministry of Labor and Social Security, *Social Security in France* (1957); International Conference of Social Work, *Industrialization and Social Work*, Proceedings of the Eighth Assembly, Munich, 1956 (1957); International Labour Office, *Approaches to Social Security* (1942), *The Cost of Social Security—1958 to 1960* (1964); Ronald Mendelsohn, *Social Security in the British Commonwealth* (1954); John Moss, *Health and Welfare Services Handbook*, 2nd ed. (1955); National Social Welfare Assembly, *Advancing Mutual Security Through International Social Welfare* (1958); George R. Nelson, *Freedom and Welfare: Social Patterns in the Northern Countries of Europe*, sponsored by the ministries of social affairs of Denmark, Finland, Iceland, Norway, and Sweden (1953); Isaac M. Rubinow, *The Quest for Security* (1934); Arnold Toynbee, *The Industrial Revolution* (first published in 1884 as *Lectures on the Industrial Revolution in England*; 1956); United Nations Bureau of Social Affairs, *Assistance to the Needy in Less-Developed Areas* (1956); United Nations, *International Survey of Programmes of Social Development* (1959); United Nations Secretariat, Department of Social Affairs, *Methods of Administering Assistance to the Needy* (1952), *Methods of Social Welfare Administration* (1950); U.S. Department of Health, Education, and Welfare, Social Security Administration, *Social Security Throughout the World* (1958); Frederick Watkins, *The Political Tradition of the West* (1948); H. C. Wilensky and C. N. Lebeaux, *Industrial Society and Social Welfare* (1958); Hans J. Morgenthau (ed.), *Welfare State and After* (1964); Margaret Gordon, *The Economics of Welfare Policies* (1963); Louis A. Ferman *et al.*, *Poverty in America* (1965); Joint Commission on Mental Illness and Health, *Action for Mental Health* (1961); James Farndale, *Trends in Social Welfare* (1965); Richard M. Titmuss, *Income Distribution and Social Change* (1962); John C. Turnbull *et al.*, *Economic and Social Security* (1962); Michael Harrington, *The Other America* (1962). (F. R. BR.)

SOCIETY ISLANDS: *see* FRENCH POLYNESIA.

SOCIETY OF JESUS: *see* JESUS, SOCIETY OF.

SOCINUS, the latinized form of the Italian Sozini, Sozzini, or Soccini, the name of two Italian religious thinkers of the 16th century, whose influence survived in Socinianism and subsequently in Unitarianism (*q.v.*).

LAELIUS SOCINUS (Lelio Francesco Maria Sozini) (1525–1562) was born at Siena on March 25, 1525, of a distinguished family of jurists. He himself was trained for the law at Padua, but his desire to trace the laws of men to their source in the law of God led him to biblical research and this led him to doubt the dogmas of the church. At the age of 21 he went to Venice, then the leading city of the reformist movement within the church, and afterward traveled extensively in Switzerland, France, England, and Holland, his family name and personal charm assuring him a warm welcome. At this stage he combined enthusiastic piety with subtle theological speculation. At the end of 1548 he settled for a time in Zürich where he pursued his studies in Hebrew and Greek. During the following year he was in constant correspondence with Calvin on doctrinal questions. In 1550 he was Philipp Melanchthon's guest in Wittenberg. Wherever he went he prosecuted his theological inquiries, being particularly concerned with such questions as the resurrection of the body, grace and predestination, the sacraments, the doctrinal basis of the original gospel, and repentance. The fate of Michael Servetus (*q.v.*) directed his mind to the Trinity. Rumours that he was a heretic began to spread, and after a warning from Heinrich Bullinger he composed a confession of faith (July 1555), which, while giving the impression of orthodoxy, yet left the door open to heretical views.

In 1556 the death of his father, who had disinherited him, and the sequestration of his share of his father's estate by the Inquisition left Laelius in financial difficulties. Though he secured diplomatic support from the courts of Vienna and Cracow, he

failed to recover his patrimony. He spent the last years of his life at Zürich and died there on May 14, 1562.

His confession of faith, letters, and two dissertations are his only extant writings. He was too speculative a thinker to reach any well-defined conclusions, but his work undoubtedly bore fruit in the fully developed system of his nephew, Faustus.

FAUSTUS SOCINUS (Fausto Paolo Sozzini) (1539–1604) was born at Siena on Dec. 5, 1539, the second child of Alessandro Sozzini (1509–41), brother of Laelius. He was brought up by his mother and grandmother and had no systematic education. Early in 1561 he went to Lyons, possibly for study. After a visit to Zürich to acquire his uncle Laelius' papers (1562), he was briefly in Geneva. On returning to Lyons, he composed his first work, an interpretation of the prologue of St. John's Gospel, in which he wrote of Christ as divine by office rather than by nature.

Toward the end of 1563 he went back to Italy and became secretary to Duke Paolo Giordano Orsini, son-in-law of Cosimo I de' Medici, ruler of Florence. He afterward regarded this period spent at the Florentine court, where he lived in outward conformity to the Catholic Church, as 12 wasted years, though he wrote there (1570) *De sacrae scripturae auctoritate*, in which he established the authority of Scripture on rational and historical grounds. After the death of Cosimo I (1574), Faustus settled for more than three years at Basel, studying Scripture. As a result of discussion with the French Protestant minister Jacques Couet (Covetus) on the question of salvation through Christ, he wrote *De Jesu Christo servatore* (completed 1578, published 1594 in Cracow), in which he denied that Christ appeased the wrath of God by suffering the penalty, through his death, that was justly due to man. Socinus taught that the Word or will of God appeared in the form of flesh, as a man. The resurrection of Christ enables his devout followers likewise to look forward to resurrection in spiritual bodies. After the resurrection, Christ ascended to the right hand of God and henceforth shared God's power. Thus Christ, though purely human in nature, was worthy of adoration as the adopted Son of God, on whom, after the ascension, God had bestowed the government of the world.

The circulation of this work in manuscript brought him to the notice of Giorgio Blandrata (*q.v.*), then court physician in Transylvania, where an eloquent Unitarian bishop, Francis Dávid, had renounced the worship of Christ. Blandrata, fearing that this step would jeopardize the religious settlement reached earlier (*see* TRANSYLVANIA), invited Socinus to visit Kolozsvár (modern Cluj, in Rumania) to persuade Dávid to accept his own doctrine of a modified invocation of Christ within a Unitarian theological scheme. Socinus' reasoning, however, was unsuccessful (autumn 1578–spring 1579).. After leaving Transylvania, Socinus went to Cracow, where he attained a predominant influence in the Minor (or antitrinitarian) church centred on the colony at Rakow. He became its accepted leader and eventually set his stamp upon its theology, although his views on baptism, which he regarded as applicable only to converts to Christianity, apparently excluded him from formal membership. He converted the Arians of Little Poland from their avowal of Christ's preexistence and persuaded most of the Lithuanian Unitarians, at a provincial synod, to accept his teaching that Christ might be invoked in prayer.

The controlling interest in the teaching of Socinus was the attainment of eternal life. The Christian religion was a divinely revealed way of attaining that life, which man, who is by nature mortal, learns through Scripture. Jesus Christ was in his nature a real man though not an ordinary man, for he was conceived by the Holy Spirit. He lived without sin and by his suffering taught men how to bear their own sufferings. The Lord's Supper is an act commemorating the death of Christ; faith is not merely belief that the teaching of Christ is true but such belief as issues in repentance for sins and in an obedience which leads to eternal life.

Socinus supported the Anabaptist Rakovians in their refusal to bear arms and held that the command not to kill admitted no exceptions. Men must, however, offer obedience to the civil government where this did not conflict with the teaching of Christ, and members of the church could act as magistrates if not required to inflict capital punishment.

To avoid persecution Socinus left Cracow in 1583 to live on the nearby estate of the nobleman Krzysztof Morsztyn, whose daughter Elzbieta he married in October 1586. Her death the following September and the later loss of his property through the action of the Inquisition at Siena (1590) were severe blows to Socinus. From 1587 to 1598 he lived in Cracow supported financially by his friends. In 1598 enraged university students attempted to martyr him, and he took refuge at Luslawice, a village east of Cracow. At the turn of the century he began to collect and revise his writings for publication, but the project was unfinished when he died on March 3, 1604, at Luslawice. His incomplete *Christianae religionis institutio* may well have been the basis for the Rakovian catechism, a full exposition of Socinian thought published at Rakow in 1605.

The importation of Socinian writings into England during the 17th century, while not the primary source of antitrinitarian thought there, was long influential upon the development of Unitarian theology, particularly in the doctrines of the person and work of Christ.

BIBLIOGRAPHY.—Two works ascribed with reasonable certainty to Laelius ed. by D. Cantimori and E. Feist, *Per la storia degli eretici italiani del secolo xvi in Europa*, pp. 56–78 (1937). For the works of Faustus see *Bibliotheca fratrum Polonorum*, vol. 1–2 (1656). Polish trans. of Faustus' letters by L. Chmaj, 2 vol. (1959). *See* also A. Gordon, "The Sozzini and Their School," *Theological Review* 16:293–322, 531–571 (London, 1879); D. M. Corey, *Faustus Socinus* (1932); D. Cantimori, *Eretici italiani del cinquecento* (1939); E. M. Wilbur, *A History of Unitarianism: Socinianism and Its Antecedents* (1945); G. Pioli, *Fausto Socino* (1952); S. Kot, *Socinianism in Poland* (1957); G. H. Williams, *The Radical Reformation* (1962); L. Chamj, *Faust Socyn* (1963). (F. KY.)

SOCIOLOGY (ARTICLES ON). Sociology as an academic discipline represents one approach to the systematic study of man and society in order to understand and, ultimately, to attempt to help determine the course of civilization and culture. The broad background of this study is discussed in CIVILIZATION AND CULTURE and, more specifically, in SOCIAL SCIENCES. The article SOCIOLOGY traces the history of its development as a science and discusses fields of specialized interest, and sociological theory, methods, and applications. As a discipline, sociology represents a way of looking at and trying to isolate and analyze significant aspects (problems) of society in order to understand them, find out what causes them, and try to modify or control them. One such problem area is outlined under POVERTY and UNEMPLOYMENT. The history and philosophy of efforts to deal with the problem are covered in PHILANTHROPY; POOR LAW; and WELFARE STATE. Programs to combat poverty are dealt with in CHILD WELFARE; SOCIAL SECURITY; SOCIAL WELFARE; UNEMPLOYMENT INSURANCE; and DISASTER RELIEF. Techniques are covered in SOCIAL SERVICE, while a number of shorter articles describe specific institutions and social inventions, as BEDLAM; CASUAL WARD; CHARITY COMMISSIONERS FOR ENGLAND AND WALES; FOUNDLING HOSPITAL; RED CROSS; and SOCIAL SETTLEMENTS.

Another problem area is that of crime and delinquency (*see* CRIME; JUVENILE DELINQUENCY). CRIMINOLOGY presents a survey of theories of causation, correction, and prevention; programs and institutions are covered in PRISON; PROBATION; PAROLE; RECIDIVISM; and HABITUAL OFFENDERS; and, for youthful offenders, REFORMATORY and the British BORSTAL SYSTEM. (*See* also CHILDREN'S COURT.)

Problems of urbanization are dealt with in CITY; METROPOLITAN AREA; HOUSING; and URBAN RENEWAL. (*See* also GARDEN CITY; NEW TOWNS.) Population problems are the subject of POPULATION; BIRTH CONTROL; and MIGRATION. (For population data *see* VITAL STATISTICS.)

Inter-group relations are the concern of CLASS, SOCIAL; SEGREGATION, RACIAL; GENOCIDE; GHETTO; MARGINAL GROUPS, and related articles. CHILD LABOUR; FORCED LABOUR; PROSTITUTION; and SLAVERY deal with the exploitation of others.

Sociology shares with other social sciences concern with man's most universal institution, the family, its history, and the changes it is undergoing in the modern world: these are analyzed in MARRIAGE and FAMILY.

Other articles of general interest include COMMUNICATION;

INVENTIONS AND DISCOVERIES; and SECRET SOCIETIES. Sociology also deals with a number of subjects which have a narrower or more specialized historic or modern interest, such as CONCUBINAGE; and NUDISM.

Each of the above references will refer the reader to related articles. Biographies of the leading personalities in each field indicate their contributions.

SOCIOLOGY is one of the several specialized social sciences. Its primary objective is to gain a knowledge of man and society insofar as it may be achieved through investigation of the elements, processes, antecedents and consequences which are involved in group living.

Sociology is differentiated from the other social sciences primarily by the basic fact with which it starts, the fundamental assumption which underlies its activities, and the problems with which it is concerned. The basic fact is the relative helplessness of the human infant at birth who is always born into, and depends for his survival upon, a social group. The fundamental assumption is that the conduct of the person—his ways of thinking and ways of acting—and the nature of the social orders—its structure, function, and values—are to be understood as a product of group life. The problems with which sociology is concerned may be stated as a long list of specific items but, more appropriately, may be generalized into broad groupings of subject matter such as those represented by social organization, social psychology, human ecology, and demography. Each one of these broad fields of sociology may be further divided into subfields and into rural and urban aspects, and all are concerned with various problems of theory and method.

As a social science, sociology, contrary to popular misconception justified in part by its early history, does not have as its objectives the determination or modification of social values, the proposal of reforms, the design or administration of welfare programs, or the direct promotion of a better social order. These are important objectives to be sure, but they are the tasks of the statesman, the administrator, the legislator, the educator, the social worker, the clergyman, the labour leader, the agitator, and the propagandist, rather than of the sociologist as a social scientist. The discipline of sociology, strictly speaking, is concerned only with the pursuit and funding of knowledge about man and society. In such knowledge, which ideally comprises generalizations drawn from empirical and verified investigation, the sociologist strives to understand and to achieve prediction of human conduct and social phenomena. Values are among the data which he studies and not the end product of his labours.

This article on sociology is divided into the following sections and subsections:

I. HISTORY OF SOCIOLOGY

The word sociology was first used by Auguste Comte (1798–1857) in 1837 in a series of lectures on which his famous work *Positive Philosophy* (6 volumes, 1830–42) was based. His coining of "sociology," a hybrid of Latin and Greek still disliked by many, has led him to be regarded by a number of writers as the founder of sociology. Yet, his conception of sociology as an all inclusive general social science—a synthesis of all knowledge about "humanity" which he used in the sense of the universal society—and his promotion of "positivism" as the foundation for the mystical religion of humanity which he proclaimed in his later life are widely divergent from contemporary conceptions and practices of sociology. Comte did make a great contribution to the social sciences in general in producing a synthesis of the social thought of the century which preceded him, beginning with the work of David Hume, in laying great stress on the application of the positive method to social phenomena—a method of observation, experimentation, and comparison—and in his evaluation of the importance of studying laws of succession and association of phenomena rather than cause. It is more accurate to say that Comte in his classification of the sciences and his description of their development made a place for and heralded the approach of sociology, rather than that he founded it.

The history of sociology, in a loose sense, began centuries before Comte, but, like all history, has no clear-cut beginning. Historians of sociology, at least those who have published histories of sociology or of social thought, have variously begun their accounts with ancient civilizations (*e.g.*, Pitirim Sorokin, Joyce O. Hertzler, or Emory S. Bogardus), with the Greeks (*e.g.*, Floyd N. House, Harry E. Barnes, or James P. Lichtenberger), or with 19th-century thought (*e.g.*, P. Barth, Albion W. Small).

If the genesis of sociology is traced to folk, philosophical, political, and legal thought about man and society, then sociology had its beginnings in the earliest records of human thought in the ancient civilizations of Egypt, Babylonia, Assyria, Persia, India, and China. If sociology is rigidly defined as a specialized empirical social science, then it may be said that it certainly did not begin before the 20th century. It is much less important, however, to fix a precise date for the beginning of sociology in the narrower sense than it is to have an understanding of what it aims to do, of the problems with which it is concerned, and of some of the main currents of thought which led to its emergence.

1. Social Thought Prior to the 19th Century.—The history of sociology in the broader sense is essentially the history of social thought through the ages—a heritage which sociology shares with all the sciences, natural and social. Although various aspects of social thought, from the earliest human records which can be traced back 5,000 to 6,000 years to the dawn of the 19th century have special implications for the development of sociology, only the briefest mention can be made of them. Ancient literature and relics contain keen insights into interpersonal relations, group life, social control, social organization, social stratification, human nature, and social institutions, including the family, the state, and property. Early social thought was largely of a common-sense nature, sporadic, traditional, and unverified, and suffused with the interventionism of the supernatural.

Systematic social thought, at least in a form which has been reasonably well preserved, originated with the Greek philosophers. Beginning with about 600 B.C. a remarkable series of Greek thinkers became preoccupied first with the world of nature and then with the world of man. The Greek philosophers, and especially Plato and Aristotle, provided penetrating analyses of the relation of man to society, of social institutions, and of social processes; contributed to important methodological developments; and originated basic elements of social theory which have remained in the focus of intellectual discussion ever since.

Despite the setting of the stage for the development of science by the Greeks, natural science did not develop until 2,000 years after Thales (*c.* 624–*c.* 546 B.C.) and social science until more than 2,000 years after Aristotle (384–322 B.C.).

By the 15th and 16th centuries, however, the combined influences of a number of events produced the revival of learning manifested

in humanism and naturalism, and the way was clear for the resumption of critical social thought. By the end of the 17th century, natural science had forged a relatively secure place for itself, and thought about the social as well as the natural world had become largely secularized.

2. 19th-Century Influences.—By 1800 many interrelated intellectual pathways led to the consideration of various aspects of man and society and a number of them contributed to the emergence of the science of sociology. The way had been paved for the application of scientific method to social phenomena (compare SCIENTIFIC METHOD for distinctions among the disciplines considered sciences) by such men as G. Vico (1668–1744) and C. Montesquieu (1689–1755); quantitative methods had been applied to population phenomena by the "political arithmeticians" such as William Petty (1623–87) and John Graunt (1620–74); and important insights into central sociological problems had appeared in the writings of the "Scottish moralists," Francis Hutcheson (1694–1747) and Adam Ferguson (1723–1816).

Major influences on the development of sociology in the 19th century were exerted by scholars who dealt with substantive and methodological problems which, at least implicitly, helped delineate a field and methods for the social sciences in general, as well as for sociology in particular. These writers may be classified into groups, not mutually exclusive, organized around leading ideas. The ideas may be broadly grouped into three categories: first, those focusing on problems which set the stage for, or themselves became, central problems in sociology; second, those emanating from biology, which exerted an important influence on early sociology; and third, those concerned with methodology and techniques of research. In addition to these ideas, the emergence of sociology was also influenced by ideas and programs of social reform, which on the one hand stimulated interest in sociology and on the other confused its role as a science.

Ideas Related to Central Problems.—The writers contributing to this category of ideas included the scholars who dealt with the nature of the historical process, the distinction between the state and society, the nature of culture, folk psychology, the analysis of self-consciousness, the influence of the physical environment, the role of "social forces" and "interests," the economic activity of man, and population problems.

The writers who inquired into the nature of the historical process sought to relate successive events and to generalize from the experience of history. They devoted considerable attention to the question of whether history could or should generalize as against reporting unique events. Important participants in this development included Kant, Fichte, Hegel, Marx, Wilhelm Dilthey, and Wilhelm Windelband.

In early social thought the "social" was largely synonymous with the "political." Preoccupation with the state precluded attention to other aspects of society. The scholars who differentiated between the state and society helped to point up the need for the study of the nonpolitical aspects of the social order. Thus they made explicit in their affirmation or opposition the need and place for a new science. They included such men as Robert von Mohl (1799–1875), H. von Treitschke (1834–96), and Ludwig Stein (1859–1930).

Also made explicit during the 19th century was the idea of culture as a thing in itself. It was emphasized that culture, both in its material and nonmaterial aspects, could be studied historically and comparatively to illuminate the nature of man and society. Early culture historians and anthropologists included Friedrich Ratzel (1844–1904), Adolf Bastian (1826–1905), Alexander von Humboldt (1769–1859), and Lewis H. Morgan (1818–81). These men, and others, collected a mass of data on the cultures of diverse people, and their efforts to compare, analyze, and explain similarities and differences helped to stimulate the sciences of anthropology and sociology.

Another group of writers were expressly concerned with "folk psychology" and with the relation of folk psychology to the individual mind. Their analysis of folk psychology and the "collective mind" led them to explanations of human nature and of individual traits as arising from culture. These writers defined one of the central problems of sociology in its contemporary sense in highlighting the dependence of a person on the social group and describing personality as a subjective aspect of culture. They included such men as Johann F. Herbart (1776–1841), the Grimm brothers Jacob (1785–1863) and Wilhelm (1786–1859), Moritz Lazarus (1824–1903), H. Steinthal (1823–99), and Wilhelm Wundt (1832–1920).

Closely related to the study of folk psychology and the nature of the mind was the analysis of the growth of self-consciousness. The description of self-consciousness as a phenomenon dependent upon society and arising in social interaction continued to play a central role in 20th-century social psychology. Contributors to this analysis included Von Humboldt, Georg Simmel (1858–1918), and James M. Baldwin (1861–1934).

Although ancient and early modern scholars were concerned with the influence of environment, various streams of thought during the 19th century led to more intensive investigation into the influence of geographic factors on social organization and human behaviour. Geographers and anthropogeographers who investigated this problem varied in their conclusions from strict environmental determinism to more modern indications of the limiting influences of physical environment. These writers helped pave the way for the science of sociology in general, and more specifically for contemporary sociological interest in human ecology and regional studies. They included such men as Carl Ritter (1779–1859), H. T. Buckle (1821–62), P. G. F. Le Play (1806–82), and Vidal de la Blanche (1845–1918).

Another central problem in early sociology originated in the writings of the scholars concerned with "social forces," "interests," and the social process. The notion of social forces appeared in the historical literature in efforts to explain events in terms of their underlying and antecedent factors. The notion, however, was also used more generally to explain basic human and group behaviour. Among the writers who sought explanations of behaviour in social forces and interests were Buckle, Ludwig Gumplowicz (1838–1909), Gustav Ratzenhofer (1842–1904), and Gabriel Tarde (1843–1904).

The development of economics during the 19th century represented the emergence of perhaps the first of the social sciences to a position of respectability and general acceptance. The description of the economic man as a rational, calculating, hedonistic person stimulated interest in individual motivation and in the analysis of rational behaviour. The emphasis placed by economic writers on the division of labour led to investigation into its relation to the social structure and process. Especially noteworthy among the writers on economic matters were Adam Smith, David Ricardo, Jeremy Bentham, and J. S. Mill. An important offshoot of classical economic thinking was that represented by critics of the new industrial order and of classical economics, manifest largely in the writings on socialism. In this literature the names of Karl Marx and Friedrich Engels were outstanding.

Closely tied in with, and contributing to, economic thinking were the discussions of "the population problem." Dramatized by Malthus, whose "law of population growth" was originally presented as a refutation of Condorcet's and William Godwin's notions of perfectionism, consideration of the relation between population and means of subsistence occupied considerable attention in the world of practical affairs, as well as among scholars. Writers concerned with population and with the antecedents and consequences of population growth included, besides Malthus, Francis Place (1771–1854), Archibald Alison (1792–1867), Adolphe Quételet (1796–1874), Marx, and Herbert Spencer.

Ideas from Biology.—Important developments in biology during the 19th century, climaxed by the theory of evolution, contributed both directly and indirectly to early thought in sociology. The influence of biology on sociological thinking may be considered under four groups of ideas; namely, the theory of evolution, the "organic analogy," the problem of "heredity versus environment," and the theories of racism.

One of the most dominant influences on 19th-century thinking, in general, was that exerted by the theory of evolution. Stimulated by Malthus' discussion of the "struggle for existence" and

with the taxonomic labours of Linnaeus available, Alfred R. Wallace and Charles R. Darwin independently published their contributions which made evolution the cornerstone of social as well as of biological thinking for decades. Evolutionism was quickly adopted by students of social phenomena as a basic explanatory principle applied to the social order, as well as to the biological. Among the writers who adopted social evolutionism were Walter Bagehot (1826–77), John Fiske (1842–1901), Gumplowicz, Ratzenhofer, and Spencer.

In addition to providing the principle of evolution, biology also exerted an important influence on the development of sociology in the writings of the scholars who compared society with an organism. Impressed by the organic character of society, a number of writers described the social order in terms paralleling the biologist's description of an organism. The organic analogy, while subject to the limitations of all analogies, did perform the function of directing attention to the organic character of society and stimulated more intensive inquiry into the nature of the structure and function of the social order. It led to consideration of those aspects of society that made the social whole more than a sum of all its parts. Among the writers to utilize the organic analogy were Comte, Spencer, Albert Schäffle (1831–1903), Paul von Lilienfeld (1829–1903), René Worms (1869–1926), and J. Novicow (1849–1912).

Influenced by the prominent place of variability in the theory of evolution, and on the basis of studies of individual differences, the attention of a number of scholars was devoted to the consideration of the relative roles of heredity and environment in human behaviour. Individual differences were recognized as attributable to both hereditary and environmental factors. Writers with a biological orientation tended to stress the greater importance of heredity: first, as the basic factor in explaining individual differences in abilities and behaviour; and second, as the basic factor underlying social change. Especially prominent in the development of these ideas were Sir Francis Galton (1822–1911) and Karl Pearson (1857–1936).

A special variation of the efforts to explain human behaviour in biological terms was found in the ideas of a number of writers on racial superiority and inferiority. On the whole, this literature, while often ornamented with the trappings of biological findings, was propagandistic in character and tended to play up the alleged superiority of the so-called Aryan or Nordic race. They are mentioned here because they stimulated a widespread controversial literature which helped to focus attention on historical and cultural, as well as biological, factors in accounting for racial as well as individual differences. Among the writers who contributed to the literature of racism were Arthur de Gobineau (1816–82), G. V. de Lapouge (1854–1936), and Otto Ammon (1842–1916). (*See* also INTERRACIAL RELATIONS: *Racialist Thinking*.)

Ideas Relating to Methodology and Techniques.—The writers briefly treated above contributed to some of the central substantive problems of emergent sociology. Equally important were the writers who helped develop a methodology for sociology. These writers may be considered in three groups; first, those who contributed to general methodology; second, those who contributed to statistical techniques; and third, those who contributed to the "social survey."

Noteworthy, first of all, were the methodological contributions of a number of historians who, in grappling with the problems of method in history, made important contributions to social science in general. F. K. von Savigny (1779–1861) stressed the methodological implications of continuity in the historical process; K. F. Eichhorn (1781–1854) called attention to the complexity of social phenomena and the importance of the idea of multiple causation; B. G. Niebuhr (1776–1831) emphasized the necessity of being critical of sources and of the role of the investigator; and Leopold von Ranke (1795–1886) gave great impetus to the habit of documentation and annotation. Also of great importance were the methodological contributions of Comte, Dilthey, and Windelband, who grappled with the general problems of social science methodology as contrasted with method in the natural sciences.

Simultaneously with the consideration of general problems of method there were important developments in the application of mathematics to the analysis of social data. Building particularly on previous work in probability theory, as well as in general mathematics, statistical techniques were evolved during the 19th century suitable for the treatment of social phenomena. Important contributors to these developments included P. S. Laplace (1749–1827), Quételet, Carl Friedrich Gauss (1777–1855), Adrien Marie Legendre (1752–1833), Francis Galton, and Karl Pearson.

Significant from the standpoint of method as well as for the substantive data it provided was the "social survey" as it developed during the 19th century. The social survey was essentially a fact-finding enterprise designed to illuminate social and economic evils in a specific area for purposes of ameliorative action. Although action oriented, the social survey provided basic data for research and stimulated widespread interest in the study of social, economic, and political problems. Some of the more important social surveys were those conducted by Le Play, Sir Edwin Chadwick (1800–90), and Charles Booth (1840–1916). The latter's famous report, *Life and Labour of the People in London* (17 volumes, 1889–1903), set a pattern for similar surveys in other countries.

Other 19th-Century Influences.—In addition to the developments outlined above, sociology was greatly influenced by the increasing attention during the 19th century to the social, economic, and political problems arising from the accelerated pace of the Industrial Revolution and urbanization. These problems were dramatized by extreme manifestations of poverty, high mortality rates, urban slums, abuses of labour, political unrest, and political corruption. Nineteenth-century critics of the industrial order, philanthropic and utopian efforts to deal with the more severe social and economic conditions, the literature on socialism and anarchism, and increasing awareness of and efforts to deal with the problems of a rapidly changing social order constituted an important part of the 19th-century climate in which sociology emerged as a distinctive field.

Pioneer Sociologists.—Nineteenth-century writers included a number of men who are generally regarded as the founding fathers of sociology. Early sociologists were not the product of any one nation. In France they included, in addition to Comte, Émile Durkheim (1858–1917), Lucien Lévy-Bruhl (1857–1939), Gabriel Tarde, and Gustave LeBon (1841–1931); in Germany, Ferdinand Tönnies (1855–1936), Georg Simmel, and Max Weber (1864–1920); in Austria, Ludwig Gumplowicz and Gustav Ratzenhofer; in Russia, J. Novicow and Maxim Kovalevsky (1851–1916); in Italy, Scipio Sighele (1868–1913) and Vilfredo Pareto (1848–1923); in England, Herbert Spencer (1820–1903), Leonard T. Hobhouse (1864–1929), and Patrick Geddes (1854–1932); and in the United States, Lester F. Ward (1841–1913), William Graham Sumner (1840–1910), Albion W. Small (1854–1926), Franklin H. Giddings (1855–1931), Edward A. Ross (1866–1951), and Charles Horton Cooley (1864–1929). Other countries also developed pioneer sociologists, but their works were, for the most part, untranslated and they did not enter into the mainstream of sociological development.

Of the pioneer sociologists, perhaps the most influential in the impact of their work on other sociologists were Durkheim, Weber, and Cooley. Durkheim provided a clear delimitation of the field of sociology as a science devoted to the study of "the social fact." Social facts, which included forms of behaviour, thought and feeling, could be studied, he emphasized, as things in themselves. The social fact was exterior to the individual, anterior to him, and exercised a constraint upon his behaviour. It was to be studied as a collective representation and not in its individual manifestations. In his *The Rules of Sociological Method* (George E. G. Catlin, ed.), he described what he regarded as the field and the method for sociology. In other publications, notably *Suicide, a Study in Sociology* (1897, translated by John A. Spaulding and George Simpson), the *Division of Labour in Society* (1893, translated by George Simpson), and the *Elementary Forms of the Religious Life* (1912, translated by J. W. Swain), Durkheim provided examples not only of the application of his methods but also of notable pioneer empirical investigations.

Weber also delimited a field for sociology. To Weber, sociology was a science devoted to the interpretation and understanding of social behaviour so as to achieve its prediction. He defined "social behaviour" as behaviour carried out according to the intention of the acting individual in reference to the behaviour of others. Sociological analysis was to be devoted mainly to getting at the intentionality of social behaviour. This was to be achieved through gaining an "understanding" of social behaviour: intellectually, if the behaviour was rational, by recognizing the logic of the means-end relationship; by means of empathy, if the act was irrational, by projecting oneself into a situation and experiencing the emotional context involved. Weber made an important methodological contribution in his description and use of "ideal type" concepts as tools of analysis against which to observe and analyze real situations. He also recognized the utility of statistical methods, especially the fundamental notions of probability and association as tools of research. Weber's consideration of bureaucracy and social stratification contributed materially to sociological investigation of these subjects. Some of his more important publications available in English are *The Theory of Social and Economic Organization* (translated by A. M. Henderson and Talcott Parsons), *From Max Weber: Essays in Sociology* (translated by H. H. Gerth and C. Wright Mills), and *The Methodology of the Social Sciences* (translated by Edward A. Shils and Henry A. Finch).

Cooley, in his doctoral dissertation *The Theory of Transportation* (1894), contributed in an important way to interest in human ecology as a field of sociology. In this work he dealt with transportation and the general economic base as a substratum of social organization. His subsequent work resulted in major contributions to general sociology and social psychology. He began with a theory of human nature and the genesis of the self, as set forth in his *Human Nature and the Social Order* (1902). His study of personality led him to the study of social structures and forms as treated in his *Social Organization* (1909). He completed what may be regarded as a system of sociology in his analysis of collective life in *Social Process* (1918). Cooley's treatment of the "primary group," the "looking-glass self," communication, and the problem of society and the individual provided subsequent sociologists with an important part of their conceptual framework. His methodological contributions lay largely in his suggestive work in ecology on the one hand, and in his description of "sympathetic introspection" as a tool of research on the other.

Mention should also be made of Small as one of the important early sociologists, particularly for his contribution to the development of sociology as an academic discipline and as a profession. On the substantive side, Small, although he changed his position in the course of his career, helped define the scope of sociology as a science; played an important part in introducing European, and especially German, sociological thinking into the United States; and laid great stress on the development of method as a prerequisite to the advance of sociology as an empirical science.

It is not possible, of course, even to summarize the work of all of the individual builders of sociology. Most of the founders had in common a penchant for system building and, no doubt under the influence of the brilliant and dramatic 19th-century scientific successes in other fields such as that represented by Darwin's theory of evolution, attempted to account for man and the social order by broad, comprehensive, and global principles. Most of the systems were characterized by universal explanatory schemes, such as "evolutionism," or by monistic explanations of human behaviour such as "imitation."

A number of sociologists subscribed to the view that sociology was essentially a normative, meliorative discipline. They engaged in great debates over nominalism and realism, as manifested in the discussions of the "atomistic" or "organic" conceptions of society, or the distinction between the individual and social aspects of the group; over the relation of sociology to values; over the priority of the study of the form as opposed to the content of individual and group behaviour; over the role of the "biological," the "psychological," and the "sociologistic" factors or levels of explanation; over the materialistic and idealistic conceptions of the historical process; over the suitability of natural science methods for the study of social phenomena and especially the applicability of quantification, or statistical techniques; over the conception of sociology as the general social science or one of the special social sciences. In the works of the pioneers these and other issues, as they bore on the emergence of a science of society, were clarified; and if the problems were not resolved, their discussion was, on the whole, exhausted, so that points of departure were provided for subsequent students.

The common and essential characteristic of these early sociologists was that they were, in the main, still social philosophers rather than social scientists. Their basic methodology was still grounded in speculative thinking rather than empirical investigation, although to be sure there were some notable exceptions (*e.g.*, Durkheim's *Suicide*, 1897, and Sumner's *Folkways*, 1906). They were concerned more with discussions of what sociology should be and should do rather than with the doing of it. They performed an important function in making explicit some of the basic issues, consideration of which was perhaps prerequisite to the emergence of sociology as an empirical science; in delimiting an area of investigation; in posing many of the specific problems of sociology; in framing a methodology and providing a number of research techniques; in contributing materially to the development of sociological concepts; and in providing important insights which influenced the course of subsequent developments in sociology.

3. 20th-Century Developments.—System building in sociology continued into the 20th century, and for the first two decades there was also a sustained interest in the broad general problems of the type indicated above. The distinctive marks of 20th-century sociology, however, lie in its delimitation as a separate social science; in the emphasis placed upon research as the primary task of sociology; in the emergent conception of theory as a prelude to, and product of, research rather than an exercise in armchair speculation; in the broadened participation of sociologists in empirical investigation; and in the increasing use of statistical and experimental or simulated experimental methods. Accompanying these developments was a great increase in the number of sociologists and in the widespread appearance of sociology in university and college curricula, especially in the United States, as a recognized academic discipline. In these developments, sociology, if it did not become a science in a rigorous sense, at least definitely moved in that direction. The great bulk of sociological publications comprised reports of empirical research, and broad social speculation ceased to be a major preoccupation of sociologists, especially in the United States.

The shift from speculative social thinking to empirical research was led and fostered by the generation of sociologists who followed the pioneers and who, for the most part, were their students and sometimes their disciples. The trend toward empirical investigation was most manifest in the United States, partly because of the availability of greater resources for research. Research grants by the large foundations (such as the Rockefeller Foundation and the Carnegie Corporation of New York), expanding research appropriations by the federal government, especially for agricultural research at the outset, and increased expenditures by action agencies for "applied research" provided the means for the conduct of empirical investigations. Research activities were stimulated, also, through the development of professional organizations whose primary objective was the encouragement of such activity. The most important of these were the American Sociological Association, founded in 1905 as the American Sociological Society, and the Social Science Research Council, which was organized in 1923. Sociological research activity, however, was by no means confined to the United States.

In the United States the transition from social speculation to empiricism was spearheaded by such men as W. I. Thomas, Florian Znaniecki, Robert E. Park, E. W. Burgess, W. F. Ogburn, Ellsworth Faris, Richmond Mayo-Smith, R. M. MacIver, L. L. Bernard, F. S. Chapin, and H. W. Odum. Not all of these men themselves pioneered in research activity, but they all showed an enthusiasm for, and interest in, research which they passed on to their students. Early 20th-century examples of empirical investigation included such works as Sumner's *Folkways* (1906), Paul U.

Kellogg's *The Pittsburgh Survey* (1909–14), C. J. Galpin's *The Social Anatomy of an Agricultural Community* (1915), and W. I. Thomas and Florian Znaniecki's *The Polish Peasant in Europe and America* (1918–21). In addition, works which, while themselves not empirical investigations, contributed inestimably to the increasing tempo of research activity included Park's essay "The City: Suggestions for the Investigation of Human Behavior in the Urban Environment" (1915) and Park and Burgess' *Introduction to the Science of Sociology* (1921). The former provided a framework for urban and ecological research. The latter brought together in a systematic fashion the insights and concepts of earlier writers set in a framework oriented to research. Pursuant to these publications, a series of monographs were published by young sociologists, trained at the University of Chicago, which constituted the first mass series of empirical investigation in sociology.

After about 1940, the relatively large volume of discrete research reports in sociology led some sociologists to feel that the pendulum had swung too far away from theory. Pure fact finding was criticized as being as unproductive as pure speculation. In consequence, there was a discernible attempt on the part of sociologists to tie their investigations to relevant theory and to direct their research activities toward the testing of explicitly stated, verifiable hypotheses.

Sociology was slower to develop in Europe than in the U.S. during the 20th century, especially in research. Up to World War II, in England a mere handful of men had claim to the designation of "sociologist," although a number of prominent scholars had contributed to the development of various aspects of sociology. The situation was similar on the continent, with the exception of France and Germany. Although an important literature was developed which contributed materially to sociological insight and theory, empirical research did not reach the stage or magnitude of development in either France or Germany that it achieved in the United States. In France, sociological activity was closely interwoven with that of anthropology and, in large part, not clearly differentiated.

After World War II, however, there was evidence of increasing interest in sociology in Europe and in other parts of the world. The number of chairs in sociology at recognized universities grew considerably. An International Sociological Association was established under the sponsorship of the United Nations Educational, Scientific and Cultural Organization in 1949, and Louis Wirth (1897–1952), of the University of Chicago, was elected its first president. (P. M. Ha.)

II. FIELDS OF SOCIOLOGY

There is no generally accepted classification of the fields within sociology. The classification used below, for purposes of convenience in summarization, is arbitrary. The fields listed are closely interrelated and not mutually exclusive. They serve to denote general areas of specialized interest for research and teaching purposes.

A. Social Organization

Social organization, the core field of sociology, is concerned with the description and scientific analysis of human organization. The study of social organization includes the full range of human groups, from the smallest—the family and the play group—to the largest and most complex, the nation-state. It also includes industrial, political, educational, and religious institutions. The term covers social stratification, social mobility, and the organization of occupations, professions, and ethnic and racial groups. It involves study of the sources of social change and analysis of social control; that is, the informal and formal rules by which men govern and regulate their social relations.

Since ancient times, men have recognized that social groups are more than the mere sum of individuals. Human groups have a character apart from the characteristics of their members. Often this has been called the moral order or the social order. Philosophers have traditionally sought to understand man's ability to produce a stable social order and the reasons for breakdowns of the social order. The modern sociologist who studies social organization is continuing this tradition of inquiry. Instead of speculation, he seeks to make use of the methods and empirical techniques of social research.

The sociological study of organization is closely related to similar research interests in other social sciences. Social anthropology (*q.v.*) generally confines its investigation to the social organization of primitive societies, but the literature of the two fields have many areas of common interest. The work of social anthropologists, particularly those of France and to a lesser extent those of England, has had a direct influence on the study of social organization in the United States. Since World War II, social anthropologists have shown increased interest in modern societies. The convergence of the two fields has been accelerated as primitive societies throughout the world have become modernized. Social organization research is linked to economics in the study of industrial organization and of economic development, especially in underdeveloped nations. Likewise, politics has been analyzed as a form of social organization under the name of political sociology and political behaviour.

The widely ranging subject matter of social organization is given unity by the major theoretical approaches to the field which find their origin in the writings of European scholars. One major stream of thought is that of the "economic determinists" who see social relations reflecting the mode of production and economic interests. While "economic materialism" is most closely associated with the name of Karl Marx, this point of view has found expression in a variety of sociologists unconnected with Marx's political philosophy. In the United States economic determinism was transformed into a broad theory of technological development by William F. Ogburn, a leading figure in the "Chicago empirical school" of the University of Chicago. In his influential book *Social Change* (1922) he developed the theory of "cultural lag," in which he saw technology changing more rapidly than other aspects of social organization—such as culture, social knowledge, and law. Therefore, he suggested, technological development unless guided would produce instability and disruption in the social order. The economic and technological factors in social organization are also prominent in the human ecology school which was developed at the University of Chicago by Robert E. Park and his associates. A later important reformulation of the technological emphasis was presented by William F. Cottrell in *Energy and Society* (1955) in which he related the forms and uses of energy to the processes of social change.

The other stream of thought is more diffuse but has its central focus in the importance it attaches to social values, social norms (social rules), and communication systems of human groups. These writers emphasized the social bonds that men develop in the pursuit of their group interests. Émile Durkheim, in his *Division of Labor in Society* (1893), expounded the basis of the "social rules" school of social organization. It is not the technology per se but the forms of social solidarity that regulate human behaviour. In the primitive society, he saw "mechanical solidarity"—a repressive form of social organization—based on a simple division of labour. In modern society, with its complex social forms, social solidarity becomes "organic"—less repressive but more unstable and weaker. Ferdinand Tönnies formulated a related theory in which he saw social organization being transformed from *Gemeinschaft* (community) to *Gesellschaft* (society). In a *Gemeinschaft*, men are organized by social rules that are traditional and effective, whereas in a *Gesellschaft* rational self-interest operates to weaken traditional bonds. Louis Wirth in his essay "Urbanism as a Way of Life" (1938) applied these concepts to the modern metropolis. Robert Redfield (*q.v.*), the social anthropologist, expressed this outlook as the folk-urban continuum: folk society represented the social order based on a simple, small, isolated, and stable organization; the urban society is the complex, large-scale, and heterogeneous human settlement (1941).

Max Weber rejected economic determinism and emphasized social values, ideology, and beliefs systems in social organization. His *The Protestant Ethic and the Spirit of Capitalism* (1930) presented the thesis that religious beliefs were important ingredients in the development of Western industrialism. Weber analyzed the status, or noneconomic, aspects of occupation—that is,

the prestige and esteem given to different types of occupations and professions. Social status for Weber was an elemental aspect of human behaviour and, together with organizational authority, accounted for much of the form of modern social organization.

Two major efforts have been made to unify these diverse theories of social organization. Talcott Parsons' *The Structure of Social Action* (1937) represents a comprehensive effort to blend the common elements in the work of Durkheim, Weber, and others who make use of social values and social norms. While Parsons' terminology has not necessarily been generally accepted, his writings gave direction to the influential "structural functional" or "functionalism" school. Parsons is concerned with a very general theory of social organization which is not bound by concrete historical periods and which is applicable to both simple and complex social groups. His key concept is the social system, the pattern of role relations and institutions that develop among men.

Harold Lasswell and Abraham Kaplan in *Power and Society* (1950) developed an alternative synthesis of social organization theory which seeks to bridge both economic determinism and social values. For these authors, social class was too broad a concept to describe modern social organization. Instead they analyzed society in terms of more specific skill groups. In their framework, power relations are central concepts.

Despite the vigorous efforts to construct theories of social organization, the major accomplishments of the field rest mainly on the more empirical research into concrete patterns of social stratification and into particular types of organization such as community organization, bureaucracy, and the like.

1. Social Stratification.—By social stratification, the sociologist means the layers of a society which come about because social groups differ in occupation, income, prestige, and style of life. All groups are characterized by their stratification. Social stratification systems are based on hierarchies of rewards and privileges as well as differences in skill, education, and power. Under the impact of industrialism, social stratification undergoes vast changes. Therefore sociologists have accumulated a considerable body of empirical research literature on patterns of social stratification and social mobility. The earliest studies of social stratification include the purely descriptive works of Europeans such as P. G. F. Le Play and Charles Booth who sought to record the income and living conditions of family groups, especially the lower classes, whose style of life was being disrupted by the advent of the factory system. In the United States the social survey movement (mentioned above), sponsored by the pre-World War I social reformers, collected descriptive materials for purposes of social planning.

But it was in the works of the founders of American sociology in the early 20th century—for example, William Graham Sumner, Charles H. Cooley, and William I. Thomas—that there emerged a general concern with understanding the origins and consequences of social stratification. William G. Sumner, who worked extensively in primitive societies (*Folkways*, 1906), was one of the first to analyze social class relations in the United States (*What Social Classes Owe to Each Other*, 1883). Charles H. Cooley in his writings on the disruption of the moral order of modern society became concerned with the impact of class conflict on social values (*Social Organization*, 1909). In order to account for the social and personal disorganization which the Polish emigrant was experiencing in the United States during the first part of the century, W. I. Thomas and Florian Znaniecki presented their classic study, *The Polish Peasant in Europe and America* (1918–21), which has become a prototype of social stratification research.

Popular ideology in the United States has tended to ignore the existence of social classes, but the depression of the 1930s brought an increased interest in social stratification and social inequality. Under the impact of economic determinist thinking, numerous studies were done seeking to describe the consequences of social change on the class structure. C. Wright Mills (1916–1962) was one of the most energetic of this group, and his book *White Collar* (1951), which describes the decline of the old independent small business class and the rise of the salaried employee, represents this approach.

Other empirical sociologists undertook to demonstrate that social stratification was not the exclusive result of economic factors. Social scientists such as F. S. Chapin developed scales for measuring the socioeconomic position of families. By applying the techniques of social anthropology, W. Lloyd Warner and his associates prepared a series of research monographs on the social stratification systems of small communities which focused on friendship, social contact and marriage patterns, and voluntary associational life. Warner's findings underline that a person's position in the social structure is not based on economic factors per se but involves popular conceptions of occupations and social prestige.

With the growth of sample surveys and public opinion polling, it became possible to extend the study of prestige and social stratification beyond the confines of particular small local communities. Paul K. Hatt, by means of the National Opinion Research Center, collected the first nationwide data on the public's attitudes toward a wide range of occupations and established the general pattern of occupational prestige in the United States. Richard Centers, using a similar methodology, described the popular images toward the social class system (*The Psychology of Social Classes*, 1949).

Research into the status aspects of social stratification has probed the importance of race, ethnicity, religion, and age as a basis for social distinctions. Edward Franklin Frazier's classic study of *The Negro Family in the United States* (1939) focused not only on the internal dynamics of the minority family but also on the position of the Negro family in the larger stratification system based on Negro-white relations. Gunnar Myrdal in *An American Dilemma* (1944; rev. ed., 1962) presented a monumental report on the social position of the Negro, while Everett Hughes, in *French Canada in Transition* (1943), presented a comprehensive study of the changing balance of social stratification in Canada between the English Protestants and French Catholic stock. The extensive literature on social stratification in the United States and elsewhere was effectively summarized by Bernard Barber in *Social Stratification: a Comparative Analysis of Structure and Process* (1957) and also by Joseph A. Kahl in *The American Class Structure* (1957).

2. Social Mobility.—A central topic in empirical studies of social stratification has been the measurement of the amount of social mobility; that is, the movement of an individual or a family from one social position to another, either upward or downward. The theoretical foundation for the study of social mobility was presented by Pitirim Sorokin (*Social Mobility*, 1927) and Joseph Schumpeter (1915). With the accumulation of material wealth in an industrial society, sociologists have addressed themselves to the question of whether the opportunities for upward social mobility have decreased. Studies of the available data for the United States since 1900 permit the conclusion that there has not been a decline in opportunities for upward social mobility. This is based on historical analysis of specific communities in, for example, Natalie Rogoff's work on Indianapolis (1953) and on the examination of particular leader groups such as businessmen (F. W. Taussig and C. S. Joslyn, 1932, and W. Lloyd Warner and James C. Abegglen, 1955). While the opportunity for mobility has not decreased, avenues have changed and higher education has become more of a prerequisite for upward mobility.

Sociologists are in agreement as to the major changes in the social pyramid during the 20th century. There has been a growth in the size and composition of the middle-income and middle-status groups, while the lower groups have remained stable or declined. The boundaries between social groups have become blurred, rather than sharper and more distinct, and within social groups there is greater social diversity. In an industrial society, upward mobility is accompanied by downward mobility for an important minority. Changing technology and old age force some groups to decline in economic and social status.

Social stratification and social mobility studies, as they relate to the organization of a whole nation-state, supply an important basis for comparative social organization research. In connection with the Subcommittee on Social Stratification of the International Sociological Association, studies in various industrialized countries

on social mobility were completed: Great Britain (D. Glass, 1954), Germany (M. Janowitz, 1958), France (Bresard, 1950), Sweden (G. Carlsson, 1958), Denmark (K. Svalastoga, 1960), Netherlands (van Tulder, 1956), and Japan (K. Odaka, 1959).

The Russian Research Centre of Harvard University studied social mobility in the Soviet Union on the basis of interviews with refugees (Alex Inkeles and Raymond Bauer, 1959). While these studies do not permit precise comparison, it appears that the United States probably has a higher rate of social mobility than other industrialized countries, although the differences are not as great as generally assumed and patterns are tending to converge among these different countries (S. M. Lipset and Reinhard Bendix, 1959). As sociologists deepen their knowledge of social mobility they become more concerned with the consequences of social mobility. Upward mobility brings social benefits in increased income, broader social perspectives, and greater participation in community and voluntary associations; it also has its disruptive consequences. Social mobility often tends to disrupt interpersonal relations and to increase marital tensions and mental illness.

3. Community.—Closely related to the study of social stratification has been the interest in community research, which has developed a continuous tradition of firsthand observation. The original empirical community study was the description of a midwestern town reported in Albion Small and George Vincent's *Introduction to the Study of Society* (1894). Since then, there has been a continuous flow of studies of small communities, local communities in metropolitan areas, and ethnic and religious enclaves. The U.S. Department of Agriculture sponsored a series of rural community studies as part of its planning activities. The community study has been taken up wherever sociological research is practised throughout the world.

Of U.S. studies of smaller communities, the most well known are *Middletown* (1929) and *Middletown in Transition* (1937) by Robert and Helen Lynd; *The Social Life of a Modern Community* by W. Lloyd Warner and P. S. Lunt (1941); and *Elmtown's Youth* by August B. Hollingshead (1949).

Paralleling the studies of smaller communities has been the work in urban sociology, guided by a powerful curiosity to understand the rich and complicated texture of metropolitan life. Many of these studies were under the direction or stimulation of the University of Chicago in the 1920s and particularly Robert E. Park. His essay entitled "The City: Suggestions for the Investigation of Human Behavior in the Urban Environment" (1915) remains a guidepost. These sociologists were in turn influenced by Georg Simmel and his famous essay on "The Mental Life of the Metropolis" (1902). There was a strong literary quality in these original studies of urban life: Harvey Zorbaugh described *The Gold Coast and the Slum* (1929), Louis Wirth, *The Ghetto* (1928), Horace Cayton and St. Clair Drake worked in the Negro community (*The Black Metropolis*, 1945), and William Whyte wrote about the Italian slum in *Street Corner Society* (1943). These efforts were directed toward describing how the struggle for economic existence in the metropolitan community influences human values. The community was seen as a social system rooted in geography and bounded by the daily transactions of men and women as they exchange goods and services.

In the sociological view, the scope of a person's community has grown and its boundaries have become imprecise as society becomes more industrialized and more impersonal. One of the consequences of advanced industrialism has been suburbanization, the movement of persons out of the central business district, and thereby an increased separation of place of work and place of residence. A major theme in community studies has been the weakening of social bonds of community life, a growth of social disorganization, as families become more mobile in their daily existence and as modern careers require frequent moves from community to community.

In the 1950s a new cycle of community studies questioned the trends toward impersonality and disruption, or at least indicated the limits on these processes. Sociologists came to see the community of residence as the locus of essential functions, such as family formation, child rearing, the development of patterns of consumption and friendship. The newer community studies emphasized factors that make for stability and integration in residential communities (Robert Angell, 1947; Morris Janowitz, 1950; and the reports of the Detroit Area Study of the University of Michigan). Likewise, there has been increased interest in the operation of voluntary associations and political institutions of the local community which help regulate social and economic change. The prototype is Floyd Hunter's study of Atlanta, *Community Power Structure* (1953), in which he analyzed the top leadership of a metropolitan area as it controlled social welfare. Interest in the mechanics of social planning and particularly in urban renewal have produced a variety of investigations into the power realities of community life. American metropolitan centres have a pluralistic base of community power, in part because of the withdrawal of the top economic leadership from day-to-day decision. As a result, there has been a growth in the importance of professional political and organization leadership in the modern community.

4. Bureaucratic Organization.—In popular language, bureaucracy denotes an inefficient agency—public or private. (Compare, however, many meanings in BUREAUCRACY.) Bureaucracies may be defined as those large and specialized organizations men create in order to accomplish a specific task—economic, political, or even cultural. Following the writings of Max Weber, the essential characteristics of a bureaucratic organization have been identified as involving four factors: specialization; a hierarchy of authority; a system of rule, including rules for recruiting new personnel; and impersonality. The sociological analysis of bureaucracy centres on the factors that create organizational cohesion and effectiveness and those that impede administrative leaders from achieving their stated objectives.

Important general social organization treatises on bureaucracy and large-scale organization have been prepared by Chester Barnard, *The Functions of the Executive* (1938), Herbert Simon *Administrative Behavior* (1947), and Kenneth Boulding, *The Organizational Revolution* (1953). These writers described the common problems of modern bureaucracy regardless of the type of organization. They probed behind the formal rules and official organizational charts into the informal life of the organization—the status systems, authority relations, communications networks and patterns of reward and discipline which determine organizational effectiveness.

Large-scale organizations can be classified on the basis of their goals and their type of function—industrial, political, military educational, religious, or "therapeutic" (mental hospitals, correctional institutions). Research into the structure and operations of industrial organization was pioneered by Elton Mayo and his collaborators. Mayo took the measurement of productivity as the point of departure, and he emerged with the importance of small-group solidarity in productivity. He extended his work to link the factory system to the larger society. These studies have been repeated and modified by numerous investigators and supply an important basis for industrial sociology. The basic outline of the Mayo school with its strong perspective toward understanding the industrial system from the standpoint of managerial problems was presented by F. J. Roethlisberger and William J. Dickson in *Management and the Worker* (1939) and by T. N. Whitehead in *Leadership in a Free Society* (1936). A broad overview of industrial organization in the United States was formulated by Wilbert Moore, *Industrial Relations and the Social Order* (1946). Russian industrial organization was described by J. S. Berliner in *Factory and Manager in the U.S.S.R.* (1957); Reinhard Bendix in *Work and Authority in Industry* (1956) undertook one of the few efforts at cross-national comparison of industrial organization contrasting the United States, Great Britain, the U.S.S.R., and East Germany.

Because of the interest of sociologists in social reform, significant research is available on correctional institutions, social welfare, and health organizations, particularly mental health. The main theme of these researches has been the organizational barriers which prevent such institutions from achieving their stated goals. Donald Clemmer in 1940 wrote on the social organization of the

prison, and Gresham Sykes subsequently analyzed the maximum security prison (*Society of Captives*, 1958). Comprehensive treatment of the mental hospital was presented by Alfred Stanton and Morris Schwartz (1954), and by Ivan Belknap in *Human Problems of a State Mental Hospital* (1956). The National Institute of Mental Health of the U.S. Public Health Service sponsored social organization research on therapeutic institutions.

Interest in the organizational features of political parties was stimulated by the writing of Robert Michels on *Political Parties* (1915), who combined political polemics with sociological research. His so-called "iron law of oligarchy" argued that political democracy was impossible because the internal organization of political parties was inevitably undemocratic. For the United States, no adequate sociological studies exist on the structure of the major political parties or on the organs of government. The literature has been limited to specific case studies, such as Philip Selznick, *The TVA and the Grass Roots* (1949), and S. M. Lipset, M. Trow, and J. Coleman, *Union Democracy* (1956), which reports on the internal political structure of a small atypical union. More comprehensive sociological studies have been prepared on totalitarian political organization: Franz Neumann, *Behemoth* (1942) for Nazi Germany, and Barrington Moore, Jr., *Soviet Politics* (1950) and *Terror and Progress* (1954) on the Soviet Union.

Emerging areas of interest in bureaucratic organizations include educational institutions, Theodore Caplow and R. J. McGee, *The Academic Marketplace* (1958), Paul F. Lazarsfeld and Wagner Thielens, *The Academic Mind* (1958); the military, Morris Janowitz, *Sociology and the Military Establishment* (1959); the government bureau, Peter Blau, *The Dynamics of Bureaucracy* (1955); and the volunteer social service association, David Sills, *The Volunteers* (1958).

5. Small Groups.—Sociologists have been traditionally interested in specific small groups such as the family, the delinquent gang, or the neighbourhood play group. They collect extensive data on changing functions of the family and assemble statistics on marriage and divorce rates. Delinquent groups were first studied by Frederic Thrasher in the period after World War I (*The Gang*, 1936). Informal groups in voluntary associations and in leisure-time activities have also been the object of research. But the field of small groups as an essential aspect of social organization emerged only after 1945 when the importance of small primary groups was rediscovered. This was a rediscovery because Charles H. Cooley in *Social Organization* (1909) had earlier indicated on theoretical grounds the importance of these groupings for the social and moral order.

Primary groups are those small, relatively persistent, face-to-face associations—for example, family, peer groups, work groups. These groups relate the individual to the large units of social organization, the stratification system, the community, and large-scale organizations. Primary groups strengthen the human social bonds and condition their moral character, since in primary groups the demands of economic and political life are adjusted to the persistent needs of men. These aspects of primary groups were set forth by Edward A. Shils in his essay "The Study of Primary Groups" (1951). The primary group concept stands as a bridge between sociology and psychology because it is a mechanism by which individuals learn to accept and incorporate the social values in their social environment. This process has come to be called "socialization" and it extends throughout the entire life of the individual.

Primary groups supply the basis for social solidarity and group cohesion in large-scale organizations. Sociologists during World War II were impressed by the importance of primary groups in military life. Soldiers fought less for ideological and nationalist reasons and more for the respect and love of their immediate comrades. In industrial research, primary group structures are related to absenteeism, morale, and productivity, in that primary group solidarity has been found to be related to high industrial morale and low absenteeism. However, primary group solidarity results in opposition to managerial rules when workers feel themselves deprived or frustrated or when trade unions become a competing basis of loyalty and social solidarity.

Primary group relations have been found to be essential in facilitating a person's assimilation into a residential community. Children are the active agents in stimulating family friendships and neighbourly contacts. These primary group contacts and informal lines of communication operate to increase participation in community affairs. The political process and the influence of the mass media also work through primary group networks.

Lazarsfeld, Berelson, and Gaudet found in their empirical study of voting behaviour in Erie County (Ohio) that the pressure and influence of primary group members operated in the case of those persons who changed their voting intentions during the political campaign (*The People's Choice*, 1944; 2nd ed., 1948).

The study of primary groups has been developed by the use of specialized research methodologies. Sociometric techniques, devices by which persons express their preferences for friends and group members, make it possible to chart primary group structures. J. L. Moreno developed the sociometric techniques in *Who Shall Survive?* (1934), and they have been applied to a great variety of organizations from schools for delinquent girls to military combat teams. Primary groups are also studied by means of laboratory replications of real-life situations (K. Lewin, R. Lippitt, and S. K. Escalona, 1940). Sociologists also emphasize the importance of direct and careful observation of small groups in their natural social environment (George Homans, 1950).

6. Social Disorganization and Deviant Behaviour.—The basic concepts of social organization have been used extensively to explain deviant behaviour of individuals. Émile Durkheim's investigation of *Suicide* (1897) set the pattern for the field of social organization. He demonstrated that this form of extreme behaviour—generally considered to be a personal and psychological event—varied in different social groups and was related to the structure of society. He discovered that suicide rates varied among different nations, different religious groups, different classes and under different economic conditions. While subsequent researches have refined his methods and conclusions, his basic theory is still accepted as relevant for understanding deviant behaviour. The sociological basis of personal disorganization was enhanced by the writings of W. I. Thomas, who saw deviant behaviour as the result of a clash between primary group needs and cultural values of the larger society. These approaches have been applied to insanity, crime and delinquency, and alcoholism, as manifestations of social disorganization. In the area of psychopathology, Ellsworth Faris and H. Warren Dunham charted the higher incidence of severe mental disorders in working class areas (1939); those findings were amplified by August Hollingshead in *Social Class and Mental Illness* (1958). Hollingshead also concluded that the lower classes had fewer available facilities for dealing with mental illness. Research into the social factors of mental illness has called into question popular views, as it has in other areas of deviant behaviour, that with industrialism there is a marked increase in personal disorganization. Herbert Goldhamer and Andrew Marshall found that in selected areas in the United States there was no increase in the rate of severe mental illness after the middle of the 19th century (*Psychosis and Civilization*, 1950). Likewise, careful comparisons of mental illness rates in modern urbanized communities and rural folk communities, such as the Hutterites, indicate little difference in the amount of mental illness, but rather differences in the type and form of illness (Joseph W. Eaton and Robert J. Weil, *Culture and Mental Disorders*, 1955).

A variety of approaches have been used to link crime and juvenile delinquency to social organization. Numerous statistical studies initiated by Clifford Shaw and Henry McKay locate the highest incidence of crimes in low socioeconomic census tracts and geographical areas. Changes in homicide rates were found closely associated with the business cycles by Andrew Henry and James Short (*Suicide and Homicide*, 1954). The subculture of criminal and juvenile offenders was extensively studied by McKay, Thrasher, A. Cohen, and L. Ohlin and R. Cloward. As in the case of mental disorder, not only is crime more heavily concentrated in the lower social strata, but the social facilities for avoiding criminal careers are less accessible in lower socioeconomic communities. Sociologists have also studied alcoholism and found

it closely related to the norms and values of particular ethnic subcultures.

7. Social Control and Social Change.—As the various aspects of social organization theory and empirical research matured, sociologists came to recognize the necessity for general synthesis that would explain the basic transformations of society. The study of a society as a whole undergoing transformation has come to be designated as the study of social control or social change. It includes both primitive society undergoing modernization and industrialized societies entering the advanced phases of automation and high-energy technology. The study of social control and social change stands at the frontier of the field of social organization.

The original social organization theorists, both of the economic determinist and the social-value schools, had implicit theories of social change. While these theories remain of importance in the intellectual history of sociology, they are considered to be too narrow and too simple to account for the complexities of social change. Karl Mannheim in *Man and Society in an Age of Reconstruction* (1940) initiated a new phase in the study of social control and social change. His writing reflected the introduction of three important trends. First, economic and technological factors in social change were extended to include the organization of science and of scientific knowledge. Under his influence a sociology of science and a sociology of knowledge developed because of the central importance of science in social change. Sociologists began to study science as an organized institution and profession, the social conditions of which bring about the development of science and the social consequences of science (Bernard Barber, *Science and the Social Order*, 1952). Second, under Mannheim's influence, sociologists broadened their concepts of social stratification to include the study of leadership and the relations of leaders to their followers. The writings of Harold D. Lasswell have been important in directing this trend. The social composition, career experiences, and perspectives of elite groups have come to be seen as important indicators of the direction of social change. (Reinhard Bendix, *The Higher Civil Servants in American Society*, 1949; C. W. Mills and Helen Schneider, *The New Men of Power, America's Labor Leaders*, 1948; Morris Janowitz, *The Professional Soldier*, 1960; Donald R. Matthews, *U.S. Senators and Their World*, 1960). Third, students of social organization have demonstrated interest in social control factors which make possible political democracy (Edward A. Shils, *The Torment of Secrecy*, 1956, and William Kornhauser, *The Politics of Mass Society*, 1959). This perspective has also influenced the study of social control and social change in the new nations.

8. Professional Practice.—Students of social organization have become concerned with the application of these findings to social policy. Traditionally, sociologists limited their work to analysis and diagnosis of social problems and the causes of social and personal disorganization. With the development of the theoretical basis of social organization a broader perspective has emerged. Sociologists have observed that modern society relies heavily on skilled personnel to deal with the problems of social control. It is dominated by professionals and by a trend toward more professionalization.

The study of professions is regarded as a legitimate object of sociological inquiry. Professions are groups of skilled specialists who are organized to regulate the services they perform for society. A distinction can be made between the free professions, like the lawyer and the private doctor, who work independently and directly serve their clients, and the bureaucratic professions, like the schoolteacher who serves in large institutions. The trend is toward the bureaucratic professional, raising new problems of self-regulation. The British social scientists A. M. Carr-Saunders and P. A. Wilson assembled in 1933 the basic documentation of the growth of the profession. In the United States both Talcott Parsons and Everett Hughes developed theoretical concepts for the analysis of professional behaviour. A series of monographs was prepared on the sociological aspects of specific professions, such as professional training, career lines, and internal organization. Among the professions that have been analyzed in some detail are the legal profession, social workers, the military, and the academic.

Sociologists have come to see the professions as a major vehicle for applying the findings of social organization research to the solution of social problems. The Russell Sage Foundation actively stimulated sociological research into the professions. The objectives are, first, to increase the self-awareness of professional groups; second, to supply professional groups with accurate knowledge of the sociological characteristics of their clients and of problems facing their profession; and, third, to assist them in evaluating the effectiveness of the professional services they render.

The application of social organization knowledge centres in the professional schools and professional training. As a result, there has been an increased movement to incorporate sociologists and students of social organization into the faculties of schools of social work, public health, medicine, education, nursing, law, and business. Sociologists also serve in professional agencies and in industry and governmental agencies as research and staff specialists. (Mo. J.)

B. Social Psychology

Social psychology is concerned with behaviour as it is influenced by the actual or implied presence of others. It is believed that the foundations for the characteristically human behaviours of being sympathetic, joking, conforming to expectations, experiencing shame or pride, and forming preferences are developed in the earliest mother-infant relations and subsequently elaborated in relations with others. The understanding of this socialization and the social functions of language and communication is a basic concern of social psychology.

In the study of attitudes the social psychologist is concerned with the processes that influence attitude formation and change. In the study of group process, the social psychologist is concerned with the consequences of attainment of group goals and the role differentiation that attends these adaptations.

In the study of social movements, leadership, interpersonal perception, social rituals, and collective behaviour, the social psychologist builds upon his basic concerns with personality development and group process. He employs experimentation, survey methods, documentary study, and controlled and participant observation. The study of morale, team efficiency, voting, and consumer preference—largely conducted by social psychologists—is one of the most vigorous areas of applied social science. For a further treatment *see* Psychology, Social. (F. L. St.)

C. Human Ecology

Among the many theoretical schemes found in contemporary sociology, one has been prominently concerned with the community. "Human ecology" has drawn upon selected concepts, assumptions, and postulates originally developed in the life sciences in order to describe and explain the structure of human communities. Although the approach was developed primarily by sociologists in the U.S., a direct counterpart emerged in the "social morphology" of the Durkheimian school in France. In fact, ecological thinking can be found in a number of disciplines, and the interests of sociologists working in both the ecological and morphological traditions clearly overlap those of economists, geographers, and anthropologists; indeed, parallels can be found in every area of scholarship that deals with the relations between man and environment.

From its inception, the unique contribution afforded by the sociological version of human ecology has stemmed from its distinctive attack upon certain problems of social organization or structure. First of all, "man" is studied in aggregates, whether large or small, and attention is focused on the structure of relationships between and among aggregates. Secondly, the "environment" is conceived in broad terms, with full appreciation of man's considerable capacity for altering his surroundings. In some instances, the ecologist's interest is in the interconnections between such familiar units as households and firms. These units are regarded as the structural cells or building blocks that make up the community, and the various interactions and linkages among them

have attracted the attention of ecologists concerned with the nature of the community's "internal" structure. At another level of aggregation, certain "external" relations between and among communities are examined, so that whole communities, in their turn, become the building blocks making up still larger systems, *e.g.*, regions and nation-states.

Within sociology, the field has had a rather curious history. The "classical" version of human ecology emerged out of the intellectual ferment generated by a series of naturalistic studies of urban life initiated in the 1920s by R. E. Park, E. W. Burgess, and their students (including R. D. McKenzie and J. A. Quinn). Their curiosity was aroused by the apparent order that somehow developed out of the diverse activities of literally thousands of individuals—individuals who had come from many places, who were largely unknown to each other, and who were each seemingly pursuing disparate and even contradictory ends. In particular, these students of city life became fascinated by the order that was manifested in space.

The original Chicago approach, which had flowered in the 1920s and the early 1930s, underwent a series of searching critical examinations in the late 1930s and the 1940s; most notable among these critiques were those by M. A. Alihan (1938) and W. I. Firey (1947), two works which gave rise to a view that came to be known as "cultural" ecology. But the classical version itself actually did not receive a fully developed statement until 1950, when Quinn's *Human Ecology* provided a synoptic review of the empirical literature, organizing a vast amount of research from many sources within the broad spatial framework sketched earlier by Park.

Space has remained a central concern to ecologists. Nevertheless, it remained for a student of Quinn and McKenzie to broaden the focus of the approach. In his *Human Ecology* (1950), A. H. Hawley defined human ecology as the study of the form and development of the community. Spatial patterns were given close attention, but the principal concern was with the structure of the community, with structure viewed as a "collective adaptation" on the part of a localized population to its total environment.

In addition to organizational configurations observable at a given point in time, ecologists have always shown keen interest in change in the internal structure of the community over time, particularly as it accompanies population growth, and in shifting relations among communities. Only a few of these broad theoretical interests have received detailed empirical treatment. Research in human ecology has been chiefly concerned with such subjects as the spatial distribution of distinctive groups (racial, ethnic, religious, and occupational) within urban areas, the location of manufacturing firms and business establishments, the relationships between workplace and residence, and the territorial division of labour within and among communities.

This brief review should suggest that structural questions are at the heart of human ecology. The traditional focus upon spatial patterns persisted in ecological research although ecological theory has been broadened—at least in the more programmatic statements—to include a systematic treatment of the spatial and temporal manifestations of community structure as simply aspects of that structure. In this view, space and time are no more than dimensions upon which structural patterns may be conveniently observed and measured. At one time, then, the ecological viewpoint was applied chiefly to the internal areal arrangement of the large city; but it has been extended over the years to deal with communities of all sizes and types and to explore their external relations with the outside world. As a consequence, human ecology promises to yield at least one kind of comparative sociology of the community. *See* further ECOLOGY: *Human Ecology*.

D. Demography

Demography, the study of population, is by no means limited to sociologists; it also attracts biologists, geneticists, medical men, public health specialists, biometricians, actuaries, statisticians, and economists. Only in the United States is technical demographic training concentrated in academic departments of sociology; in other parts of the world, demographers are likely to receive their training in one of the other fields.

Demography has been formally defined as "the study of the size, territorial distribution, and composition of population, changes therein, and the components of such changes, which may be identified as natality, mortality, territorial movement (migration), and social mobility (change of status)" (P. M. Hauser and O. D. Duncan, "Overview and Conclusions," in Hauser and Duncan, eds., *The Study of Population: an Inventory and Appraisal*, p. 2, University of Chicago Press, 1959). Each term in the definition denotes a topic that is subject to sociological study, and the following paragraphs set out some of the ways in which the main problems of population are studied.

Demographers deal with population structures and processes. The first three facets of population—size, distribution, and composition—are basically static aspects, in that census enumerations, whether complete or partial (via sampling), yield a "snapshot" as of a given point in time.

(1) Population *size* in relation to the means of subsistence has long been regarded as "the population problem." At least since the days of Malthus, Western writers have attempted to deal with the balance between human numbers and the capacity for their support. Sociological contributions to an understanding of this complex problem have consisted largely of showing how variable elements of culture (such as technology, patterns of social organization, and legal or traditional codes of conduct) affect the balance between population and resources. At least a few sociologists have also been concerned with the implications of population size for patterns of social organization and forms of interaction.

(2) Territorial *distribution* and redistribution of population requires the study of human numbers in relation to area. In its simplest form, the problem resolves into an examination of population density. Sociologists have long been concerned with the distribution of population between rural and urban areas. Urbanization is viewed as a process involving population concentration, the appearance of large and densely settled agglomerations, and a shift in the balance between rural and urban sectors. Explanations of this process have been advanced by sociologists and others, and a number of scholars have also come to be interested in the redistribution of population in and around urban areas; in the U.S., for example, this has given rise to studies of "suburbanization," while in the United Kingdom the emergence of "conurbations" has attracted considerable attention. Again, the components of population change must be examined, for both migration and natural increase (the balance between births and deaths) may bring about changes in the distribution of population between areas. Finally, sociologists have studied the effects of urbanization, the consequences of depopulation of certain areas, and the implications of varying density levels for patterns of social organization and forms of social interaction (*e.g.*, in the city as contrasted with the countryside).

(3) The *composition* of population refers simply to the characteristics of human aggregates, of which any number may be distinguished, ranging from those found in classifications based on purely biological traits such as blood types, through racial and ethnic subdivisions, to such social and economic attributes as marital status and occupation. Questions of population "quality," or the relative predominance of desirable and undesirable physical or social traits, also entail a compositional view of population. There are numerous sociological analyses of the "functional" composition of population, as in studies of the occupational and industrial structure, changes in the labour force, and shifts in the level of formal education; these materials yield basic data on social structure. In addition, "biological" composition is relevant to sociologists. For example, the simple age-sex composition of a population—which reflects past variations in natality, mortality, and migration—sets one of the important conditions to which that population's social organization must adapt. Sociologists have also come to be concerned with the "aging" of population, at least from the standpoint of the social problems it may engender. Changes in age composition are particularly amenable to mathematical analysis (*see* POPULATION: *Knowledge About Population*).

The demographic *components of change* are events that produce

alterations in population size, distribution, or composition. In studies of population size (*e.g.*, growth) or of the redistribution of population between areas, the demographer need only be concerned with natality, mortality, and migration. The population of the world as a whole is, of course, affected only by the balance between births and deaths (natural increase or decrease), but studies of population change in particular parts of the world have always had to take account of the effects of net migration as well. The population of the New World, for example, grew in two ways —by a surplus of births over deaths, and by a surplus of in-migrants over out-migrants. Similarly, analyses of redistribution within countries (*e.g.*, urbanization) have had to establish the levels of natality and mortality in rural and urban areas as well as to determine the number of migrants flowing between them.

The study of changes in population composition is often complicated by the fact that some types of composition are subject to change via "social mobility," as well as by births, deaths, and migratory movements. Changes in age-sex composition in a given period can be fully explained by reference to these last three components, but the marital composition of a population, at least over a brief time span, can obviously be altered considerably without any births, deaths, or migrations if a substantial number of persons marry or secure a divorce. Although not ordinarily so labeled, these individual shifts in status are forms of social mobility in the generic sense suggested in the definition of demography set out above. The four components are taken up below as subjects of sociological study.

(1) *Natality*, or fertility, is of interest in its own right, aside from its importance as a component of population change. Sociologists have been analyzing trends in the level of natality ever since reasonably reliable data on births in particular areas began to become available in the 19th century. Trends in countries with different socioeconomic systems, and with dissimilar cultural traditions, have been compared, as have the trends in rural and urban areas within countries. An area of sociological inquiry into demographic behaviour that has received an enormous amount of attention is that of fertility "differentials"; in addition to variations by place of residence, sociologists have examined birth rates according to such indicators of socioeconomic status as education, occupation, and income, and according to ethnic background and religious affiliation. Much of this research is social-psychological in orientation, entailing a search for the attitudes and values that might predispose individuals toward exercising control over the number and spacing of births.

(2) *Mortality* has proved to be somewhat less interesting to sociologists, despite its importance as a component of population change; this area of research has been left to others—biostatisticians, epidemiologists, and actuaries. The gross trends in the death rate are fairly well known, and some sociological attention has been given to the analysis of rural-urban and socioeconomic differentials in mortality. In particular, "social class" differentials have been subjected to close scrutiny in Great Britain. Falling death rates in underdeveloped areas, particularly after World War II, played a major role in the explosive growth of world population and drew the attention of sociologically oriented demographers to the causes and consequences of variations in mortality.

(3) If mortality has been relatively neglected by sociologists, the same cannot be said for *migration*, or territorial movement. In the modern period, following the emergence of the nation-state, students of migration have come to distinguish between "international" and "internal" movement, with the latter term referring to circulation within politically bounded areas. Sociologists have been interested in the size, direction, and composition of migrant streams, in examining the selectivity of migration, in establishing the causes of migration, and in assessing the consequences of migration for the individual migrant and for the societies and communities of origin and destination. Although internal rural-urban movement drew most sociological attention, other migrant streams have been examined. Within urban areas, for example, residential shifts between inner and outer parts of the built-up area received rather detailed treatment in sociological literature. Finally, motives for migration of all types have been sought by sociologists working within a social-psychological framework.

(4) *Social mobility*, or change of status, has received a great deal of attention by sociologists, who ordinarily use the term to refer to movement between ranked statuses, such as occupations of different prestige. However, only a few mobility problems have been examined from a demographic perspective, *e.g.*, the relation between occupational mobility, on the one hand, and migration and fertility, on the other. Accessions to and separations from the labour force have been long-standing concerns of those who deal with manpower problems, and sophisticated methods have been employed, including the construction of tables of working life, a variant of life table techniques. Still another area is that of the family cycle. The processes of family formation and dissolution (including marriage, divorce, separation, widowhood, etc.) are changes of status that have attracted research attention. Since marriage is an event that can be dated, it is amenable to treatment in terms of the stable population model; as a result, the connection between nuptiality and fertility has been increasingly subjected to mathematically sophisticated treatments, to the benefit of both the sociology of the family and formal demography.

The study of population starts from an intrinsically quantitative base, and the quantities dealt with—human numbers—are vital facts in every field dealing with the human species. Indeed, the utility of population data has come to be widely appreciated by those engaged in policy formulation and decision making, who have made increasing use of demographic work over the decades. In sociology, the fundamental importance of population lies in the impact of demographic factors upon the structure and functioning of community and society. Sociologists have a strategic role to play in providing a fuller understanding of the social determinants of demographic structures and processes.

See also POPULATION; VITAL STATISTICS. (L. F. Sc.)

E. Rural and Urban Sociology

Rural sociology, which developed out of interest in the study of rural life and rural social organization, was recognized as a specialized branch of sociology by 1916, when the meetings of the American Sociological Association were devoted to the central topic of rural sociology. Its counterpart, urban sociology, which developed as a specialization somewhat later out of concern with the sociology of urban phenomena and urban communities, was similarly recognized by the Association in 1925.

Rural Sociology.—Beginning about the middle of the 19th century, with the Industrial Revolution well under way, there was interest in and recognition of the contrasts between rural life and life in the urban industrial centres. However, it was not in Great Britain, where the Industrial Revolution began, but on the continent that an interest developed in the life of rural people. Frédéric Le Play (1879) was among the early students of peasant life with a particular interest in the peasant family. His classic case studies were among the early systematic accounts, although they tended to idealize peasant life based upon the solidarity and stability of the family. Many decades later Carle C. Zimmerman (1935) followed Le Play's family case study approach in his characterization of rural life in problem areas of the United States.

There was also an interest in the social and political aspects of peasant society in Europe. Karl Marx was concerned with the relationship of the peasant to the land and with the development of an agrarian theory consistent with his socialist doctrine. Weber was also concerned with the landholding and communal aspects of peasant life. Thomas and Znaniecki's later study (1918–21) of the Polish peasant in Europe and America is one of the classic studies of peasant society, one in transition from an old established rural culture to a new and urbanized culture. Robert Redfield (1930, 1941) contributed the concept of the "folk society" in contrast to urban society, based upon his studies in Central America. He described the changes in kinship and institutional structures and changes in beliefs and in manners of dress and behaviour as one moved from the isolated village to the urban centre. Oscar Lewis (1951), studying the same villages about two decades later, found tension, factionalism, and other evidence of disorganization.

Another major development in the study of rural society con-

cerned the village. Sir Henry Maine (1871) compared village communities in Eastern and in Western society, distinguishing between types of villages according to the nature of the bonds which held them together. E. Dwight Sanderson's extensive study (1932) of the rural communities around the world was a major contribution to the understanding of the many forms of settlement and structure of rural life and how they were influenced by geographic, economic, and cultural factors. The contrast between closely settled villages, line villages, and scattered farm settlement has been of particular interest to rural sociologists in Europe and in America.

The difference in the structure and content of rural and urban society attracted the interests of sociologists in Europe and in the United States. Thus, rural society, according to Tönnies (1926), is characterized as a society based upon kinship and a commonality of interests (*Gemeinschaft*, community), and the city as a society in which specialized interests dominate (*Gesellschaft*, society). Durkheim (1893) provided another basis of distinction between the small community in which relationships are based upon kinship and homogeneity (mechanical solidarity) from that in which there is a specialization of relationships united through larger institutional structures (organic solidarity). J. H. Kolb (1946), drawing from Cooley, described the greater emphasis upon primary rather than upon secondary groups in rural society. Relationships are more personal, intimate, and face-to-face, as in the rural neighbourhood, in contrast to the more formalized relationship in the larger community. Charles Loomis (1950) drawing from Tönnies attempted to classify groups within rural society along a continuum from the highly *Gemeinschaft* Amish society to the *Gesellschaft*-like government agencies serving rural society.

As the tools and the methods of rural sociology have become more refined, an interest developed in the structure of the rural community. Josiah Galpin's classic study *The Social Anatomy of an Agricultural Community* (1915) led to a definition of the rural community in ecological terms showing the interdependence of the town with its rural hinterland in an area of scattered farmsteads. Kolb (1959) and others in Wisconsin followed that tradition and showed the interdependence not only of the farm population with those living in the towns but also the interdependence of the farm population with the towns of various sizes within an area. For example, the small towns serve as social centres and as the location of the school, or church, or country store, while the larger centres provide more specialized materials, goods, and services for the area. Kolb and his students showed how the neighbourhood based upon local institutions, mutual aid, and interaction had special significance in the American rural scene, in which agricultural villages common in Europe and other countries were absent.

Improved means of transportation and contact in rural areas has led to the structuring of social relationships according to like social and economic interests. The social clique has replaced the neighbourhood in rural society, in which social class becomes the basis of social interaction, rather than proximity or kinship. Social cliques become important in the communication of information, for the control of institutions, and for the socialization of the young. Along with this, the development of cooperatives, farmers' organizations, and agricultural agencies provides more effective channels of information, influence, and services for the farmer. This change in the structure of rural society is described in A. Gallaher's restudy (1961) of a central U.S. community, studied earlier by James West (1945); in A. Vidich and J. Bensman's *Small Town in Mass Society* (1958); in the proceedings of the Rural Sociological Society in 1961, *Our Changing Rural Society;* and in the proceedings of the First World Congress of Rural Sociology, in 1964, published in *Sociologia Ruralis* (1964). While rural communities have struggled to maintain their local institutions and customs, they have become increasingly dependent upon governmental agencies, which in turn have implanted the more universal and secular standards of the larger society upon them.

As urban centres have grown, rural sociologists have become more concerned with settlement patterns around urban centres. They have shown how the fringe areas of urban centres tend to attract people employed in the cities but having rural interests and background. These urban areas attract those interested in cheap land and more space. W. A. Anderson and Olaf Larson (1953) showed how the nature of social participation and social relationships in the fringe areas are influenced by both rural and urban interests and attitudes. Later studies concerned the influence of industrialization upon rural communities. The fringe area beyond the suburbs is one in which rural institutions exist alongside urban developments, usually with little public planning or control. Rural sociologists studying the fringe areas have been more concerned with the changes occurring in rural institutions, while others have been concerned with the problems of delinquency, political attitudes, and governmental structures. O. D. Duncan and A. J. Reiss (1956) compared the characteristics of rural and urban centres of various sizes showing how size and type of centres affect the characteristics of their populations (*see* also *Demography*, above).

Hundreds of descriptive studies of rural communities contributed greatly not only to knowledge of rural life but also to knowledge of social structure and process in general. Studies such as those by Kulp (1935) of the Chinese peasant, T. Fukutake in Japan (1963), C. M. Arensberg and S. T. Kimball in Ireland (1940), S. C. Dube in India (1958), E. C. Banfield in Italy (1958), Donald Pierson in Brazil (1951), Orlando Fals-Borda in Colombia (1955), Horace Miner in French Canada (1939), W. M. Williams in England (1956 and 1963), and others are sources on the values, institutional patterns, social change, and other processes in rural life throughout the world. Anthropologists made many of these studies using a detailed and descriptive approach; rural sociologists tended to investigate more specific issues, such as kinship, institutional patterns, and the processes of change in rural communities. The scattered community studies resulted in few generalizations, however, about social structure and processes in rural life. More systematic studies such as those sponsored by the Bureau of Agricultural Economics of the United States Department of Agriculture during the 1940s under Carl C. Taylor and the study of four regions of Puerto Rico by Julian Steward (1956) described the regional variations in rural life patterns, and in particular their reaction to various agricultural programs.

The selectivity of rural-urban migration and its consequences both for the rural community and for urban areas were among the first topics of concern for rural sociologists in the United States and in Europe. Studies showed, for example, that young people from small farm and tenant families tend to migrate to the cities at high rates. In the United States females migrated at a younger age than did males, and Negroes more than white people. With improved transportation, sons and daughters of families in the less productive rural areas tend to go to the city to work but to maintain their ties to their families for social as well as economic reasons. This pattern is similar to that of the stem family in Europe, described by Le Play, in which family members working in the cities do not sever ties with the home family.

While the characteristics of rural-urban migration in Western society have been described extensively, the underdeveloped countries, as in some parts of Latin America, present new patterns of rural-urban migration. Whole families move and take their rural culture with them. Without status or security on the land, they move to the city in the hope of a better means of subsistence. They inhabit the slum areas mushrooming around the industrial centres and participate little in the life of the city.

The choice of education, occupation, and residence by rural youth is one of the more recent topics of study in rural areas. The interrelationship of occupational and educational aspirations, choice, and attainment as related to spatial and social mobility has been extensively studied in the United States by W. H. Sewell and others. The disadvantaged position of rural youth due to their lack of educational opportunities, lack of knowledge of urban occupations, and the lack of parental and community support has been well documented. Although consolidation and improvement of rural schools helped overcome the disadvantages, a wide gap remains between the potential of rural born and reared youth and their occupational attainments. The proportion of illiterates in underdeveloped societies is, as would be expected, always greater in the rural than in urban areas.

A social psychology of rural life began with the studies of J. M. Williams (1925). Williams showed how the attitudes of rural people in a rural community were influenced by the nature of its environment, its frontier character, and its cultural heritage. The fatalism, skepticism, frugality, and self-reliance are due in part to the uncertainties of the natural forces upon which they depend and to the puritanical tradition from which they come. W. A. Anderson (1947), Edgar Schuler (1938), Howard W. Beers (1941), and others investigated rural persons' attitudes toward rural life, toward agricultural programs, and toward their future, which vary by social class, by tenure, and by other characteristics. Not until the studies by Leland Stott (1939) and Ray Mangus (1948) on the personality characteristics of children and W. H. Sewell (1952) on the socialization of children were studies directed specifically toward personality characteristics and their differences between rural farm and town people. They helped to explode the myth that rural farm children rate higher in what are considered desirable personality traits than those from towns and cities.

Social and technological change began to receive attention by rural sociologists in the 1950s. Charles Hoffer in Michigan (1942), B. Ryan and N. Gross in Iowa (1950), E. Wilkening in North Carolina (1950), H. F. Lionberger in Missouri (1955 and 1957), and C. P. Marsh and A. L. Coleman in Kentucky (1954) were among the first to attempt to study the factors affecting the adoption of innovations such as hybrid corn, improved varieties of seed, fertilizer, and various other techniques. Lionberger (1961) and E. Rogers (1962) attempted to summarize the findings of those studies and contributions to the development of a body of generalizations on the communication of agricultural information and the adoption of innovations in agriculture. Institutionalized sources of information, for example, were found to be most effective for innovations that constitute rather marked changes in farming operations and for persons of middle and upper socioeconomic status. Personal contact of friends, neighbours, and relatives remained most important for farmers with low education and with relatively few contacts outside their own locality. The distinctive characteristics of the "innovator," the "early adopter," and other persons were also described. Studies in other countries by A. W. van den Ban (1958), B. Benvenuti (1961), G. Jones (1960), S. A. Rahim (1961), and F. Fliegel and F. Oliveira (1963) extended to other cultures the generalizations about technological change in agriculture.

Among later developments is the study of social and economic change in underdeveloped areas. Since the majority of the people of underdeveloped areas are rural, the understanding and solution of their problems require knowledge of the structure and cultural features of rural society. Anthropologists have studied the ways in which their integrated social and cultural systems affect receptivity to change and in turn are affected by change. They have pointed out the need for considering the communities and the cultures in these countries as total entities in which there is an interdependence among the technological, economic, social, and psychological factors. Sociologists have been more concerned with the problem of how specific types of change occur and how they are affected by both internal and external forces. The study of community development has been the point of departure for many studies of rural people, their institutions, and programs of change in India, Pakistan, and the Philippines.

In addition to general treatment of the sociology of rural life in the United States (listed in the bibliography) studies sponsored by the U.S. Department of State in Latin America include N. L. Whetten's *Rural Mexico* (1948) and later his *Guatemala: the Land and the People* (1961); Lowry Nelson's *Rural Cuba* (1950); Olen Leonard's *Bolivia, Land, People and Institutions* (1952); Carl C. Taylor's *Rural Life in Argentina* (1948); and T. Lynn Smith's *Brazil: People and Institutions* (1954). Similar studies in other parts of the world have included Irwin Sanders' *Balkan Village* (1949); G. F. Rivera and Robert T. McMillan's *The Rural Philippines* (1952); Arthur F. Raper's *The Japanese Village in Transition* (1950); and Arthur F. Raper and M. Jensen's *Rural Taiwan (Formosa): Problem and Promise* (1953). As the involvement of the United States in World War II and then in reconstruction and development increased during the 1940s and 1950s the assistance of rural sociologists was sought for work in predominantly rural countries. At the same time an interest in rural life and institutions developed in Europe, particularly in the Netherlands under the influence of E. W. Hofstee, who made numerous demographic and sociological studies of rural life, summarized in his *Rural Life and Rural Welfare in the Netherlands* (1957). In England H. E. Bracey's *English Rural Life* (1959) filled a gap in the knowledge of English rural institutions and processes. Peter von Blanckenburg contributed a sociological account of rural life in Europe in his *Einführung in die Agrarsoziologie* (1962). Akshaya R. Desai's *Introduction to Rural Sociology in India* (1953) was the first attempt to bring together materials on the villages and institutions of rural India.

Recent works in the United States emphasize the nature of rural life in a highly urbanized society in which the farm population makes up no more than 7% of the total. Farming is becoming a large-scale business, and rural areas are invaded by city workers, recreation seekers, and conservation interests. In the United States, *Rural Sociology*, the journal of the Rural Sociological Society, completed its 30th year of publication in 1965 and has broadened its coverage to include materials from many countries. Others include *Sociologia Ruralis*, initiated by the European Society of Rural Sociology in 1957; *Quaderni di Sociologia Rurale*, published by the Italian Society for Rural Sociology, beginning in 1961; *Wieś współczesna*, published in Poland since 1956; *Journal of the Pakistan Academy for Rural Development*, since 1959; and the *Revue des Études Rurales*, beginning in 1959.
(E. A. W.)

Urban Sociology.—Sociological interest in urban phenomena began to grow in the latter part of the 19th century when a number of important studies were undertaken in England and the United States. The classic English study of this period was the 17-volume work of Charles Booth, *Life and Labour of the People in London* (1889–1903), an empirical investigation of the state of the poor in that city. The most impressive U.S. study of the period was Adna Weber's investigation of comparative urbanization worldwide.

Systematic development of urban sociology first occurred in the department of sociology at the University of Chicago through the work of Robert E. Park and Ernest W. Burgess and their students, beginning in 1920. Viewing the city as a social laboratory for the investigation of human behaviour, they set a generation of students to investigate problems of human behaviour in the urban environment. The most systematic statement of this approach to investigation, referred to as the Chicago school of urban sociology is found in Robert Park's essay "The City as a Social Laboratory" (1915) and a later one by Louis Wirth, "Urbanism as a Way of Life" (1938). Scott E. W. Bedford's *Readings in Urban Sociology* (1927) and Nels Anderson and Edward C. Lindeman's text, *Urban Sociology* (1928), were the earliest attempts at textbook treatment.

Ways of Viewing Cities.—Cities are social facts of many dimensions, and they have been described and analyzed from many points of view other than the sociological. One way to define the sociological interest in cities is to examine some of the other views.

Men have moralized about cities throughout the history of urban settlement. The social philosopher treats cities (and city life) as objects of human preference, as moral entities. Cities are judged by men as good or bad, as parasitical or fecund entities. To the poet Shelley, "hell is a city much like London," while to the philosopher Aristotle, "the good life can only be lived in the city." The moralist's view of the city presupposes a set of values in terms of which the city is judged as to its desirability or undesirability as a place of human existence. The urban sociologist, on the other hand, seeks a factual description and analysis of cities and city life.

The political scientist views the city as a legal corporation or as a form of government. His concern is with development of urban forms of government, problems of governing the city, and the reshaping or planning of the city through means available to government. The political scientist's view often is close to that of

the moralist since he is concerned with matters of policy, of how government may deal with the problems of urban living.

Historians for the most part are interested in particular cities in history, usually the great cities, their desire being either to illuminate a historical period through the history of a city or to describe cities as history has made them.

The urban geographer and the human ecologist to some extent have shared a point of view in their study of urban phenomena. The urban geographer tends to treat the city as a focal point in the occupation and utilization of the earth by man. Interest in urban phenomena is dictated by an area viewpoint, by such questions as the location, distribution in physical space, and changes in size, location, and internal structure. The city is viewed in terms of the relationships between man and his territorial environment.

The interests of urban sociologists to some extent overlap all of those described above, but urban sociology has a distinct focus as well. The urban sociologist generally views the city as a form of human community in which there are particular ecological forces of integration, in which human beings acquire certain behaviour patterns as a result of association with one another, and in which institutions and forms of social organization give to human life a characteristic aspect called urban. The urban sociologist regards the city and its civilization as a distinctive kind of integration of human activity and as a major source for the initiation and control of social life in societies.

No single frame of reference or theory for the study of urban phenomena characterizes the discipline of urban sociology. There are, rather, three major approaches to the investigation of urban phenomena used in description and analysis. These are the ideal-type community approach, the urban trait-complex approach, and the rural-urban continuum approach; they are sometimes combined.

Ideal-Type Community Approach.—The ideal type is a construct obtained by abstracting the characteristics of an object or institution to their logical extreme and perfection. The ideal-type method is used in urban sociology to describe communities. Usually a number of characteristics describe the ideal type of urban community, but the more characteristics included in the type the less closely do empirical cases of communities approximate it. Some ideal-type approaches posit polar conceptions of communities. As mentioned above, Redfield considered folk and urban societies to represent polar types. He defined the folk society as a small, isolated, nonliterate, homogeneous society whose members have a strong sense of group solidarity. The opposite characteristics define the urban society. Durkheim regarded as polar the society based on mechanical solidarity and that based on organic solidarity; Maine and Tönnies utilized somewhat similar distinctions in their descriptions of forms of human community.

Common to all these ideal-type approaches to human community is use of the types to test hypotheses about human behaviour in a community or society. The ideal form of the community is seen as a source or cause of variation in human behaviour and organization. Redfield, for example, saw the decline in isolation and homogeneity of the community as factors in the secularization and individualization of behaviour of its members.

Trait-Complex Approach.—The trait-complex approach to study of urban communities is closely related to the ideal-type approach, the major difference being that the former uses empirical rather than logically pure attributes in defining a community. The empirical attributes are usually viewed as causally connected. For some sociologists a single variable is seen as the causal or generating variable in the trait complex. P. A. Sorokin and C. C. Zimmerman considered the single generating variable to be occupation: the agricultural occupation defines the rural community, while nonagricultural occupations define other forms of community. The occupational variable is seen in turn as generating other differences among communities. The most important of these usually considered by urban sociologists who use this approach are the size of community, density of settlement, homogeneity-heterogeneity of the population, social differentiation and stratification, mobility of the population, and its system of social interaction. The problem of rural or urban sociology then is to see whether these differences between rural and urban worlds produce differences in social institutions and organizations, cultural and social processes, vital processes, and the psychology of its resident population.

A related approach in urban sociology is that of investigating the extent to which concentration of population tends to give rise to the social phenomenon of the city as a way of life. The two features—concentration of population and the city as a way of life—are allowed to vary independently. The question then raised is, to what extent is the urban way of life limited to an urban population (one concentrated in physical space)? Investigators with this point of view do not see a one-to-one correlation between an urban population and an urban way of life. Concentration of population is seen as giving rise to certain features of social organization that are urban in character. But these features, once established, can be carried beyond the confines of the city so that it is possible to speak of rural populations as highly urbanized.

Rural-Urban Continuum Approach.—The third major approach in urban sociology posits a rural-urban community continuum in which there is a continuous gradation from rural to urban such that all human communities can be placed empirically at some point on the continuum. The definition of rural and urban polar ideal types of communities often implies such a continuum. A major difficulty in this approach is that while a given characteristic of a community may vary from rural to urban, all the characteristics of the community do not vary together with the same degree of urban or rural quality; thus it is difficult to place a given community empirically on a rural-urban continuum.

Factors Studied.—The factors investigated by urban sociologists are chosen from four major sets of community characteristics that define the community as (1) an ecological structure; (2) a unique demographic structure; (3) a characteristic form of social organization; and (4) characteristic forms of social relationships. Each of these characteristics of an urban community defines a focal concern of some urban sociologists.

The human ecologist studies the urban community from the perspective of community structure. The community is defined as the structure of relationships through which a localized population meets its daily requirements. The urban ecologist studies the ways in which the urban community is differentiated and organized, particularly in its spatial organization. He is thus concerned with such problems as the segregation of persons and institutions in space, their concentration and redistribution, and he tries to account for the patterns formed; for example, residential patterns of ethnic groups, business establishments, or occupation groups. At the same time ecologists are interested in the temporal organization of cities, the rhythm and tempo of an urban population. They therefore try to account for such things as the movement of persons between residence and workplace and the patterns of such movement. Finally ecologists are interested in accounting for the distribution of urban communities in space and for the movement of people and the organized relationships that are established among them. They are particularly interested in studying the functional relationships among urban communities and their hinterlands, the exchange relationships among them, and their bases for functional specialization.

Closely related to the interest in ecological organization is the interest in community demographic structure, considered by some ecologists to be only an aspect of the ecological approach. The two most important community aspects investigated in this approach are size and density of settled area. The principal problem for demographic investigators is exploration of the aggregative aspects of the city, such as its population composition and its vital processes of population growth and mobility.

The city often is viewed as a unique form of social organization or social system, the normative organization of the urbanized society and the structures developed for social interaction in cities then becoming the major objects of investigation. Some of the aspects of urban social organization most commonly studied include economic organization, system of social stratification, normative integration, kind and degree of social cohesion among the

residents, and nature of social control. Of particular concern in these investigations is what is uniquely urban about these institutional arrangements and their organization. Are they the peculiar products of urbanism?

Cities also are seen as generating characteristic forms of social relationships. Some sociologists see cities as particularly productive of anonymous and impersonal relationships, persons meeting one another in segmental or highly specialized social roles rather than as full personalities. Collective behaviour in the mass or crowd also is thought of as more distinctively urban. The urban sociologist therefore turns his attention to the kinds of social relationships developed in daily living, the role requirements of an urban dweller, and the ways in which the human personality meets the requirements of urban living. Particular attention also is given to the way in which deviating behaviour is structured in the city, whether it is productive of mental illness, or organized delinquent gang warfare and organized crimes, of dissident social movements, and so on.

Urban v. *Rural*.—Urban sociology is predicated upon the assumption that there are clearly distinguishable differences between urban and rural ways of life finding expression in different technologies, divisions of labour, social organizations and institutions, personalities, and even social problems. Historically the country and the city are viewed as representing different stages of civilized life. Contemporaneously they reflect different modes of social existence. Some sociologists, however, question this assumption, maintaining that the proper study of sociology in this respect is the territorial organization of human groups into communities. Whether or not urban sociology will remain a field in a developing discipline is problematic. There are, however, certain problems dealt with in urban sociology to which the discipline of sociology will undoubtedly recur.

Process of Urbanization.—One of these is the general problem of population distribution and redistribution. An important aspect of population redistribution is the process of urbanization, a process of population concentration such that the ratio of urban people to the total population in a territory increases. Both the size of individual urban centres and the number of urban centres in a territory can increase without urbanization taking place, since only when a larger proportion of the inhabitants come to live in cities is urbanization said to occur. There are four major factors which sociologists use to account for urban growth and urbanization. Their effect has been intensified since the Industrial Revolution of the 19th century.

First of these factors is that of greater productivity per agricultural worker, such that the agricultural system is capable of producing a surplus of food to supply urban populations and also to permit labour to be withdrawn from food production and used in the production of capital goods and services. Technological changes in agricultural production are of course the most important factor changing productivity per agricultural worker. The shift of labor from agricultural to nonagricultural occupations results from the greater income elasticity of demand for industrial as compared with agricultural goods.

The second major factor in the trend to urbanization is a direct consequence of the Industrial Revolution, giving rise as it did to the modern factory system and the industrial city. The invention of efficient techniques for converting energy in fuels and the derivative development of mass production made possible a large number of specialized cities with large populations within an area. Specialization in industrial production leads in fact to high interdependence not only between rural and urban peoples but also among cities; continued industrial production in almost any city depends upon continued production in many other cities.

The increased efficiency of the technology of transportation and communication is a third major factor in urbanization. Because of their specialization and their integration with both a rural and urban hinterland, cities become highly dependent upon trade. Increase in the efficiency of long-distance transportation therefore has a powerful effect in stimulating urban growth, since it makes possible increased trade through access to more and larger markets. Changes in local transport technology are also important in urban growth since they make possible the rapid intracity movement of persons and goods.

The demographic transition accompanying the Industrial Revolution has been a fourth major factor in urban growth. The technology of sanitation and medicine brought with it sharp decrease in mortality, and since countries undergoing urbanization are characterized by high birth rates the result is a phenomenal population growth. The expanding population supplies the urban industrial labour force.

The rate of urbanization has varied both historically and contemporaneously in different countries of the world. Prior to 1800 there were no highly urbanized countries since the cities required large rural populations for their support. From 1850 to 1900 rapid urbanization occurred in countries undergoing the Industrial Revolution, and by the middle of the 20th century the rate of urbanization was seen in many countries to be much faster than in the 19th, because of developing technology. Although highly urbanized areas are generally highly industrial areas, urbanization is not a simple function of industrialization.

Study of Community.—A second major problem area to which the discipline of sociology undoubtedly will recur is the relation of community to territorial location and its relations to the structure of social systems. The principal concern here lies with the study of persons in interaction in territorial locations. The subfields of rural and urban sociology have tended to subvert an interest in the study of community phenomena, so that a comparative community framework of inquiry did not emerge until the middle of the 20th century.

There are three major approaches to the study of community which overlap with the approaches to the field of urban sociology. These may gradually take the place of the earlier concerns with a rural and an urban sociology.

The first of these approaches is based on viewing the community as an ecological system. Ecologists assume that community structure manifests itself in a spatial and a temporal pattern. The ecologist's interest in community centres around generalizations about the structural features of communities and how these change in response to external conditions. The second major approach which has characterized the work of social anthropologists in particular, views the community as a microcosm of the larger social system. Each community is described and analyzed in terms of all of the structural and functional features of the larger society since it is viewed as a territorial microcosm of the larger society. Each community thus is viewed as having a stratification system, a power structure, educational, religious, economic, and political subsystems, and so on. The community viewed as an aspect of structure of all social systems is a third major approach to community study. Populations are investigated in relation to their territorial location with special reference to the social relations that arise between them as a result of territorial location. Sociologists who follow this approach hold that a community system differs from other systems in that locality is a datum in the integration of the system.

When the field of urban sociology emerged within the discipline of sociology, it began as a perspective for viewing all aspects of social life. The city was not only a distinct way of life or form of community but a social laboratory within which all aspects of social life could be examined. With the view that the city was a social laboratory for the investigation of human behaviour, many investigations were carried out within the context of an urban community and the conclusions were thought of as referring to urban phenomena only because of the urban locus. As the field of urban sociology develops, this point of view gradually is being replaced by one based on a comparative sociology of the community.

As these changes occur, many of the problems that were first treated within an urban or rural sociological framework have emerged as separate fields apart from the discipline of urban sociology. Among these are the fields of social stratification, industrial sociology, and human ecology. With the continued emergence of new sociological fields, much of what was traditionally treated within the discipline of urban sociology no longer is rationally included within it.

(A. J. Re.)

III. SOCIOLOGICAL THEORY

1. Introduction.—Sociology as a discipline is defined primarily in terms of a conceptual scheme; its principal organizing focus lies in its theory. This theory, however, clearly stands in an early stage of its development, with only a broad consensus on the principal frame of reference, with uneven degrees of clarity and precision of conceptualization in its various parts, with only spotty formulations of analytical uniformities in the relations among variables, and with uneven achievement of empirical verification of such propositions as we have.

The following condensed review of the main structure of sociological theory is confined to an outline of the definition of major theoretical problem areas and of salient concepts developed in relation to them. It remains mainly at the categorial level of exposition, including the outlining of a number of important classifications. A number of important substantive propositions of general significance that could have been stated are omitted here because of lack of space: to state them with sufficient generality and still make their empirical relevance clear would have entailed an exposition too long for this article.

Furthermore, this article consciously adheres to the terms habitually used by the author, who cannot presume to speak for his profession as a whole. There is as yet no clear-cut consensus on sociological theory as a total scheme, as contrasted with the elements of theory used by many practising sociologists. The elements have been very imperfectly codified. The degree to which the following outline will prove to be accepted for the future can be determined only by a long process of critical work. Considerable modification is to be expected as a matter of course; that there is now a broadly acceptable general framework somewhat along these lines seems to be a reasonable presumption.

2. Theory of Social Systems.—The focus of sociological theory is in the theory of social systems, which may be conceived as the systems organized about—and generated by—processes of interaction among living organisms. Clearly the overwhelming interest is in the interaction of human beings, but certain aspects of sociological theory are continuous over wider reaches of the organic world, thus making contact with the study of animal behaviour. The most distinctively human features of interaction are functions of the presence of language and, hence, communication as a symbolic process. Indeed communication at these levels is the principal content of social process for human beings.

The role of symbolization brings social system theory into particularly close relationship to the theory of culture, which may be thought of as concerned with that aspect of action which is organized about the meanings of cultural patterns and their genesis and conditions of maintenance and change. Sociological and cultural theory are both closely interdependent and interpenetrating over a considerable area, but it seems important to treat them as analytically distinct. Among the social sciences, anthropology, particularly in the United States, is especially concerned with the analysis of culture, and, of course, a variety of other disciplines are concerned with analysis of cultural meanings as such from the point of view of their content, for example, jurisprudence, logic, literary criticism, and others. (*See* further CIVILIZATION AND CULTURE: *Concepts of Civilization and Culture* for distinctions in anthropological and other usage.)

The special significance of culture for the social system, in turn, has an important bearing on the special character of the relationship between the social system and the personality of the individual. By this we mean the organization of the behavioural system linking the biological organism with the social and cultural systems which, from a regulatory point of view, come to constitute the most important parts of the environment in which the person acts. This organization must be consonant with the functional requirements of symbolic interaction, which is accomplished through the internalization of some of the pattern elements of cultural systems, and with the modes of organization of social systems as internalized objects and patterns. The discovery of such internalization of cultural patterns and social objects was one of the most essential conditions of the development of both sociological and personality theory. The most important work converging on this theme was the sociological analysis of Émile Durkheim, the psychology of Sigmund Freud, and the "social psychology" of Charles H. Cooley, G. H. Mead, and W. I. Thomas.

Sociology as a theoretical science does not deal with the whole of the social system as delineated by its mutual relations to the cultural system, the personality, and the organism of the individual —and through the latter to the physical environment. Its concern may be defined as the nature and operation of more or less institutionalized patterns of culture in regulating or controlling, and thereby partially determining, the actions and relationships of individuals. Its special interest is thus on the normative aspect of the social system. The more specifically operative aspects that are distinctively social are ordinarily called economic and political and are the foci of the two prominent social science disciplines carrying those names.

Though analytically distinct, the sociological aspect of social systems is so closely interdependent with the economic and political that the disciplines on each side of this line must be very closely related to one another, so much so that only gradually has it become possible to draw theoretically acceptable distinctions. In the areas ordinarily treated as the concrete concern of economics and political science, the sociologist is more concerned with the institutional framework within which economic and political processes occur and with the internalization of this framework in personalities, whereas the economist and political scientist are more concerned with the specific interplay of "forces" by which economic production, on the one hand, and political decision making, on the other, are brought about.

It goes almost without saying that sociological theory, in its special concern with the normative components of social interaction, attempts to maintain a scientific point of view. The question of the orientation of the sociological observer to these normative components is inherently complex. As scientist he cannot "take them for granted" as directly relevant to him, either in the sense of personal commitment or that of personal rejection. This complex relation, particularly for the study of the observer's own society, is a special case of the very general problem of the relation of the scientific observer to his subject matter.

Like all theories dealing with living systems, the systems with which the sociologist deals are empirically open systems. This has been implied in the stress above on the intimacy of the relations between social and cultural systems, and that between social and personality systems, and again in relation to the biological organism. Such openness does not, however, erase the significance of the analytical distinctions or the empirical boundaries. There is an increasing tendency to treat such issues in terms of a logical model of input-output relationships. "Adjacent" systems thus become mutually dependent for meeting each other's "needs," or functional requirements, through their mutual outputs. The most important historic model of this relation is the economic conception of the division of labour and the mutual dependence resulting from it, which was adapted for sociology above all by Durkheim.

If boundaries are to be maintained, and with them the relevant specification of output to adjacent systems, there must be some modes of normative control over the boundary processes themselves and over the processes internal to the system on which the interchanges of inputs and outputs are dependent. For social systems this system of normative control is precisely what sociological theory attempts to analyze.

The concept social system is an analytical concept which is used on a wide range, from macroscopic to microscopic. At the latter end a small experimental group or a nuclear family can be defined as a social system. In the middle, large formal organizations or associations, local communities, and so on, may be treated as social systems. The concept society is generally used to designate the type of social system which shares a general common culture; often, but not always, language; nationality; and religion; and which is politically organized with reference to a territorial area. But even intersocietal, or more familiarly inter"national," relations can be analyzed as constituting social systems.

3. Structure and Process.—In common with other theoretical schemes in empirical fields, sociological theory divides its con-

cern, on the one hand, between problems of structure and of process and, on the other, between problems of the conditions of stability and of change. The structure of social systems, we have suggested, is composed of patterns of culture. These may be classified under four headings as follows:

(1) Expectations of role-performance for the individual. These are located specifically in contexts of interaction with other individuals and hence cannot be defined without reference to the complementary patterns of expectation for the performance of those with whom the individual of first reference interacts. Any given role is thus, as R. K. Merton has put it, part of a role-set. Since the role-performances of individuals are the sociologically relevant units of interaction process, at this level there must be a network of normative expectations for interacting units. Some degree of "meshing" of such expectations is a condition of minimal stability for such a system.

(2) The organization of plural role-units to constitute collectivities. A collectivity is the aspect of this organization that is relevant to the processes of making decisions to attain goals. It is the political unit of social structure par excellence. The normal individual has plural roles, of course, and is a member of plural collectivities. The concept of collectivity need not as such imply "formal organization," though this is characteristic of the larger and more durable cases. The collectivity is the primary agency of the performance of the social system as a system.

(3) The patterning of rights and obligations among role and collectivity units in relation to norms defining desirable modes of performing functions in the system and avoiding or minimizing dysfunctions. The primary function of these norm components is definition of the situation. The prototype is the legal system in an advanced society, but other components of the system of social control belong in this category. The standards for the definition of legitimate function and for allocation of resources in the system are the focal structures.

(4) The patterning of commitments to values. By values of a social system we mean conceptions of the desirable type of system of reference, for example, a politically organized society. As institutionalized in a social system its values are to a sufficient degree internalized by its members. These values define only the types of system and are not, as such, adequate guides to action because they do not take account of differentiations of function or particularities of situation. Values for a higher-order system are specified to the level of each of its subsystems on the basis of the general rule that, so far as the system is integrated, the pattern tends to remain constant through the various levels, but the content changes to meet the requirements of each subsystem.

All four of these categories are structural in the sense that, over certain ranges of variation, their principal properties may be presumed to be constant, and processes can thus be analyzed relative to them. Empirically, of course, they change, but theoretically the problem of analyzing such change must be distinguished from that of analyzing process in a given structural setting.

The relation of structure to process in social systems generally is mediated through the system of norms (category 3 above). The primarily relevant types of norms in this connection are those that may be called integrative standards. These regulate the evaluation of contributions to the functioning of the system. The most important are solvency as the standard regulating economic productivity, effectiveness as regulating goal attainment, solidarity as regulating the integration of the system, and integrity as regulating the maintenance of the patterns of commitment.

Solvency can serve as an illustration. In a highly differentiated society virtually all units in the social system are enmeshed in the marketplace, on which they are dependent for access to resources and very generally also for mediating relations to the recipients of outputs. There is then the imperative of balancing money expenditures by money income. For the business firm and the normal family household, this is achieved through the money proceeds of operations—the general alternative is liquidation or, for the household, "relief." In the very large category of units to which this version of the standard does not apply, there must be specific ways of securing funds to meet operating expenses, either through contributions or through some form of compulsory levy. So far as such supplementing of money proceeds occurs, it means in general that the unit is "subsidized" from the surplus economic productivity of other units.

The solvency aspect of the operation of economic standards operates within a framework of institutional norms, of which contract and property are the most important parts. Contract is the institutionalization of the establishment of rights and obligations by voluntary agreement among units of the system, individuals or collectivities, *i.e.*, where the relationship is neither ascribed nor compulsory. There are, in turn, two main types of contract: (1) that which establishes directly reciprocal relations between units of essentially the same type, *e.g.*, two firms and (2) the contract of employment, through which the services of an individual are made available to an organization.

The institution of property has historically concerned rights in the use, control, and disposal of physical objects of possession including animals, and on occasion human beings. With the development of market systems, and the more complex ramifications of the monetary medium, financial assets have come increasingly to be the most important part of the property system, a tendency that has been greatly accentuated since the legal development of the corporation.

The institutional framework of contract and property meshes in turn with that of formal organization and the institution of authority, since physical resources, controlled through money alone, and human services, through the money to pay for them and through the contract of employment, are the essential resources for achieving collective goals. For their use to be made effective, however, they must be organized, and authority, with the use of power as a medium, is an essential component of organization in this sense.

Power is involved not only with the use of authority in formal organization but also with the "consent" side of collective action, in selection of leadership and the granting to leadership of authority to act. In turn these elements are related to the exercise of influence—without the coercive sanctions which are the hallmark of power—in mobilizing support for collective action, and in "defining the situation" for those who assume responsibility.

Process in social systems is, then, a matter of interchange of inputs and outputs. These interchanges essentially are processes of communication through which decisions concerning collective action, and the normative framework within which it takes place, are arrived at, and resources are allocated among the units involved.

The essential categories of internal societal output—that is, between subsystems of the social system—are, first, economic production (which is not technological or physical production as such but what economists call "value-added"), which constitutes the fluid resources available; second, collective decisions, which form the basis for the commitment of such resources to specific functions or goals of the subsystems; third, contributions to the integration of these decisions and the corresponding allocations of resources among the subsystems; and, fourth, the normative commitments of the units of the system, ultimately its individual members.

Where social systems are sufficiently differentiated, the genesis and interchange of output products cannot occur without the operation of generalized media of interchange. Again economic phenomena in the concrete sense provide a model in that money is the prototypical case of such a medium of interchange. It is clearly not itself an "output" since it has no "utility." Quite strictly it is a kind of "language" by which the economic meanings of goods and services are coded and transmitted. In the political sphere the corresponding mechanism is power, which is here conceived not simply as capacity to coerce, though coercion is very germane to it, but as capacity to make binding and authoritative decisions and commitments. Not only the power of the "executive" to make binding decisions but also administrative authority and political support through exercise of the franchise belong in the general category of power. We also consider influence, in the socially integrative sense, and the development of normative com-

mitments to be further cases of generalized media for the "control" of social action.

The processes that the sociologist must take into account are not exhausted by the internal interchanges and the mediating mechanisms just reviewed, but must also essentially include interchanges over the boundary of the social system itself with adjacent systems. The essential interchanges are, first, those involving institutional values and the cultural organization of knowledge; second, those involving participants of the system. Empirical knowledge organized and codified at cultural levels is one of the most essential resources for the functioning of social systems, but the output of socially relevant problem statements and the organization of inquiry are equally essential. In relation to the individual the essential input is "motivational commitment" to the performance of social roles. The corresponding output consists of socially meaningful rewards to the individual.

4. Stability and Change.—Both internal interchanges and interchanges across the boundaries of a system can be viewed in either of two ways. On the one hand, it is through these processes that social systems are maintained as going concerns. This is the focus of the problem of the conditions of stability of a system. In sociology as in other theoretical sciences, the problem of these conditions is often formulated in terms of the concept of equilibrium. Equilibrium may be static, concerned with maintaining the system unchanged, or moving, *i.e.*, maintaining an orderly pattern of change. In either case the conception is a device for analyzing the consequences of changes in conditions as "disturbances," which set in motion counteracting forces tending to the restoration of equilibrium. Of course whether or not equilibrium will in fact be restored is for each case an empirical question. The "tendency to equilibrium" is a very generalized hypothesis for the analysis of states of social systems and may or may not be validated for a particular case, or class of cases. Invalidation, however, because of the relative precision of formulation of the problems of the relations between factors of stability and of instability, can be highly useful in throwing light on just what factors account for the instability.

In content the most important sociological concept which bears on the relation between stability and change, between equilibrium and disequilibrium, is that of institutionalization. Institutionalization is the state of relation between the normative patterning of a social system—its values and norms in the senses outlined above —and the other essential components in the determination of the state of such a system. Normative patterns are institutionalized insofar as they are integrated with the nonnormative factors in such a way as to maintain a condition of relative equilibrium. Institutionalization in this sense is not a matter of presence or absence, but of degree. In addition to the "firmness of institutionalization," however, the probability of maintenance of equilibrium is of course a function of the intensity of pressures to change.

Internally the focus of the problem of maintenance of institutionalization lies in the relation between the economic and political interests of units and subsystems of the social system, and the normative factors. The most general concept for discrepancies in this relation is strain. The consequences of strain in turn may be subclassified as involving tendencies to deviance, in the sense that units have interests in violating known and acknowledged normative expectations, and tendencies to conflict, in the sense that the interests of different units or classes of units come to be opposed to each other if there is no adequate mechanism for reconciling conflict. The ultimate in conflict is the polarization of the system in terms of two conflicting sets of units. A third aspect of strain is what, following Durkheim, has been called anomie, which is lack of clarity in the normative "definition of the situation" so that it is not so much a matter of actors not wanting to do what is expected, but not knowing sufficiently concretely what in fact is expected. Finally there may be strain because of uncertainty as to the cultural acceptability of the normative patterns themselves, *i.e.*, in value terms, of their legitimacy. These are all clearly analytical conceptions, and in most concrete cases of strain all four types of components are involved.

These internal foci of strain must be related to the problems of the exogenous conditions of equilibrium. As factors in change, these concern the cultural basis of the values which are institutionalized, the motivational commitments of participants, and the adequacy of the cognitive conceptions of the system and its relation to its environment. If stability is to be maintained, all of these factors must be integrated in terms of procedural mechanisms, by which conflicts and strains can be "handled."

In dealing with change, it is essential to discriminate among a number of types. One pair of such types concern what we may call the "progessive-regressive" axis. This relates to the level of adaptation on which the system stands in relation to its environment. In these terms the focus is on capacity to "cope" with the exigencies of the environment, which in turn means command of resources. The attainment of higher levels of adaptation involves, internally, processes of differentiation in structure, of extension of participation of units in the system—which in one aspect is quantitative growth—and upgrading in the level of expectations, and in the level of generality of the normative structures.

Regressive change is the obverse of this. It is frequently dealt with under the heading of "disorganization," though this term also often designates some of the concomitants of progressive change. The distinction may be a relative matter, in that progressive change always means the disorganization of some subsystems institutionalized at earlier stages of a process. Regressive change involves shrinkage in the resource base and decrease in the fluidity of resources. It involves, frequently, increasing particularity and rigidity, rather than increasing universalism and flexibility, which in our terms is normative downgrading. It often means decrease in scale and breakup of more extensive systems into segments.

The second major axis of social change concerns the pattern of the normative culture itself, notably the value system. Whereas progressive and regressive change concern differences of level along a given directional line, value-change is change of direction of development. The values of a system may orient its functioning more in the direction of concern with and active control over its external environment, or more in that of internal concerns—with the maintenance of harmonious integration or with the integrity of its commitments to a more static pattern of life. It may also vary in terms of whether the main orientation is more instrumental or more concerned with attainment of immediate goal-states. Thus the Communist societies are heavily oriented to attainment of the specifically proclaimed goal of Communism, whereas the "liberal" societies are more concerned with a general pattern of development, without specifically conceived goal-states. Both types, however, are predominantly activistic in orientation to the environment.

All of the factors that in any way influence social action can in principle influence value-change. In the normative aspect of the mechanisms that control the operation of a social system, however, the value system is the highest structural component in the order. Its change constitutes, therefore, a "bottleneck" such that only when other possibilities of adjustment have failed does this pattern itself tend to change. If the change is not to be of a regressive, disorganizing type, then there must be some promulgation, in a more or less "charismatic" manner, of a directional pattern for the change. This may be couched in what we call ideological terms at secular levels, or it may go to the higher orders of cultural orientation we ordinarily call religious.

Since a change in value-pattern is by definition a radical change, it is unlikely that the "promulgation" of new values will come from "official" quarters; *i.e.*, elements who do so by virtue of their institutionalized statuses. The context may or may not, however, be revolutionary in a political sense; charismatic religious movements, for example, have frequently led initially to withdrawal of their adherents from commitment to the existing order rather than to immediate attempts to change it.

However this may be, promulgation of new values is by no means enough. The "promulgators" must not only proclaim a message; they must create a movement of dedicated followers. The task of this movement, very generally, is simplified by a

process sometimes called dedifferentiation, by which many features of the older system that need not be in conflict with the new values are "suppressed" in favour of concentration on the "essentials," *i.e.*, goals that can be "single-mindedly" pursued against serious opposition. Following this there must be a process of progressive institutionalization of the structural components necessary for stability under longer-run and more complex conditions.

5. Concept of Evolution.—It seems likely that a more general theoretical analysis of change in social systems, with special reference to their sociological components, will fall into the framework of the concept of evolution. Earlier attempts in this direction have for the most part been disavowed, but there has been a revival of interest in the problems and increasing recognition of the importance of the questions. No clearly articulated new theory is yet available. It seems clear, however, that it will have to give special consideration to the "modern" type of society, sometimes called "industrial," with its developed monetary and market economy, its bureaucratic and more or less "democratic" polity, its universalistic system of norms, particularly at the legal level. The differences between "capitalistic" and "socialistic" versions will probably prove to be secondary variations. As Max Weber a generation ago said in the introduction to his *The Protestant Ethic and the Spirit of Capitalism*, this "modern" type seems to be of universal-historical significance.

It is also clear that such a theory will have to maintain close articulation with conceptions of the theory of cultural evolution and of biological evolution. Developments in the latter field seem to be such as to make this articulation substantially easier than it would have seemed to be on the basis of the 19th-century biology.

(T. PA.)

IV. METHOD

Throughout his existence man has been observing his own actions and those of his fellows, with the aim of understanding and predicting the behaviour of his social environment. The same aim is held by sociology, the only distinguishing feature of which is the systematic nature of its observations and inferences, codified as its "method" and its "theory." Thus the development of sociology as a social science is coincident with the development of its methods and its theory.

Two obstacles have hindered the development of the social sciences, relative to the natural sciences. In the first place, social phenomena seem fundamentally more complex, that is, less fully determined by a few easily observable factors. This complexity, or relative unpredictability, makes far more difficult the task of identifying relations between causal factors and their consequences. Secondly, experimentation, which has been the most powerful of scientific methods in natural sciences, is severely limited in the social sciences. Only in restricted ways can man carry out experiments on his fellows purely for the sake of knowledge.

Economics has in large part bypassed the problem of complexity by abstracting one element of the functioning of society—valued objects that are produced, exchanged, and consumed—and then assuming that behaviour toward these objects is governed by rational maximization of value. Psychology has in large part bypassed limitations on experimental control by expanding its interests to include animal behaviour, and using animal experimentation where experiments with man are difficult or morally impossible. Sociology has been unable to use either of these approaches to any great degree.

There have been three major stages in the development of sociological methods. It is a curious attribute of sociological method that none have been discarded as new types of data arise; the new techniques are merely added to the existing ones. The major reason is that each newly-developed method is applicable only to a part of the problems which sociologists attack, and the previous methods must still be used for the remaining problems.

The three stages can best be characterized as the use of existing data from natural situations, the purposive collection of data from natural situations, and the control of situations via experiments, to create certain kinds of data not forthcoming in natural situations. Crosscutting this classification is an important distinction between two kinds of data: quantitative, usually obtained through counting of one sort or another, and qualitative, which include no numerical quantities. The more recent developments include a larger component of quantitative data than earlier ones, and a number of quantitative techniques in design and analysis have arisen.

1. Use of Existing Data.—Early research in sociology differed from ordinary experience only in its systematic comparisons. The data were ordinarily obtained from diverse sources: early censuses and precensus government sources; accounts of primitive societies reported by explorers, travelers, and early anthropologists; and other similar observations. Often those observations contained no numerical data but required strategic contrasts and comparisons, as in the work of Herbert Spencer (1820–1903), Ferdinand Tönnies (1855–1936), Max Weber (1864–1920), and William Graham Sumner (1840–1910), and in the early work of Émile Durkheim (1858–1917). The method differs little from the principal methods available to historians today and is still the only method available to contemporary sociologists for most problems requiring historical data.

In a few instances, the data included counts and allowed numerical comparison. Examples occurred in the later work of Durkheim (*Suicide*), who used government statistics to relate suicide rates to various factors in the social structure. Among present-day sociologists, demographers and human ecologists use such data almost exclusively. With the expansion of census data and other counts of socioeconomic events, the potentialities of such data for sociological analysis expanded enormously. For analysis of these data, the methods of statistical analysis discussed under survey research below are applicable; in addition, some methods have been particularly designed for demographic and ecological research.

2. Collection of Data for Sociological Purposes.—As in any science, the exclusive use of data collected for other purposes places a severe limitation on research. Even in the early days of sociology, before 1900, a few researchers, with a particular problem in mind, systematically collected their own data. Frédéric Le Play (1806–82) collected detailed data on family budgets of different types of workers; Charles Booth (1840–1916) used observation and interview to study social life among the working classes of London.

In more recent years, many "community studies" have been carried out by a sociologist or group of sociologists coming into a community as temporary residents, observing, interviewing, and, through the use of detailed field notes, attempting to analyze and then "resynthesize" the social life of the community. One of the best examples of such research is one of the earliest: Robert and Helen Lynd's *Middletown* (1929); another excellent example is William Foote Whyte's *Street Corner Society* (1943). This work is largely qualitative, employing few numbers in the analysis but attempting to lay out the functioning of the community as a system of behaviour. This approach has refused to sacrifice the complexity of a social system for the apparently firm conclusions of quantitative comparisons. As a result, few methodological rules have developed, and the success of the method continues to depend largely upon the researcher's observational skill and his ability to fit together his information so as to mirror the community as a social system. Such work still constitutes a major segment of empirical research in sociology. One of the few methodological treatises for these techniques is Social Science Research Council Bulletin #53, *The Use of Personal Documents in History, Anthropology, and Sociology* (1945).

A very different strategy is employed in survey research, a method that has become widespread. Survey research developed from opinion polling in the 1930s and, like opinion polling, has had the following characteristics, all contrasting with the qualitative studies of social systems discussed above: (1) a focus upon a narrow segment of life, such as voting behaviour, husband-wife relations, consumer behaviour, or attitudes toward public affairs; (2) a systematically selected sample of persons, using the theory of sampling as developed in statistics; (3) a systematic interview

or questionnaire, with a predetermined set of questions; (4) relatively greater emphasis on attitudes than on behaviour; and (5) use of quantitative statistical comparisons in analysis of the data. In choosing this strategy, survey research has gained the important advantage of quantitative evidence, and a resulting precise level of confidence, for its statements. It has largely given up the ability to make statements about a functioning social system, for it samples individuals randomly from this system, and it samples a narrow slice of attitudes or behaviour in interviews. It is likely, however, that the great expansion of applied sociology in recent years has been principally due to the development of survey research. The potential response of a market to a new product, how voters feel about an issue, why families move, who watches various television programs, and a host of similar questions and problems are particularly amenable to survey techniques.

The choice of precision rather than complexity in survey research has led to increasing codification of methods. Research design and sample design, questionnaire construction, coding procedures, and analytical methods have been developed and communicated fully enough to make this class of social research less an art and more a science.

The analytical methods used in survey analysis have several special properties. They are first of all attuned to the state of sociological knowledge and are designed to give evidence for propositions of the sort, "As X increases, Y increases." There is seldom a well-developed theoretical structure for which parameters are to be estimated. Rather, statistical tabulations are ordinarily designed to test the null hypothesis of no relationship between X and Y. Secondly, the variables or attributes may characterize units at one of several levels: individuals, relationships, groups, organizations, social systems. There are special methodological problems at each level (*e.g.*, the problem of aggregation when measurements are on individuals but the proposition refers to organizations) and further problems which arise when the proposition itself refers to two levels. Finally, the attributes which characterize individuals are frequently dichotomies (*e.g.*, vote versus non-vote) or ordered variables, rather than continuous variables. Thus the descriptive statistics developed for continuous variables, such as multiple regression, least squares, and product-moment correlation, have limited applicability in survey analysis. Instead, the less well-developed area of attribute statistics is applicable.

Modifications of survey research have extended its range to the study of organizations and loosely structured social systems (mentioned below).

3. Experimentation.—In some areas of sociology, experimentation is practicable. Numerous studies of small groups have been carried out experimentally: studies of communication patterns, of leadership and morale, of group decision making, of problem solving, of social influence, and a host of others. In contrast, problems which imply long-standing relations, powerful social ties, well-developed norms or status systems, or the interplay of several institutions appear less amenable to experimentation.

As with survey research, codification of techniques has resulted in communicability of method. Statistical techniques of design and analysis, such as analyses of variance, are usable in experimentation; in addition, special observational techniques have been developed to record interactions and other events.

In some cases, research that is a cross between field research and experimentation is possible. In industry, experiments on work groups within the framework of the ongoing social system of the factory have been carried out. As another example, a summer camp was subdivided into two distinct and competing groups. An embedding of experiments within ongoing social systems provides promising possibilities for research.

4. Developments.—Probably the most important development taking place in sociological method is the modification of survey research and statistical analysis for the study of social organizations and social systems. This development applies quantitative rigour to the complex sociological problems of the functioning of a social system. Such modification involves changes in sample design, in questionnaire design, and in analytical methods. Changes in sample design take into account the structure of the organization or system and choose strategic samples at points in that structure. The sample becomes no longer one of individuals, but a sample of relations, or of elements in a structure. Sampling of role relations in a formal organization or "snowball sampling" in a system of informal relations are two types of this sampling.

Changes in questionnaire design complement the sampling changes in two ways: the questionnaire is fitted to the structural location of the individual being interviewed, so that foreman and worker, mother and child, are asked not wholly identical questions, but also complementary questions, and other questions specific to their positions; and questions are asked which refer to specific social relationships and to the social context as felt by the respondent. In these two ways, a questionnaire is designed to measure an element in a social structure rather than an individual considered as a separate entity.

The methods of analysis appropriate to these changes are designed to illuminate the functioning of a system (although in most such analyses, there is no formal model for the system). This involves the use of different descriptive statistics (*e.g.*, intraclass correlation, measures of pair-similarity or pair-complementarity). It also requires a step beyond an ordinary analysis that examines discrete propositions. It requires a synthesis of the separate relationships to mirror the functioning system, a task which, in qualitative studies of communities or other social units, is carried out by the artistry of the observer. In such analysis-and-synthesis, electronic computers play a special role, for the data from the social structure can be stored in a computer's memory in such a way as to preserve the structure. As a consequence, quantitative analysis of the behaviour of the system, rather than of the behaviour of individuals, has become feasible.

5. Special Techniques.—Besides the methods discussed above, there are some special techniques available for certain problems. One of these is known as "sociometric method." All members of a group are asked questions about their relations (actual or desired) with other members, resulting in a complex network of directed relations. Several types of analysis have been carried out with these networks, and, in particular, mathematics appropriate to such networks (matrix algebra, graph theory, and combinatorics) has been used to attack questions of cycle length, clique formation, and others.

A second special technique is used to abstract and classify meaningful content from communications, such as newspapers, books, television, movies. The method, known as "content analysis," makes possible quantitative comparisons of communication content and has been used for such diverse purposes as international images, political campaign issues, advertising themes, and newspaper bias in news presentation.

A final special technique is the use of mathematical models in certain areas of sociology. Such models have been used principally for providing measures of structural properties of social networks or kinship structures, and measures of attitudes or values. In a few cases, a mathematical model has been developed for representing a particular social process, such as occupational mobility or value change. (JA. S. C.)

V. APPLIED SOCIOLOGY

Sociology had increasing application to the solution of practical problems, especially after the second decade of the 20th century. The findings of sociological research have been utilized, particularly in such fields as criminology, social work, education, race relations, planning, government administration, marketing, communication, propaganda, public opinion polling, social psychiatry, industrial relations, and marriage and family counselling. Most sociologists continued in teaching and research activities at universities and colleges.

A number of sociologists, however, utilized their training as professional personnel in fields such as those listed above as well as in the conduct of applied research activities for agencies in these and other fields.

See also references under "Sociology" in the Index.

BIBLIOGRAPHY.—For a general introduction to sociology *see:* Leonard

Broom and Philip Selznick, *Sociology* (1955); W. F. Ogburn and M. F. Nimkoff, *Sociology*, 3rd ed. (1958); Kingsley Davis, *Human Society* (1949); George A. Lundberg, Clarence C. Schrag, and Otto N. Larsen, *Sociology* (1958).

For the history of sociology *see:* Harry E. Barnes (ed.), *An Introduction to the History of Sociology* (1948); Georges Gurvitch and Wilbert E. Moore (eds.), *Twentieth Century Sociology* (1945); Louis Wirth, "The Social Sciences," in Merle Curti (ed.), *American Scholarship in the Twentieth Century* (1953); Floyd N. House, *The Development of Sociology* (1936); Pitirim Sorokin, *Contemporary Sociological Theories* (1928). References will be found in these volumes to the works of the pioneer sociologists.

For current research reports *see* sociological journals, published in the U.S. unless otherwise specified: the *American Sociological Review* (organ of the American Sociological Association); the *American Journal of Sociology; the British Journal of Sociology; Social Forces; Sociology and Social Research; Social Problems; Sociometry; Rural Sociology; Population Index; Population Studies* (British). *See* also *UNESCO Bibliographies in the Social Sciences* (1951) and *Current Sociology* (prepared by the International Sociological Association, published by UNESCO). (P. M. Ha.)

Social Organization: General: Kingsley Davis, *Human Society* (1949); Émile Durkheim, *The Rules of Sociological Method*, 8th ed. (1938); Robert Merton, *Social Theory and Social Structure* (1949); R. E. Park and E. W. Burgess, *Introduction to the Science of Sociology* (1921); Edward A. Shils, *The Present State of American Sociology* (1948); Talcott Parsons, *The Social System* (1951); Max Weber, *The Theory of Social and Economic Organization*, tr. by A. M. Henderson and Talcott Parsons (1947). *Bureaucratic Organization:* S. N. Eisenstadt, *Bureaucracy and Bureaucratization* (1958); Alvin Gouldner, *Patterns of Industrial Bureaucracy* (1954); James G. March and Herbert A. Simon, *Organizations* (1958); Robert K. Merton *et al.* (eds.), *Reader in Bureaucracy* (1952). *Community:* Jessie Bernard, *American Community Behavior* (1949); Paul K. Hatt and Albert J. Reiss (eds.), *Cities and Society* (1957); Robert E. Park, *Human Communities* (1952). *Deviant Behavior:* Milton L. Barron, *The Juvenile in Delinquent Society* (1954); Marshall B. Clinard, *Sociology of Deviant Behavior* (1957); Thorsten Sellin, *Culture Conflict and Crime* (1930). *Family:* Ernest Burgess and Harvey J. Locke, *The Family: From Institution to Companionship* (1945, 1950); Carle C. Zimmerman, *Family and Civilization* (1947). *Age Groups:* S. N. Eisenstadt, *From Generation to Generation* (1956). *Minority Groups:* E. Franklin Frazier, *Black Bourgeoisie* (1957); George Eaton Simpson and J. Milton Yinger, *Racial and Cultural Minorities* (1953). *Small Groups:* A. P. Hare, E. F. Borgatta, and R. F. Bales (eds.), *Small Groups: Studies in Social Interaction* (1955); Georg Simmel, *The Sociology of Georg Simmel*, tr. and ed. by K. H. Wolff (1950). *Stratification:* Reinhard Bendix and Seymour M. Lipset (eds.), *Class, Status and Power: a Reader in Social Stratification* (1953); T. H. Marshall (ed.), *Class Conflict and Social Stratification* (1938); Leonard Reissman, *Class in American Society* (1959). (Mo. J.)

Human Ecology: M. A. Alihan, *Social Ecology: a Critical Analysis* (1938); D. J. Bogue, *The Structure of the Metropolitan Community: a Study of Dominance and Subdominance* (1949); E. W. Burgess (ed.), *The Urban Community* (1926); O. D. Duncan and A. J. Reiss, Jr., *Social Characteristics of Urban and Rural Communities, 1950* (1956); O. D. Duncan *et al.*, *Metropolis and Region* (1960); W. I. Firey, *Land Use in Central Boston* (1947); M. Halbwachs, *Population and Society: Introduction to Social Morphology*, tr. by O. D. Duncan and H. W. Pfautz (1960); A. H. Hawley, *Human Ecology: a Theory of Community Structure* (1950); R. D. McKenzie, *The Metropolitan Community* (1933); R. E. Park, *Human Communities: the City and Human Ecology* (1952); R. E. Park, E. W. Burgess, and R. D. McKenzie, *The City* (1925); J. A. Quinn, *Human Ecology* (1950); L. F. Schnore, "Social Morphology and Human Ecology," *The American Journal of Sociology*, vol. 63, no. 6, pp. 620–634 (May 1958); L. F. Schnore, *The Urban Scene: Human Ecology and Demography* (1965).

Demography: G. W. Barclay, *Techniques of Population Analysis* (1958); D. J. Bogue, *The Population of the United States* (1959); K. Davis, *The Population of India and Pakistan* (1951); United Nations, Department of Social Affairs, Population Division, *The Determinants and Consequences of Population Trends* (1953); L. I. Dublin, A. J. Lotka, and M. Spiegelman, *Length of Life: a Study of the Life Table*, rev. ed. (1949); H. T. Eldridge, *The Materials of Demography: a Selected and Annotated Bibliography* (1959); P. M. Hauser, "Demography in Relation to Sociology," *The American Journal of Sociology*, vol. 65, no. 2, pp. 169–173 (September 1959); P. M. Hauser and O. D. Duncan (eds.), *The Study of Population: an Inventory and Appraisal* (1959); L. F. Schnore, "Social Mobility in Demographic Perspective," *The American Sociological Review*, vol. 26, pp. 407–423 (June 1961); M. Spiegelman, *Introduction to Demography* (1955); C. Taeuber and I. B. Taeuber, *The Changing Population of the United States* (1958); I. B. Taeuber, *The Population of Japan* (1958); W. S. Thompson, *Population and Progress in the Far East* (1959); P. K. Whelpton, *Cohort Fertility* (1954). (L. F. Sc.)

Rural and Urban Sociology: A. L. Bertrand *et al.*, *Rural Sociology: an Analysis of Contemporary Rural Life* (1958); H. E. Bracey, *English Rural Life: Village Activities, Organisations and Institutions* (1959); Otis Dudley Duncan and Albert J. Reiss, Jr., *Social Characteristics of Rural and Urban Communities, 1950* (1956); J. H. Kolb and Edmund de S. Brunner, *A Study of Rural Society* (1946); Herbert Kotter, *Landbevölkerung im Sozialen Wandel* (1958); C. P. Loomis and J. A. Beegle, *Rural Sociology: the Strategy of Change* (1957); Henri Mendras, *Sociologie Rurale* (1957); Lowry Nelson, *Rural Sociology* (1955); Walter Slocum, *Agricultural Sociology* (1962); T. Lynn Smith, *The Sociology of Rural Life* (1953); P. A. Sorokin, C. C. Zimmerman, and C. J. Galpin (eds.), *Systematic Sourcebook in Rural Sociology* (1930–1932); Lee Taylor and Arthur R. Jones, Jr., *Rural Life and Urbanized Society* (1964); Peter von Blanckenburg, *Einführung in die Agrarsoziologie* (1962). (E. A. W.)

Charles Booth, *Life and Labour of the People in London* (1889–1903); Adna Weber, *The Growth of Cities in the Nineteenth Century* (1899); Robert E. Park, "The City: Suggestions for the Investigation of Human Behavior in the Urban Environment," *American Journal of Sociology*, vol. 20 (March 1915); Louis Wirth, "Urbanism as a Way of Life," *American Journal of Sociology*, vol. 44 (July 1938); Scott E. W. Bedford (ed.), *Readings in Urban Sociology* (1927); Nels Anderson and Edward C. Lindeman, *Urban Sociology* (1928); S. R. Angelucci *et al.*, *The World's Metropolitan Areas* (1959); Otis Dudley Duncan *et al.*, *Metropolis and Region* (1960); Robert M. Fisher (ed.), *The Metropolis in Modern Life* (1955); Paul K. Hatt and Albert J. Reiss, Jr. (eds.), *Cities and Society* (1957); Amos Hawley, *Human Ecology* (1950); "The Role of Cities in Economic Development and Cultural Changes; Proceedings of a Conference Held at the University of Chicago, May 24–26, 1954," *Economic Development and Cultural Change*, vol. iii, nos. 1 and 2; Rupert B. Vance and Nicholas Demerath (eds.), *The Urban South* (1955). (A. J. Re.)

Theory: Howard Becker and Alvin Boskoff (eds.), *Modern Sociological Theory* (1957); Reinhard Bendix, *Max Weber: an Intellectual Portrait* (1960); Edgar F. Borgatta and Henry J. Meyer (eds.), *Sociological Theory: Present-Day Sociology from the Past* (1956); Charles H. Cooley, *Social Organization* (1909, 1962); Lewis A. Coser and Bernard Rosenberg (eds.), *Sociological Theory: a Book of Readings* (1957); Émile Durkheim, *The Division of Labor in Society*, tr. by George Simpson (1933, 1964) and *The Elementary Forms of the Religious Life*, tr. by J. W. Swain (1912, 1961); Sigmund Freud, "Inhibitions, Symptoms and Anxiety" in *The Standard Edition*, vol. xix (1st German ed., 1926); Llewellyn Gross (ed.), *Symposium on Sociological Theory* (1959); George C. Homans, *Social Behaviour* (1961); M. J. Levy, Jr., *The Structure of Society* (1952); R. M. MacIver, *Social Causation* (1942); G. H. Mead, *Mind, Self and Society*, ed. by C. W. Morris (1934); Robert K. Merton, *Social Theory and Social Structure* (rev. and enlarged ed., 1957); Robert K. Merton, Leonard Broom and Leonard S. Cottrell, Jr. (eds.), *Sociology Today* (1959); C. W. Mills (ed.), *Images of Man: the Classic Tradition in Sociological Thinking* (1960); Vilfredo Pareto, *The Mind and Society*, ed. by Arthur Livingston, 4 volumes (1935); Robert E. Park and Ernest W. Burgess, *Introduction to the Science of Sociology* (1921); Talcott Parsons, *The Structure of Social Action* (1937) and *The Social System* (1951); Talcott Parsons *et al.* (eds.), *Theories of Society* (1961); William I. Thomas, *Social Behavior and Personality. Contributions of W. I. Thomas to Theory and Social Research*, ed. by Edmund H. Volkart (1951); N. S. Timasheff, *Sociological Theory: Its Nature and Growth* (1955); Max Weber, *The Religion of China: Confucianism and Taoism*, tr. and ed. by H. Gerth (1951), *Ancient Judaism*, tr. and ed. by H. Gerth and D. Martindale (1952), *The Religion of India: the Sociology of Hinduism and Buddhism*, tr. and ed. by H. Gerth and D. Martindale (1958), and *The Theory of Social and Economic Organization*, tr. and ed. by Talcott Parsons and A. M. Henderson, with an introduction by Talcott Parsons (1947; reprinted 1957). (T. Pa.)

Method: G. W. Barclay, *Techniques of Population Analysis* (1958); Social Science Research Council Bulletin No. 53, *The Use of Personal Documents in History, Anthropology and Sociology* (1945); C. Selltiz *et al.*, *Research Methods in Social Relations*, rev. ed. (1959); F. F. Stephan and P. J. McCarthy, *Sampling Opinions* (1958); H. Hyman *et al.*, *Interviewing in Social Research* (1954); H. Hyman, *Survey Design and Analysis* (1955); P. F. Lazarsfeld and M. Rosenberg (eds.), *The Language of Social Research* (1955); H. Zeisel, *Say It with Figures*, 4th rev. ed., (1957); H. Blalock, *Social Statistics* (1960); G. Lindzey (ed.), *Handbook of Social Psychology*, vol. 1 (1954); B. Berelson, *Content Analysis in Communications Research* (1952); J. S. Coleman, *Introduction to Mathematical Sociology* (1964). (Ja. S. C.)

SOCIOLOGY, ANIMAL. The qualitative distinction between the sociology of man and that of other animals is not sharp, but since other animals lag far behind man in the development of abstract thought, in symbolic communication and in cultural institutions, there are very large quantitative differences. The distinctions are by no means absolute, however, and in many relations comparisons are continually being made between the social life of human and nonhuman animals. Indeed, the social life of animals is of interest to many only because of possible comparisons and, more especially, of contrasts with that of man. On the other hand, there are persons for whom the various patterns of behaviour of other animals are of prime importance in their own right. This animal-centred focus is often found among observers

of the group life of animals; it is the dominant attitude among many bird watchers. Yet even the most extreme zoophiles are usually aware of some of the human aspects in the social behaviour of other animals.

Ethologists, zoologists who study the behaviour of animals in their natural environments, claim that man can learn much about himself by objectively observing the societies of other animals. Sociologists, however, often maintain that studies of animal societies have little or no relevance to the study of man's society. The dispute quickens when discussion revolves about "aggression," which ethology regards as inherent to man as well as to other animals. The point is of particular interest because of man's potential for destructiveness and his unique collective expression of aggression, war (see *Agonistic Behaviour*, below).

All too often reports concerning animal sociology are given as though the animals observed were feathered or four-footed men. They are given the thoughts, feelings, and intentions of men. Such anthropomorphic accounts are misleading. Man is the only animal we can know both objectively and subjectively. Except in the figurative sense, no person has ever been a worm, a fish, a snake-in-the-grass, or even an ape. We have no inside information concerning the feelings or the outlook of such animals. The valid interpretation of their behaviour, social behaviour included, is difficult and often impossible except from the point of view of the outsider. This article attempts to be carefully objective. The story remains interesting even under such conditions.

The sociality of animals is of course an outcome of their entire individual repertories of behaviour patterns, which are discussed in the article ANIMAL BEHAVIOUR. This article is divided into the following sections:

I. SOCIAL SYSTEMS

A. GENERAL FEATURES

1. Concepts and Definitions.—A social system is a communication system. It consists of senders and receivers whose roles may be interchangeable, and of signals that produce reliable and consistent responses. Such signals as posturing, facial gesture, utterance, or physical contact provide the binding power of social systems. They will be discussed in the broader treatment of communication that follows.

There are generally three approaches to the study of communication. The first (syntactic) involves the signals themselves, which are studied as physical phenomena; *e.g.*, communicative sounds are analyzed and described in terms of their frequency, amplitude, and timing. The second approach (semantic) involves the meaning of a signal (this path is of course closed to the student of animal communication since one cannot approach the question of meaning on a subjective basis as in human communication studies). The third approach (pragmatic) involves the determination by inference of the significance of a signal to the communicant; it is based on the nature of the receiving animal's response. Models or artificial stimulus inputs can be employed to determine accurately the signal responsible for a particular response.

There are different degrees of sociality, which may be defined either by the amount and kind of information being exchanged or, more conventionally, by the form in space and time of the social system. An aggregation may be distinguished as a group of interacting organisms that associate either because of the existence of some particular environmental circumstance, such as a localized food source, or because organisms merely have temporarily come together by chance. An aggregation has no definite constitution or structure and disbands readily when conditions responsible for the original grouping no longer hold. This is in contrast to a true social organization in which the members of a group actively seek to maintain a relationship. The maintenance of a social relationship indicates some degree of specialization and interdependence by the members of that social system.

At different times in the life cycle of an individual, different forms of sociality may be exhibited. For example, a songbird such as the American robin passes through a fall migratory phase during which it forms large flocks of mixed sexes. In the spring, however, the males become intolerant of one another and maintain and defend territories, on which pairing, mating, and rearing of the young take place. At the end of each breeding season the birds again exhibit a flocking tendency.

2. Types of Interaction.—In the classical sense, social behaviour implies the relationships found within the life cycle of a single species, but under the definition proposed here all types of interaction will receive attention. There are interspecific (between species) as well as intraspecific (within species) relationships.

Interspecific Relations.—These may be conveniently categorized on the type of influence one species has on another. In a neutral relationship no obvious influence is exerted by one species upon another. If both organisms have a negative action on each other, competition obtains. Symbiotic relationships are those in which one or both members of an interaction system derive benefit. If one benefits and the other is unaffected, commensalism exists. If both members of the association benefit, the relation is mutualistic. An amensal system occurs when one organism negatively affects a second without being affected itself. Parasitic or predatory conditions exist when one member benefits to the detriment of the second (*see* also SYMBIOSIS).

Such interspecific systems may be exceedingly complex. Certain small fishes exhibit a "cleaning" symbiosis with other larger species of fishes. The cleaner removes parasites from the host fish, which recognizes the cleaner by its characteristic movement and colour markings. Some other fishes have evolved colour and behaviour mimicry and resemble cleaners. These fishes are, however, predators that bite small pieces from the host fishes.

A discussion of social interaction includes a consideration of ecological communities, natural associations of plants and animals forming an interdependent entity in nature. Leaving aside the plant-animal relationships, the investigation of animal populations alone has yielded many insights into the interdependency of living things. It has been demonstrated that the survival of a given population depends on the maintenance of an optimal density or group size. The rate of development of sea urchin eggs is a function of their density: there is an optimal group size above and below which the rate of cleavage of the developing egg is affected



BY COURTESY OF A. L. COHEN

Slime molds transforming (left) from individual cells to (centre) "slug" stage and finally (right) to a fruiting body may represent a primitive level of sociality

adversely. It is equally interesting to note that members of a given population may condition the environment in such a manner as to promote survival of other members. This type of conditioning is apparent in colonies of one-celled organisms (bacteria, protozoa) as well as in mammals. An active rat colony is continually excavating burrows that serve for several generations. A definite survival advantage is conferred on young rats that mature in a colony containing a number of ready-made tunnels. The burrow is an artifact passed on to future generations in a manner analogous to the passage of artifacts within a human cultural system.

Such observations gave rise to the formulation of the principle of proto-cooperation. This concept implies the existence of a nonspecific, derived mutual benefit from the interaction of organisms which is basic to the evolution of any ecological community. On the other hand, any population of organisms is necessarily dependent on a certain level of available resources. Ultimately, this means that a spatially restricted population must be limited in its growth by the negative interaction of its members who compete for these resources. These biological ideas influenced social thought through the philosophic writings of Herbert Spencer, whose interpretation of Darwinism promoted competition ("survival of the fittest") as the dominant evolutionary force, and through the social idealism of P. A. Kropotkin, whose concept of mutual aid emphasized cooperation. Both are extreme views of nature, whose reality is neither total savagery nor complete solicitude. Behavioural mechanisms of a competitive nature exist side by side with those mechanisms generally termed beneficial or cooperative. This idea will subsequently be dealt with at greater length.

Intraspecific relationships probably began with the slime molds, which bridge the gap between independent existence and multicellularity. Multicellular organisms, interdependencies of cells, led imperceptibly to social organizations, interdependencies of individuals. The cellular slime molds exhibit a life cycle that includes a free-living and a colonial stage. During the free-living stage these animals behave as amoebas, feeding by means of pseudopods and reproducing by division. At high population densities the amoebas aggregate into a sluglike form that moves as a unit. This interdependent mass of amoebas is formed by means of chemical communication; the "slug" produces a substance known as acrasin, which induces free amoebas to follow up the gradient and add themselves to the group. Within the mass of cells, differentiation and specialization occur. These primitive forms are of interest to animal sociology in that their "slugs" may be the first evolutionary attempts leading to multicellular organisms. It has been proposed that most multicellular plants arose from such a process of aggregation but that most multicellular animals arose from the subdivision of a single-celled multinucleated organism. This may in part explain why a multicellular animal body achieves complex levels of diversity and interdependence among organ systems and tissues far beyond that displayed by plants. If higher plants arose from clumping of individual cells, this process may well have set initial limits on the degrees of tissue diversity and interdependence.

Reproduction plays an important role in the establishment of social relationships. Organisms such as the amoeba, which reproduce asexually by simple binary fission, are totally independent of others of their kind for reproduction. Other unicellular organisms such as paramecia are dependent to a degree on each other to sustain reproduction. A species of *Paramecium* may include several different strains that must exchange genetic material at some specified interval following a series of asexual divisions. This process of exchange of genetic material is known as conjugation, one of the many forms of reproductive interdependence that has evolved in the animal kingdom. Sexual reproduction is of nearly universal occurrence among multicellular animals. Special sex cells (egg and sperm), containing one-half of the normal genetic material, are produced. These cells, or gametes, unite to form a zygote with the restored amount of genetic material. The zygote then undergoes a series of divisions and becomes a multicellular organism. (*See* also REPRODUCTION.)

Sexual reproduction has come about by the evolution of interdependence resulting from a differentiation of roles. Unless environmental variables regulate gamete discharge completely (as occurs among fishes), selection favours the evolution of a communication system whereby the activities of the reproducing pair become synchronized to ensure continuation of the species.

Parental care of the eggs or young has evolved independently in many animal taxa and further fosters socialization, in this case a parent-young (or parent-egg) interaction system. Three general types of families may be recognized, based on studies of fish: a father family, in which the eggs or young are defended by the male alone; a mother family, in which the female leads, nourishes and protects the young; and finally the parental family, a combination, in varying degrees, of both types of parental care throughout the rearing of young.

More complex types of intraspecific sociality include groups of animals of both sexes and of several age classes. If these groups are not simply aggregations, they show some cohesion over a period of time and tend to coordinate their activities. Such groups may form by two general methods of recruitment: internal, by birth, or external, from genetically unrelated individuals. Family bands tend to be interdependent, cohesive, and closed to external recruitment; typically they consist of several generations of both sexes. Herds, schools, and flocks are generally formed by both internal and external recruitment; a herd, for example, may be a temporary grouping actually composed of family subgroups.

The maintenance of cohesion in social groupings involves several basic mechanisms. Synchronization of movements or activities is, in part, an aspect of social facilitation, the tendency for animals in groups to exhibit an intensity of behaviour different from that they show in isolation. For example, individual chickens will eat more food when grouped than when isolated. Flocking in birds or schooling in fishes results in part from instinctive mechanisms promoting a mass duplication of activities such as moving, settling, and feeding. Cohesion is also facilitated by a following response which involves a tendency to follow a moving member of the group. A member that initiates the movement regularly is a leader. Quite often in such mass movements as fish schools, loose bird flocks, or locust swarms, a true leader is lacking and the leading edge of the group is in constant flux. On the other hand, herds of red deer or horses often have a leader that consistently initiates movement within the group.

Cohesive social groupings involve an interchange of stimuli that promote group integration. Special grooming ceremonies among mammals and among birds permit an exchange of stimulti that maintains contact and familiarity. Special sounds, such as the call notes of flocking birds or choruses of frogs and crickets, serve as attractant stimuli necessary to group integration. Even aggression, popularly thought of as detrimental to group formation, has a definite and well-established social function. Aggression between members of a species (intraspecific aggression) serves to (1) promote group integrity by excluding aliens; (2) strengthen pair bonds by excluding other potential mates; and (3) encourage spacing, thus minimizing competition for available food and shelter. Interspecies aggression, on the other hand, establishes food webs of predators and competitors. Within a social group, aggression is generally controlled by the maintenance of a dominance hierarchy.

B. Methods of Communication

A social signal is derived from some neutral source that was originally not solely communicatory in function. Natural selection favours the enhancement of such a neutral source if it conveys some information. Most visual displays originated from several neutral sources. For example, before moving, most animals have a set of preparatory activities termed intention movements. A bird crouches and then raises its wings before flying. These simple preparatory movements may become selected for in a population if there exists some advantage for the group to know the intention of the sender. A simple crouch may later be used in response to a certain social situation. Such behaviour may undergo natural selection until the population develops a ritualized bowing motion.

In a similar manner responses of the autonomic nervous system to various stimulus situations may be selected for—to produce ritualized secretion by skin glands, ritualized defecation or urination, and ritualized breathing or vocalization. The process of ritualization involves an enhancement of the original movement, sound, odour, or posture. Elaborate structures and colour patterns are all derived from the action of natural selection to produce a more conspicuous effect. In some cases a movement may undergo selection until it is so stereotyped that its origin becomes obscured.

1. Chemical Signals.—Since the chemical sense is one of the most widespread sensory modalities in the animal kingdom, it is not surprising that chemical communication is an all-pervading phenomenon. Chemical particles are detected by a variety of sensory devices in different animals.

Lower animals, such as protozoans and flatworms, detect chemical changes in their environment by means of osmotic fluctuations. Social insects, such as ants, bees, and termites, depend on chemical

RUSS KINNE FROM PHOTO RESEARCHERS

PERSONAL CONTACT OF GROOMING BETWEEN BABOONS HELPS MAINTAIN STRONG SOCIAL BONDS

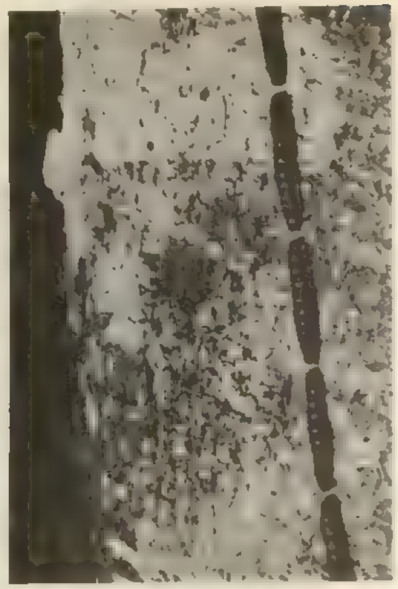
glandular exudates that guide the colony (*see* Social Insects). Very often volatile chemicals called scents are used to attract sexual partners. The male silkworm moth responds selectively to the airborne particles secreted by a virgin female of its species; he tends to fly upwind on perceiving the particles and, by seeking out the greatest intensity, eventually finds the female.

Vertebrates, which have taste buds and nostrils, can detect flavours and odours (actually chemical particles) that can convey a variety of meanings. Dogs and wolves stake out territories by urinating or defecating at "scent posts." When a member of a minnow school is injured or killed, special glands in the skin release a substance termed *schreckstoff* ("alarm-substance") which warns other minnows of danger. Sharks may be driven into a feeding frenzy when they detect blood in the water.

2. Visual Cues.—Visual signals require a photoreceptive sense organ. Unlike chemical signals, visual signals are direct, rapid, and localized. Such signals are produced in nocturnal insects or deep-sea fishes by bioluminescent organs. In day-active animals visual signals are produced by stereotyped movements or postures that are easily recognized. Often colours and special structures, such as crests or manes, further enhance the display movement. Fishes and octopuses possess special cells known as chromato-

(LEFT) JOE VAN WORMER FROM PHOTO RESEARCHERS, (RIGHT) KITCHEN-KINNE FROM PHOTO RESEARCHERS

FLOCKING GEESE (LEFT) AND SCHOOLING FISH (RIGHT) ARE SYNCHRONIZED GROUPINGS LACKING A TRUE LEADER



JOHN GERARD FROM NATIONAL AUDUBON SOCIETY, (RIGHT) JANE BURTON FROM PHOTO RESEARCHERS

Chemical communication binds (left) termites in a rigid caste colony and (right) provides scent trails for processionary caterpillars

phores that allow rapid colour changes that convey a subtle form of visual communication. Seasonal plumage or hair colour changes in birds and mammals achieve similar effects.

The use of models in the experimental determination of visual communication systems in birds and fishes has met with much success. Through such techniques it has been found that animals often respond to a very small but relevant portion of the total visual pattern that humans perceive. For example, a young herring gull pecks at the bill of its parent and thus induces it to regurgitate food, which is then eaten by the young bird. Studies of the pecking response, using cardboard models, show that the head shape and bill structure have little to do with releasing the response by the young. The chicks respond best to an object that resembles a long, thin stick striped at the top with red bands on a light background; such an artifact that produces a greater response than a natural stimulus is said to be a supernormal stimulus for releasing a response. In this example, the striped stick must be held vertically near the chick and moved slightly to produce an optimum pecking response. The model is really an exaggeration of the relevant stimuli present in an adult gull's bill, a red spot on the tip of the lower mandible that serves to release the pecking response. From the chicks low level of view the bill resembles a long, thin object in the same manner that the supernormal stimulus does.

Just as certain animals respond to a small part of the total available sensory input, others respond to a complex of stimulus configurations. Many birds, fish, and mammals have graded displays that change with the motivational state of the sender. For example, in the chaffinch, as the tendency for attack increases the bird shows a progressive raising of the wings and an increased flexing of the legs. Cats may show a graded series of postures from a crouch with flattened ears and slitted eyes to a standing posture with erect ears and open eyes. All intermediate variations indicate mixtures of tendencies to attack and flee. The more complex and subtle the social interaction the more highly graded is the signalling mechanism, indicating that learning has probably played an important role in associating the appropriate responses with a given expression. A complex set of sense organs is required for perceiving the relevant stimuli and a complex nervous system is necessary for associating the signals with the responses.

3. Sounds.—Auditory communication is very common in terrestrial vertebrates and invertebrates. It has great importance for aquatic vertebrates such as fishes and mammals. The chief advantage of auditory communication over visual communication is that it can be performed out of the direct line of sight of another animal. As with visual communication, auditory communication can give information concerning the sex, age, species, motivational state, and individual identity of the sender. Sound messages can promote aggregation (mating song of crickets) or dispersion (distress call of birds). The use of the tape recorder and sound spectrograph has permitted accurate analyses and physical descriptions of auditory signals as well as experimental verification of the role of sound in animal communication.

4. Contact.—Tactile communication is the direct contact between two individuals. A contact ceremony is often quite ritualized in that certain areas of the body are primarily involved in an initial contact. A set of stereotyped contact postures are developed; in mammals these generally include naso-anal, naso-nasal, and naso-genital contact patterns. Chemical, visual, auditory, and tactile communication may all be involved simultaneously.

See also PSYCHOLOGY, COMPARATIVE: *Man Compared with the Lower Animals* for a survey discussion of aspects related to communication.

II. INTERSPECIFIC SOCIALITY

As indicated earlier, animals occur in communities, predictable species complexes within habitats characterized by specific climatic, geochemical, and botanical features. Members of a community are continually interacting and thus exchanging information. By their interaction, community members exert selective pressures on each other that confer a survival value to certain characteristics within a given population.

1. Predator and Prey.—A predator and its prey form a very special case of interspecific communication. A predator is faced with the problem of locating appropriate prey and, conversely, a prey is faced with detecting a predator and either escaping or defending itself. Occasionally, the most conspicuous anatomical features of a species are the direct result of selection for defense against predators. The armadillos are a case in point. The most heavily armoured armadillos (*Tolypeutes*) roll into a ball and thus present an impregnable sphere to the predator. The other genera behave differently when attacked. The *Chlamyphorus-Chaetophractus* complex, as well as the genus *Priodontes*, has flat carapaces that provide adequate protection when the animals dig a shallow depression into the ground to avoid predators. The genus *Dasypus*, lightly armoured and not so well adapted for digging relies on its long legs and flees into dense vegetation, where its plates serve to ward off rough branches. In the family of armadillos, then, the behavioural traits of rolling into a ball, digging into the ground, and running to cover are differentially correlated with differences in the thickness of the armour, the length of claws, and the length of limbs. All these conspicuous features have resulted from selective pressures exerted by predators.

Both interspecific and intraspecific sociality may be promoted by the action of predators. Colonial life permits many sense organs to be employed in the detection of predators. If warning cries are selected for, many colony members, without actually perceiving a predator, can be made aware of its presence. This is the protection afforded a flock or a school. Furthermore, sudden scattering of flock or school members confuses the predator, imparing its chances for a successful kill.

Solitary animals frequently evolve cryptic postures or camouflage patterns in response to the selective pressures exerted by predators. Freezing in place is a response that renders prey less conspicuous because of its immobility. Camouflage aids concealment because either the markings blend with the background or the patterns of line and colour break up the outline of the animal.

2. Model and Mimic.—If an animal is distasteful or possesses a potent weapon such as a sting, it is generally avoided by experienced predators. Such unpleasant prey often assumes a striking colour or a conspicuous pattern (warning colours). Species

resembling noxious types with warning coloration share the immunity from predation conferred on the noxious species.

Such model-mimic systems need not, however, evolve only from the discrimination practised by a predator. In the case of cleaning symbiosis discussed previously, the host fish and the cleaner fish have evolved elaborate interspecific communication systems. Communication involves special soliciting postures by the host and specific coloration and movement patterns by the cleaner. In several cases, a small predatory fish that resembles a cleaning species can approach a prey fish and feed at least once with impunity. The hosts, however, soon learn to discriminate between the true cleaner and the mimic predator, but a definite selective pressure exists to favour the maintenance of the mimic since naive host fish can still be approached by it. This points up an unsolved problem with model-mimic systems: if the mimic achieves a population density sufficient to cause the selecting species to discriminate between the model and the mimic, the mimic could be exterminated. (*See* also MIMICRY.)

3. Competitors.—Within a natural community, different species compete for some resource in restricted supply. Selection tends to reduce competition by producing populations that utilize the available resources in specialized ways, thus producing diversity among members of the same community. In fact, members of the same species compete with one another more strongly than they do with members of different species. The wood warblers of the eastern United States provide an example of such specialization: different species found in the same community tend to forage in different zones of the trees. Some, like the Cape May warbler, forage in the crowns of trees; others, like the myrtle warbler, forage in the lower limbs. The effect is reduction of interspecific competition.

A definite dominance order may exist among several species in a community that compete for a single limited resource. In arid regions of southwestern California, small mammals of the chaparral community compete for sources of water. During the dry season the succulent prickly pear cactus is the common source of water for several rodent species. The desert wood rat nests in prickly pear patches and aggressively defends them against other wood rats and other rodent species including the agile kangaroo rat and several species of deer mice. A definite limit is placed on the population density of the associated mammalian species by means of the dominant status that the wood rat holds.

4. Mixed Groups.—Numerous vertebrate species form social systems that exhibit some cohesion and permanence over time. These social structures include mixed flocks of birds, mixed herds of hoofed mammals, and mixed schools of fishes. Bird flocks of mixed species have received the most study.

In the New World tropics honeycreepers, blue tanagers, and green tanagers typically form mixed flocks. Such social groupings have selective advantages including an enhanced ability to detect predators and to locate food. Within a mixed flock, several species are nuclear in that they consistently form a flocking association, while other species may be occasional joiners and are hence considered attendant. The nuclear species are generally of two types, termed passive and active. The passive species exhibit pronounced gregariousness and tolerate the proximity of the active, nuclear species that are motivated to join and synchronize their movements with the passive nuclear species. Because of the survivorship benefits resulting from mixed flocking, selection favours seasonal or permanent plumage changes that reduce conspicuous differences between nuclear species in mixed flocks.

III. INTRASPECIFIC SOCIALITY

A. Reproductive Association

Social groups may arise from the processes of reproduction. Sexual behaviour demands the formation of mass mating groups, temporary or permanent pair bonds, leks, or harems. Parental care results in the various forms of family groups. The progeny themselves may constitute a social group. As a result of asexual reproduction in protozoa, the newly formed individuals may constitute an aggregation because of the limited nature of suitable habitat. Young spiders emerging from an egg case may constitute a temporary aggregation because many eggs are laid in one spot. On the other hand, complex, cohesive organizations may arise from family groups as a result of a division of labour and a sustained interchange of positive stimuli among the group members. The complex insect social organizations of the Hymenoptera and termites are derived from family groups with a permanent division of labour based on a caste system (see *Insect Colonies*, below).

1. Mating and Courtship.—Certain environmental cues serve as stimuli to prepare the sexes for reproduction. In the temperate zone the reproductive periods are cyclic and correlated with seasonal climatic changes. Increasing light with the onset of spring can serve to stimulate the hypothalamic part of the brain via the optic nerves in ferrets and many species of birds. Such hypothalamic stimulation results ultimately in the production of pituitary hormones that induce production of mature sex cells. Light functions as a necessary stimulus to reproduction in cattle and sheep, but here the shorter day lengths of late summer rather than the longer day lengths of spring are the trigger.

Environmental stimuli may be quite precise in their effects. Palolo worms spawn annually in the lagoons of the South Pacific only in response to the last quarter of the October–November moon. Such synchronized spawning involves virtually the whole adult population of these marine worms. They swim to the surface in writhing masses. The posterior segments containing the sex organs are detached from the body and subsequently discharge their gametes into the water. In habitats such as the desert, selection has favoured very restricted stimulus inputs as triggering mechanisms. Arid-adapted birds such as the zebra finch are primed to respond to rainfall by an almost immediate onset of courtship, nest building, and sexual behaviour. Such an adaptation permits the rearing phase to follow the growth and seed production of arid-adapted plants, which are especially vital as a food source.

Environmental control of reproduction is stressed in the cases above, but in addition most vertebrates and all terrestrial-breeding invertebrates are generally dependent on an interchange of stimuli between the pair to promote mating. The communication system leading to the sexual act is termed courtship.

Courtship displays serve to attract mates, ensure species specificity in mating, arouse the mate, and synchronize activity. To guarantee species specificity, the form of the display and movement patterns tend to be unique. Often a chain reaction of stimuli is involved in which a signal from the male triggers a signal from the female, which in turn serves as the next initiatory stimuli for the male, and so on. Many courtship movements reflect an ambivalence that indicates origin from intention activities caused by the simultaneous arousal of attack, escape, and sexual drives within the interacting animals (*see* COURTSHIP, ANIMAL).

Courtship and related activities sustain physiological arousal demanded of reproductive activities. In a bird such as the ring dove courtship activities stimulate the pituitary to secrete follicle-stimulating hormone, which acts on the ovary to produce estrogens. Estrogens produce a sensitization of the brood patch on the breast of the bird. Brood-patch sensitization promotes nest-building activity, which in turn stimulates ovulation and subsequent progesterone secretion by the ovary. Finally, progesterone further sensitizes the brood patch and promotes broodiness, which leads the female to incubate the eggs.

Courtship and mating behaviour involve pair formation. The pairing relationship may be brief and terminated at the conclusion of the sex act, or it may involve a long-term interchange of stimuli between the partners, a pair bond. In pair bonds both parents are generally involved in the care of the eggs and young.

2. Mating Systems.—Several different mating systems can be distinguished by consideration of the numbers and sex classes of the animals involved and the duration of the pair bond. Promiscuity is a common form of mating behaviour in which no permanent pair bonds are formed. Monogamy exists when permanent pair bonding occurs. Polygyny involves one male and several females in the mating system. Two variations, lek breeding and harem formation, may be defined. A lek system involves a group of males that assemble on a traditional breeding ground. A dominance order is established with one aggressive male attracting visit-

ing females and performing most of the mating. Lek breeding has been noted in birds such as the sage grouse and cock-of-the-rock and in mammals such as the Uganda kob. In the harem system, a male defends a female group with which he has an exclusive mating right. Polyandry, a mating system involving one female and several males, is relatively rare; it occurs temporarily within certain primate societies such as that of the howler monkey.

3. Family Groups.—Parental care is well developed in some species of insects, crustaceans, and vertebrates. The parent and the young or egg clutch form a system that involves a constant interchange of stimuli. Such an interaction system is exemplified by the army ant *Eciton hamatum*. The eggs of a new brood exude substances that render them attractive to the workers. As a result the eggs are licked and carried continuously during the nomadic phase of the ants' activities. As the eggs develop, first into larvae and then into pupae, they continue to provide a focus of attention for the workers. During the pupal phase the army ant troop settles in a protected area until the pupae hatch. The newly emergent young workers trigger a renewal of wandering activity in the ant troop and the newly laid eggs for the subsequent brood are again taken up and transported during the wandering phase of the troop's life. Thus the alternation between the stationary and wandering phases of the army ant is a direct function of the stage of development of the brood.

In a similar fashion a mother rat and her brood portray a communication system of interdependence. The newborn rat is dependent on appropriate stimuli from the female for its initial development. The rat pups are attractive to the female, and at birth she licks the amniotic fluid and blood from their bodies. Lactation is sustained by the tactile stimuli of the nursing young. At the same time the turgid organs are relieved by the nursing activities of the young. The young rat will not exhibit a normal elimination reflex unless the female licks the anal-genital area. One cannot avoid the conclusion that the mother-young relationship involves an interchange of mutually interdependent stimuli.

B. Early Experience and Socialization

Research has established the importance of the social environment to the development of behaviour during maturation. The classical technique has been the isolation experiment, in which young animals are deprived of the animate and inanimate stimuli normally encountered in the course of their development in nature. Several important generalizations relevant to social behaviour have been derived from these experiments. These include the concept of a critical period, during which socialization takes place, and a realization that species differ in the amount and quality of stimuli necessary to produce normal adult behaviour.

1. Critical Period.—Imprinting is a type of rapid learning that occurs during a discrete period in the life of an animal. The classical examples of imprinting were derived from studies of ducks and geese. In a typical case of imprinting, a young gosling hatches and within 24 hours is capable of following the parent. During the process of following the parent, the gosling learns the visual cues and calls of the adult bird and can discriminate between its own parent and a strange adult goose. The process of imprinting on adult characters carries over into the young bird's adult life. Not only are the individual stimulus patterns of the parent learned but also the general stimulus configuration of a goose. Thus, the young gosling imprinted on its parent will in adult life seek out geese of its own species.

Goslings raised by different species of geese or by man will imprint on the foster parent. Such imprinting is virtually irreversible and the young bird when it reaches maturity will select its foster parent type as a sexual companion. Imprinting can take place only during a critical period of a few hours' duration. If imprinting by means of following a parent or foster parent is denied the young bird, it will be incapable of forming permanent adult liaisons with members of its own species.

A similar phenomenon is encountered in hoofed mammals in which the young exhibit a following response directed toward the parent. In the case of sheep, the mother is maximally receptive to learning the odour of her lamb during the first few hours after its birth. The time interval is critical, because if the ewe is denied access to the lamb for several hours after its birth and then is confronted with the lamb she is prone to refuse it. The lamb apparently learns the visual, chemical, and auditory characteristics of the ewe over a period of several days.

In mammals such as the dog which are helpless for some time after birth, there are longer periods of adjustment to society. Social behaviour develops from the point of full awareness to final weaning. Puppies reared in isolation never form close attachments with humans or other dogs (*see* Dog: *Development of Behaviour*).

Some species such as the golden hamster are dependent on a minimal stimulus input for the

REPRODUCTIVE ASSOCIATIONS: (ABOVE) A HAREM OF SEA LIONS; (RIGHT) A FAMILY OF MULE DEER WITH AN EXPERIENCED DOE LEADING THE GROUP; (BELOW) THREE SAGE GROUSE COCKS IN COURTSHIP DISPLAY BEFORE ONE HEN ON A STRUTTING GROUND

(TOP LEFT) C. T. HOTCHKISS FROM NATIONAL AUDUBON SOCIETY; (TOP RIGHT) JOE VAN WORMER FROM PHOTO RESEARCHERS, (BOTTOM) BILL GABRIEL FROM NATIONAL AUDUBON SOCIETY

normal development of adult social behaviour. Hamsters removed from the nest at three days of age and reared in isolation by artificial feeding and incubation techniques appear little affected in their development of adult patterns of aggressive, courtship, and sexual behaviour.

2. Effect of Social Deprivation.—Higher primates, including man, are strongly dependent on a social milieu for normal development. A young rhesus monkey reared in total isolation from other monkeys, with only a crude model consisting of a nursing bottle fixed to an upright wire body with a false face, learns to associate the model with food and with protection from frightening stimuli. When two models are supplied to the monkey, one with a wire body and the other covered with cloth, the infant chooses the cloth model over the wire model. This preference is shown even though the infant can feed only off the wire model. The young seeks a certain type of tactile stimulation from the model that becomes associated with comfort.

Infant monkeys raised with models exhibit abnormal behaviour when placed in groups. As juveniles they show a tendency to withdraw and rarely exhibit normal sexual behaviour. If, however, such isolated young are allowed to form play groups with other isolated infants at about six months of age, their adult behaviour is not so deviant. Play-group therapy initiated after one year of age, however, is less effective in promoting normal social behaviour.

Likewise, children reared in foundling homes where a minimum of social stimuli are present exhibit behavioural pathologies of an irreversible nature. Motor development as well as intellectual development is slow. After three years of age such institutionalized children are generally hopelessly retarded. It appears that infant higher primates have periods of socialization during which social stimuli are imperative for initiating learning processes that carry over into adult social life. (*See* PRIMATES: *Socialization.*)

Periods of socialization generally occur during the transition phases between the infant stage and juvenile stage of maturation. In higher vertebrates such transition phases are characterized by play behaviour. In play behaviour no ultimate resolution occurs, but rather components of functional adult patterns appear. Incipient food seeking, fighting, mating, and flight are displayed. Play groups, including littermates or the parent, are a source of the stimuli necessary for normal maturation (*see* PLAY, ANIMAL).

C. FLIGHT AND FIGHT

1. Agonistic Behaviour.—Agonistic behaviour includes flight, threat, aggression, and fighting. Aggression may be defined as the act of initiating and sustaining an attack. Fighting results from an attack and may be defensive or aggressive in character. Flight involves the avoidance of an attack by running away. Threat is generally a ritualized display involving mixed elements of aggression and flight.

Aggression can be elicited by a variety of stimuli. In some cases the stimulus releasing an attack is quite specific. A red breast induces an attack response by male chaffinches, and similarly a red belly induces attack by male stickleback fishes. Aggression can be enhanced by learning. A mouse may be trained to fight in a specific situation if it is always allowed to win; conversely, it may be trained to avoid or flee if it always loses in a staged encounter. Aggressive behaviour, with a neurohormonal basis, is ultimately determined by heredity. Man has taken advantage of certain hereditary tendencies to fighting in fish, fowl, and dogs, selecting them for his own purposes.

The control of agonistic behaviour is reflected in a variety of behavioural mechanisms. Many animals defend an area or territory around a set of vital resources. In some cases a territory may involve only the nest and a feeding area, but in all cases the living space is divided for best utilization of available resources.

Agonistic behaviour may be divided into two broad categories: interspecific and intraspecific. These categories reflect entirely different selective pressures. In interspecific relations of a predator-prey or competitive nature, selection often favours mechanisms of communication entirely different from those it favours in intraspecific relations. Thus, an herbivore such as the European rabbit has evolved definite flight or freezing patterns when it perceives a sudden stimulus contrast or the odour of a predator. On the other hand, the rabbit is highly territorial and during the breeding season the males fight vigorously with one another over territories containing females in reproductive condition. Very definite communicatory mechanisms are selected to promote intraspecific attack, threat, and submissive behaviours.

BY COURTESY OF HARRY HARLOW, PRIMATE LABORATORY UNIVERSITY OF WISCONSIN

SEPARATED AT BIRTH FROM ITS REAL MOTHER, A YOUNG RHESUS MONKEY SHOWS MORE FONDNESS FOR A COMFORTABLE CLOTH "MOTHER" EVEN THOUGH FOOD IS AVAILABLE ONLY ON THE WIRE "MOTHER"

Most aggressive displays are ritualizations of basic behaviour patterns that originally reflected ambivalence. When an animal is aroused to fight it generally passes through a preliminary phase during which it has a simultaneous tendency to attack and to flee. It may then show intentions to approach or to avoid, or it may alternate between rushing and withdrawing. Generally, the autonomic nervous system is activated, hormone levels and blood distribution are altered, and hair is raised in a definite pattern. These concomitants of agonistic arousal have provided the substrate upon which selection has acted to produce ritualized threat displays. Such displays frequently suffice to warn away an intruder without the necessity of energy expenditure in an actual fight.

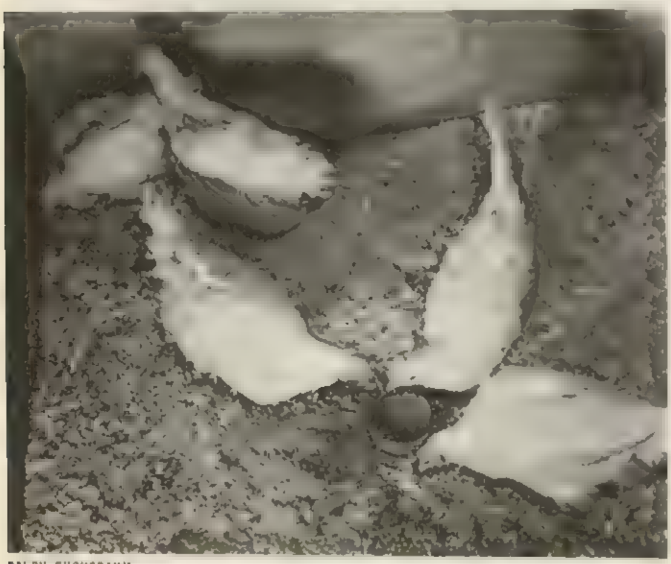
Selection has also favoured the evolution of ritualized combat, which involves a test of strength rather than a fight to the death. Deer mice generally fight with their bellies pressed together; the animals roll about on the ground but do not bite. Norway rats and certain mice exhibit a preliminary fighting phase involving a side display and pushing or upright postures and striking with their forepaws before actually chasing and biting. Deer, cattle, and antelope have antlers or horns that are employed in intraspecific fighting and to a much lesser extent in interspecific defense against predators. The fighting technique in hoofed mammals has involved selection for both ritualized contests and defensive patterns to avoid serious wounding.

Finally, selection has favoured the evolution of submissive displays to inhibit attack by members of the same species (conspecifics). An animal may flee from an attacking conspecific, as is the

case in many birds. In the much studied peck-order hierarchy of chickens, once the degrees of dominance are established, relations among members of the flock are relatively stable and aggression is reduced. In species possessing death-dealing structures such as fangs and claws, bodily harm to conspecifics can only be avoided if a definite submissive posture is evolved that inhibits further attack. This is especially true in species that live in a semi-permanent or permanent social grouping. Konrad Lorenz pointed out the efficiency of such a mechanism in dogs and wolves. When a defeated animal rolls on its back, exposing its vulnerable throat and belly, the posture inhibits further attack by a dominant (*see* also DOG: *Behaviour*). The point is well made that a weapon evolves together with a defensive system. If aggression is allowed unlimited expression, the survival of a species is placed in jeopardy.

2. Value of Aggression.—Aggressive behaviour is not pathological. A species is always in a dynamic state with its environment and the extent to which any agonistic behaviour is displayed is a function of the long-term needs of the species population. Aggression permits the spacing of a population and the establishment of hierarchical relationships. Seasonal fluctuations in aggressive behaviour correlated with breeding is one well-known example of the adaptive nature of aggression.

Aggressive behaviour displayed by man has its physiological and economic bases; however, mankind has invented a new level of aggressive expression termed warfare. War is possible only when a social system is developed enough to support economically the needs of a group that does not contribute to the primary production



RALPH BUCHSBAUM

AGONISTIC BEHAVIOUR: (ABOVE) CHICKENS FEED ACCORDING TO THEIR POSITION IN THE PECK ORDER: (BELOW) WINNING BLACK MOUSE IN A DOMINANT STANCE EVOKES A SUBMISSION REACTION FROM A WHITE MOUSE

of food and raw material. This nonproductive unit can then be geared to perform actions of offense and defense. (Only in the societies of slave-making ants is there an analogous phenomenon of warrior castes supported by a social mass of worker ants in a home colony.)

As technology advances, so also, unfortunately, does it produce tools of greater destruction, operable at distances more and more removed from their effects. Man loses the passion of immediacy, which can move him to violence or to mercy, depending on his victims' responses. He becomes a remote killer, deaf and blind to the mercy-releasing attitudes of his victims. With calculation, unconstrained by the results of his action, he can wreak incredible destruction, of a magnitude beyond the reach of any other animal. This stands in marked contrast to the personal involvement of early man in warfare or the aggressive interactions of other animals. His only hope is in cultural control of aggression, which his expanding technology has not given him time historically to develop. Perhaps the development of "ultimate weapons" by man will force him to adopt "ultimate peace" as the only cultural alternative to his own imminent extinction.

D. Invertebrate Societies

1. Simple Groups.—The simplest form of invertebrate social organization results from an attraction among the members of the group but it lacks any cohesion or coordination of movement. Cockroaches, attracted by the odours of other roaches, regardless of sex or age class, form compact groups in cracks. Such a grouping has been termed an assembly. Assemblies may be seasonal and include such social phenomena as swarming, sleeping, and hibernating groups.

Simple coordinated social groups are those in which the members regulate their behaviour with respect to one another. These include groups arising from reproduction and groups requiring some form of coordination, as among migratory insects that move *en masse* in a directed fashion. Insects that habitually sing as a chorus are able to synchronize their output and form a singing male group with the males spaced out in a predictable fashion.

Pair groups and family groups represent either temporary or long-term social organizations, the primary motivations being sexual drive and solicitation for young.

2. Insect Colonies.—Complex coordinated social groupings exhibit some division of labour outside the reproductive processes. Included under this criterion are such activities as group nest-building, group food storage, and group defense of an area. Such complex integration of activities has reached its culmination in the social insects, especially Isoptera (termites) and Hymenoptera (ants, bees, and wasps).

All the complex societies found in the Isoptera and Hymenoptera are derivations of a family group. Such societies evolved by a series of steps because of the selective advantage accruing from increased parental care.

In the higher termites, the ants and bees, complex societies have evolved independently that show remarkable similarities. Neuter castes have evolved and thus the reproductive success of the founding female (or, in the case of termites, the founding pair) is a function of the success of the whole colony as an integrated unit. This leads to the sometimes held concept of the colony as a supraorganism (see *Extensions and Analogies*, below). A termite, bee, or ant colony is analogous in a functional sense to an organism. Its reproductive success is a function of the total integration of the activities of descendants from a single genetic endowment. *See* SOCIAL INSECTS.

E. Vertebrate Societies

Vertebrate societies, except for man, have never achieved the degree of interdependence found in insect societies. This is primarily because each member of a vertebrate society is a potential reproductive. Neuter castes have not evolved; as a result, vertebrate societies do not approximate the level of a supraorganism but rather remain composed of equipotential individuals.

1. Fishes.—Several social phenomena are exhibited by the teleost or bony fishes including the ability to hold territories and

form dominance hierarchies, family groups, and schools. Many species of fishes form seasonal or permanent groups called schools, which exhibit simple coordination and some cohesion. If a schooling fish is isolated, it will exhibit a rapid series of movements about its living space that increase the probability of its locating other individuals. A fish employs visual, chemotactic, and vibrational cues to locate a school and if isolated will readily join any group passing near it. Many fishes, however, respond selectively when offered a choice between a school of its own species or a school of a different species. Aggressive behaviour is reduced in any school member, but in seasonally social fish the onset of the breeding season results in the breakup of the group and the establishment of individual territories.

Schooling fishes regulate their behaviour with respect to one another so that during linear or turning movements a definite synchrony is maintained, a manifestation of social facilitation. There seems to be a distinct survival advantage provided by schooling, since it evolved independently many times within the teleosts. Certainly there is a mechanical advantage inherent in swimming in a school. The fishes forming the leading edge of the school expend the greatest energy in overcoming the displacement drag of the water; following fishes have a considerably reduced displacement problem. A large school also may locate a food source more easily than a single fish; if one part of a school finds food, the school tends to turn in on the source. Predatory fishes often hunt in schools and encircle schools of the prey fishes. In addition to these selective advantages for schooling one primary advantage stands out: the school is a protective device. A school of fishes, when attacked, responds in a manner that promotes survival; the response corresponds to the adaptive type of the fish. Pelagic fishes, which wander over the open ocean, have several responses to predators including streaming to and fro, which distracts the predator, leaping into the air as a group, forming a dead space around the predator by avoiding it, and inducing avoidance through social facilitation in peripheral school members. Semi-pelagic schooling fishes are generally active in the vicinity of reefs and when attacked scatter as a group into the protection of offshore rocks. Reef fishes generally have special individual hiding places to which they flee individually when the school is attacked. Schooling fishes that are bottom feeders disperse but remain close to the bottom when attacked.

The teleosts have a range of reproductive adaptations with parallel and convergent evolutionary trends. Although many species lay vast quantities of eggs that are fertilized externally, there are definite trends toward laying a smaller number of yolky eggs that require greater parental care. The evolution of internal fertilization with the ejection of fertilized eggs by the female is a step toward the evolution of live bearing (viviparity), which results in the production of fully developed young with greater chance of survival.

Parental care, occurring in the majority of bony fishes, involves constructing a nest (cavity, bubble raft, water weed), guarding the eggs, or carrying the eggs or young in the mouth. The role of parental care may fall to either parent or to both; in general the bony fishes are characterized by father families, in which the male constructs a nest and subsequently defends the egg clutch within it.

The parental care system is adaptive. Darters of the genera *Etheostoma* and *Percina* display two levels of parental care with their adaptive correlates. *Percina caprodes* spawn on the gravel bottoms of small streams. The sexes do not differ conspicuously at the time of reproduction and the male does not defend a definite territory. A male will pair with a female and defend her, but only during the actual course of spawning. The eggs are laid in the gravel in small groups of 3 to 15. A female may spawn with several different males until as many as 2,000 small eggs have been laid. There is no parental care. In contrast, the sexes of *Etheostoma maculatum* show conspicuous differences, the male being brightly coloured and larger than the female. He actively defends a small territory that includes the potential site of egg deposition. The female remains paired with the male and lays eggs singly until about 400 have been deposited. The eggs are large and yolky and the small clutch is defended, cleaned, and fanned by the male. This comparison indicates several important generalizations. Increasing parental care is accompanied by a decrease in egg production. Division of labour in parental care involving defense of a territory is generally accompanied by a strong sexual dimorphism, with the male becoming brightly coloured and developing conspicuous accessory structures advertising his maleness.

Similar evolutionary trends have been described within the family Cichlidae. The more primitive genera exhibit little dimorphism, lay large numbers of eggs, and lay the eggs on the bottom. The pair bond is persistent and the pair cooperatively tend and guard the eggs. From such a parental family several genera have evolved that mouth brood. Mouth brooding involves the uptake of the eggs by one parent into its mouth immediately after spawning. With the evolution of mouth brooding there has been a concomitant increase in egg size and decrease in the total number of eggs laid. Either the male or the female is the mouth brooder, depending on the species; with this division of labour selection has favoured sexual difference in coloration and body size.

In summary, teleost fish exhibit schooling and to some extent territorial and parental care behaviours. The complexity of their social organizations does not exceed that found in the parental family and even with the complex forms of parental care there seems to be no permanent cohesion of the family group. (*See* also FISH.)

2. Amphibians and Reptiles.—Among amphibians complex parental-egg interrelationships may exist, but no special interdependent social unit exists beyond the level of parental care. However, amphibians do exhibit special forms of sociality termed breeding assemblies. These involve homing after a dispersal phase to specific breeding areas such as ponds or streams, participation in choruses by male frogs, and inclusion in breeding swarms of mixed sexes as typified by some species of newts.

Eggs are guarded by males of the salamander genera *Cryptobranchus* and *Hynobius*. The female tree salamander (*Aneides lugubris*) guards her clutch also and several generations may jointly occupy the same cavity in a rotten log. In the midwife toad (*Alytes obstetricans*) the male carries the eggs wrapped about his body. The female Surinam toad (*Pipa*) transports her eggs in special skin pockets on her back. The eggs are encapsulated at the time of spawning after they have been fertilized and pushed into her vascular skin by the male. The male of a Chilean frog, *Rhinoderma darwini*, takes the newly laid clutch into his mouth and forces the eggs into his vocal sacs. The eggs are transported in this moist environment until the time of hatching. Viviparity has seldom evolved in the amphibia, but the salamander *Salamandra maculata* and the frog *Nectophrynoides vivipera* are notable exceptions that bear the young alive. (*See* also AMPHIBIA.)

In general, reptiles are characterized by a lack of complex social organizations. Territorial defense by males is common in the day-active lizards. In species such as *Sceloporus occidentalis*, territorial display and mating rituals are highly developed. Parental care, aside from nest building, is sporadic and extends little beyond guarding the nest. Special forms of social behaviour include mass hibernation in traditional dens by many species of temperate zone snakes. In addition, communal nesting sites are employed by the European water snake (*Natrix natrix*) and some geckoes. Reproduction is completely by internal fertilization, with definite courtship ceremonies preceding copulation. Egg laying is the primitive mode of reproduction, but many species have evolved ovoviviparity, in which the eggs develop within the female until the moment of hatching. A few species of reptiles are truly viviparous and nourish their young by means of a placenta. These include the night lizard (*Xantusia vigilis*) and several species of skinks. In *Xantusia* the female has been observed to break the membranes and pull at the emerging young.

3. Birds.—The social life of birds has been an intensive area of investigation for decades. Conspicuous social phenomena include flocking, pair bonding, complex parental care, territorial defense, colonial nesting, and communal nesting.

Flocking behaviour results from response mechanisms that promote group synchronization and integration. Call notes promote contact while in flight or when dispersed and feeding in dense

vegetation. Social facilitation, or the readiness to duplicate the activities of other individuals, such as straight flight, turning flight, landing, or feeding, is essential in maintaining group synchrony. At the same time, birds very often exhibit mechanisms that promote dispersion. These include such behaviour patterns as individual distance, which may vary according to whether the birds have just arrived on a perch (arrival distance), become settled (settled distance), or recently flown off (departure distance). Individual distance results in part from a tendency to approach or to flee from an approaching conspecific.

Many bird species form flocks outside the breeding season but maintain territories during the rearing phase of the life cycle. Very often the form of the flock reflects the type of social organization at the time of reproduction. Among the African finches, *Euplectes afra* is characterized by a strong territorial defense during the breeding season, whereas *Quelea quelea* breeds in dense colonies in which only the nest entrance itself is defended against conspecifics. Outside the breeding season *Q. quelea* forms dense, synchronized flocks, whereas *E. afra* flies in loose flocks.

Colonial nesting is a common phenomenon among birds. Pairs of birds build separate nests that they defend against conspecifics. All the nests are located close together in some restricted area. Colonial nesting is common among seabirds that occupy a small island or rock cliff for nesting year after year. Communal nesting occurs when two or more females lay eggs in the same nest and share the incubation duties.

The use of one nest by several females has evolved convergently in several families. In the family Cuculidae, which is characterized by the nest-parasitic cuckoo, the subfamily Crotophaginae portrays a series of forms culminating in the communal nesting anis. All crotophagines live in colonies of about 12 individuals. The colony members defend a group territory that includes feeding, nesting, and roosting areas. The primitive species of the Crotophaginae, *Guira guira*, forms pair bonds but the nest defense is weak. This behaviour may have been the prelude to communal nesting. The genus *Crotophaga* is characterized by weak pair bonding and polygamy. The females do not defend their nests against one another, and in *C. ani* several females deposit eggs in the same nest. Colonial nesting does not lead to communal nesting in other bird species. It would seem that communal nesting is facilitated by a breakdown in pair bonding. Factors contributing to loose pair bonding may include physiological changes in the female such as spontaneous ovulation, which reduces the necessity for sustained or prolonged courtship. On the other hand, communal nesting cannot occur unless the species is already a colonial nester. A factor that contributes to colonial nesting is the absence or shortage of available nesting sites. Seabirds are very often colonial nesters on islands or cliffs, but a strong pair bonding prevents communal nesting. In the case of Crotophaginae the colonial nesting habit was probably reinforced in the savannah habitat, the centre of origin for the subfamily. Widely spaced groves of trees favoured colonial life, but the loss of pair bonding by the hypothesized physiological changes within the female promoted communality.

All birds are oviparous and exhibit complex forms of parental care. Parental care includes nest building, incubation, guarding the nest and young, and feeding the young. One or both parents may be involved in the rearing process.

Primitively, birds laid large eggs with copious food reserves. The young were precocial when hatched and needed no parental feeding. The large egg required a long incubation period with a long attentive period by the adult. Advanced birds have departed from the primitive pattern in several ways. The perching birds (*Passeriformes*) are generally small and as a result lay small eggs with a small amount of nutritive yolk. The young are more dependent when they hatch and must be provided with food.

The division of labour between the sexes in parental care runs the gamut from total female care to total male care. In the ostrichlike birds (ostrich, emu, etc.) the male assumes most parental care duties while in most other birds the common trend is for the female to assume the major role in parental care in varying degrees among the different families.

4. Mammals.—The defining feature of mammals is the presence of female mammary glands that produce milk. For this reason parental care of a mother-family type is universal. The simplest set of social relations involves a brief pairing by the male and female, separation of the male and female, and, at birth, the formation of a female-young social unit. Dispersal of the young at maturity coincides with the breakup of the family group. In contrast to birds, mammals very rarely display pair bonding or the inclusion of the male in the role of parental care; such behaviour has evolved only among some primates and carnivores.

Outside the social units resulting from reproduction, mammals exhibit the following social structures: colonies of individuals typified by ground squirrels, communal roosts and hibernating assemblies of bats, herds of grazing hoofed mammals, schools of cetaceans, and hunting packs of certain carnivores. The phenomena of territoriality, dominance hierarchy formation, individual distance, and social facilitation are demonstrable in many mammalian groups.

LOUIS RENAULT FROM PHOTO RESEARCHERS

A COLONY OF CORMORANTS CROWD THE ISLAND OF SAN LORENZO, OFF PERU; THESE BIRDS PAIR AND THEN DEFEND THEIR NESTING TERRITORY

Insectivores.—The order Insectivora includes species of shrews, the males of which exhibit a dispersed spatial distribution except during temporary pair formation, the female and the litter constituting the only social groupings. An exception to this general rule is the least shrew, which apparently does form small family groups consisting of a male, female, and the young. Some insectivores, such as the solenodon or the striped tenrec, may form family groups similar to that described for the least shrew. In general these small mammals do not form complex cohesive societies. Pair bonding is limited and adult male antagonism coupled with the antagonism of the female at parturition prevents stable group formation.

Rodents.—Rodents exhibit a wide variety of social organizations. The European hamster is typically an asocial mammal. The adults defend their individual burrows against intruding conspecifics and the only stable social grouping is the female with her litter. Pairing is very brief, although the male may indulge in a prolonged period of courtship in order to gain entry into the burrow of the female. The male employs chemical communication by marking objects with a flank gland secretion in the territory of the female.

The Norway rat is a more tolerant species. Adults may den alone at low population densities but they are able to form family bands that utilize a common foraging area and burrow system. Strange rats are attacked when they wander into the group territory. Special communicatory mechanisms have evolved to promote contact familiarity including a ritualized crawling over and crawling under behaviour shown by rats belonging to the same family group.

The prairie dog is a highly social species that lives in colonies, each composed of subgroups termed coteries, which include an adult male, several adult females, and their young. Coterie members defend a group territory against strangers. A variety of contact postures promote and reinforce recognition among coterie members. The beaver is also highly social, and lives in a family group composed of an adult male and female with the current litter and the young of the previous year. The beaver family utilizes a communal lodge and foraging area. Their group activities result in complex artifacts such as dams, food caches, and lodges. Apparently each beaver family is a closed social unit, and intruding conspecifics are repulsed.

Ungulates.—As in all mammals, the basic social unit is the female and her infant. Most young ungulates are able to follow the female shortly after birth. Within a few hours the female recognizes her young, and in the course of a few days the young animal knows its mother. Some ungulate females, such as the moose, are asocial during the rearing phase, but many, such as the red deer, exhibit strong herding tendencies.

In many ungulate species, including the red deer and the bighorn sheep, the adult males form loose herds separate from the female herds. Other species, such as the bison, may tend to form unisexual subgroups but may also form a large herd of mixed sexes. The size of the herd is related to the habitat; *e.g.*, dense forests or mountainous areas prevent the formation of large groups. The largest herds are formed by those ungulate species adapted to an open savannah or grassland habitat.

During the breeding season the males tend to become antagonistic toward one another. Many ungulate species exhibit a pronounced sexual dimorphism, with the males bearing conspicuous antlers or horns, which are used to intimidate or to fight. The male fights lead to the establishment of dominance or to the defeat of rivals, thus conferring on the dominant a mating priority.

The type of mating system in part depends on the size of the female group. Solitary ungulates form brief pair associations; however, species that typically form large herds of females and their offspring usually exhibit generalized mating systems. In species like the red deer, typified by extreme male-male antagonism during the rut, one male attaches himself to a female group and establishes a harem. If the males are more tolerant, as among bison, and are loosely attached to a female herd, the males will establish a dominance hierarchy with the dominant male accomplishing most of the breeding.

A mating system found in the Uganda kob involves dominance and the defense of a territory. The adult males have a traditional breeding ground where they establish and maintain small territories. The receptive females individually approach this breeding ground and join a male, generally one whose territory is centrally placed.

Ungulates and rodents have evolved social systems significantly more complex than those typified by most of the smaller insectivores. Living in herds, colonies, or communal family bands correlates with a shift from high energy diets including meat, nuts, and seeds to low-energy diets composed mainly of grasses. The shift to grass as the main foodstuff has been accompanied by major changes in dental patterns and in the digestive tract. The eating of grass is generally accompanied by a loss of the food-caching instinct and the adoption of behaviour patterns ensuring survival in an open habitat. Colonial life in the prairie dog and herd life in the bison and other plains-dwelling ungulates is an overall adaptation to promote warning and escape from predators. Herding or colonial life is possible in the grass eaters because the environment can support many more cellulose feeders per acre than it can animals utilizing high energy food sources such as seeds or, in the case of carnivores, other animals.

Carnivores.—Terrestrial carnivores are typified by a high potential for social interaction. As young animals, carnivores engage in complex play activities involving elements of sexual, prey-catching, and agonistic behaviour. As adults, however, many carnivores are solitary. This is especially true for those species that hunt by stealth, such as the cats. On the other hand, canids such as wild dogs, wolves, and coyotes, which must hunt by cooperatively running down game in a semi-open habitat, typically live in social groupings. It is noteworthy that the cheetah, unlike other cats, is adapted for hunting in a typically canid fashion and has evolved the tendency to hunt in small family groups.

All terrestrial carnivores produce helpless (altricial) young. They are reared in a protected spot and are generally immobile during their first few weeks of life. Group-hunting canids exhibit the phenomenon of provisioning whereby the male aids the female in bringing food to the den site for the young. Provisioning occurs in species that exhibit pair bonding or that live in small groups. Group life for a carnivore is possible only if the species is mobile enough to forage over a large area. The habit of food provisioning acquires a distinct survival advantage for the species, especially when the pregnant female is confined while giving birth.

Some examples for the major social systems found in the terrestrial carnivores may be listed as follows: species showing pair bonding with provisioning at parturition and the formation of small hunting groups (wolf, coyote, red fox); species utilizing a common nesting site for several generations (European badger); species living in loose hunting groups with little pair bonding (lion); species generally solitary except for a brief pairing period (bears, most cats, and most weasels); species forming a foraging unit of females with their young (coatimundi).

The aquatic carnivores include otters and pinnipeds (seals, sea lions, and walruses). The pinnipeds forage at sea but return to land or ice to breed, to molt, and to rear their young. Dense colonies may be formed during the breeding season, especially in species that are migratory and have a traditional, insular breeding ground. Two general types of social organization during the breeding season may be discerned. The first system is characterized by pronounced male-male antagonism and the formation of harems. Such polygynous breeding systems are typical of the elephant seal, fur seal, and gray seal. On the other hand, seals of the family Phocidae, including the common harbor seal, do not form harems but rather exhibit a promiscuous breeding system. Species having a harem system during the breeding season also exhibit a pronounced sexual dimorphism; the males may be two to three times as large as the females.

F. Primate Societies

Primates have exhibited several independent adaptive radiations during their evolutionary history. These include the evolution of the lemurs on Madagascar, the radiation of the platyrrhine mon-

keys in South America, the radiation of the catarrhine monkeys in the Old World, and the evolution of the anthropoid apes, including man.

1. Lemurs.—The lemurs exhibit a diversity in social structure that parallels certain adaptive trends already discussed. The mouse lemurs of the subfamily *Cheirogalinae* are nocturnal forest dwellers adapted to feeding on insects, small vertebrates, and some vegetable matter. These species have two to three altricial young and typically do not form pair bonds or permanent social groupings. In this way they exhibit tendencies parallel to those found in the insectivores. The larger lemurs of the family *Indridae* are vegetarians that produce one young having a prolonged period of development and maturation. Some indrids are diurnal and others nocturnal or crepuscular (twilight-active). All species appear to live in family groups. The true lemurs (family *Lemuridae*) are vegetarian and crepuscular. As with the indrids, they produce one semi-precocial young with a prolonged maturation period. They typically live in troops of adults of mixed sexes.

2. Higher Primates.—The higher primates, including the larger monkeys and the apes, are all characterized by a rich social life. The infants undergo a relatively extended period of maturation accompanied by a highly social interaction with their peers and with adults of both sexes; however, there is still much diversity in the exact form of the social organization. Communication mechanisms include visual and auditory signal systems. The higher primate facial expressions and vocalizations permit a graded series of expressions with a concomitant complexity of information transfer.

The social composition of the different groups varies. Some marmosets, the lar gibbon, and the night monkey live in family groups consisting of one adult pair and their juvenile or infant offspring. Species such as the spider monkey and the sacred baboon may live in loose troops composed of more cohesive subgroups. The subgroups are usually (1) females with their young accompanied by one adult male; (2) groups of adolescents of both sexes; or (3) all male subgroups. Species such as the howler monkey, the chacma baboon, and the rhesus monkey live in large troops of adults, juveniles, and infants. Some tendency toward the formation of unisexual or age-specific subgroups may be shown, but these species tend to be quite cohesive in their movements.

All primates have a predictable home range where they forage and sleep. The howler monkey, gibbon, and rhesus monkey tend to defend these areas against intruders. The chacma baboon, however, exhibits no overt agonistic defense and yet maintains a core area in its home range that is relatively free from invading conspecifics. The chimpanzee appears to exhibit no defense of a specific area but wanders freely in its home range, mingling with other chimpanzees and then separating. The cohesion of groups may be pronounced as in the case of the howler monkey, gorilla, and baboon; on the other hand, the chimpanzee exhibits little cohesion and groups are constantly forming and splitting. As a consequence, the chimpanzee troop is very permeable to outsiders, but the howler monkey and baboon societies remain essentially closed to infiltrators.

In the gorilla troop the oldest adult male initiates movement and controls the duration of rest periods. To some extent the dominant adult male baboons initiate and direct movements, although a baboon troop on the move exhibits a pattern that places the dominant males in the centre of the group. The actual leading edge of the troop is composed of the young males, who, consequently, are the first to contact a dangerous situation and give warning. Thus the function of vigilance and composition of the vanguard falls to the younger males, although the dominant males control the general direction and rate of movement.

Assertion of dominance may be weak and scarcely noticeable in species such as the Indian langur; on the other hand, dominance may be overly expressed in the rhesus or baboon troops. Dominance results from a long process of learning. Adult males are dominant over females and younger males; among the adult males as well as the adult females a definite hierarchy exists. A dominant animal has freedom of access to food and mates, thus a definite advantage exists for those attaining a superior social status. Reproductive control is achieved through dominance. In howler monkey, baboon, and macaque troops an estrous female typically consorts with several males, although the dominant frequently has access to a female at the height of her estrus period.

Primates demonstrate several trends in the evolution of sociality. The higher primates have departed from the patterns of nocturnality, omnivorous diet, and the production of rapidly maturing altricial young. Instead, one finds in the higher primates diurnality, a vegetarian diet, and the production of a single young capable of clinging to the mother but undergoing a prolonged period of maturation. The long period of maturation allows for an extended period of socialization. Young primates appear to be dependent on a rich social environment for the normal development of their patterns of behaviour. Experiments involving isolation of young rhesus monkeys demonstrated the complete lack of normal sexual behaviour in monkeys reared alone. Daily play with other juveniles compensates, in large part, for the effects of rearing in the absence of a mother (*see* PRIMATES: *Socialization*).

In spite of their social life and dependency on a rich social environment, primates do not exhibit group hunting activities or provisioning for the female. Food sharing is almost unknown except between the female and her infant. Coordinated activities are demonstrated by male baboon groups during defense against predators, or by male howler monkeys in their vocalization ceremony used to establish the bounds of their territory against neighbouring troops. Selective advantages promoting group life thus appear to be related primarily to defense against predators, to the maintenance of a group territory, or to the acquisition and maintenance of access to estrous females.

Selective pressures promoting social organization vary depending on the overall adaptive syndrome of the species, and variations are quite common. The marmosets exhibit pairing tendencies and the male carries the young thus participating directly in parental care. This adaptation may have resulted from the peculiar energy demands on a very small species faced with the problem of continually transporting the young during the rearing phase. The energy requirements of the lactating female may be such that to be relieved of the burden of transporting the young is of extreme selective advantage. The obvious selective advantage promoting group cohesion in the baboon and macaque is the transition from an arboreal to a terrestrial life with the attendant dangers of increased predation pressure. Group defense of the troop by the adult males confers a definite survival advantage.

3. Men.—Man's early pre-australopithecine ancestors were probably organized into social bands similar to that of the baboon. At this time man was an herbivore-omnivore foraging on the ground. At the australopithecine stage man had transformed his economy and become a hunter. (The contemporary mammalian economy that most nearly approximates australopithecine man's economy is that of the pack-hunting canids. The canids typically show group hunting and provisioning of the female. For similar economic reasons australopithecine man must have begun to develop a similar type of social system.) Somewhere between the level of australopithecine man and pithecanthropus man there developed a tendency to establish permanent family subgroups within a hunting group and to form cultural control systems of kinship, taboo, and religious symbolism.

Primate species such as the Japanese macaque may exhibit learned traditions of food selection that are passed on from generation to generation in the same manner that cultural traditions are passed on in human societies. While man's cultural attributes and use of tools has its roots in the behaviour patterns of his subhuman predecessors, he has ordered them and recorded them as no other species has.

Man has achieved a fantastic degree of complexity in his modern societies. His division of labour and interdependence have proceeded by two methods: (1) the creation of a fixed caste system exemplified by the guilds of renaissance Europe or the social structure of India in the last century; (2) the erection of a series of economic divisions with a free choice of occupancy. When Western society opted for a social philosophy based on equal opportunity for all mankind, it effectively banished the notion of fixed castes

and permitted a great mobility among the members of its society. Stability in Western society was conferred by the great complexity and division of labour demanded by technological advances. Such complexity provides an internal compensation analogous to compensation mechanisms found in the physiology of whole animal organ systems.

Yet even with such interdependence and specialization in modern society, persons remain independent units of reproduction and enjoy a social mobility conducive to individual development and freedom.

IV. Extensions and Analogies

Great care must be exercised in forming judgments about probable social behaviour in untested situations and especially about inner social feelings of animals under any conditions. It is unsafe to assume, without direct information, that one order of birds will show social patterns closely resembling those of some species belonging to a different order. It is still more hazardous to try to estimate the degree and kind of sociality across wider taxonomic gaps. For example, pending direct inquiry, it is unsafe to predict that individual-to-individual social hierarchies exist among sharks from the fact that such social rank-orders exist among many species of bony fishes.

Still greater caution must be exercised, again pending direct information on the precise point at issue, before making extrapolations between human sociality and the social life of other species. Men may and do disregard or compensate for inconveniences based on position in the social "peck-order," but we cannot judge therefrom that other animals do likewise. Despite the need for great restraint in such matters, stimulating cross comparisons can be made that help to integrate animal sociology into a unified whole despite the marked differences between the social relations of man and those of other animals.

1. Organism and Supraorganism.—The analogy between the biological processes within the organism and the sociological processes within society long ago attracted attention. Thomas Hobbes' *Leviathan* and the less cumbersome comparisons of Herbert Spencer in his discussion of the analogies between the organism and society are matters of historic interest only, but they did more or less vaguely foreshadow the modern realization, particularly by biologists, that similarities exist between individual organisms and organized social groups. Since even ecological communities, as well as more closely knit societies, behave as more or less integrated entities, they are often regarded by biologists as representing a set of emergent levels concisely designated as supraorganisms.

Biologists do not stress the resemblances between societies and a closely coordinated organism such as an insect or a man. They do see a few resemblances between the social supraorganism and simpler invertebrates. Sponges, for example, often have one part of the individual colony working in direct opposition to another when the cooperation of both is needed for efficient operation of the whole. Starfish, in righting themselves, and even in walking across a smooth surface, often tear off some of their tube feet that were working in opposition to the movement of the whole.

Society, human society as well as an insect society based primarily on inherited behaviour patterns, shows many pertinent comparisons with these more poorly integrated organisms, and yet primitive organisms and primitive or advanced societies are somewhat integrated wholes. The chief value inherent in the application of the supraorganismic concept to animal communities or societies lies in the suggestion of analogies. Similar mechanisms of integration that are in many cases not homologous become apparent through comparisons made within such a unified outlook.

Insects and mammals stand near the apex of their respective evolutionary branches of the animal kingdom. The similarities between the social life developed in each class are mainly matters of convergence and analogy. To the extent that the sociality of each develops from the cooperative tendencies of their common ancestors, their sociality is vaguely homologous. In many ways this is not a minor point. Even the sociality of man in all probability developed from basic drives that pervade all higher vertebrates. While such drive-systems in man are emancipated in part from their original stimulus fields, some continuity still remains.

2. Subsocial and Social Levels.—For general and comparative sociology, the chief interest is not between extrapolations from the organismic to the supraorganismic level in organization—from the biological to the social, so to speak—but rather in the permissible extrapolations from the subsocial to the truly social levels and in the opposite direction as well. Here again arises the problem as to where truly social life begins.

A part of the difficulty involved in answering this question may be avoided if it is recalled that many different levels of social organization exist and that these usually overlap. Among all the groups that the biologist regards as somewhat more social than is usually implied by more vague ecological relationships, the following may be recognized: (1) Those that show their social habit merely through the toleration of the close proximity of other similar animals in the same restricted space; these may exist without any positive mutual attraction and represent the toleration level of sociality. (2) Those groups that react more or less definitely as units—the group-integration level. (3) Those that show physiological division of labour. (4) Those that show morphologically distinct castes, each associated with some phase of the division of labour (the highest insect level). (5) Those that are organized to a large part by tradition and by symbolic communication and that frequently react to abstractions (the human level).

In presenting this abbreviated list of recognized social levels, there is no intention of assuming that the more complex types of social life evolved through the less complex ones. Such may or may not have been the course of social evolution. The insects do not stand in the direct line of man's ancestry. Their state of social development represents an insect evolution of sociality, not a stage toward that found in man.

Animals on the higher planes of sociality continue to show certain of the group attributes characteristic of the lower levels. Group survival values have been demonstrated throughout the whole series and apparently extend to the threshold of primitive life. Extrapolations forward from the lower to higher social levels can be made with more confidence than those in the opposite direction. It can be predicted, from studying the social survival values of primitive organisms, that, other conditions being equal, groups of more complex animals will have certain cumulative and emergent survival values as groups. It is much more hazardous to try to project highly developed human social traits, like social consciousness, back into the less complex social levels reached by other animals.

3. Conclusion.—A survey of animal societies indicates that tendencies toward cooperation and competition coexist as polar opposites within all phyletic levels of interorganismic activity. Which of these two tendencies is the more fundamental? Considering the evidence available, contrary to Herbert Spencer and the conclusions of social Darwinism, the cooperative forces equal the competitive in biological and sociological importance.

A survey of evolutionary trends indicates that organisms, societies, and ecological communities have tended to become more complex, interdependent, and integrated. The greater biological stability imposed by increasingly complex chains of organismic interaction confers tremendous survival value. However, this trend toward greater integration and complexity does not mean the loss of the capacity to express aggression or compete through ritualized aggressive mechanisms. Thus man, when viewed as an animal, must be considered as having both a natural capacity for socialization and at the same time the potential for expressing no inconsiderable amount of aggression. A rational consideration of man's nature should, then, take into account all of man's natural potential rather than seek to deny the multipolar nature of his biological background.

Bibliography.—W. C. Allee, *Cooperation Among Animals with Human Implications* (1951); R. Ardrey, *The Territorial Imperative* (1966); J. T. Bonner, *Cells and Societies* (1955); R. G. Busnel, *Acoustic Behaviour of Animals* (1963); J. B. Calhoun, "The Social Use of Space," in *Physiological Mammalogy*, ed. by W. Mayer and R. Van

Gelder, pp. 1–187 (1963); C. R. Carpenter, "Societies of Monkeys and Apes," *Biological Symposia*, vol. 8, pp. 177–204 (1942); N. E. Collias, "The Analysis of Socialization in Sheep and Goats," *Ecology*, vol. 37, pp. 228–239 (1956); J. H. Crook, "Flock Organization in Birds," in *Current Problems in Animal Behaviour*, ed. by W. H. Thorpe and O. L. Zangwill (1961); I. Eibl-Eibesfeldt, "The Fighting Behavior of Animals," *Scient. Am.*, vol. 205, no. 6, pp. 112–122 (December 1961); J. F. Eisenberg, "The Behavior of Heteromyid Rodents," *Univ. Calif. Publs. Zool. 69* (1963) and "The Social Organization of Mammals," in *Handb. Zool. 8* (10) (7) (1965); W. Etkin, *Social Behavior and Organization Among Vertebrates* (1964); A. M. Guhl, "The Social Order of Chickens," *Scient. Am.*, vol. 194, no. 2, pp. 42–46 (February 1956); H. F. Harlow, "Development of Affection in Primates," in *The Roots of Behavior*, ed. by E. L. Bliss, pp. 157–166 (1962); H. Hediger, *Wild Animals in Captivity* (1950); R. A. Hinde and N. Tinbergen, "The Comparative Study of Species Specific Behavior," in *Behavior and Evolution*, ed. by A. Roe and G. G. Simpson, pp. 251–268 (1958); D. Isaac and P Marler, "Ordering of Sequences of Singing Behaviour of Mistle Thrushes in Relationship to Timing," *Anim. Behav.*, vol. 11, pp. 179–188 (1963); M. Keenleyside, "Some Aspects of Schooling Behaviour of Fish," *Behaviour*, vol. 8, pp. 183–248 (1955); S. C. Kendeigh, "Parental Care and Its Evolution in Birds," *Illinois Biol. Monogr. 22* (no. 1–3) (1952); J. A. King, "Social Behavior, Social Organization and Population Dynamics in a Black-Tailed Prairie-dog Town in the Black Hills of South Dakota," *Contr. Lab. Vertebr. Biol. Univ. Mich.*, no. 67 (1952); A. Kortlandt and M. Kooij, "Protohominid Behaviour in Primates," *The Primates, Symp. Zool. Soc. Lond.*, no. 10, pp. 61–88 (1963); P. A. Kropotkin, *Mutual Aid, a Factor in Evolution* (1917); D. S. Lehrman, "On the Organization of Maternal Behavior and the Problem of Instinct," in *L'Instinct dans le comportement des animaux et de l'homme*, pp. 475–520 (1956); K. Lorenz, *King Solomon's Ring* (1952) and *On Aggression* (1966); P. Marler, "The Logical Analysis of Animal Communication," *J. Theoret. Biol.*, vol. 1, pp. 295–317 (1961); T. E. McGill, *Readings in Animal Behavior* (1965); M. Moynihan, "The Organization and Probable Evolution of Some Mixed Species Flocks of Neotropical Birds," *Smithson. Misc. Collns.*, vol. 143, no. 7 (Publ. 4473) (1962); A. Murie, *The Wolves of Mount McKinley*, U.S. Govt. Printing Office, Washington, D.C. (1944); J. A. Oliver, *The Natural History of North American Amphibians and Reptiles* (1955); W. Pfeiffer, "Alarm Substances," *Experientia*, 19:113 (1963); A. Portmann, *Animals as Social Beings* (1961); *Biological Symposia*, vol. 8 ed. by R. Redfield (1942); O. W. Richards, *The Social Insects* (1961); G. B. Schaller, *The Mountain Gorilla: Ecology and Behavior* (1963); T. C. Schneirla and J. S. Rosenblatt, "Behavioral Organization and Genesis in the Social Bond in Insects and Mammals," *Am. J. Orthopsychiat.*, vol. 31, pp. 223–253 (1961); J. P. Scott, *Aggression* (1958); J. P. Scott and M. V. Marston, "Critical Periods Affecting the Development of Normal and Maladjustive Social Behavior in Puppies," *J. Genet. Psychol.*, vol. 77, pp. 25–60 (1950); C. H. Southwick (ed.), *Primate Social Behavior* (1963); Centre Nationale de la Recherche Scientifique, Paris, *Structure et physiologie des sociétés animales* (1952); N. Tinbergen, *The Study of Instinct* (1951) and *Social Behaviour in Animals* (1953); F. Walther, "The Mating Behavior of Certain Horned Ruminants," *Int. Zoo Yb.*, no. 3 (1961); S. L. Washburn and I. DeVore, "Social Behavior of Baboons and Early Man," in *Social Life of Early Man*, ed. by S. L. Washburn, pp. 91–105 (1961); W. M. Wheeler, *Social Life Among the Insects* (1923) and *The Social Insects* (1928); V. C. Wynne-Edwards, *Animal Dispersion in Relation to Social Behaviour* (1962). (J. F. E.)

SOCOTRA (SOKOTRA; SUQUTRA), an island in the Indian Ocean, about 150 mi. (240 km.) ENE of Cape Guardafui, Somali Republic. Socotra is a dependency of the Mahri sultanate of Qishn and Socotra which forms the easternmost state of the British Protectorate of South Arabia, previously known as the Aden Protectorate. The island is about 70 mi. (110 km.) long, has an average breadth of more than 20 mi. (30 km.) and an area of about 1,200 sq.mi. (3,100 sq.km.). It is the most easterly of a group, of which Socotra and the islets of Samhah and Darsa (called the Brothers) are on united coral banks 10 to 30 fathoms deep. 'Abd al Kuri, 60 mi. from Socotra, is separated from them by a seavalley of 100 fathoms, and there are also two rocky islets, 13 mi. from 'Abd al Kuri, called Ka'l Fir'awn.

The summits of Socotra, Jabal Haggier (4,656 ft. [1,419 m.]), Samhah (2,440 ft. [744 m.]), and 'Abd al Kuri (1,670 ft [509 m.]) are among those land surfaces which have been longest, if not always, above water. There are wild asses of Nubian origin, humpless cattle, resembling small Alderneys, and sheep apparently of degenerate European type. Of the flora, dragon's blood (*Dracaena*), myrrh, frankincense, and cucumber trees are most famous. The affinities of flora and fauna are not only Asian and Ethiopian, but Mascarene, South African, and Antipodean-American.

History.—The name Socotra, and its Greek and Roman form, Dioscorides, is usually traced back to the Sanskrit Dvipa-Sakhadara, "island abode of bliss." It has been identified with the Egyptian Pa-anch, the Panchaia of Virgil, as well as the Isle of the Blest where Gilgamesh obtained a tree of life. The quest for immortality may have been the motive in the South Arabians' first use of incense and is reflected again in the legend of the phoenix, which has been associated with Socotra, as has that of Castor and Pollux. Dragon's blood is still called *dam al-'akhwein*, "the blood of the two brothers," the dragon and the elephant. Iskuduru, a country conquered by Darius the Great of Persia, may have been Socotra, and it was probably visited, with Punt, by the Egyptians in search of myrrh and frankincense. The *Periplus of the Erythraean Sea* (1st century A.D.) mentions a foreign population of Arabs, Indians, and Greeks and says that Socotra traded with Al Mukha in Yemen and with the Tamil country and Broach in India. The inhabitants of Socotra were for long Christians, probably Monophysites, and in the reign of the Roman emperor Constantius II the ambassador and bishop Theophilus Indus included Socotra in his diocese. Cosmas Indicopleustes (6th century) thought the Ptolemies had placed Greeks there. St. Francis Xavier, visiting Socotra after 1541, believed the inhabitants to be Nestorians. The last recorded traces of Christianity were in the 17th century. The Portuguese occupied Socotra from 1507 to 1511 when it reverted to the Mahri sultan.

In 1834 the British made an agreement for storing coal in Socotra and negotiated for the purchase of the island. A detachment of troops was sent but withdrawn when the sultan firmly refused to sell. In 1838 he proposed to hire the island to the British but the capture of Aden made this unnecessary. In 1876 the sultan agreed not to cede to any other power and by treaties of 1886 and 1888 accepted British protection for the whole sultanate. In 1954 an advisory treaty was signed between the sultan and the British government.

Population, Administration, and Social Conditions.—The population of Socotra is estimated at 12,000, the capital Hadibu (formerly called Tamrida by Europeans) may have a population of about 1,000 out of the 4,000 or 5,000 settled inhabitants of the coastal villages. The pastoral nomads of the interior possibly number rather more. The Socoteris are largely of Mahri origin, but are different in manners from the Mahris of the mainland. The Socoteri language also differs from Mahri, but both belong to the Hadara group, spoken by Veddoid tribes called Ahl al-Hadara, identified with the Hadoram of Genesis.

Socotra is directly ruled by the Mahri sultan. Its status as a possession of that chief is very old The *Periplus* describes it as being subject, by some ancient right, to "the king of the frankincense country," which included the Hadhramaut, the Mahra country, and probably the Qara country (now in Oman), which produced most of the incense.

The sultan's constitutional position in his own country differs greatly from that in his dependency Historically he is not recognized, nor does he claim any right to rule in the former, while in the latter his power is absolute. The Mahri nation which includes the Mahri tribe, is matrilineal and divided into four *'useba* (each consisting of a number of clans), and some independent clans, not all true Mahris. Of these latter the "royal" clan, Bin'Afrar, claims common origin (with the Beni'Afif sultans of Lower Yafa' in the western area of the Protectorate of South Arabia) from the Himyarite god Yafa'. There are two branches of the "royal" clan, the Sa'd bin Towar and the 'Amr bin Towar. The sultan belongs to the former, and recent sultans have preferred to live at Hadibu while the head of the latter represents him on the mainland at Qishn with a cadet of the branch at Sayhut. The Mahri "system of anarchy" depends closely on recognized relationships and functions of the *'useba* with each other and the Bin'Afrar clan. In interclan functions the principal *'useba* is the Bin Sar which includes the Bin 'Ali Muqaddam clan. The *muqaddam* (chief) determines disputes between any two clans and arranges truces. A man declared guilty of murder by the *muqaddam* during a truce should be slain by his clansmen. An appeal lies from the *muqaddam* to the sultan but neither has jurisdiction in internal clan affairs. In external affairs the Bin Boqi bin Ahmed *'useba* is responsible for defense against outside aggression. This

originated with an invasion of Kathiris in the 15th century when the sultan brought in Yafa'i mercenaries to expel them. The Bin Boqi thereafter guaranteed Mahraland against foreign invasion and still receive an annual retainer of 480 thalers. They also receive 120 thalers from the sultan of Muscat for a guarantee of nonaggression.

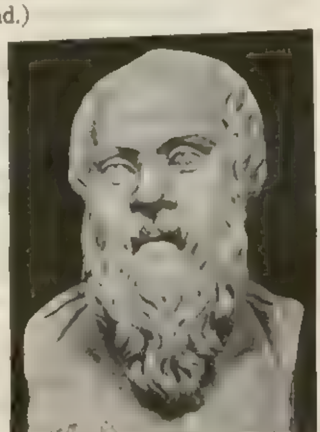
The sultan exercises his direct, but somewhat primitive, rule in Socotra with the aid of a vizier and the headmen of the settlements and pastoral clans. Though the tribesmen in the Mahra country are, like all tribesmen in the Protectorate of South Arabia, entitled to bear arms, those in Socotra apparently never have been, but the sultan has a few armed retainers, mostly of African slave descent, to execute orders. Tribal law applies, but there is a *qadi* (judge) and the *shari'ah* (Islamic law) is sometimes enforced.

The Economy.—About half the settled inhabitants of the coastal villages are fishermen and part-time pearl divers. Others have gardens but few domestic animals. The small foreign community are almost all traders. The pastoral nomads own about 20,000 cattle and more than 1,000 camels. They make the highly valued Socotran ghee, collect dragon's blood, aloe juice, and frankincense, and also civet from trapped animals. They grow a little tobacco, Eleusine millet, and inferior dates. Hadibu is the largest settlement and principal anchorage. No anchorages are safe, however, during the southwest monsoon. Qadhub has a large Royal Air Force airfield established in 1942 for antisubmarine operations. (W. H. Is.)

SOCRATES (*c.* 470 B.C.–399 B.C.), the great Athenian philosopher. Since he was 70 years old when he was put to death in 399 B.C., he must have been born in or about 470, ten years after the Battle of Salamis. His father, Sophroniscus, was a friend of the family of Aristides the Just (the tale that he was a sculptor first appears in the 3rd century B.C. in Timon of Phlius and seems to be only a misinterpretation of a playful remark in Plato). The philosopher's mother, Phaenarete, acted as a "midwife," but no inference as to social status can be founded on this.

The memoir writer Ion of Chios mentioned meeting Socrates at Samos in the company of the philosopher Archelaus, the Athenian pupil of Anaxagoras, presumably during the military operations of 441–440; and the connection between the two men is also asserted by Aristoxenus (*q.v.*), while the doxographical tradition based on Theophrastus calls Socrates the "disciple" of Archelaus. Plato, Xenophon, and Aeschines of Sphettus agree in depicting him as intimate with the leading figures of the Periclean circle (Aspasia, Alcibiades, Axiochus, Callias). Xenophon concurs with Plato in saying that he was well versed in geometry and astronomy, and this representation agrees with the narrative of Plato's *Phaedo* and the Aristophanic burlesque of *The Clouds*. Socrates must already have been a conspicuous figure at Athens when Aristophanes and Ameipsias both made him the subject of their comedies in 423, and since they made a special point of his neediness he had probably suffered recent losses. (The marked poverty of his old age is said in Plato's *Apology* to have been caused by his preoccupation with his mission to mankind.)

Socrates was married, apparently late in life, to Xanthippe, by whom he left three sons, one an infant. Xenophon speaks of her high temper; there is no evidence that she was a "shrew"; the sons, according to Aristotle, proved insignificant.



BY COURTESY OF THE SOPRINTENDENZA ALLE ANTICHITÀ DELLA CAMPANIA, NAPLES

BUST OF SOCRATES

Socrates' record for endurance was distinguished. He served as a hoplite, perhaps at Samos (441–440), and, during the Peloponnesian War, at Potidaea (432–430; where he saved the life of Alcibiades), Delium (424), and Amphipolis (422, unless he went on the expedition of 437–436). In politics he took no part, knowing, as he told his judges, that office would mean compromise with his principles. Once at least, in 406–405, he was a member of the boule (*q.v.*) of 500; and at the trial of the victors of Arginusae, he resisted, at first with the support of his colleagues, afterward alone, the unconstitutional condemnation of the generals by a collective verdict. He showed the same courage in 404, when the oligarchy of the Thirty Tyrants in Athens, wishing to implicate honourable men in their proceedings, instructed him and four others to arrest Leon, one of their victims: Socrates disobeyed, and says, in Plato's *Apology*, that this might have cost him his life but for the counterrevolution of the next year. (For the background to these events *see* GREECE: *History*.)

In 399 Socrates was indicted for "impiety." The author of the proceedings was the influential Anytus, one of the two chiefs of the democrats restored by the counterrevolution of 403; but the nominal prosecutor was the obscure and insignificant Meletus. There were two counts in the accusation, "corruption of the young" and "neglect of the gods whom the city worships and the practice of religious novelties." Socrates, who treated the charge with contempt and made a "defense" which amounts to avowal and justification, was convicted, probably by 280 votes against 220. The prosecutors had asked for the penalty of death; it now rested with the accused to make a counterproposition. A smaller, but substantial, penalty would have been accepted, but Socrates took the high line that he really merited the treatment of an eminent benefactor, maintenance at the public table. He consented only for form's sake to suggest the small fine of one *mina*, raised at the entreaty of his friends to 30.

The claim to be a public benefactor incensed the court, and "death" was voted by an increased majority, a result with which Socrates declared himself well content. As a rule at Athens the condemned man "drank the hemlock" within 24 hours, but in the case of Socrates the fact that no execution could take place during the absence of the sacred ship sent yearly to Delos caused an unexpected delay of a month, during which Socrates remained in prison, receiving his friends daily and conversing with them in his usual manner. An escape was planned by his friend Crito, but Socrates refused to hear of it, on the ground that the verdict, though contrary to fact, was that of a legitimate court, and must therefore be obeyed. The story of his last day, with his drinking of the hemlock, has been perfectly told in the *Phaedo* of Plato, who, though not himself an eyewitness, was in close touch with many of those who were present.

Main Sources of Information.—Socrates wrote nothing; therefore our information about his personality and doctrine has to be sought chiefly in the dialogues of Plato (*q.v.*) and the *Memorabilia* of Xenophon (*q.v.*). Both men were nearly 45 years younger than Socrates, and could therefore speak from firsthand knowledge about only the last 10 or 12 years of his life.

Xenophon, whose relations with Socrates seem not to have been close, has even been suspected of deriving much of his material from Plato's dialogues. His admitted deficiencies in imagination and capacity for thinking do not make him the more faithful exponent of a philosophical genius. We need also to discount Xenophon's apologetic purpose. His most valuable statements are those which appear most at variance with his main thesis that the prosecutors of Socrates were mistaken from their own point of view.

Plato's more vivid picture has been suspected on the ground that Plato used Socrates as a "mouthpiece" for speculations of his own. What this really means is that the "ideal theory" or doctrine of "forms" expounded in the *Phaedo* is held to have been originated by Plato after the death of Socrates. There are serious reasons for denying this assumption though they have not yet convinced all scholars; in any case it is a *petitio principii* to employ it, without investigation, as an argument to discredit Plato's testimony. Xenophon's silence at most only proves that Socrates did not converse on such matters with him.

In some important respects Plato's testimony is confirmed by the remains of Aeschines of Sphettus. *The Clouds* of Aristophanes yields valuable information about Socrates in his middle 40s,

though allowance is made for the work's character as a burlesque. It should be compared carefully with the autobiographical statements put into the mouth of Socrates in the *Phaedo* (*Great Books of the Western World*, vol. 7, pp. 240 ff. [96*a*–100*a*]). These are not "contemporary evidence," but they are clearly meant to express Plato's bona fide belief about his master's intellectual history.

Personal Characteristics.—Though Socrates was a good fighting man, his outward appearance was grotesque. Stout and not tall, with prominent eyes, snub nose, broad nostrils, and wide mouth, he seemed a very Silenus. But, as his friends knew, he was "all glorious within," "the most upright man of that day" (Plato, *The Seventh Letter* [324*e*]). His self-control and powers of endurance were exemplary; "he had so schooled himself to moderation that his scanty means satisfied all his wants."

But Socrates was no self-tormenting ascetic; he "knew both how to want and how to abound," and could be the soul of the merriment at a gay party. He had no sympathy—this was a main point in the *Telauges* of Aeschines—with the slatternliness of his friend Antisthenes or the godly dirtiness affected by "Pythagorists." There was nothing of the complacent self-righteousness of the Pharisee, nor of the angry bitterness of the satirist, in his attitude toward the follies or even the crimes of his fellowmen. It was his deep and lifelong conviction that the improvement not only of himself but of those with whom he might have to do was a task laid upon him "by God"; but the task was not to be executed with a scowling face and an upbraiding voice. Like St. Francis Xavier, he thoroughly understood how important it is to one who would win men's souls to be "good company." Conscious of his own infirmities, he felt a real and profound sympathy for those who had not learned to master their frailties and passions.

Socrates was a true patriot, and his devotion to Athens was made only the more evident by his conviction that he could best prove it by setting his face resolutely against the attractions of specious and popular, but deadly, false theories of public and private morality. When the city brought him to trial and threatened him with death his sense of civic duty forbade him to withdraw into exile before the trial, or to accept the opportunity of escape during his unforeseen imprisonment. It was his very patriotism that made him an unsparing critic of the "democracy," which means, in Friedrich Nietzsche's phrase, "one flock and no shepherd," and so led directly to the accusation that proved fatal to him.

Nothing was more marked in his character than an unusually keen sense of humour, an appreciation of the comic in human nature and conduct that protected him at once against sentimentality and against cynicism. This is what his opponents in Plato call his "irony," and treat as an irritating affectation. "Intellectually the acutest man of his age, he represents himself in all companies as the dullest person present. Morally the purest, he affects to be the slave of passion" (W. H. Thompson, *The Phaedrus of Plato*, 1868). No doubt, in part this irony was "calculated"; it "disarmed ridicule by anticipating it." But its true source is the spontaneous sense of "fun" which makes its possessor the enemy of all pretentiousness, moral or intellectual, in himself and in others. And it is certain that, though the purity of Socrates is beyond question, he really had an ardent and amorous temperament; a fragment of the *Alcibiades* of Aeschines confirms Plato's representation on this point.

Religion.—Socrates was clearly a man of deep piety with the temperament of a mystic. Like other educated men of his age, he regarded mythology, with its foolish or immoral tales about gods, as a mere invention of the poets. But he found it easy to combine his own strong belief in God, the all-wise and all-good ruler of the world, with the view that in practice we could worship God in the way prescribed by "the usage of the city." God's existence is shown, he held, not only by the providential order of nature and the universality of the belief in him but by warnings and revelations given in dreams, signs, oracles. The soul of man partakes of the Divine; the concluding pages of Plato's *Apology* prove that Socrates had a strong belief in the soul's immortality. (Xenophon for apologetic reasons is silent on the point, but has reproduced the argument in the dying speech of his Cyrus in the *Cyropaedia*. Aristophanes, too, makes Socrates combine the parts of "infidel" physicist and hierophant of a mysterious private faith, and in *The Birds* [1553, *seq.*] represents him as presiding at a fraudulent séance.) He was regular, says Xenophon, in prayer and sacrifice, though he held that since only the gods know what is good for us, our prayer should simply be "give me what is good"; we must not dictate the form the blessing should take. It is clear from Plato that Socrates was deeply influenced by Pythagorean and Orphic religious ideas, though he regarded the ordinary Orphic mystery-monger with healthy contempt.

The evidence that Socrates had a markedly "mystical" temperament is abundant. Plato tells of his curious "rapts," in one of which he stood spellbound for 24 hours in the trenches before Potidaea, and there seems to be an allusion to this singularity in *The Clouds* (171, *seq.*).

The accounts of the philosopher's "Divine sign" tell the same story. This, according to Plato, was a "voice" often heard by Socrates from childhood. It forbade him to do things, but never gave positive encouragement. (Xenophon, who makes more of the matter, says, less probably, that it did give positive directions.) Plato treats the "voice" very lightly; by his account, it merely gave prognostications of good or ill luck, and the occasions of its occurrence were often "very trivial." Thus it was neither an "intuitive conscience" nor a symptom of mental disorder, but an "interior audition," a "psychic phenomenon" of a kind not specially uncommon.

Mode of Life.—Socrates' whole time seemed to be spent "out of doors," in the streets, the marketplace, and, more particularly, the *gymnasia*. He cared little for the country and rarely passed the gates. Though he frequented by choice the society of lads of promise, he also talked freely to politicians, poets, artisans about their various callings, their notions of right and wrong, the familiar matters in which they might be expected to take an interest. The object of all this, he says in the *Apology*, was to test the famous Delphic oracle that had pronounced him the wisest of men. It is clear from the *Apology* that the oracle had made this declaration, no doubt because the Delphic authorities knew from the form of the question what answer was desired. The presupposition of the *Apology* is that this happened before Socrates had become conscious of his mission to his fellowmen; even at that early date, it is implied, he had the highest of reputations in circles interested in wisdom. Moreover, the Eleatics from Megara and the young pupils of the Pythagoreans from Thebes and Phlius who were attached to Socrates must have formed their connection with him before the Peloponnesian War.

Acutely sensible of his own ignorance, Socrates set himself to convict "the god" of falsehood. But when experience showed that those who thought themselves wise were unable to give any coherent account of their wisdom, Socrates had to admit that he was wiser than others, just because he alone was aware of his own ignorance. This account is plainly tinged with the usual "irony." Socrates did not take Apollo and his oracle very seriously. But that he was quite serious in believing himself charged with a mission, not from "Apollo" but from God, to preach to his fellowmen the supreme importance of knowledge of what is for the soul's good is proved by his declaration that he was more than ready to face instant death rather than to neglect his commission. The poverty in which this mission had involved him and the austerity of the rule of life which it entailed were notorious.

Summer and winter, Socrates' coat was the same; he had neither shoes nor shirt. "A slave who was made to live so," the sophist Antiphon said, "would run away." This self-imposed life of hardships was the price of his spiritual independence. His message was variously received. Some of those whose false pretensions were exposed by his trenchant criticizing regarded him with ill will; many thought him an officious busybody. Among the younger men, many merely thought it good sport to see their elders silenced. Others (Xenophon says that this was the case with Alcibiades and Critias [*qq.v.*]) deliberately attached themselves to him for a time "for private ends," believing that to learn the secret of so acute a reasoner would be the best preparation for success

in the law courts, the council, and the assembly. Others sincerely hoped by associating with him to become good men and true, capable of doing their duty by house and household, by relations and friends, by city and fellow citizens.

Finally, there was an inner circle who entered more deeply into Socrates' principles and transmitted them to the next generation. But these were not "disciples" united by a common doctrine. Socrates finally repudiated all claim to have "disciples." The bond of union was a common reverence for a great man's intellect and character. It was, in the main, this group who were collected around Socrates on the day of his death; many of them, for instance Euclid of Megara and the young Theban Pythagoreans, Cebes and Simmias, were foreigners from states which had been enemies of Athens in the Peloponnesian War.

The Accusation and Its Causes.—The explanation of the attack made on Socrates is simple. He had been on terms of close friendship with the two men whose memories were most obnoxious to the democrats: Critias, the fiercest spirit among the extremists of the "Terror" of 404; and Alcibiades, whose self-will had done so much to bring about the downfall of the Athenian empire. The charge of "educating Alcibiades" was made prominent in the pamphlet written a few years after the trial by the sophist Polycrates, in justification of the verdict. More than half a century later, the orator Aeschines reminds his audience that Socrates had been put to death because he was believed to have educated Critias. In point of fact, it was absurd to make Socrates responsible for the ambitions of Alcibiades, and, as he reminded his judges, he had disobeyed an illegal order from Critias and his colleagues at the risk of his life. But it is natural that he should have had to suffer for the crimes of both men, the more so because he had been an unsparing critic of democracy and of the famous democratic leaders and furthermore had not, like the advanced democrats, withdrawn from Athens during the "Terror."

Socrates was, in fact, suspected of using great abilities and gifts to pervert his younger associates from loyalty to the principles of democracy, and the convinced democrats who had recovered the city in 403 were unwilling, as J. Burnet has said, "to leave their work at the mercy of reaction." That they took no steps for four years is probably explained by the state of complete confusion and congestion into which the disorders of 404 had thrown the law courts. The motives of Anytus, an upright, unintelligent democrat, are thus quite explicable. From his point of view, Socrates would be at the best a moderate oligarch, and democrats who remembered the career of Theramenes (*q.v.*) could not be expected to make a fine distinction between the moderate oligarch and the traitor.

The real grounds for the attack could not be disclosed in the indictment, since the amnesty which had terminated the struggle of 404–403, and of which Anytus himself had been a main promoter, covered all offenses committed before the archonship of Eucleides (403). Hence the charge had to be couched in the form of a vague accusation of "corruption of the young." Probably for the same reasons Anytus was ashamed to appear as the principal in the matter and put forward the obscure Meletus, who might venture on "indiscretions" more openly. If Meletus was the same person who prosecuted Andocides (*q.v.*) in 400 on the same charge of "impiety," and if, as is not unlikely, he is the real author of the speech *Against Andocides* ascribed to Lysias, he must have been a half-witted fanatic, and this may explain why the charge of irreligion was added. The real nature of this "irreligion" appears never to have been explained. Xenophon suggests that the allusion was to the "Divine sign" but this cannot be correct. It is clear from Plato's *Apology* that Meletus said nothing about the "sign" at the prosecution, and that Socrates is speaking with his "usual irony" when he pretends to guess that the mention of "religious novelties" in the indictment referred to it. In the *Apology*, Socrates says that the prosecution is, no doubt, relying on memories of Aristophanes' *The Clouds*, where he had been made to talk "atheism" as part of the burlesque on men of science.

But there must have been more behind the charge, and it seems likely that Burnet is right in reminding us that the prosecution of Andocides revived the old scandal of the "profanation of the mysteries," which had thrown Athens into a ferment on the eve, 416–415, of the Sicilian expedition. The two chief victims, Alcibiades and his uncle Axiochus, were both among the intimates of Socrates, and there is reason to think that others of his friends were affected. If this is what lay behind the charge, it can be understood why its real meaning seems never to have been explained. In view of the terms of the amnesty, the matters in question were not really within the competency of the court.

Socrates himself, in the account of Plato, who was present at the trial, treats the whole matter with contempt. His defense consists in narrating the facts of his past life, which proved that he was equally ready to defy the populace and the Thirty in the cause of right and law, and in insisting on the reality of his mission from God and his determination to discharge it, even at the cost of life. The prosecutors had no desire for blood. They counted on a voluntary withdrawal of the accused from the jurisdiction before trial; the death penalty was proposed to make such a withdrawal certain. Socrates himself forced the issue by refusing at any stage to do anything involving the least shade of compromise. The prosecution had raised the question whether he was a traitor or, as he held himself to be, an envoy from God; Socrates was determined that the judges should give a direct verdict on the issue without evasion. This is what makes him a martyr, but also what forbids us to call Anytus a murderer.

Doctrine and Method.—Socrates was a man of the Periclean age, and the Periclean age witnessed one of the periodical "bankruptcies of science." Cosmological speculation, which had been boldly pursued from the beginning of the 6th century, seemed by the middle of the 5th to have led to a chaos of conflicting systems, each of which could establish only one point, that all its rivals were wrong. Parmenides (*q.v.*) of Elea had apparently cut away the ground from science by showing that the real world must be quite unlike anything our senses reveal to us, and that, consequently, the method of cosmology, interpretation of the world by analogies from familiar sensible experiences, is inherently fallacious. His pupil Zeno (*q.v.*) of Elea seemed to have shown that even the postulates of mathematics are mutually contradictory.

Science, then, seemed impossible, and that was why the ablest men of the generation before Socrates, such as Protagoras and Gorgias, turned away from the pursuit of it and tried to find a use for the intellect in professions that concerned themselves, not with the discovery of truth, but with making a success of human life (*see* SOPHISTS). "Probability is the very guide of life," and in practical matters "useful" points of view may be attainable, even if scientific certainty is beyond our reach. According to the narrative of the *Phaedo*, Socrates, as a young man, began with an experience typical of his age. He was enthusiastically interested in "natural science," and familiarized himself with the various current systems, being especially interested in the contrast between the old Milesian type of cosmology with its flat earth and the Italian type with its spherical earth. He was also interested in the mathematical puzzles raised by Zeno about "the unit" (*i.e.*, the problem of continuity). He discovered, to his distress, that though each authority was quite sure that the views of the others were wrong, none of them could give any proof that his own were right.

There was a complete lack of critical method. For a moment Socrates hoped to find salvation in the doctrine of Anaxagoras (*q.v.*) that "mind" is the source of all cosmic order, since this seemed to mean that "everything is ordered as it is best that it should be," that the universe is a rational teleological system. But on reading the book of Anaxagoras, he found that the philosopher made no effective use of his principle; the details of his scheme were as arbitrary as those of any other.

The Socratic "Hypothesis."—After this disappointment, Socrates decided that he had "no head for physics" and must fall back on his own mother wit. Accordingly he resolved from then on to consider primarily not "facts" but *logoi*, the "statements" or "propositions" that we make about "facts." His method would be to start with whatever seemed the most satisfactory "hypothesis," or postulate, about a given subject, and to consider the consequences that follow from it. So far as these consequences proved

to be true and consistent, the "hypothesis" might be regarded as provisionally confirmed; if they were false or mutually inconsistent it would be discredited. But it would have to be a strict rule of method not to confuse inquiry into the consequences of the "hypothesis" with proof of its truth. If the question of its truth were raised, the issue could be settled only by deducing the initial "hypothesis" as a consequence from some more ultimate "hypothesis" which both parties to the dispute were content to accept. The method, still familiar to us as that of true science, is manifestly suggested by reflection on the antinomies of Zeno, whom Aristotle called the creator of "dialectic" and whom Plato, in the *Parmenides*, afterward described as meeting Socrates in the youth of the latter.

The Doctrine of Forms.—According to Plato, Socrates next proceeded to take it as his own fundamental "hypothesis" that every term (such as "good," "beautiful," "man") which has a single unequivocal denotation directly names a single selfsame object of a kind inaccessible to sense perception and apprehensible only by thought. Such an object Socrates calls an *idea* or *eidos*, that is, a form (*q.v.*). The sensible things on which we predicate beauty, goodness, humanity, have only a secondary and derivative reality. Strictly speaking, we must not say that they *are* this or that, but only that they *become* this or that for a time, in virtue of the temporary "presence" to them of the corresponding form, or, as it is also expressed, in virtue of their "participation" in the form. A sensible thing, in fact, is simply a temporary complex of forms.

In the *Parmenides* of Plato Socrates is made to expound this doctrine to the great philosophers Parmenides and Zeno as his solution of the standing puzzle of the one and the many. This is the doctrine of forms as it is stated in the *Phaedo* and *Republic*. Though it is quite different from the version of the doctrine ascribed by Aristotle to Plato, scholars in the 19th century usually assumed that it is an earlier version of that doctrine consciously devised by Plato after the death of Socrates. The chief argument for this view is based upon the observation of Aristotle that Socrates rightly "did not separate" the universal from the particular (*Met. M*, 1078 *b*30) as, it is apparently implied, Plato did. It is, however, not clear that Aristotle means by this, what he never says expressly, that Socrates did not teach the doctrine ascribed to him in the *Phaedo*. He might equally mean that the doctrine of the *Phaedo* does not itself involve the kind of "separation" of the universal from the particular to which he objects in what he describes as the Platonic theory, and, since the *Phaedo* is one of the Platonic dialogues to which he most frequently alludes, it is strange that he should never have said that it misrepresents the historical Socrates on a capital point, if he really thought so.

On the other side, the doctrine is expressly said in the *Phaedo* to be a familiar one which Socrates "was always" repeating, and it is hard to believe that Plato could have made such a statement about a speculation of his own, especially as most of the personages of the *Phaedo* were certainly still alive long after the dialogue was written. It is hard to see what could be the point of such a mystification, and harder to understand how its author could have expected it to be successful. Of course, we can be guided only by considerations of probability. If we think the probabilities are against the view that credits Plato with deliberate mystification, we must be prepared to admit the possibility that he is also reproducing the thought of Socrates in the further development of the *Symposium* and *Republic*, where we hear of a supreme Form, that of Beauty, or Good, the vision of which is the far-off goal of all intellectual contemplation. We may fairly suspect that the thought of Socrates is undergoing development in the mind of Plato, but it will be natural to regard the development as, in the main, unconscious, and to recognize that no complete separation of the Socratic and the Platonic in the result is possible. (*See* also PLATO.)

Logical Method.—It is certain that on the logical side the thought of Socrates proceeded "as if" the doctrine of forms expounded in the *Phaedo* were its point of departure. Both Plato and Xenophon bear out the remark of Aristotle that Socrates may fairly be credited with two things, "inductive arguments" and "universal definitions" (*Met. M*, 1078 *b*27). The "universal definition" is an attempt to formulate precisely the meaning of a universal significant predicate, that is, to apprehend what the *Phaedo* calls a form; and it is from the practice of Socrates, who aimed at the clarification of thought about the meaning of moral predicates as the first indispensable step to the improvement of practice, that the theory of logical division and definition, as worked out in Plato's later dialogues and the logical treatises of Aristotle, has arisen.

The "inductive arguments" mean the characteristic attempts to arrive at such formulations by the consideration of simple and striking concrete illustrations familiar to us from both Xenophon and Plato, the perpetual arguments about "shoemakers and carpenters and fullers," which the fashionable speakers in Plato profess to think vulgar. Induction, on this view of it, is not regarded as a method of proof. Its function is that of suggestion; it puts the meaning of a proposed "definition" forcibly and clearly before the mind. The justification of the definition, then, has to be sought in a consideration of the satisfactoriness of the "consequences" which would follow from its adoption. Socrates himself sought for his "definitions" principally in the sphere in which he was most interested, that of conduct, private and public. As Aristotle says, he concerned himself with the "ethical," character and conduct, not with "nature" at large. This is what Cicero means by saying that he "brought down philosophy from heaven to earth."

Ethics and Politics.—Before Socrates, cosmology had been the chief topic of interest; after him, the central problem of philosophy was to formulate a rule of life. With him the "practical use of reason" comes by its rights. In this respect Socrates stamped on philosophy a character which it has never lost. The main outlines of his philosophy of conduct are fortunately quite certain, and could be discovered if we had no more material than the Platonic *Apology* and the *Memorabilia* of Xenophon. As the *Apology* tells us, the specific message from God which Socrates brought to his fellowmen was that it is the great business of life to practice the "care" or "tending" of one's "soul," to "make one's soul as good as possible," and not to ruin one's life, as most men do, by putting care for the body or for "possessions" before care for the "soul."

The thought which is here fundamental is that of the "soul" or *psyche* as that which is most truly a man's self. In Greek literature, down to the end of the 5th century, we can trace two main senses of the word *psyche*. In the first place, *psyche* means "the breath of life" which a man parts with in dying. It is this which, in popular superstition, is left as a mere "ghost," or "shade," when the man "himself," his body, has perished. In earlier Ionian science this is identified with the "air" which a man inhales so long as he is alive. Secondly, in circles influenced by the Orphic religion, the soul is thought of as something which has a destiny beyond the grave, but this, too, is something different from the self. It is a sort of stranger inhabiting the body, but having little to do with the conduct of normal life. It "sleeps while the body is active, but wakes when the body sleeps," and reveals itself chiefly in dream and trance. From the beginning of the 4th century we find *psyche* coming at last to mean what "soul" means to us, the normal waking personality, the seat of character and intelligence, "that," as Socrates says in Plato, "in virtue of which we are called wise or foolish, good or bad," and as this usage of the word first appears in writers whom we know to have been influenced by Socrates (Isocrates, Plato, and Xenophon), we may fairly ascribe it to his influence. The thought now works out thus. The soul *is* the man (in the later Academic formulation a man *is* "a soul using a body").

Our happiness or well-being, then, depends directly on the goodness or badness of the soul. It is no happiness to possess health, or strength, or wealth, unless we know how to *use* these advantages rightly. If we use them wrongly, they will only be so many means to misery. The reason so few achieve happiness is not that men do not wish to be happy. No one ever wishes for anything but true good, that is, true happiness, but men miss their happiness, in spite of the universal wish for it, because they do not *know* what it is. They mistake for real good things which are not really good (*e.g.*, unlimited wealth or power).

In this sense, "all wrong-doing is involuntary." The first and fundamental requisite for happiness, then, is that men should *know* true good and not confuse it with anything else. The good state of the soul is precisely that state in which it never makes the mistake of taking anything to be good when it is not really good. To "make one's soul as good as possible" thus means to attain the knowledge of good which will prevent us from *using* strength, health, wealth, opportunity, wrongly. If a man has this knowledge, he will always act on it, since to do otherwise would be to prefer known misery to known happiness, and this is impossible. "*All* the virtues are one thing," *knowledge* of good, and all "vice" is one thing, ignorance of true good.

"Popular" goodness—what passes currently as virtue—is mostly illusory, because it is mainly a matter of habit, not of assured conviction about good. It breaks down under temptation; but if a man really knew that, for instance, to commit a crime is worse than to suffer loss or pain, or death, no fear of these things would lead him to commit the crime. The professional sophist, again, claims to be able to teach "goodness," but the claim is shown to be unfounded by the very fact that the sophist treats "goodness" as though it were a neutral "accomplishment" that might be conveyed by mere instructions. Now an accomplishment, or "art," can always be put to either of two uses, a good or a bad, as the physician, for instance, can use his professional knowledge to cure or to kill.

Knowledge of good is the one knowledge of which it is impossible to make an ill use; the possession of it is a guarantee that it will always be used aright. Thus Socrates becomes, as against the relativism of Protagoras, the founder of the doctrine of an absolute morality based on the conception of a felicity which is the good, not of Athenians or Spartans, or even of Greeks, but of man as man. It is not in virtue of our allegiance to a particular city, nor even of our place in a particular historical civilization, but in virtue of our universal humanity, that we have the task of "making the soul as good as possible," or, as Socrates also said, in language influenced by Pythagoreanism, "making it like God."

Politics, from this point of view, does not differ in principle from ethics. The business of the statesman also is the "tending" of souls, though his task is to aim at making, not only his own soul, but the souls of all his fellow citizens "as good as possible." The knowledge of good is also the "royal" science or science of governing, the foundation of all statesmanship. The radical vice of ancient democracy, according to Socrates, is that of not demanding evidence of any special knowledge in its leaders; it suffers the destinies of society to be in the hands of men without true insight. Partly this means that by not demanding intellectual qualifications for office, democracy surrenders the control of affairs into the hands of men with no adequate expert knowledge.

But this is only a minor part of Socrates' indictment. His main criticism is that though in some departments, at least, the democracy refuses to take the advice of anyone but a qualified expert, on the question of the morality and justice of a proposed policy it treats any one citizen's opinion as of equal value with another's. Even a Themistocles or a Pericles plainly had no knowledge of true statesmanship, as we see from the fact that they neither taught the principles of it to their sons nor had them taught these principles by others, and if we look at the actual achievements of these men we can see that they were, so to say, good "body-servants" of the "people"; they gave it the things that tickled its taste, such as a navy and a commerce; they were no "physicians of the body politic," for they did not promote "righteousness and temperance," the spiritual health of the community. That is, they measured national greatness by wealth and empire, not by character. According to Plato, Socrates maintained that he himself, who abstained all through from active politics, was the one Athenian of the time who deserved the name of statesman. He deserved it because he understood, as the men of action did not, that national, like individual felicity, depends on the knowledge of good which inevitably leads, where it is possessed, to the action which makes the soul "as good as possible."

Plato's *Republic* may fairly be said to be, on its political side, a picture of the life of a society in which the whole system of social and economic life is based on this Socratic conviction that "politics" is the application to the community at large of the principle that knowledge of the absolutely good is the necessary and sufficient condition of well-being.

How far any of the special regulations of the *Republic* embody actual convictions of Socrates is more than we can say, though it is significant that the *Aspasia* of Aeschines represents Socrates as maintaining one of Plato's "paradoxes," the capacity of women for war and politics.

The Socratics.—The thought of Socrates has, in the main, been made fruitful for subsequent ages by being taken up and continued in the lifework of Plato. A more temporary influence was exercised by certain other members of the group of Socratic men whom it has become customary to speak of as the "minor" Socratics. The most important of them are Antisthenes of Athens and Euclid of Megara, with whom the Cynics and the Megarian School (*qq.v.*) of the 4th and early 3rd centuries are historically connected. With them it is usual also to mention Aristippus of Cyrene, often still spoken of as somehow connected with the Cyrenaics (*q.v.*) of the early 3rd century.

It is probable, however, that the current accounts exaggerate the closeness of the connection between these men and the later schools. Aristippus figures in Xenophon's *Memorabilia* simply as a luxurious and refined man of the world who makes it his rule of life to extract personal enjoyment from existence, sitting loose to all attachments which might interfere with his ease and not allowing himself to take root anywhere. The later anecdotes about him bear out this representation. There is no good evidence that he had a philosophy or originated a school. Aristotle ascribes no doctrine to him and never mentions a "Cyrenaic" school of hedonists, though he could hardly have avoided doing so in his discussions of hedonism in the *Nicomachean Ethics* if he had known of one. Plutarch expressly describes the Cyrenaics as contemporaries of Epicurus, and all the names of members of the school known to us belong to the time of the successors of Alexander the Great. The one point of doctrine common to them appears to have been that they rejected the notion of a good more permanent than the pleasure of the moment. The supposed connection of Aristippus with them seems to be based on a confusion with his grandson of the same name, who, according to Eusebius, reduced his grandfather's practice to theory.

Euclid of Megara was a friend both of Socrates and of Plato, who temporarily took refuge with him after the death of Socrates, and, at a later date, dedicated the *Theaetetus* to him. All that we know of his teaching is that he held to the monism of Parmenides, maintaining that nothing is real except "the One," which is also called "wisdom," "intellect," and "God" (Diogenes Laërtius, ii, 106). The mention of "wisdom" as a synonym for "the One" seems to reveal the influence of Socrates. Most of our notices of Megarians deal with men of a later time, Eubulides, a contemporary of Aristotle, Diodorus Cronus, and Stilpo. These men were pugnacious formal logicians famous for their rejection of the notion of "possibility" which is so fundamental in the Aristotelian philosophy. According to them, nothing is possible except the actual. Aristotle resented the criticism so keenly that "sophist" in his terminology appears to be regularly equivalent to "Megarian logician." It is not clear how these "eristics" or disputators are connected with the monism of Euclid. There are reasons for supposing the puzzling antinomies of Plato's *Parmenides* to be a parody of Megarian logic.

Antisthenes was a friend of Socrates of long standing, with a marked individuality of his own, and a voluminous writer much admired for his style. He does not appear to have been a "disciple," though he was personally attached to Socrates, and particularly admired his strength of will and mastery of his passions. In philosophy he is chiefly known for two things, his denial of the possibility of making judgments in which the predicate and subject terms are nonidentical, and his insistence in ethics on the simplification of life by the reduction of our wants to an indispensable minimum.

In virtue of the latter he was commonly regarded as the founder of Cynicism and it is certain that he personally influenced the famous Diogenes and that the later Cynics were in the habit of regarding him as a model man. But it is not clear either that the Cynics of the 4th century were a "sect" or "school" in any real sense of the words, or that the nickname "dog" was ever given to anyone before Diogenes. It was believed in later antiquity that there was a personal feud between Antisthenes and Plato, and it seems certain that one of the works of Antisthenes, called *Sathon*, was a virulent personal attack on Plato. But the ingenuity spent in the 19th century on discovering polemical allusions to Antisthenes in Plato's dialogues seems to have been mostly wasted. According to Plato, the logical paradox that "contradiction is impossible" was maintained by numerous persons in the days of Socrates. Hence it seems unreasonable to detect special allusions to Antisthenes in the frequent references to this paradox in the dialogues.

See also ETHICS, HISTORY OF; PHILOSOPHY, HISTORY OF; and references under "Socrates" in the Index.

BIBLIOGRAPHY.—For general histories of Greek philosophy, which necessarily provide accounts of Socrates, *see* the *Bibliography* of the article PHILOSOPHY, HISTORY OF; for more specialized accounts of the background *see* J. Burnet, *Greek Philosophy: Thales to Plato* (1914) and *Early Greek Philosophy*, 4th ed. (1930). The article PLATO analyzes the Socratic dialogues in the course of the text and covers them also in its *Bibliography*, the editions and commentaries of J. Burnet being especially valuable. For Xenophon's evidence *see* XENOPHON.

Individual studies of Socrates include G. Zuccante, *Socrate* (1909); H. Maier, *Sokrates, sein Werk und seine geschichtliche Stellung* (1913); A. Busse, *Sokrates* (1914); A. E. Taylor, *Varia Socratica* (1911), *Socrates* (1933, and numerous reprints); A. K. Rogers, *The Socratic Problem* (1933); A. J. Festugière, *Socrate* (1934); G. Bastide, *Le Moment historique de Socrate* (1939); R. Kraus, *The Private and Public Life of Socrates* (1940); A. Banfi, *Socrate* (1943; new ed. 1963); O. Gigon, *Sokrates, sein Bild in Dichtung und Geschichte* (1947); V. de Magalhães Vilhena, *Le Problème historique de Socrate* (1952); F. M. Cornford, *Before and After Socrates* (1932); R. L. Levin and J. Bremer, *The Question of Socrates* (1961). For special aspects see J. Burnet, "The Socratic Doctrine of the Soul," *Proceedings of the British Academy* (1915–16); A. E. Taylor, "Plato's Biography of Socrates," *Proceedings of the British Academy* (1916–17); H. von Arnim, *Xenophons Memorabilien und Apologie des Sokrates* (1923); C. Pellegrino, *Se Socrate e Platone conobbero la Bibbia* (1936), with supplement *Echi biblici nel pensiero greco* (1938); M. Landmann, *Der Sokratismus als Wertethik* (1943); L. Nelson, *Socratic Method and Critical Philosophy* (1949); W. K. Richmond, *Socrates and the Western World: an Essay in the Philosophy of Education* (1954); C. J. Classen, *Sprachliche Deutung als Triebkraft platonischen und sokratischen Philosophierens* (1959); A. H. Chroust, *Socrates, Man and Myth: the Two Socratic Apologies of Xenophon* (1957); R. Guardini, *The Death of Socrates*, Eng. trans., new ed. (1962); L. Versényi, *Socratic Humanism* (1963).
(A. E. TA.; X.)

SOCRATES (*c.* 380–*c.* 445), ecclesiastical historian, was a Christian lawyer, born and bred in Constantinople, who supplemented the latter part of the ecclesiastical history of Eusebius of Caesarea (*i.e.*, from A.D. 306 to 324) and continued it to his own day. His work extends to A.D. 439; it is arranged in seven books corresponding to the emperors ruling in the Eastern Empire (Constantine I, Constantius, Julian and Jovian, Valens, Theodosius I, Arcadius, Theodosius II). The extant history is a second edition. Socrates says that in the first two books he had previously followed the ecclesiastical history of Rufinus (A.D. 403) too closely, but that he had rewritten them after having become acquainted with the writings and documentary collections of Athanasius.

He had little knowledge of the Western church, but had access to good written sources for the history of the church in the Eastern Empire; these he supplemented from oral tradition and, where possible, eyewitness accounts. Besides Eusebius, Rufinus and Athanasius he drew upon the important collection (no longer extant) of conciliar proceedings made by Sabinus, bishop of Heraclea. An unsettled problem is the extent of his debt to the lost history of Gelasius of Caesarea (d. 395). Socrates was sympathetic toward, though not himself a member of, the schismatic Novatianist church in Constantinople and derived information from its records and traditions. With regard to eyewitnesses, however, his precepts were better than his practice, for he could not resist an anecdote; as a result, the earlier part of his ... where he is dependent almost entirely on written evide... trustworthy than the later.

Straightforward and sensible, but without technical skill ... torical investigation, he made some bad mistakes, especially chronology (*e.g.*, over the Council of Sardica). He had no ... found insight, but his layman's outlook is interesting: a healthy anticlericalism (he dislikes hierarchical pride and ... tentiousness), tolerance of divergent ecclesiastical ... position to coercion in religion, an attitude to heresy ... liberal for his time (though often lacking in theol... tion) and a recognition that church history must be se... general context (though here also his perception of p... better than his practice). Loyalty to Constantinople may accou... for his attacks upon the Alexandrians Theophilus and Cyril ... his judgment of John Chrysostom, bishop of Constantinople, is reserved and partly unfavourable.

Whatever his faults, Socrates provided what remains an indis... pensable framework for the ecclesiastical history of the period he covers, and preserved many documents of the first ... His history served as the foundation of a parallel work by his contemporary Sozomen (*q.v.*) and was later excerpted, together with the histories of Sozomen and Theodoret, to make the Latin ... *toria Tripartita* of the monk Epiphanius, from which the ... church of the West took much of its knowledge of early Christianity.

BIBLIOGRAPHY.—Edition by R. Hussey, 2 vol. (1853), reprinted by W. Bright, 2nd ed. (1893); Eng. trans. by A. C. Zenos in H. W... P. Schaff (eds.), *Nicene and Post-Nicene Fathers*, 2nd ... (1891). *See* also F. Geppert, *Die Quellen des Kirchen...* *Socrates Scholasticus* (1898).
(S. L. ...)

SODA, a term applied somewhat loosely to a number of ... mon compounds of sodium—chiefly sodium carbonate, ... the anhydrous form (soda ash), the monohydrate (soda ...) and the decahydrate (washing soda or sal soda). Sodium ... bonate is known as baking soda, and sodium hydroxide as ... soda. Carbonated water (water charged with carbon dioxide) is inappropriately called soda water. For the important role of soda compounds in industrial processes, *see* ALKALI MANUFACTURE.
(A. E. H.; ...)

SODALITE, a member of the group of rock-forming ... comprising the isomorphous species sodalite, haüynite ... and lazurite. Sodalite (so named because it contains soda) occurs as well formed colourless crystals in the ejected limestone ... of Monte Somma, Vesuvius. In the nepheline-syenite of ... non, in Ontario, bright sky-blue material has been quarried ... use as an ornamental stone. Haüynite, or haüyne (named ... R. J. Haüy), occurs as bright blue crystals and gr... lavas of Vesuvius, Rome, the Eifel, etc. Noselite, or ... found as grayish crystals in the sanidine bombs of ... Lazurite is an important constituent, together with some haüyn... and sodalite, of lapis lazuli (*q.v.*).

The chemical composition of the members of the group is as follows:

Sodalite	$Na_8(Al_6Si_6O_{24})Cl_2$
Hauynite	$(Na,Ca)_{4-8}(Al_6Si_6O_{24})(SO_4)_{1-2}$
Noselite . . .	$Na_8(Al_6Si_6O_{24})SO_4$ ...
Lazurite . . .	$(Na,Ca)_8(Al_6Si_6O_{24})(S,SO_4)$ ...

Crystals usually have the form of the rhombic dodeca... are often twinned with interpenetration on an octa... They are white or often blue, in colour and have a vit... The hardness is 5 and the specific gravity 2.2–2.4.

These minerals are characteristic constituents of ... rich in soda and they also occur in metamorphic li...
(L. J. S.; ...)

SODALITIES OF OUR LADY, a Roman Cath... association (Lat. *sodalis*, "companion"). The first Sod... Lady was established at the Roman College in 1563 ... Leunis, a Flemish Jesuit. The first rule of the sodalities ... ardent devotion to the Mother of God of all members ... a deep interior spiritual program and, as far as their ... conditions permit, enter into any and all works for the ... temporal, and social welfare of others. Sodalities were ...

[illegible] Pope Gregory XIII and later by other popes [illegible] XII ratified all previous commendations and [illegible] [illegible] ments, he directed sodalists to become persons of [illegible] who ardently long to live a more perfect [illegible] respond to the relief of those in grave spiritual and material need.

[illegible]ies are organized at all levels, from children to [illegible] persons. There are more than [illegible] [illegible] World Federation of Sodalities. The world [illegible] first congress in Rome in 1954. The secretariat for the [illegible] States National Federation of Sodalities is in W[illegible] the international secretariat is in Rome. (D. J. W.)

SODA WATER: *see* SOFT DRINKS.

SODDY, FREDERICK (1877–1956), British scientist, [illegible] laureate in chemistry, 1921, who was among the first to [illegible] the existence of isotopes and gave them that name, was [illegible] Eastbourne, Sussex, on Sept. 2, 1877, and was educated at [illegible] College, the University College of Wales, Aberystwyth, [illegible] Merton College, Oxford.

[illegible] Soddy was demonstrator of chemistry at [illegible] University, Montreal, Que., where he worked with [illegible] Rutherford. Afterward he worked under Sir William [illegible] at University College, London [illegible] University of Glasgow. In [illegible] he became professor of [illegible] at Aberdeen University and next [illegible] professor of chemistry at Oxford.

With Rutherford, Soddy developed the disintegration theory of [illegible] radioactive elements. He concluded (1913) [illegible] should exist in two or more forms [illegible] atomic weights but are indistinguishable [illegible]; these forms he called "isotopes." He was elected a [illegible] of the Royal Society in [illegible] awarded the Nobel [illegible]. He died on Sept. 22, 1956, at Brighton.

SODEN, HERMANN, FREIHERR VON (1852–1914), [illegible] biblical scholar, author of a new theory of [illegible] the New Testament, was born in Cincinnati, O., Aug. [illegible] was educated at the University of Tübingen. He [illegible] of Dresden-Striesen in 1881 and [illegible] Jerusalem Church in Berlin. [illegible] in the University of Berlin [illegible] appointed extraordinary professor of [illegible] on Jan. 15, 1914.

[illegible] most important book is *Die Schriften des neuen Testaments in ihrer ältesten erreichbaren Textgestalt hergestellt auf Grund ihrer Textgeschichte* (1902–13).

[illegible] introduced, besides a new notation [illegible] *H* and *I* [illegible] he recon[illegible] *I-H-K* [illegible] to have [illegible] Clement [illegible] *I-H-K* [illegible] the New Testament [illegible] applications of readings so remarkable that [illegible].

[illegible] even among biblical scholars. (K. La. X.)

SÖDERBERG, HJALMAR EMIL FREDRIK (1869–[illegible]) [illegible]

[illegible]

SÖDERBLOM, NATHAN [illegible] Nobel [illegible] land [illegible] and Oriental [illegible] Ritschl's writings [illegible] the [illegible] with an open mind and [illegible] teaching of Luther. [illegible] chaplain to the Swedish legation in Paris he was appointed professor of theology at Uppsala [illegible], where he had an invigorating effect on the students and [illegible]. From 1912–14 he [illegible] chair at Leipzig. [illegible] appointed archbishop of Uppsala and primate of [illegible].

[illegible] the first World [illegible] 1925 which [illegible] united with the [illegible] World Council of [illegible]

[illegible]

SÖDERMANLAND, [illegible] east [illegible] 2,61 sq.mi. (6,831 sq.km.). The relief is broken [illegible]

the Södertälje ship canal to the east of it; this bypasses Stockholm in linking Lake Mälar with the Baltic. Although not a rich district, it gains through being crossed by trunk railways and roads.

Nyköping, the county town, lost iron-ore shipment traffic to Oxelösund, which is less icebound. (A. C. O'D.)

SÖDERTÄLJE, a town of Sweden, in the *län* (county) of Stockholm, lies between a bay of Mälaren Lake and the Baltic Sea, there connected by the Södertälje Canal (1819), 20 mi. SW of Stockholm. Pop. (1960) 33,014. In and around the town are St. Ragnhild's Church (believed to date from 1100), the town hall (1735), Tveta Church, Gripsholm Castle, and an open-air museum on the Torekällberget (hills). The town was founded in the 10th century A.D. and damaged by fire in 1390, 1650, and 1719. In the 19th century it was a well-known spa and bathing resort; it is now dominated by its industries, which include the manufacture of trucks and buses, metal goods, milk separators and milking machines, medicines, light concrete, and tobacco products. The town is a rail and road junction on the routes from Stockholm to Malmö, Göteborg, and Oslo, Nor. (F. Vo.)

SODIUM is a member of the group of chemical elements known as the alkali metals (other members are lithium, potassium, rubidium, cesium, and francium). Metallic sodium is a soft, silvery white material that tarnishes rapidly in air because of the formation of a film of the oxide. Sodium is very widely distributed in the form of its compounds, of which ordinary salt, or sodium chloride, is the most familiar. The chemical symbol for sodium is Na (from the Latin *natrium*), its atomic number is 11, and its atomic weight is 22.9898. Sodium has only one stable and naturally occurring isotope, of mass number 23, but radioactive isotopes with mass numbers 20, 21, 22, 24, and 25 have been prepared. Sodium-22, a positron emitter with a half-life of 2.6 yr., is often used as a radioactive tracer for sodium. Sodium-24 (a beta emitter with a half-life of 15 hr.) is, because of its short half-life, procurable only by laboratories reasonably close to the production machines.

The Metal.—Many of the physical and chemical properties of sodium may be explained by the electronic configuration of the sodium atom: $1s^2, 2s^2, 2p^6, 3s^1$. The inner ten electrons constitute a rare gas configuration and are not involved in the usual chemical reactions of sodium. Only the single 3s electron acts as a "valence electron"; it is removed very easily to yield an Na^+ ion. In the metallic state the valence electrons of the assembled sodium atoms can migrate in an electric field; hence the metal exhibits a very high electrical conductivity. Since there are not enough valence electrons to bind the atoms firmly to their neighbours in the solid state, sodium is soft (it may be readily sliced with a knife) and it has a low melting point (97.83° C) and a low heat of fusion (0.622 kg.cal/mole). Sodium boils at 883° C (with a heat of vaporization of 22.2 kg.cal/mole) to give a vapour consisting largely of monatomic sodium. A small fraction of the atoms in the vapour are dimerized as Na_2 molecules, which have a dissociation energy of 19 kg.cal/mole. The standard potential of a sodium electrode in aqueous solution is 2.714 volts relative to the standard hydrogen electrode:

$$Na = Na^+_{(aq)} + e^- \qquad E^\circ = 2.714 \text{ volts}$$

This potential places sodium among the most powerful reducing agents in aqueous solution and explains why sodium reacts vigorously with water to give hydrogen and a solution of sodium hydroxide:

$$Na + H_2O \longrightarrow Na^+ + OH^- + \tfrac{1}{2}H_2$$

The apparent radius of the sodium ion in its compounds is 0.98 Å (Ångström units), or 0.98×10^{-8} cm.

The ionization potential of sodium (the energy required to completely remove the 3s electron from the gaseous atom) is 5.138 electron volts. However, there are many lower energy levels to which the valence electron can be excited, corresponding to excited states of the sodium atom. Spontaneous transitions among these energy levels give rise to the emission spectrum. A very prominent pair of lines in the spectrum (the Fraunhofer D line) at 5896 and 5890 Å accounts for the bright yellow colour of both the sodium flame and the sodium vapour lamp. The flame colour is the basis of a very sensitive analytical test for sodium.

Sodium forms an amalgam with mercury with almost explosive violence. A sodium-lead alloy is often used in place of metallic sodium as a reducing agent or a drying agent. A eutectic alloy of 22.7% sodium and 77.3% potassium (called "NaK") is liquid down to −12.5° C. Sodium, like the other alkali metals, has the remarkable property of dissolving without appreciable reaction in solvents such as liquid ammonia and amines to form highly conducting blue solutions. In these solvents the sodium ionizes to the Na^+ ion and the blue solvated electron.

The first isolation of sodium (by Sir Humphry Davy in 1807) was achieved by the electrolysis of fused sodium hydroxide. A modification of this method, the Castner process, was long used as a commercial method for preparing the metal. Most sodium is now prepared by the Downs process, in which a fused mixture of sodium chloride and either calcium chloride or sodium carbonate is electrolyzed to yield chlorine as well as the metal. Pure molten sodium chloride is not used because at the melting point of sodium chloride (801° C) the vapour pressure of sodium metal is dangerously high and sodium is appreciably soluble in the fused salt. With the salt mixtures, the electrolysis cells may be operated in the vicinity of 600° C.

Metallic sodium is used in the manufacture of the peroxide (Na_2O_2), cyanide (NaCN), amide ($NaNH_2$), hydride (NaH), and borohydride ($NaBH_4$), all of which are of technical importance. It is extensively used as a reagent in synthetic chemistry. The sodium-lead alloy is used in the manufacture of tetraethyl lead, a common constituent of antiknock gasoline. Sodium, like the other alkali metals, possesses to a marked degree the photoelectric effect (emission of electrons when exposed to light) and was once used in the manufacture of photoelectric cells. On account of its relatively high heat capacity and high heat conductivity sodium is used as a heat transfer medium. In some types of atomic reactors, molten sodium is used to transfer heat from the core of the reactor to a water boiler for steam generation.

Compounds.—Because of its intense chemical activity sodium is never found as the metal in nature; in the form of compounds however, it is widely distributed. The earth's crust contains about 2.8% sodium in the combined form, placing it sixth in order of abundance among the elements. Sodium chloride is exceedingly abundant and widespread, being the chief salt present in sea water besides occurring throughout the world in extensive deposits that probably owe their existence to the evaporation of prehistoric seas. Sodium chloride comprises about 80% of the dissolved matter in sea water. (For additional information on sodium chloride, *see* SALT; SALT DOME.)

Sodium carbonates are also widely dispersed in nature, forming constituents of many mineral waters and occurring as the principal saline components in natron or trona lakes, as efflorescences in Lower Egypt, Iran, and China, and as trona (urao) in Mexico, Colombia, and Venezuela. The solid crusts found at the bottom of the salt lakes of the Araxes plain in Armenia contain about 16% of carbonate and 80% of sulfate. In Colombia there occurs a double salt, $Na_2CO_3 \cdot CaCO_3 \cdot 5H_2O$ known as gaylussite. Enormous deposits of carbonates, mixed in some cases with sulfate and with chloride, occur in Wyoming, California, and Nevada. Vast areas of the steppes in Hungary contain sodium carbonate in the soil.

Natural sulfate occurs in an anhydrous condition as thenardite, Na_2SO_4, in Tarapacá province, Chile, and in the rock-salt deposits at Espartinas near Aranjuez, Spain. The world's largest known deposits of natural sodium sulfate are located in Saskatchewan. Hydrated sulfates occur at several localities in Spain and at Mühlingen in the Aargau canton of northern Switzerland; copious deposits of glauberite, the double sulfate of sodium and calcium are found in the salt mines of Villarrubia in Spain, at Stassfurt, Ger., and in Tarapacá, Chile, etc. A native nitrate of sodium is obtained in great abundance in Tarapacá and also in the province of Atacama in Chile; it was formerly imported into Europe in enormous quantities as cubic nitre for the preparation of saltpetre. Cryolite, a fluoride of aluminum and sodium, is mined extensively

in Greenland and elsewhere for industrial purposes.

These minerals form the principal natural sources of sodium compounds—the chloride as rock salt and in seawater being of such predominating importance as to outweigh all the others. But it is questionable whether, taken altogether, the mass of sodium they represent is as much as that disseminated throughout the rocky crust in the form of sodium feldspar (*i.e.*, as silicate of sodium) and in other sodium-containing rocks.

In its compounds, sodium always exists as a +1 ion. Its salts are generally very soluble in water; three exceptions are the fluosilicate, Na_2SiF_6; the antimonate, $NaSb(OH)_6$; and the zinc uranyl acetate, $NaZn(UO_2)_3(C_2H_3O_2)_9 \cdot 6H_2O$.

Sodium hydride, NaH, is a crystalline substance obtained by the direct reaction of hydrogen with sodium at 250° to 350° C. It burns when heated in air and reacts vigorously with water, giving one mole of hydrogen per mole of hydride:

$$NaH + H_2O \longrightarrow Na^+ + OH^- + H_2$$

The analogous reaction of sodium hydride with alcohols is valuable for the preparation of alkoxides:

$$NaH + ROH \longrightarrow NaOR + H_2$$

Sodium hydride reacts with methyl borate at elevated temperatures to form sodium borohydride, $NaBH_4$:

$$4NaH + B(OCH_3)_3 \longrightarrow NaBH_4 + 3NaOCH_3$$

Sodium borohydride is stable in alkaline aqueous solutions and is useful as a powerful reducing agent. Sodium borohydride is an important starting material in the preparation of diborane, B_2H_6.

Sodium monoxide, Na_2O, may be prepared by direct oxidation in the presence of excess metal. Reaction in an excess of oxygen yields chiefly the peroxide, Na_2O_2. In the industrial preparation of the peroxide, the oxidation proceeds stepwise, as follows:

$$2Na + \tfrac{1}{2}O_2 \longrightarrow Na_2O$$
$$Na_2O + \tfrac{1}{2}O_2 \longrightarrow Na_2O_2$$

If sodium peroxide is heated to about 450° C. in 300 atmospheres of oxygen, the superoxide, NaO_2, is formed:

$$Na_2O_2 + O_2 \longrightarrow 2NaO_2$$

The peroxide and superoxide are powerful oxidizing agents. Sodium peroxide reacts with carbon dioxide to liberate oxygen:

$$2Na_2O_2 + 2CO_2 \longrightarrow 2Na_2CO_3 + O_2$$

The last reaction found application during World War II and thereafter for the revivification of air in confined spaces such as aboard submarines and also aboard airplanes at extremely high altitudes.

Sodium hydroxide, NaOH, also called caustic soda, is manufactured by the electrolysis of a solution of sodium chloride, a process that results in the formation of the hydroxide and hydrogen at the cathode and chlorine at the anode. The overall cell reaction is

$$2H_2O + 2Cl^- = 2OH^- + H_2 + Cl_2$$

The electrolysis cell is constructed so as to prevent the interaction of the hydroxide and the chlorine, which give hypochlorite. In the Castner-Kellner process, a mercury cathode is used so that a sodium amalgam is formed. The cell is constructed so that the amalgam reacts with water to form a very pure hydroxide solution, which is then evaporated to give the dry product. (*See* also Alkali Manufacture: *Electrolytic Production of Caustic Soda.*)

Sodium amide, $NaNH_2$, is prepared by the direct reaction of sodium with ammonia:

$$Na + NH_3 \longrightarrow NaNH_2 + \tfrac{1}{2}H_2$$

Sodium amide is used in the preparation of sodium cyanide by reduction with carbon at red heat:

$$NaNH_2 + C \longrightarrow NaCN + H_2$$

Sodium amide finds considerable use in synthesis as a reagent for the introduction of amino groups into molecules, for dehydration and for dehydrohalogenation. Sodium amide reacts in liquid ammonia with nitrous oxide to form sodium azide, NaN_3:

$$2NaNH_2 + N_2O \longrightarrow NaN_3 + NaOH + NH_3$$

Sodium azide has the remarkable property of decomposing cleanly to its elements at 300° C:

$$NaN_3 \longrightarrow Na + \tfrac{3}{2}N_2$$

The manufacture of sodium carbonate, Na_2CO_3, also called soda ash, is treated under Alkali Manufacture: *Solvay Process of Soda Ash Manufacture.* The anhydrous salt is a colourless material having an alkaline taste and reaction due to the hydrolysis of the carbonate ion. When heated, the bicarbonate is converted to the carbonate:

$$2NaHCO_3 \longrightarrow Na_2CO_3 + CO_2 + H_2O$$

Of the sodium silicates the most important is the mixture known as water glass, which is formed by calcining a mixture of white sand, sodium carbonate and charcoal, or by dissolving silica in sodium hydroxide under pressure. It is a colourless sirupy liquid that is used in certain printing processes, as a cement for glass, porcelain, etc., as a fireproofing agent, and as a water softener.

The manufacture of common soda glass principally involves the high-temperature reaction of sodium carbonate, calcium carbonate and sand to give a glass of the following approximate composition: SiO_2, 71%–78%; Na_2O, 12%–17%; CaO, 5%–15%; Al_2O_3, 1%–4%. Such glass has a remarkable permeability to sodium ions. At room temperature, the mobility of sodium ions is sufficiently great in some glasses to permit the manufacture of a glass electrode for determining sodium ion concentrations in aqueous solutions.

See also references under "Sodium" in the Index.

Bibliography.—W. M. Latimer and J. H. Hildebrand, *Reference Book of Inorganic Chemistry*, 3rd ed. (1951); M. Sittig, *Sodium, Its Manufacture, Properties and Uses* (1956); J. Kleinberg *et al.*, *Inorganic Chemistry* (1960). (W. L. Jo.)

SODOMA (Giovanni Antonio Bazzi, known as "Il Sodoma") (1477–1549), leading Italian painter of the early 16th century in Siena, was born at Vercelli, in Piedmont, in 1477. From 1490–97 he was apprenticed to G. M. Spanzotti, a minor Piedmontese artist, but was afterward much influenced by the great masters of the High Renaissance, the style of Leonardo da Vinci (whose work Sodoma must have seen in Milan and Florence) and later that of Raphael being particularly decisive in determining his mature style. He was invited to Siena in 1501 and subsequently spent the bulk of his working life there. He died in Siena on the night of Feb. 14–15, 1549.

Sodoma's earliest works of repute were painted between 1505 and 1508 in the Benedictine monastery of Monteoliveto Maggiore, near Siena, where he completed the series of scenes from the life of St. Benedict left unfinished by Luca Signorelli in 1498. As a result of these he was invited to Rome by the celebrated Sienese banker Agostino Chigi, and was employed by Pope Julius II in the Stanza della Segnatura in the Vatican. Although he was superseded by Raphael in 1509, some of his ceiling decoration remains, and Raphael included Sodoma's portrait as well as his own in the "School of Athens" painted in this same room. Sodoma's masterpiece in Rome is the fresco decoration of Agostino Chigi's bedroom in the Villa Farnesina, where his "Marriage of Alexander and Roxana" (probably 1511–12) vies as a decorative achievement with frescoes by the school of Raphael in the same villa.

Sodoma's career spanned the period of transition from High Renaissance to Mannerism, and his work echoes these main trends. He had, however, a peculiar gift for suggesting the sensuous beauty of the human form (a characteristic that probably endeared him to the Sienese), and an exaggerated, almost mystical, emotionalism that anticipates one aspect of the Baroque. The St. Catherine frescoes in S. Domenico, Siena (begun in 1526), the "Christ at the Column" and the "Descent into Limbo" (both mature works in the Pinacoteca Nazionale, Siena) show Sodoma's art at its

best and most typical, as also does the St. Sebastian (Pitti Gallery, Florence), where his considerable feeling for landscape can also be appreciated. Works in Pisa Cathedral relate to Sodoma's visit there in 1540 and are among the few documented works of his late years.

Vasari, who disliked Sodoma, makes the most of the unpleasant sobriquet by which he was known from 1512 onward. From what is known of Sodoma's life, however, the nickname is more likely in the first place to have been the result of a joke, but it was adopted by the artist himself, and is the name by which he is now generally known.

See R. H. H. Cust, *Giovanni Antonio Bazzi* (1906), who quotes relevant documents; L. Gielly, *Giovan-Antonio Bazzi* (1911).

SODOM AND GOMORRAH, two of the five "cities of the plain" in the Dead Sea region, the others being Admah, Zeboiim, and Zoar (Bela) (Gen. 13, 14, 19). Sodom and Gomorrah, cities of legendary wickedness, were destroyed by a rain of "brimstone and fire" (Gen. 19:24). Their most likely location is beneath the shallow waters of the southern end of the Dead Sea (*q.v.*). Several minor streams which flow into the Dead Sea from the east could have provided water for their support. The level of the sea has risen in historical times, owing to deposits of silt and to other factors. Confirmation of heavy settlement of population in this general location is provided by Bab edh-Dhrâ, which was once a place of pilgrimage and feasting located about 500 ft. above the Dead Sea Peninsula (Al Lisan) and was used by the inhabitants of the cities below. Jabal Usdum ("Mt. Sodom"), at the southwest end of the sea, reflects Sodom's name. Josephus (*Jewish War* iv:8:4) locates Zoar definitely at the Dead Sea's southern end. (E. D. Gr.)

SODOMY: *see* Homosexuality.

SOENDA ISLANDS: *see* Sunda Islands and Strait.

SOEST, a town of West Germany in the *Land* (state) of North Rhine-Westphalia, Federal Republic of Germany, lies 30 mi. (48 km.) E of Dortmund by road. Pop. (1961) 33,304. Soest is in the middle of the fertile loess-covered tract of land known as the Soester Börde extending south from the Lippe River. The Grosser Teich ("great pond") in the town centre is fed by warm springs; an affluent of the Lippe rises there. The town is among the most charming in Westphalia, with a profusion of flowering trees, numerous timbered houses, and many old buildings. The remains of the surrounding walls include the Osthofentor gatehouse (1526). A 12th-century Roman Catholic cathedral (St. Patroclus) adjoins the Baroque town hall (1713), whose municipal archives include letters by Luther and Melanchthon and an important early Protestant theological library. The Wiesenkirche ("church in the fields") is one of the most beautiful Gothic churches in Germany and its pleasing stained-glass window (*c.* 1530) of the Last Supper is typically Westphalian. There was severe bomb damage in World War II, but restoration has taken place. The town is the market for the farm products of the Börde, and its industries include the production of sugar, storage batteries and lamp bulbs, wire, machinery, road making equipment, and clothing.

Soest was first mentioned (as Sosat) in documents in 836 and received its first charter in 1144. Until 1444 an appanage of the archbishops of Cologne, it became one of the most important Hanseatic towns. Its 12th-century code of municipal law (*Soester Schrae; jus susatense*) was one of the earliest and best.

SOEST, a municipality in the province of Utrecht, Neth., lies 28 mi. (45 km.) SE of Amsterdam by road. Pop. (1960) 28,370. It includes Soestdijk, where there is a royal residence; Soesterberg, with a military airfield; and Soestduinen. Soest town, founded in 1029, has a Gothic church dating from about 1400 and restored in 1959. Formerly an agricultural centre, the town is now largely residential, with many inhabitants commuting to the large towns, but has expanding light industries. (R. T.)

SOFA: *see* Chair and Sofa.

SOFIA (Sofiya), capital city of Bulgaria and centre of the *okrug* (district) of Sofia, is also an *okrug* in its own right. It is situated on two tributaries of the Iskur River, 1,800 ft. (550 m.) above sea level. Its climate is temperate continental and the temperature averages 1.9° C (35.4° F) in January and 20.2° (68.4°) in July. Pop. (1962) 713,308, not including suburbs.

Sofia is divided into six districts radiating from the centre of the city. There are some historical monuments, including the church of St. George, parts of which date from the 3rd century, which has been heavily restored, as also has the church of St. Sophia; and two mosques, one of which houses the archaeological museum. The domed Alexander Nevski cathedral was built in gratitude to the Russians who liberated Bulgaria in 1877–78. Notable also are the monument to the Russian army of liberation, the headquarters building of the Bulgarian Communist Party, and the mausoleum of the statesman Georgi Dimitrov (*q.v.*). As the centre of Bulgaria's cultural life Sofia houses the Kliment Ochridsky University, the Bulgarian Academy of Sciences, and the National Library, Theatre, Opera, and Art Gallery. There are many general education and vocational schools, health centres, and hospitals. Considerable rebuilding was done in Sofia after World War II. Traces of air raids in 1943 and 1944 were eliminated, new public buildings, houses, housing estates, hospitals, and schools were built, and parks and gardens were laid out. In 1961 a general plan was adopted for the development and reconstruction of the city. The chief industries are engineering and metallurgy, textiles, and food processing, with the manufacture of furniture, footwear, chemicals, and rubber goods. Thermal power stations supply the city with electricity and heat.

Sofia is the most important communications centre of Bulgaria, and is linked by road and rail to Belgrade, Istanbul, Athens, and Bucharest. There are regular air services to the larger towns of the country and to the principal cities of Eastern and Western Europe, as well as to Africa and Asia. City transport services are provided by a network of trams (streetcars), buses, and trolley-buses.

Sofia, known to the Romans as Serdica, grew and flourished during the reign of the emperor Trajan (2nd century A.D.) and until the 4th century. In the 5th century, as a province of the Byzantine Empire, it was devastated by the Goths, the Huns, and other peoples. Under Justinian (6th century) Serdica once more became an important town in the Eastern Empire and the original church of St. Sophia, which later gave its name to the town, dates from this time. The town was occupied by the Slavs in the 7th century. During the reign of King Krum in the 9th century Serdica was included in Bulgaria, and under the name of Sredets it was second in importance to Turnovo, capital of medieval Bulgaria.

It was under Byzantine rule from 1018–1185 and it was captured by the Turks in 1382. Many of the ancient buildings were destroyed and the town took on an Oriental appearance. In 1878 the Russian troops liberated Sofia and the constituent assembly which met in Turnovo in February 1879 chose the town as the capital of Bulgaria. This, together with its geographical situation, helped its rapid political, economic, and cultural development.

Sofia District (Sofia Okrug) has an area of 2,811 sq.mi. (7,281 sq.km.); pop. (1963 est.) 329,900. The city of Sofia lies on a plain surrounded by mountains. To the south are the Vitosha and Lyulin national parks. The district is famous for its mineral water springs and baths which are to be found in Sofia itself and in the suburbs of Gorna Banya, Knyazhevo, and Ovea Kupel, and the villages of Bankya and Pancharevo.

The church at Boyana, 6 mi. (10 km.) S of Sofia, is famous for its 13th-century wall paintings. The Iskur Dam lies 24 mi. (39 km.) SE of Sofia and Pancharevo Lake is at a distance of 9 mi. (14 km.). The waters of the dam are used for irrigation and power. (L. Di.)

SOFTBALL is one of the most popular participant sports in the United States. Though there are conflicting claims of invention and origin, it is generally agreed that softball, closely akin to baseball, developed from a game called "indoor baseball" which was first played in the gymnasium of the Farragut Boat Club in Chicago, Ill., in 1887. The game was introduced to the playgrounds and parks of Chicago and Minneapolis and later promoted throughout the country as a desirable outdoor game to be played in small public park and school playgrounds. It became known

in the United States by various names such as kitten ball, mush ball, diamond ball, indoor-outdoor and playground ball. Since there were wide variances in playing rules, size and type of playing equipment, and lack of uniformity as to dimensions of the playing field, a need developed to secure a uniform code of rules. In certain states, particularly Minnesota, Wisconsin, Florida and Colorado, the game rapidly progressed; playing rules were made more uniform and improved equipment was developed by manufacturers.

At the National Recreation Congress in Springfield, Ill., in 1923, a rules committee was appointed by Joseph Lee, a United States pioneer in recreation. The Lee committee was later enlarged to form the International Joint Rules Committee of Softball, which came to include representatives of a number of organizations which promote and sponsor softball. The committee was formed to gain general acceptance of and to publish and circulate a standard set of rules; to study the game in order to recommend changes in rules that would benefit the largest number of players; and to cooperate with manufacturers of sporting goods to secure standard and satisfactory equipment.

The Amateur Softball Association of America, organized in 1932, came to be the recognized governing agency for promotion and control of organized national competition. It became a member of the Amateur Athletic Union, the U.S. Olympic Commission, the National Recreation Association, and the International Softball Federation. Through its metropolitan, state and regional organization, it assumed responsibility for directing thousands of teams in competition, culminating in world championship tournaments for men and women, which have been held without interruption from 1933.

The International Softball Federation acts as a clearing centre for the softball organizations of more than 20 countries, including Australia, Japan, Argentina, England, and Canada.

Basically, the fundamentals of softball are the same as those of baseball (*q.v.*). Batting and fielding strategy are similar, but since the game is played on a much smaller area the action is much faster. It seldom requires more than an hour to complete seven innings for an official game.

A distinctive technique in softball is the underhand pitching. Special skills were soon developed in this art and pitchers developed baffling windups, followed by a release and throw of the ball that travels with a speed comparable to that of the overhand regulation baseball.

The regulation playing field for softball should have a clear, unobstructed area within a radius of 200 feet from home plate between the foul lines. It includes a diamond-shaped area with 60-ft. base lines. The pitching distance for men is 46 ft. and for women 40 ft. Bats must be round, not more than 34 in. long and not more than 2¼ in. in diameter at the largest part. The official softball is a smooth-seam ball, not less than 11⅞ in. nor more than 12⅛ in. in circumference, weighing between 6 and 6¾ ounces.

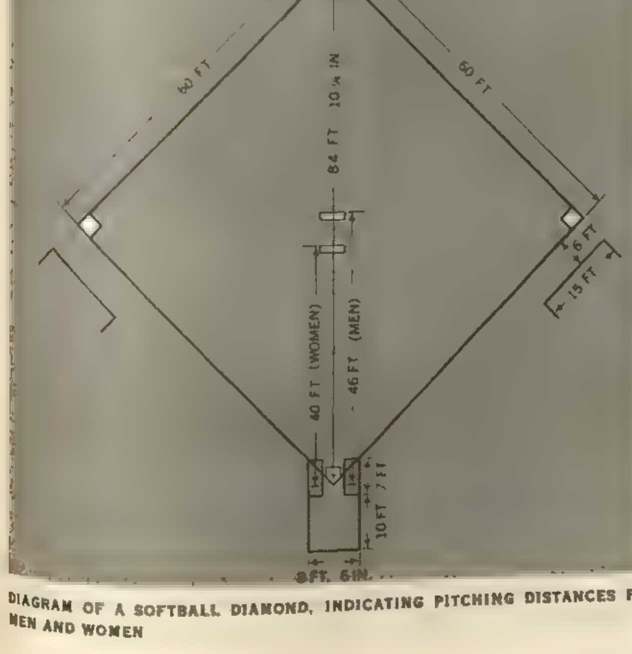


DIAGRAM OF A SOFTBALL DIAMOND, INDICATING PITCHING DISTANCES FOR MEN AND WOMEN

In summary, the rules for softball differ from those of baseball in the following respects:

(1) Softball base lines are 60 ft. long whereas baseball base lines are 90 ft.

(2) Softball pitching distance is 46 ft. compared with baseball's 60 ft. 6 in.

(3) The official softball is heavier and larger than a baseball.

(4) The official softball bat is shorter and lighter than a baseball bat.

(5) Softball games consist of seven innings whereas baseball games are nine innings.

(6) The chief difference in the pitching rules is that in softball the ball is delivered by an underhand motion to the batter with a follow-through of the hand and wrist past a straight line of the body before releasing the ball. The pitcher must pause for one second at the conclusion of his windup before delivering the ball.

(7) Base stealing is permitted in both games, but in softball the runner must keep contact with the base until the pitcher releases the ball on delivery to the batter.

Principal reasons for the great popularity of softball are that the game is played with speed, that it can be played at relatively low cost for equipment and uniforms, that it can be and is played by children and young and old men and women, and it has great appeal to spectators. Because of the smaller playing area, softball fans are continually thrilled by hard drives, bare-handed stops and fast throws. Close decisions are the rule on most plays. It gained wide acceptance, moreover, as an outstanding girls' and women's team sport; hundreds of women's softball teams are members of leagues which compete for honours in a world's championship tournament for women.

Softball also demonstrated its value as an informal recreational activity for industrial and business, church and school groups, at picnics and outings, and on neighborhood vacant lots and playgrounds.

One variation of the game, called slow-pitch, is played with a 14- or 16-in. ball on a diamond with 50-ft. baselines. Teams are organized for league play. (A. T. N.)

SOFT DRINKS, or nonalcoholic beverages, may be divided into two classes: carbonated beverages (sometimes called aerated waters) and still beverages (those without carbonation). With the exception of club soda or sparkling water, carbonated drinks contain sugar (or other sweetening) and edible acids, as well as artificial flavours or natural flavours derived from fruits, nuts, roots, herbs or other plant sources. Still beverages generally contain natural or artificial fruit flavours.

History.—Carbonated beverages and waters developed from European attempts to imitate the popular and naturally effervescent waters of famous springs, with primary interest in their reputed therapeutic values. The effervescent feature of the waters was recognized early as most important. Jan Baptista van Helmont (1577–1644) first used the term "gas" in his reference to the carbon dioxide content. Gabriel Venel referred to "aerated water," confusing the gas with ordinary air. Joseph Black named the gaseous constituent "fixed air."

Numerous reports of such experiments and investigations were included in the *Transactions* of the Royal Society of London in the late 1700s, including the studies of Stephen Hales, Joseph Black, David Macbride, William Brownrigg, Henry Cavendish, Thomas Lane, Joseph Priestley and others. Duplication of the waters of Bad Pyrmont, a spa in Germany near the city of Hameln, was one of the major objectives, probably because of the popularity of the spa and also because of the relatively simple nature of the mineral content of the waters there.

Priestley, who obtained "fixed air" for his experiments from a brewery near his house, published *Directions for Impregnating Water with Fixed Air* in 1772. For his reports concerning "fixed air" and the mixture of waters with it, but with special emphasis on the medicinal values shown by him to exist in artificially car-

bonated water (but since found to be illusory), he was awarded the Copley medal by the Royal Society of London in the following year.

Meanwhile, other European scientists were active in the field. Studies by Torbern Bergman, a Swedish chemist, were published in the reports of the Swedish Royal Academy of Science in 1775. Carl Wilhelm Scheele of Sweden was similarly engaged. Antoine Lavoisier, in France, identified the "fixed air" of Priestley as a combination of carbon and oxygen which he called *gas acide carbonique*. John Mervin Nooth, in England, developed special apparatus for the preparation of small quantities of the effervescent waters. His report appeared in the 1775 *Transactions* of the Royal society.

Improvements in the Nooth device were made by Jean Hyacinthe de Magellan in 1777, and in 1781–83 the English chemist Thomas Henry described apparatus for the production of carbonated waters on a commercial scale. Many others followed, and factories and bottling plants were opened in Geneva, Switz., Paris, London, Dublin, Ire., Dresden, Ger., and several other European cities during the period 1789–1821.

U.S. investigators also were interested in carbonated water and its health values. As early as the summer of 1807, in New Haven, Conn., Benjamin Silliman of Yale college began producing bottled soda water on a commercial scale, and he opened a public establishment for dispensing it. In Philadelphia, artificially carbonated waters were bottled and sold commercially by Joseph Hawkins, with machinery of his own invention. The firm was known as Shaw and Hawkins. Hawkins was issued a U.S. patent for preparation of imitation mineral waters in 1809, the first of record in this field.

The addition of flavourings to soda water to form the effervescent type of soft drink or carbonated beverage which became so widespread in America began early. Its exact origin is obscure, however. In 1768 Richard Bewley, in England, had introduced a form of soda water which he described as "mephitic julep," to be taken "with a draft of lemonade." He had followed a similar suggestion of Macbride concerning the medical efficacy of a mixture of fresh lime juice and alkaline salts taken as an effervescent drink.

Production Methods.—The basic ingredients in soft drinks are water, carbon dioxide, sugar, flavouring and, sometimes, artificial colouring.

Water, although most often taken from a safe municipal water supply, usually is given further processing to ensure uniformity of the finished product, because the amount of impurities in the municipal supply may vary from time to time. In some bottling plants the water-treatment equipment may consist simply of a sand filter to remove minute solid matter and an activated-carbon purifier to remove colour, chlorine and other tastes or odours. In most plants, however, water is treated by a process known as superchlorination and coagulation. In this process the water is exposed for two hours to a high concentration of chlorine as well as to a flocculant that removes such organisms as planktons (minute plants and animals); it is then passed through a sand filter and activated-carbon purifier.

Carbon dioxide gas gives the beverage its sparkle and tangy taste, and prevents spoilage. While it has not been proved that carbonation offers a direct medical benefit, carbonated beverages are used to alleviate post-operative nausea when no other food can be tolerated, as well as to ensure adequate liquid intake.

Carbon dioxide is supplied to the soft drink manufacturer in either liquid or solid ("dry ice") form, maintained under high pressure in heavy steel containers. As the pressure is released, the material changes into the gaseous form, in which it can be dissolved in water in quantities that vary with temperature and pressure. The amount of gas the water will absorb increases as the pressure is increased and as the temperature is decreased. Carbonation (of either the water or the finished beverage mixture) is effected by chilling the liquid and cascading it in thin layers over a series of plates in an enclosure containing carbon dioxide gas under pressure.

Flavouring sirups are made from sugar that is delivered from the refiner either in granulated form or as a 67% or 76% solution known as liquid sugar. The sugar is dissolved or diluted with processed water, and the desired flavouring substances are added. The resulting mixture is made more palatable by the addition of edible acids, principally citric acid, and its colour may be heightened by the addition of caramel (burnt sugar) or certified food-colouring materials. Large quantities of saccharin and other sweeteners are used for low-calorie beverages.

Two methods of producing the finished beverage are in use: either the sirup may be mixed with carbonated water prior to bottling or a precise amount of sirup may be measured into each bottle, after which the bottle is filled with carbonated water. In either case, the sugar content—51%–60% in the sirup—is reduced to 8%–13% in the finished beverage. The blending of sirups, mixing with carbonated water and filling of containers is carried out almost entirely by automatic machinery.

Most soft-drink producers use bottles of the returnable type, which reduce packaging costs, although some are using nonreturnable, "one-trip" bottles and metal cans. Before filling, returnable containers are thoroughly flushed with hot solutions of caustic soda or other alkalies for a minimum of five minutes, and then rinsed with potable water.

The production of still beverages involves ingredients and techniques similar to those used for carbonated beverages. However, since they lack the protection against spoilage afforded by carbonation, still beverages usually are pasteurized, either in bulk or by continuous flash pasteurization prior to filling or in the bottle. Some still beverages are sold in concentrated, frozen form, and these do not require pasteurization; however, they must be kept frozen until used. Since they are not under pressure, still beverages may be packaged not only in bottles and cans but also in cardboard cartons.

See also GINGER ALE. (JN. J. R.; H. E. K.)

SOGA, the name of one of the great families of Japanese history, the first to usurp power and thus initiators of the tradition of dual government in Japan. In addition, Soga Iname was primarily responsible for introducing Buddhism into Japan, the first Buddhist temple in that country being in his home. Although descended from a collateral branch of the imperial family the Soga's history is obscure until Iname became minister in A.D. 536. His son Umako also became minister, and, beginning with the death of the emperor Bidatsu in A.D. 585, executed a series of maneuvers that culminated in the enthronement of his niece as empress Suiko in 592. Between that year and 643 the Soga held undisputed power in Japan and could claim three emperors and one empress as their creations. In 645 Soga Iruka and Soga Yemishi were assassinated by the Fujiwara family, who were to succeed them as regents, and with these murders Soga power ended. *See* also JAPAN: *History: The Clan Period.*

BIBLIOGRAPHY.—G. Sansom, *A History of Japan to 1334* (1958); J. Murdoch, *A History of Japan*, vol 1, ch. 4 (1926); R. Ponsonby-Fane, *History of the Imperial House of Japan* (1959). (J. A. HA.)

SOGA (BASOGA, WASOGA), an east African people of the Interlacustrine Bantu group who inhabit the area east of the Nile between Lakes Victoria and Kyoga. Although sharing a common basic culture, the Soga were divided among several small states Bugabula Bulamogi, Busiki, Luuka, Kigulu, Bukoli, Bugweri Butembe, Bunya, and Bunyuli. In economy, technology, and religion they resembled the Ganda (*q.v.*). In each state there were several dozen patrilineal clans, one of which held the rulership. The northern states were often tributary to the Nyoro (*q.v.*), while the Ganda dominated the south.

Since 1894 the Soga formed an administrative district of the Uganda (*q.v.*) protectorate which achieved full independence in 1962. In the 1960s they numbered about 600,000. *See* also BANTU (INTERLACUSTRINE). (L. A. FS.)

SOGDIANA, the ancient name for a region of Central Asia centring on the fertile valley of the Zeravshan, in the modern Uzbek Soviet Socialist Republic. The vast remains at Samarkand suggest that the region has been an important centre throughout history. The earliest settlement there belongs to the Iron Age of northeast Iran, in the first half of the 1st millennium B.C. Shortly

afterward, Sogdiana passed under Achaemenid rule and appears as a satrapy of Darius I. It was later attacked by Alexander the Great and may for a time have been included in the Bactrian Greek kingdom, until the invasions of Sakas and Yüeh-chih (*q.v.*) in the 2nd century B.C. Sogdiana remained a prosperous centre of population down to the Mongol invasions. Under the Samanid dynasty (9th–10th centuries) it became for a time the eastern focal point of Islamic civilization. (F. R. A.)

SŌGI (1421–1502), Buddhist monk and greatest master of *renga* (linked verse), the supreme Japanese poet of his age. Nothing definite is known about his career before 1457, but he was born of exceedingly humble stock. His later writings suggest that after serving as a Zen monk in Kyōto he became in his 30s a professional *renga* poet. In earlier times his humble birth might have militated against his career, but in the 15th century, when violent upheavals shook the traditional social order (*see* JAPAN: *History: The Development of Feudalism*), it was not an insuperable handicap. His teachers included not only provincial *renga* masters but court nobles, and though this training undoubtedly benefited his poetry it also exerted an inhibiting influence. Sōgi's own selection of his best work shows him at his most ingenious, in the aristocratic tradition; but his modern reputation is primarily accounted for by the deeply moving vein of his simpler and more personal poems.

Sōgi is known as a traveler-poet. His life for 40 years was divided between the capital and the provinces. From 1466 to 1472, a period when warfare ravaged Kyōto, he lived mainly in eastern Japan, associating with local potentates and composing manuals of *renga* for them. His return to Kyōto in 1473 ushered in his most fruitful period. His residence became the centre of literary activity in the city, and he compiled *Wasuregusa* and other collections of his poetry. In 1480 he made a journey to Kyushu (recorded in his *Tsukushi no Michi no Ki*, Tsukushi being a poetic old name for Kyushu), not as a wandering priest in the traditional manner but as a celebrity, feted everywhere by his admirers.

Sōgi's reputation derives mainly from two *renga* sequences, *Minase Sangin Hyakuin* (1486; see *Bibliography*) and *Yuyama Sangin Hyakuin* ("One Hundred Poems Composed by Three Poets at Yuyama"; 1491); in each of these, three poets, led by Sōgi, took turns at composing short stanzas ("links") to form a single poem with many shifts of mood and direction. This expression of a multiple stream of consciousness was especially congenial to Japanese during the 15th century, when isolation and loneliness were the rule and *renga* gatherings, in which even the common people participated, provided rare opportunities for intellectual and literary companionship. Sōgi left over 90 works including the *renga* anthology *Shinsen Tsukubashū* (1495), diaries, works of poetical criticism, and manuals.

See Kenneth Yasuda, *Minase sangin hyakuin: a Poem of One Hundred Links Composed by Three Poets at Minase*, Eng. trans. (1956); Donald Keene (ed.), *Anthology of Japanese Literature* (1955).
(DD. K.)

SOGNEFJORD, an inlet of the west coast of Norway, the country's longest and deepest fjord, penetrating 114 mi. (183 km.) eastward into the mainland and reaching 4,291 ft. (1,308 m.) in depth. Sognefest lies at the entrance, which is sheltered by the Solund Islands. For the first 50 mi. (80 km.) from the sea the flanking mountain walls are virtually unbroken; beyond this, deep, narrow, secondary fjords penetrate them: Fjaerlands-, Sogndals-, and Lusterfjord to the north, Årdalsfjord to the east, Laerdals- and Aurlandsfjord to the south, which extend north toward the Jostedalsbre and Jotunheimen and south to the Hallingskarvet. At each fjord-head lies a small delta laid down by the stream that issued from the retreating glacier which formed the fjord (*q.v.*). This flat land is usually the site of the principal settlement because elsewhere the ground rises steeply from the water's edge. Branching from the Aurlandsfjord is the Naerøyfjord with the Naerøydalen, a steep valley of great beauty, at its head leading to the Stalheim Pass. Høyanger and Årdal are sites of big aluminum plants. The Sognefjord district is rich in traditions from the Viking and pre-Viking ages. The first settlers in Iceland came from this district. (L. H. HG.)

SOGN OG FJORDANE, a *fylke* (county) of western Norway. Area 7,154 sq.mi. (18,530 sq.km.). Pop. (1960) 96,045. In the north the Nordfjord penetrates 56 mi. (90 km.) eastward and in the south the Sognefjord (*q.v.*) 114 mi. (183 km.) into mountainous regions. Between these fjords the district of Sunnfjord is indented by several short individual fjords. The Jostedalsbre, the largest icefield in Europe, extends northeast-southwest from the head of Nordfjord nearly to the middle part of Sognefjord. In the island-studded coastal district, the economy is based on fisheries with farming as a subsidiary occupation. On gentle slopes along the fjords and in the valleys beyond the fjord-heads agriculture predominates. Fine orchards are characteristic of the inner and dry fjord areas. The magnificent scenery of these districts attracts many tourists. Abundant hydroelectric power has given rise to aluminum and other electrometallurgical industries. (L. H. HG.)

SOHO, a district in the City of Westminster, London, Eng., bounded by Oxford Street (N), Charing Cross Road (E), Coventry Street (S), and Regent Street (W); Piccadilly Circus is at the southwestern corner. A plan of 1585 shows the area as fields which in 1643 just came within the fortifications of London hurriedly put up against the royalists; a fort stood near the northern end of the present Wardour Street. Gregory King surveyed the area for building and in 1681 built King's, later Soho, Square, the district soon becoming fashionable. Foreign tradesmen, especially French Huguenots, finding few opportunities in the City, settled in Soho. In 1700 there were about a dozen French churches near and in Soho, but by the early 20th century the French had been displaced by, among others, the Italians whose quarter was originally in Clerkenwell. Soho has many famous names linked with it, including G. J. Casanova, William Blake, who was born in Broad (now Broadwick) Street, and Karl Marx; William Hazlitt lies buried in St. Anne's Church gardens not far from the bankrupt Theodore, king of Corsica. Soho, no longer a community, is known chiefly for its continental restaurants and food shops; Berwick market is within its bounds, and clothing firms, film companies, and publishers, especially of music, are established there. The name (*So ho* in ratebooks of 1632) is an old hunting cry.

SOIGNIES (Flemish ZINNIK), a town in the province of Hainaut, Belgium, 10 mi. (16 km.) NE of Mons, dominated by the towers of the collegiate church (10th–13th centuries), one of the country's best-preserved Romanesque buildings. Pop. (1961) 10,874. There was a Roman castrum (fortress) at the site, and the town was said to have been founded in the 7th century by St. Vincent who built a monastery there. The Old Cemetery, with some 16th-century tombs, is now a public park and contains a Romanesque chapel which now houses an archaeological and lapidary museum. Carboniferous limestone quarries are worked nearby. A hollow-glass plant operates in the town.

SOIL. The rocks and minerals of the earth's surface were the starting materials from which soils originated. Exposure to the elements and volcanic and tectonic action has resulted in the disintegration of these rocks and minerals to give a more or less unconsolidated residue over the earth's surface which is called the regolith. Organisms grew in and on the regolith, and their partly decomposed carbonaceous compounds were added to the mineral mixture. Insoluble residues were left in place. Soluble compounds were moved towards the surface, in extremely dry areas, and down and possibly into drainage waters, in humid areas. When these processes were allowed to operate without further geologic mixing for long periods of time, the regolithic material differentiated into an orderly sequence of layers, or horizons. These chemically and biologically differentiated top layers of the regolith are the soil.

Soil science and the use of soils will be discussed under the following headings:

Additional information will be found under Land Reclamation; Irrigation; Rotation of Crops; Fertilizers and Manures; Liming.

I. SOIL ORIGIN AND DEVELOPMENT

The surface of the regolith seldom remains in the place where it was first formed; during its millions of years of existence it has been moved from place to place by wind, flowing water, or advancing glaciers. Soil is any part of the regolith that has been in one place for a sufficient length of time and has been acted upon by natural physical, chemical, and biotic forces of sufficient intensity to develop detectable surface and subsurface layers called soil horizons. Soil-horizon differentiation in the regolith involves the accumulation of organic materials in the shallower horizons and the movement of dissolved mineral materials from certain horizons into neighbouring horizons (*see* fig. 1).

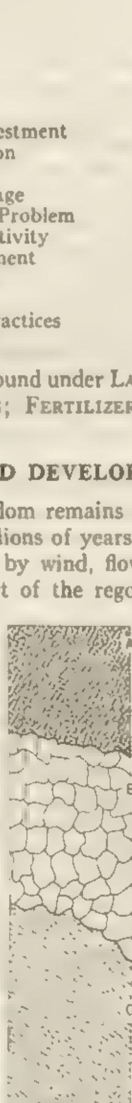


FIG. 1.—SCHEMATIC REPRESENTATION OF WEATHERING IN THE REGOLITH OF THE EARTH'S CRUST AND SUBSEQUENT SOIL-PROFILE DEVELOPMENT

The regolith was formed by the physical weathering or disintegration of massive rocks and minerals and by the chemical weathering or decomposition of some of the minerals in the parent rock. Rocks disintegrate or fracture into smaller and smaller particles due to alternate expansion and contraction resulting from heating and cooling, freezing and thawing, wetting and drying. Chemical weathering of rocks results from reactions of component minerals with oxygen and water and with carbonic, nitric, and sulfuric acids of atmospheric and biological origin.

Soils develop from the regolith by a continuation of physical, chemical, and biotic actions and reactions similar to those that originally formed the regolith and by a differentiation of the weathered regolithic material into a profile of horizons. For regolithic material to be classed as a soil, its surface must be differentiated from deeper material into from one to four possible master horizons designated from the surface downward as A, B, and C horizons and D layer. Master horizons are usually several inches thick (*see* fig. 2, 3 and 4). If minor differentiations occur within the master horizons, the subhorizons are designated by appropriate subscripts such as A_1, B_4, C_3, etc.

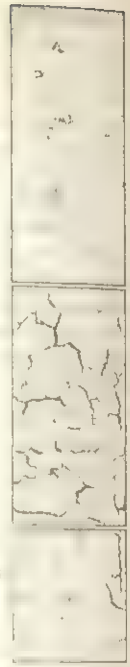


FIG. 2.—A TYPICAL SOIL PROFILE

Horizons become differentiated during soil development in the regolith because conditions vary with depth. In humid regions physical and chemical reactivity and biotic activity are much more intense in the surface than in subsurface materials. Dispersed and dissolved products are moved from the surface horizons to the subsurface horizons or they are leached completely from the developing soil profiles. In desert regions chemical and biotic soil-forming processes are much less intense than in humid regions, and the zone of greatest chemical and biotic activity is usually within one of the subsurface rather than surface horizons. There is seldom enough water in desert soils to move products of weathering from one horizon to a lower one. As a result of the above processes each horizon of a soil profile differs from the one above or below by one or more characteristics, such as colour, chemical composition, arrangement of individual particles, or size distribution of constituent particles.

The A horizons are the horizons of the surface layers of soil profiles. They contain more organic matter than the other horizons except in certain desert soils in which the production, preservation and accumulation of organic matter may be favoured somewhat more by conditions found lower down, in the B horizons. The A horizons are the most weathered and leached of the various horizons. In the soils of the cold humid temperate regions, weathering and leaching have left residues in the A horizons which are high in silica and low in iron and aluminum oxides and other sesquioxides (*see* also *The Chemical Properties of Soils* below). The opposite is true of the soils of the humid tropics in which soil-forming processes have left residues in the A horizons which are high in the sesquioxides and low in silica (*see* also Laterite).

The B horizons of the soils of the cold humid temperate regions may accumulate all or only part of the sesquioxides removed from the corresponding A horizons. In cases where only part of the sesquioxides has been deposited in the B horizons the remainder has been removed from the soil by drainage waters. Intermediate climates between the cold humid temperate regions and the humid tropical regions produce soils with A and B horizons having intermediate silica or sesquioxide residues and accumulations.

The humid temperate climates are conducive to the formation in the A and B horizons of a group of very finely divided hydrated iron and aluminum silicates called clay minerals. The B horizons of soils in these regions have been enriched by clay minerals carried by percolating water from the A horizons.

The C horizons of soils represent the types of material from which the A and B horizons were formed. These materials are

called the parent materials of the soils of which they are a part and their characters are determined by the rocks from which they were originally formed by weathering processes.

The D layers are beneath the soils and are not a part of the soil profiles. They are strata having properties much different from the overlying C horizons. They are strata, which, by virtue of their properties and shallow position, exert a significant influence on the properties of the soils above.

Some of the master horizons may be missing from soil profiles. The A horizons may have been removed from some soils by erosion. Other soils may be too young and too feebly developed to show B horizons. In still other soils C horizons may be absent because the original regolith was so shallow that it became completely occupied by the A and B horizons. In opposite cases where the regolith is extremely deep, D layers usually are not recognized because they do not influence the properties of the overlying soils.

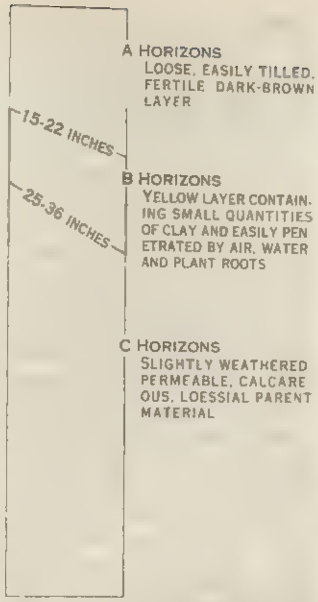


FIG. 3.—TYPICAL FEATURES OF PROFILES OF SOIL WITH HIGH PRODUCTIVE CAPACITY

Soils may also develop from thick layers of organic materials found in the bottoms of naturally or artificially drained swamps or lakes occurring in regions with cold temperate climates. These organic materials accumulated, before drainage, from dead water plants and other organisms. These organisms were only slightly decomposed due to relatively low temperatures and a lack of oxygen at the bottom of the swamps and lakes. After the organic sediments become exposed to air by drainage, they differentiate into horizons and develop into organic soils. The A horizons are first formed by partial oxidation of the organic compounds and slightly mineralized fractions are left as residues. As organic soils become older they become more and more mineralized to greater and greater depths.

The factors that influence the rate and kind of soil development include: the nature of parent material, climate, biological activity, topography or drainage, and the age of the developing soil profile. These factors, through combinations of many intensities, provide many environments and many possible variations in rate and kind of soil development. This can be illustrated by studies made on soils developed from regolithic materials deposited approximately 2,000, 10,000, and 200,000 years ago, respectively. The 2,000-year-old deposits are Indian mounds and the other two are continental glacial deposits, all occurring in humid temperate regions. The 2,000-year-old deposits show little if any soil-profile development; the intermediate regolithic materials have developed weak soil profiles; and the 200,000-year-old deposits have developed well-defined soil profiles. In contrast to this slow development in the temperate climates, in the tropics it is possible to find 50-year-old deposits of regolithic materials that have developed perceptible soil horizons.

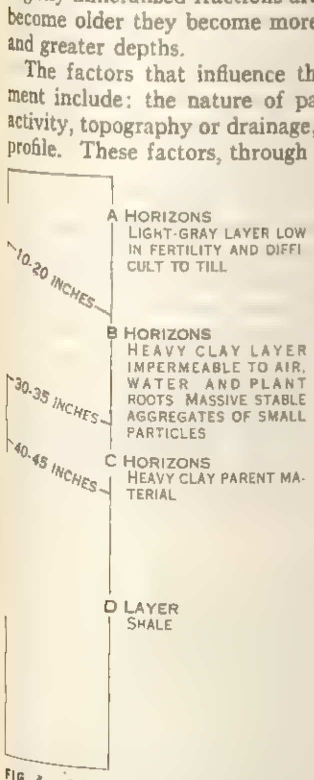


FIG. 4.—TYPICAL FEATURES OF PROFILES OF SOIL WITH LOW PRODUCTIVE CAPACITY

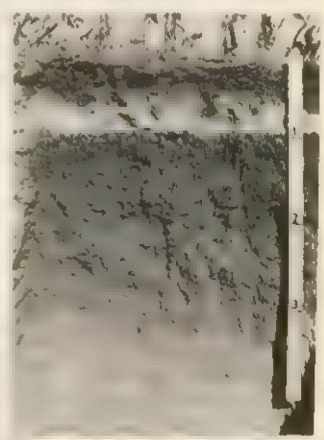
Variations in age, parent material, and environment have been responsible for the development of many kinds of soils on the earth's surface with many different physical, chemical, and biological properties and with wide ranges in productive capacities. These soil characteristics have been influential in determining the kind of native vegetation best able to survive in various parts of the world. They are also influential in determining the methods man must adopt in using and managing soils for crop production.

For more detailed discussion of the geologic processes involved in soil development *see* GEOCHEMISTRY; GEOLOGY: *Physical Geology;* HYDROLOGY; SEDIMENTARY ROCKS. (J. E. GG.)

II. SOIL CLASSIFICATION

Although soil covers the land as an almost continuous blanket, it differs greatly in character and in usefulness to man from region to region, from farm to farm, and even within single fields. Soil scientists, engineers, geologists, farmers, and others have named and classified different kinds of soil, each group according to its special knowledge and interest.

The early attempts to classify soil were based on its products rather than on its intrinsic properties, because men were interested in the fruits of the soil, not in the soil itself. Even educated people of the 18th and 19th centuries associated the soil with the lowly position held by those who tilled it, and few found it worthy of study. The infant study of geology was a reputable subject for scientists of those times, however, and European geologists were the first to make serious attempts to classify soil. To the geologists, however, the soil was merely a kind of disintegrated rock, and they classified it accordingly. This was the principal concept expressed in soil classification throughout most of Europe and the U.S. and Canada from about 1850 to 1920. Terms like granite soils, transported soils, and residual soils came from classifications by geologists.



BY COURTESY OF SOIL CONSERVATION SERVICE OF THE U.S. DEPARTMENT OF AGRICULTURE

FIG. 5.—A VERTICAL SECTION THROUGH A CLAYPAN SOIL SHOWING THE LAYERS DEVELOPED DURING SOIL FORMATION (SCALE IN FEET)

1. Development of Soil Science.—While soil classification was dominated by geologic concepts throughout most of Europe, a group of Russian scientists led by V. V. Dokuchaiev began to study the soil itself, not as decomposed rock but as a natural substance warranting special attention. They found that different kinds of soil had distinctive layers or horizons, and that the kind and arrangement of these horizons at a given place were determined by climate, vegetation, topography, and age as well as geologic material. Using this concept of dynamic soil formation, Dokuchaiev proposed the first classification of soil as an independent natural body in 1886.

The language barrier between Russia and the rest of the world prevented dissemination of the Russian concept of soil for many years. *The Great Soil Groups of the World and Their Development* (1908), by Konstantin Dimitrievich Glinka (*q.v.*) of the Leningrad Agricultural Institute, was translated into German in 1914 and, 13 years later, in 1927, into English. This work, when it finally became known, revolutionized soil classification throughout the world. Under its influence, European soil scientists emphasized concepts of soil development in their classification systems. The system of A. A. J. de Sigmond, published in 1933, was based largely on selected chemical properties acquired during soil formation. That of W. L. Kubiena, published in 1950 as *Diagnostics and Systematics of the Soils of Europe,* depended mainly on theories of the changes that occur in soils with time under different conditions. Most of the best-known studies dealt with the major genetic kinds of soil of large areas and little with the details of soil properties important within farmers' fields.

In the United States, as in Europe, the concepts of geology dominated soil classification until the Russian work became known. From 1860 to 1900, E. W. Hilgard developed a concept of soil similar to that of Russian scientists and published an accurate picture of soil properties, correlated with types of vegetation, later to become major bases of soil classification. Few understood Hilgard's concepts, however, and much of his knowledge went unnoticed.

The first major lasting contribution to soil classification in the United States came through the establishment of the soil survey in 1899. This work involved making maps on which areas of the different kinds of soils of farmers' fields were shown for predicting crop adaptation, productivity, and land values. Different kinds of soils, called soil series, were named after places where they were first found, such as Norfolk (Va.) and Hagerstown (Md.). Each series was divided on the basis of texture into units called soil types, which became the basic units of soil classification for practical application to the problems of fields and farms. Thus while the genetic kinds of soil of extensive regions were being classified as great soil groups in Russia, the soils of small areas important to the individual farmer were being classified as soil types and series in the United States.

2. Systems of Classification.—After the German edition of Glinka's work was published in 1914, C. F. Marbut, chief of the Division of Soil Survey of the United States, who translated the work into English, combined the Russian concept of great soil groups with the U.S. concept of soil types and series into a single system of classification useful for describing and naming soils of both extensive areas and individual fields. Marbut's system was presented to the First International Congress of Soil Science in 1927 and received worldwide recognition.

Knowledge about soils increased rapidly after 1930, and many schemes for classifying soils were proposed in different countries. In spite of modification of Marbut's system in 1938 and again in 1949, the soil survey staff of the United States Department of Agriculture under the leadership of C. E. Kellogg concluded that a completely new system was required to incorporate new knowledge and to eliminate the confusion of names and poorly defined classes that had developed. In 1952 the soil survey staff began the first of seven trial systems, each of which was circulated among leading soil scientists throughout the world for trial under their conditions and was revised to accommodate their findings. Linguists developed a terminology based on Latin and Greek and translatable directly into related languages. The seventh approximation was a system based on measurable soil properties rather than on theories of soil formation, although characteristics used to distinguish different kinds of soils reflected major concepts of soil development. The system was presented at the Sixth International Congress of Soil Science in 1960 in tentative form. It was adopted as the official system of the National Cooperative Soil Survey of the United States in January 1965.

In addition to the taxonomic classification of soil, mapping units of soil surveys were grouped in many ways to show such interpretations as crop adaptation, single properties, productivity, erosion hazard, and need for fertilizer or lime. The land capability system of the United States Soil Conservation Service, designed as a guide for planning use and management of farmers' soils, was one of the best known of many such interpretive groupings. Comparable but less-known groupings were developed for such special purposes as irrigation, drainage, and land appraisal.

The units of classification and mapping of detailed soil surveys have also been interpreted for engineering uses. In addition, engineers developed independent systems of classifying soil material according to texture and other properties important in engineering construction. Some confusion of names resulted because the engineering soil classifications deal mainly with soil as a material for construction (*see* SOIL MECHANICS), while those of soil scientists deal with it as a substance formed by natural processes.

(M. G. C.)

3. Soil Survey.—The purpose of a soil survey is to provide an inventory of the soil resources of a given area. First, the soils must be classified. This is done by making a field study at numerous locations of the surface and subsurface layers, samples of each being sent to the laboratory for physical and chemical examination. Next, a map is prepared showing the districts covered by the soil groups recognized. Finally, a written report is prepared giving general information about the area, descriptions of the soils, and suggestions for their use and management. Since 1935, soil surveyors have used aerial photographs in making soil maps.

In the United States, where soil surveys were begun by M. Whitney of the Department of Agriculture in 1899, the Soil Conservation Service of that department includes among its tasks responsibility for the National Cooperative Soil Survey. This produces maps and reports in collaboration with the state agricultural experiment stations and other state and federal organizations. Most United States surveys follow county boundaries, and include not only soil descriptions but land-capability groupings, types of agriculture, climate, and other significant information. In Canada, the soil survey is led by the Soils Research Institute of the national Department of Agriculture in association with the Experimental Farms service and the universities.

Soil maps based on texture (sandy, clayey, loamy, etc.) were included in the agricultural surveys of Great Britain made in the early 19th century, and some similar information was given on the early Drift maps of the Geological Survey. Chemical and mechanical analyses were included in the surveys made about 1900 by C. M. Luxmore and T. B. Wood and also by Sir A. D. Hall and Sir E. J. Russell, who based their *Report on the Agriculture and Soils of Kent, Surrey and Sussex* (1911) on the geological origin of the soils. These methods have been very generally replaced by mapping based on the forms of profiles (profile morphology), pioneered by Dokuchaiev in Russia. Surveys in Great Britain were established in 1939 and placed under the aegis of the Agricultural Research Council, London, in 1946. A soil map of Europe based on profile morphology was published in 1937 for the International Society of Soil Science, the coordinator being H. Stremme (Danzig); it was revised in 1964 under the auspices of the Food and Agriculture Organization of the United Nations (FAO).

J. A. Prescott (Waite Agricultural Research Institute) made the first soil map of Australia in 1931; much detail has since been added. In Africa, attention is concentrated on agriculturally important land, information from aerial, ecological, and reconnaissance soil surveys being used to fill the intervening spaces. G Milne made the first soil map of East Africa in 1936.

Basic information from soil surveys is important in planning land use in developing countries, and is provided by such organizations as FAO (Rome) and the Ministry of Overseas Development (London).

(ER. W.; E. J. R.; D. A. O.)

III. SOIL TESTING AND ANALYSIS

In agriculture, relatively simple soil analyses, commonly referred to as soil tests, are used extensively to evaluate nutrient availability in soils from a particular field or an individual farm and to predict whether increases in crop yield or quality will result from applications of lime or fertilizer. The development of these simple tests occurred largely after 1930. Previous attempts to relate nutrient availability in soils to the total amount of a nutrient element present in the soil or to that present in the soil water were not successful. Only when testing procedures were developed to remove and measure certain readily available portions of the nutrient elements present could definite relationships between the soil-test results and plant availability of the nutrients be established.

Following World War II interest in soil testing increased rapidly. The United States, most of the European countries and many other countries developed or expanded soil-testing laboratories to analyze soil samples from farmers' fields and advise the farmers as to the lime and fertilizer practices best suited for their soil conditions. While the majority of the testing is done for field soils, some laboratories make special tests for soils used in greenhouses or for other special uses. Most soil-testing laboratories are operated by agricultural colleges or universities or by governmental agencies. However, commercial laboratories, especially in the United States, offer this service to farmers.

Individuals may buy small soil-testing kits to test their own soils although this procedure is not generally recommended by soil scientists.

The first step in soil testing is the collection of the sample in the field. If soil conditions in the field vary because of different types of soil or different previous treatments, each condition should be sampled separately. Usually 15 or more small samples from a uniform area are collected and thoroughly mixed to provide one representative sample.

In the laboratory different tests are made for different nutrient elements. The amount of lime that should be applied is usually estimated from the pH (a measure of the acidity) of the soil sample. While tests for phosphorus and potassium are made almost universally, the testing procedures used in different laboratories are extremely varied. Solutions of acids, bases or salts of various concentrations or mixtures of these reagents may be used to extract the phosphorus or potassium from the soil. Then the phosphorus or potassium in the extract is determined. Certain reagents may provide reliable tests of nutrient availability in some soils but not in others, so different reagents are used in different laboratories.

Chemical soil tests for nitrogen are less commonly used. Most of the nitrogen used by plants is made available by microbiological decomposition of organic matter. Tests to measure the rate at which nitrogen is released from the organic matter have been developed but have not been widely adopted. Some laboratories test for calcium and magnesium in soil extracts similar to those used for potassium. Many soils contain soluble salts in amounts that are detrimental to plants. Most laboratories test for soluble salts by measuring the electrical conductivity of a soil-water suspension or of a water extract of the soil.

The boron removed in hot-water extracts of soil samples is determined in some laboratories in areas where boron deficiencies occur. Adequate soil tests for most other minor elements, such as manganese, copper, zinc, molybdenum, iron, etc., have not been developed or used extensively.

For advising farmers on the use of lime and fertilizer, the soil-test results must be interpreted differently for different crops, soils, and environmental conditions. Laboratory tests can provide only a relative measure of nutrient availability in the sample that is tested, not the actual amount of a nutrient element that will be available to plants during a growing season. Furthermore, different plants have different nutrient requirements. Various environmental conditions and soil properties influence nutrient availability to plants growing in a field, but do not influence the soil-test results. Because of this, the soil-test results must be interpreted differently for these different conditions. The interpretation must be based upon relationships developed between the soil-test results and the results obtained in fertilizer experiments conducted in the field under each of the different conditions.

Where adequate relationships have been developed, soil tests are useful guides in advising farmers on the use of lime and fertilizer. Much research is continually devoted to the development of better testing methods and to improving the relationships between the laboratory and field results for various soils and crops.

(Jn. J. H.)

IV. PHYSICAL NATURE AND PROPERTIES OF SOILS

The soil, in its relation to plants, provides a medium for root development and supplies nutrients for plant growth. To a considerable extent, the ability of soils to perform these functions is dependent upon their physical character and condition, which have much to do in determining moisture, air, and temperature relationships. Hence, the physical make-up of soils, together with their resultant properties, exert a controlling influence on chemical reactions and biological processes.

1. Component Parts of Soils.—Soil materials in a solid state—soil minerals and organic matter—constitute the soil proper; liquid and gaseous components, namely water and air, in the interstices between the solid materials, fluctuate continually, and frequently with conditions independent of the soil. The organic fraction varies widely, but seldom exceeds five per cent by weight, except in soils formed largely from plant remains. Organic soils, however, are of secondary importance to mineral soils.

The mineral constituents are extremely variable in size, shape, and chemical composition. The size of particles is one of the most significant characteristics. Water absorption, air movement, rate of solution, and ease of tillage are but a few of the numerous things affected by particle size. By mechanical analysis, which separates the particles into size groups known as separates, a quantitative statement of soil texture, illustrated in Table I, can be obtained. Results of a mechanical analysis do not include either organic matter or particles larger than two millimetres.

Table I.—*Quantitative Statement of Texture by Mechanical Analysis*

Name of separate	Diameter limits (mm.)	Amount present (percent)
Very coarse sand	2.0 –1.0	2.3
Coarse sand	1.0 –0.5	6.7
Medium sand	0.5 –0.25	8.2
Fine sand	0.25–0.10	9.4
Very fine sand	0.10–0.05	12.5
Silt	0.05–0.002	44.8
Clay	Below 0.002	16.1

Each separate is characterized by definite physical properties, due chiefly to the size of the particles composing it. As the particles become smaller, their total surface area, hence their effectiveness in transmitting properties to soils, increases greatly. Particles less than 0.001 mm. in size, especially those making up the colloidal or gluelike portion of the clay, possess active internal as well as external surface, and therefore are much more reactive than the other clay and larger particles. Because of their cohesive nature, clay colloids play an important role in the arrangement of particles.

Every soil is a mixture of separates, hence the physical properties exhibited are determined by the proportions of separates present. A group of soils having the same range in particle size and physical properties based on texture constitute a soil class. Soils in which the properties of no one separate predominate belong in one class. This class, known as loam, contains more than 35% total silt and clay with less than 50% and 27% of silt and clay, respectively; the permissible amount of sand would be too low to exert a dominant influence on the physical properties exhibited. When the silt content exceeds 50% in a mixture with less than 27% of clay, the properties of silt predominate and the soil is classed as a silt loam. The transition from loam to silt loam, or between other soil classes, is not abrupt but gradual. Variation in properties of soils within a class is indicated by descriptive terms as light, friable, and sandy; or heavy, plastic, and clayey.

For practical purposes, soils of the common classes can be grouped as follows: coarse textured soils, the sands and sandy loams; medium textured, the loams and silt loams; fine textured, the clay loams and clays.

Even though mineral particles are the very foundation of soils, they always contain some organic matter; its accumulation is an integral part of soil development. Humus, the partially decayed organic matter accumulated in soils, is a dark-coloured, structureless material, highly colloidal in nature. Unlike colloidal clay, the organic or humus colloids are not very cohesive and plastic; in general they tend to make sandy soils more cohesive and clayey soils less plastic. The efficacy of organic matter in changing soil properties depends on the amount present and the textural make-up of the soil. Thus, in determining soil class, it is necessary to take cognizance of the influence of organic matter in modifying physical properties associated with the mineral particles.

2. Structure and Its Practical Significance.—Structure deals with the arrangement of soil particles which occur as individuals, as groups or aggregates, or as a mixture of the two. In coarse-textured soils, the particles exist and function largely as independent units, but in fine-textured soils different-sized particles are bound together with colloidal materials. The aggregate produced may be either a dense, continuous mass, exemplified by a clod, or an open, porous cluster typified by a granule. With greater pore space and effective surface, the granule represents the more

desirable structure. These aggregates also vary in size, shape, and stability. A composite of the single grain and aggregate structures occurs in medium-textured soils. These are neither too open and friable, nor too compact and plastic; their physical condition is generally favourable for plant growth, and is easily maintained.

Structural differences result from variations in the kind and amount of colloidal material present. In the absence of soil colloids, the single-grain arrangements predominate, while aggregate formations are never well developed. These colloidal materials, including various clay minerals, oxides and hydrates of iron and aluminum, together with soil humus, are the products of different chemical and biological processes involved in soil development. Factors and conditions influencing soil development naturally would affect not only the formation of colloids, but also their distribution and behaviour in soils. As a result, marked differences in the natural structure of soils have been developed. Structural differences of virgin soils, also of the undisturbed substrata in cultivated soils, are of great value in any scientific study and classification of soils; too, they are of practical significance because of their relation to drainage, aeration, root development, moisture retention, and soil erosion.

When farmed, many soils soon lose much of their original aggregate structure. This change occurs primarily in the surface layer due to more rapid organic matter depletion, together with excessive and often improper handling. But, unlike texture, structure can be and frequently is changed; in fact, the main purpose of most tillage operations is to improve structure. Any control over many of the properties of a soil is possibly only if its physical condition can be modified.

3. Porosity and the Weight of Soils.—Porosity means total pore space, the soil volume not occupied with solid materials, varying from 35% to 50% in dry soils. As texture becomes finer, the size of individual pore spaces decreases but total pore space increases. Any change in structure, or organic matter, also affects the size of pores and porosity. Both are reduced by compaction, often desirable in sandy soils. Porosity is increased in fine-textured soils by proper tillage and more organic matter because of their influence in developing a granular structure in silty and clayey soils.

When compared on an equal volume basis, mineral soils when dry weigh more than water; organic soils, however, weigh less. The numerical ratio between the weights of equal volumes of dry soil and water represents the bulk density (or volume weight) of the soil. This fluctuates chiefly with differences in porosity, but is influenced also by the particle density (or specific gravity) of the various constituents in different soils.

Particle density is also a comparison of the weights of equal volumes of dry soil and water, but differs from bulk density since the equal water volume is determined by displacement and represents soil volume without pore space. Particle density is practically a constant for most mineral soils, with values of 2.65 to 2.70. Dark-coloured soils often are lighter, due to more organic matter; organic soils have particle density values from 1.5 to 2.0, varying with their mineral content. In comparison with bulk density, particle density of soils is relatively less important. Bulk density usually ranges from 1.0 to 1.5 for fine-textured soils, and from 1.2 to 1.7 for coarse-textured soils. Changes in physical condition of the surface layer of soils under cultivation cause wide variation in bulk density. In all soils it naturally tends to increase with the deeper layers, especially the subsoil.

When the bulk density is known, the weight of dry soil per cubic foot, or per acre foot, or of any fractional part, can be computed readily. The average weight of dry soil per acre 6½ or 7 in. deep, often called the "plowed section," is commonly given as 2,000,000 lb. Sandy soils are somewhat heavier, while organic soils weigh only about 1,000,000 lb. per acre 7 in. deep. These weight figures, either the averages or those computed directly from the bulk density, are useful in calculating the amount of soil water, the supply of organic matter or the quantities of plant-food elements present.

4. Soil Moisture and Its Relation to Plants.—Soil moisture, as here discussed, is exclusive of the water in chemical combinations occurring in certain soil minerals and organic compounds. The relative position of moisture with reference to the soil particles means that it is attracted unequally by the particles and acted upon by different forces. In consequence, soil moisture shows marked differences in its movement and availability to plants.

When air-dry, a soil usually contains only hygroscopic moisture; that is, moisture which adheres to soil surfaces in very thin films, moves by diffusion and is unavailable to plants. Colloidal materials are chiefly responsible for hygroscopic capacity. Hence, the capacity is never great in sandy soils, which are usually deficient in organic matter and have a low clay content.

Moisture, in excess of hygroscopic capacity, is retained in soil against gravity through the combined forces of adhesion, cohesion, and surface tension. This moisture represents the capillary form. It exists as films around and wedges between particles; moves as a film due to surface tension; and is available to plants. The innermost portion of capillary moisture, however, is held so tenaciously as to preclude true film movement, or its use by plants. Thus the soil contains both hygroscopic and inner capillary moisture when permanent wilting of plants growing on that soil occurs.

Gravitational water, known also as free, excess, or surplus water, is extremely variable, but is, in general, regarded as the difference between precipitation and a soil's capacity to absorb and hold water. Its presence in soils, contingent upon adequate drainage and a water supply in excess of capillary capacity, causes a deficiency of soil air and free oxygen, resulting in conditions which are inimical for seed germination, root development, and other desirable soil processes. Ordinarily, gravitational water is more important as a reservoir to replenish useful capillary moisture than as a direct source of water to plants.

The ability of soils to hold capillary water is determined by their porosity and adsorptive power. As the particles become smaller, or the organic content higher, or the structure more granular, the capillary capacity increases, due chiefly to the influence of clay or humus colloids, and or of the fact that more effective surface is exposed. The moisture content of soils, at any particular time, represents the balance between additions and removals of water. Natural precipitation, the main source of soil water, is supplemented in certain areas by irrigation, and at times by water brought up from deeper soil layers by capillarity; however, this process is generally too slow to be of real significance in supplying the water needs of plants. Removals of water from soils result through evaporation at or near the soil surface, transpiration from plant surfaces, percolation downward due to influence of gravity, and surface runoff. Loss by runoff decreases the potential supply of soil water.

Since water is often the chief limiting factor, even in humid regions, for crop production, every effort should be made to increase and conserve the supply of available soil water. Cultural and tillage practices that increase water absorption reduce surface runoff; moreover, these same practices control weeds, thus eliminating useless transpiration. Evaporation losses may be reduced through the use of windbreaks, and of mulches which may consist of either a protective cover of straw, boards, plastic or other materials, or a layer of dry soil. The soil mulch, however, is effective in saving water only when the water table is within capillary reach of the surface; but under this condition, drainage to remove the excess water usually is more urgent than mulching to conserve water.

Of the soil water absorbed by plant roots, only a small fraction is retained by the plant; an enormous quantity is removed from it by transpiration. This varies widely with different plants and climatic conditions, and is controlled indirectly by securing efficient use of the water transpired. Control on many soils, therefore, may mean the use of fertilizers to meet some nutrient deficiency, or it may mean other practices such as drainage, liming or proper tillage to improve productivity. Thus the transpiration per unit of dry matter produced is decreased. (*See* also IRRIGATION.)

5. Soil Air and the Need for Aeration.—The pores in soils are occupied by air or water or both, the proportion of air to water decreasing rapidly as the particles become smaller. Differences in texture, structure, and organic matter affecting porosity of soils also influence their capacity to hold water. Any increase in soil water decreases the amount of soil air. Thus, in poorly drained soils, and to a lesser extent in fine-textured soils having a compact rather than a loose condition, a deficiency of air exists.

Soil air, like the atmospheric air, consists of a mixture of gases, but it contains considerably more carbon dioxide and slightly less oxygen than does atmospheric air. With the gradual, but continual, depletion of oxygen and accumulation of carbon dioxide due to chemical and biological activities in soils, the need for soil ventilation becomes more urgent.

Aeration, the process by which stagnant soil air is replaced with fresh air, occurs naturally as a result of changes in temperature, variations in barometric pressure, movement of soil water, and other processes. In practice, however, one of the chief controls over aeration is exerted by regulating soil-moisture conditions, particularly through drainage. Tillage may be used either to lessen or promote aeration since sandy soils, usually too well aerated, need to be compacted, whereas clayey soils, frequently too compact and often crusted over, should be stirred to facilitate aeration.

6. Soil Temperature: Its Importance and Control.—Seed germination, plant growth, and, in fact, all soil reactions proceed through a wide range in temperature but each has an optimum range where it is carried on most efficiently. Successful crop production is no less dependent upon proper soil temperature than upon suitable atmospheric temperatures.

Direct solar radiation is the chief source of heat to soils. Differences in texture, structure, and organic matter that determine the moisture capacity of soils influence their ability to absorb and transmit heat. Dry soils, due to air in pores, have higher resistance to heat transfer than moist soils. Water is a better heat conductor than air, hence as soil air is replaced by water as the contact between particles, conductivity is increased; additional water, however, would retard the rate.

Heat generally moves from the surface downward as atmospheric temperature increases and from lower soil layers upward as it decreases. Changes in soil temperature lag considerably behind and are much less variable than those for atmospheric temperatures. Daily fluctuations in soil temperature are much greater in summer than in winter.

Differences in temperature of soils for the same degree of latitude are caused by variations in colour, moisture content, vegetative cover, and direction of exposure. The relation of colour to heat absorption is well known; but darker colours in soils are due chiefly to organic matter, which increases the amount of soil water. Only when the surface becomes dry, would the dark-coloured soil likely be warmer than the light-coloured soil.

Dry soil has a low specific heat, only one-fifth that of water. The greater amount of heat needed to raise the temperature of water causes soil temperature to decrease as soil water increases. Thus, wet soils are cold and slow to warm up in the spring.

A bare, uncropped soil heats more rapidly, attaining a higher temperature in summer than a similar soil protected by vegetation. The vegetation acts as an insulator, reducing both absorption of heat and loss through radiation. When protected by vegetation, leaves, or snow, heat is retained better and the soil remains warmer during the winter. In the Northern Hemisphere soils with southern exposure, receiving more heat per unit area, have higher temperatures than those with northern exposure, provided other conditions are similar.

Control of soil temperature is a matter of soil-moisture control. Excess water must be removed, if possible, by drainage. This will reduce evaporation, thereby conserving heat, and lower specific heat thus decreasing the amount of heat necessary to raise the soil temperature.

7. Soil Colour and Its Significance.—Colour provides information as to the conditions of and forces active in soil development, and represents an invaluable aid in soil classification. Since different colours are indicative of variations in conditions which influence productivity, colour is used as an important indicator of the agricultural value of soils. This use, however, necessitates that colour throughout the entire vertical section of soil be observed and interpreted in terms of soil properties affecting crop production.

Soil colours are of three main sources: decomposed organic matter, certain iron compounds, and other soil minerals such as quartz, kaolin and mica. Soils low in organic matter with iron absent or unoxidized are coloured by soil minerals. Their colours are often dulled or concealed by organic matter or iron compounds, or both. Much of the humus, the most stable portion of the soil organic matter, is a dark-brown or black residue coating the mineral soil particles. Under conditions of poor drainage and limited aeration, the reduced forms of iron together with anaerobic decay of organic matter favour the development of gray and drab coloration in soils.

Without good drainage and aeration, the iron could not be oxidized, nor the hydrated iron oxides produced to give the red and yellow colours. Red- and yellow-coloured soils are invariably low in organic matter; but when these colours, due mainly to hematite and limonite, are mixed with black from humus various shades of brown result.

With few exceptions, soil colour involves mixtures and is concerned with tints and shades rather than pure colours, so that ordinary colour charts are of little or no value in determining colour in soils. For this purpose standard soil colour charts, specifically made to include the colour range encountered in soils, are generally used.

8. Soil Tilth and Tillage.—The physical condition of the soil when considered in its relation to plant growth is known as tilth; its development must necessarily consider the requirements of the crop to be grown. Soil tilth, therefore, is more inclusive than structure, although the arrangement of the individual or aggregate structure particles is the fundamental basis of tilth in soils. While the textural make-up of soils cannot be changed practically, it is feasible to modify, through tillage practices, the structural condition of the surface layer.

Tillage, including any operation that stirs, inverts, fines, aggregates, or firms the soil, is used not only to prepare a satisfactory seedbed but also to maintain optimum growth conditions throughout the season. While plowing is generally regarded as the basic tillage practice, many soils, especially those of coarse and medium texture, often can be prepared adequately by use of a disk harrow. Frequently much of the desirable aggregate structure in the finer-textured soils is destroyed by plowing too much, or when too wet. It is imperative to avoid all tillage practices on heavy, plastic soils when wet; otherwise the soil runs together and upon drying is in a cloddy condition, which may persist for several seasons. Volume changes which accompany moisture variations due to alternate wetting and drying break down the clods and tend to encourage the formation of granules; this effect is usually more pronounced in the presence of organic matter since it counterbalances the deleterious effect of an excessive amount of clay colloids on structure. The general lack of colloidal materials in coarse-textured soils prevents puddling, or even desirable aggregation, hence the possibilities of structural modifications are limited.

Cultivation, the intertillage given a crop, is used primarily to destroy weeds. This practice also modifies structure of the soil to the depth cultivated, thus affecting aeration, water absorption, and evaporation. Conservation of moisture due to the formation of a mulch is of secondary importance to that due to the control of weeds. Furthermore, when cropped, the top portion of the surface soil is more valuable as a feeding zone for plant roots than as a soil mulch. (D. C. W.)

V. CHEMICAL PROPERTIES OF SOILS

The upper portion of the earth's crust, to a depth of a few miles, has been estimated to consist, to the extent of about 99%, of 10 chemical elements. They are, in decreasing order of abundance, oxygen (about 46.6%), silicon (27.6%), aluminum (8.1%), iron (5.1%), calcium (3.6%), sodium (2.8%), potassium (2.6%),

magnesium (2.1%), titanium (0.6%), and phosphorus (0.12%). All of the remaining elements, some 85 that are found in soils, constitute less than 1% of the earth's crust. Oxygen, the most abundant element, makes up by volume over 90% of the rocks and minerals, mainly as oxides of silicon, SiO_2, and aluminum, Al_2O_3. The 10 elements mentioned above are often referred to as the macroelements while the others like boron, cobalt, copper, manganese, molybdenum, and zinc are considered microelements. The primary rocks and minerals formed during the cooling of the magma are dominantly composed of the macroelements as alumino- and ferro-silicates of magnesium, potassium, calcium, and sodium. The microelements are most often found as occluded materials, generally as oxides, within the igneous rock formations. Igneous rocks account for approximately 95% of the earth's crust. Sedimentary materials are responsible for the remaining 5%. (*See* also ELEMENTS, CHEMICAL; GEOCHEMISTRY; SEDIMENTARY ROCKS.)

In the weathering of rocks containing primary or unaltered silicates, hydrolysis plays a dominant role. In hydrolysis, or decomposition by the union with water, the bases are split off as soluble compounds, such as hydroxides, carbonates, and bicarbonates. These, in a humid climate, are carried downward and out of the soil by percolating water, unless stopped by other chemical reactions which change them to less soluble forms. The insoluble residue from this weathering of the primary silicates is modified into secondary silicates of aluminum and iron which contain lesser amounts, or in some cases none, of the strongly basic elements, calcium, magnesium, potassium, and sodium. These secondary silicates are the clay minerals of soils, the principal component of the clay portion.

In addition to the silicate clay minerals, there occur in the clay size-range simpler compounds, the hydrated oxides of silicon, aluminum, and iron, that have split off from the primary minerals in the course of weathering. Along with these changes, organic matter accumulates from partial decomposition of plant and animal materials, so that the chemical components of the soil may be grouped in a general way into organic substances and inorganic or mineral substances.

In the following discussion, the chemical characteristics of the surface soils of the humid and semihumid temperate regions will be considered. These are the regions that contain most of the earth's cropped land and that support most of the world's human population. A brief discussion will then follow, indicating variations in surface soil character in some other climatic regions.

1. Primary Minerals.—The mineral constituents of soils may be grouped on a basis of size of particles. Sand and coarser particles, together with the coarser part of the next finer grade, silt, are essentially primary minerals. The silt size is between 2 and 50 microns (μ) in diameter ($1\ \mu = \frac{1}{1{,}000}$ mm. or $\frac{1}{25{,}000}$ in.). Primary minerals are defined as those which have undergone no appreciable weathering change. In soils they include, among others the feldspars, the pyroxenes and amphiboles, the micas, and quartz. The silts also contain the simple oxides of the microelements, apatites, and carbonates if they are present.

The primary minerals generally account for approximately 70% of the soil mass although individual soils may vary widely in this respect. Most of these minerals are very resistant to weathering, or decomposition, undergoing profound chemical change only through geologic periods of time. The primary minerals and that portion of the soil materials classified as silts, or coarser, contribute little to the dynamics of soil chemistry.

2. Secondary, or Clay Minerals.—Below the silt size-range is the clay fraction ranging in size from 2 μ down. The clay portion for the most part is composed of complex aluminum and iron silicates derived from the weathering of rocks and primary silicate minerals. The secondary or clay minerals, particularly those below the 1 μ size-range, are colloidal in size and present an enormous surface area for any given mass. Their reactions, therefore, involve surface energy and surface properties. In soils the soluble, chemically active substances accumulate at, and usually react with, the surfaces of these particles resulting in a great intensification of chemical processes around these minerals. Soil chemistry is essentially colloid, or surface, chemistry, and is controlled and conditioned by the clay minerals.

The structure of the clay minerals determines to a large degree their reactions in soils. Their basic structure centers around two crystal structural groups: (1) the silicon tetrahedron, a symmetrical configuration where a silicon atom is equidistant from 4 oxygen, or hydroxyl, atoms; and (2) the aluminum octahedron, a symmetrical configuration where an aluminum atom is imbedded in, and equidistant from, 6 oxygen, or hydroxyl, atoms. Under certain conditions a trivalent aluminum atom may replace the tetravalent silicon in the tetrahedron. Similarly divalent magnesium or iron often replaces the trivalent aluminum in the octahedra. In some clay minerals the lattice, or repeating structural sheets that make up the crystal, is composed of 2 silicon tetrahedra layers joined together by an aluminum octrahedron layer. Such minerals are called 2:1 lattice minerals. In others the lattice is made up of one silicon tetrahedron layer and one aluminum octahedron layer. These are called 1:1 lattice minerals. When ion replacements occur in the lattices such as aluminum for silicon or magnesium and iron for aluminum, a positive charge deficit is created in the mineral's crystal structure. The clay minerals are, therefore, negatively charged and may be considered as solid polyvalent anions. Most clay surfaces will also have "spots" of positive charge due to broken chemical bonds. In effect, parts of the surface may act as anions while other, much smaller, parts may act as cations.

The most common clay minerals found in soils are montmorillonite, illite, kaolinite, vermiculite, and chlorite. Montmorillonite is a 2:1 lattice mineral whose successive lattice sheets are bonded with varying amounts of oriented water molecules. Soils containing this mineral tend to shrink and crack when dry and become slick and sticky when wet. Montmorillonite is the principal clay mineral found in "gumbo" soils. Illites are 2:1 lattice minerals whose lattice sheets are bonded with varying amounts of potassium. The mineral is essentially nonswelling and imparts excellent tilth to soils. Illites can absorb, or "fix," both potassium and ammonium ions in forms that plants cannot feed upon. Kaolinite is a 1:1 lattice type mineral, nonswelling, and imparts good tilth to the soil. It has the highest anion exchange capacity of all the clay minerals. Both vermiculite and chlorite are 2:1 lattice type minerals but differ in the way the lattice sheets are held together. While widely distributed in soils, they seldom dominate the other clay minerals.

3. Ion Exchange.—The soil property known as ion exchange resides chiefly in the clay minerals. Ions are atoms or atom clusters that bear an electric charge. Those with a positive charge are cations while the negatively charged ions are anions. Anions and cations tend to unite with each other forming chemical compounds known as salts. When a salt dissolves in water a portion dissociates, or separates, into the constituent ions. The sum of the positive ion charges always equals the sum of the negative charges. The clay minerals are essentially solid surfaces with many negative charges and only a few positive charges. These must be balanced by adsorbed ions of opposite charge and attached through ionic bonds to the clay surface. The adsorbed ions can be removed and replaced by other ions of similar charge giving rise to the phenomenon known as ion exchange. Since the clay minerals are highly negatively charged, the adsorption of cations, and cation exchange, dominate their behaviour. A cation exchange reaction may be illustrated as follows:

$$\text{Clay} = \text{Ca} + 2\,\text{KCl} \rightleftharpoons \text{Clay} \langle^{\text{K}}_{\text{K}} + \text{CaCl}_2$$

Adsorbed ions are not removed from the soil by pure water and are, as a consequence, protected in the soil against excessive loss in drainage waters through leaching. This fact is of great importance in the practical maintenance of soil fertility because large quantities of available plant nutrient ions are thus preserved in the soil. The cations, however, are easily and instantaneously replaced from the clay by other cations in the soil solution. Not all cations are retained, or adsorbed, by clays with equal force. In general, the smaller the ionic radii of the ion and the greater

its charge, the more strongly it is adsorbed by the soil. The tightness with which cations are held by the soil, in decreasing order, is as follows: hydrogen > calcium > magnesium > potassium > sodium. Their relative abundance in the exchangeable form in mature soils of the humid regions is usually in the same order. In general, the laws of mass action govern the distribution of cations between the clay surfaces and the soil solution.

The capacity of different clay colloids and the soil itself to adsorb and hold cations is a fairly definite quantitative value, usually expressed in milligram-equivalents per 100 grams of sample, abbreviated m.e. The cation exchange capacity of pure clays may range from 10 to 200 m.e., while that of soils will usually vary from about 2 to 40 m.e.

Anions like phosphate, borate and molybdate may be adsorbed on the clay surfaces where positive charge sites occur. Anion exchange capacity in soils is usually very small as compared to cation exchange capacity, and is extremely difficult to measure quantitatively. Anions like the nitrate, chloride and sulfate are not adsorbed at all by the clays and are easily leached from the soil. The loss of cations from a soil through leaching can occur only if the soil solution contains salts or acids or anions that are not adsorbed by the clay minerals.

4. Soil Reaction and Acidity.—Inasmuch as acid hydrogen cations, as well as metallic or basic cations, can be adsorbed by the clay minerals, it follows that a hydrogen saturated soil is acid. The degree of acidity depends mainly on the ratio of exchangeable hydrogen ions to basic ions held on the clay surfaces, and on the hydrated aluminum oxide, or aluminum hydroxide, gels in the soil, which have both slightly acidic and slightly basic properties. Soils in humid regions gradually become more acid whether under natural or cultivated conditions, because calcium and other cations are gradually replaced from the colloids and removed from the soil by leaching. The replacing ion is chiefly hydrogen of carbonic acid and other acids from decomposing soil organic matter. The intensity of soil acidity is expressed by the term pH followed by a numeral. (*See* Hydrogen Ions.) The pH value of soils varies with the degree of saturation of the exchange capacity with bases, in contradistinction to hydrogen. Thus the base saturation of neutral soils (pH 7) varies around 75%, the remaining 25% being hydrogen. Virgin soils, *i.e.*, those that have never been cultivated, in humid temperate regions range from about pH 4.8 (strongly acid) to slightly alkaline, approximately 8.3, the latter value representing equilibrium with solid calcium carbonate. In some arid soils higher pH values are encountered, the high alkalinity being due to soluble carbonates of the alkali metal, sodium (see *Alkali and Saline Soils* below).

Soil reaction is important agriculturally because crop plants, as well as others, vary widely in their requirements and tolerances to soil acidity. Alfalfa and many clovers require essentially neutral soils, having a minimum pH of 6.2 to 6.5. Cranberries, at the other extreme, prefer acid soils with pH values around 4. The acidity, or hydrogen, in acid soils is easily neutralized by applying limestone. Neutralization of the hydrogen occurs according to the following reaction: $Clay = H_2 + CaCO_3 \rightarrow Clay = Ca + H_2CO_3$. Since carbonic acid is not stable it decomposes, $H_2CO_3 \rightarrow H_2O + CO_2$, giving rise to water and carbon dioxide. The soil hydrogen is thus converted to water, a neutral product, and removed from the soil.

5. Organic Matter.—All soils on which vegetation has occurred will contain varying amounts of organic matter. It is formed from, and composed of, plant, animal, and microbial tissues and their decomposition products. The stable portion of organic matter, or that which has undergone decomposition to a stage where the original source material is indistinguishable, is called humus. Humus is composed of a series of highly reactive brown to black coloured substances, usually of high molecular weight and containing nitrogen. The dark colour in soils, which is often associated with high productivity, is usually due to their humus content. As much as $\frac{1}{3}$ to $\frac{1}{2}$ of the humus may be humic acid, an alkali-soluble, acid-insoluble material. The remaining material is made up of amino acids, 10–25%, carbohydrates, 10–30%, fats, waxes, and resins, 2–6%, and other miscellaneous organic products.

Humus is quite resistant to decomposition and decay in soils. Many of its organic constituents are polar, or charged, thus adding to the soils ion exchange properties, especially the cation exchange capacity. Humus reacts with divalent cations, especially Ca, Mg, Mn, and many of the microelements, to form chelated compounds and salts. It also reacts with the soil clay minerals and becomes adsorbed on the clay surfaces. Such compounds, along with the gels, gums, and certain polysaccharides formed during the first decomposition stages of fresh plant residues, form protective coatings around the clay particles. These organic coatings decrease the stickiness of the clay particles and cause them to form stable aggregates. This gives rise to soils with excellent structure, and physical characteristics conducive to easy tillage. Soil tilth and resistance to erosion are closely related to the amount of organic matter the soil contains.

Soil organic matter contains about 5% nitrogen, 1% phosphorus, and 0.5% sulfur. All of the nitrogen in surface soils, except for a very small amount of ammonium held as fixed nitrogen in the illitic clay minerals and small amounts of unadsorbed nitrates present in the soil solution, resides in the organic matter. Over 30 amino nitrogen compounds have been definitely identified in humus, of which aspartic acid, serine, threonine, glutamic acid, glycine, alanine, valine, lysine, and the leucines are probably the most important. The nitrogen originally came from the air, first becoming incorporated into leguminous plants through symbiotic microbial fixation and later becoming a part of the humus through legume decomposition. (*See* Nitrogen, Fixation of.) If no fertilizer, or synthetically fixed, nitrogen is added to a soil, its ability to supply nitrogen to growing crops depends largely on the quantity of legumes grown and the continued decomposition, and loss, of the original soil organic matter supply. Although as much as 50% of the total phosphorus in surface soils is organic, such as phytin, the year-to-year quantity released during decomposition is relatively unimportant as a source of phosphorus for crops. Much of the infertility and poor tilth in soils that have been cropped for ages may be associated with their lack of soil organic matter.

6. Plant-Nutrient Supplying Ability.—Plants require for normal growth and reproduction the elements carbon, oxygen, hydrogen, nitrogen, phosphorus, sulfur, potassium, calcium, and magnesium in variable but fairly large amounts. In addition, iron, manganese, zinc, copper, boron, and molybdenum are required in small amounts, and are designated as trace or micronutrients. For some crops sodium, chlorine, cobalt, and vanadium may be required.

With the exception of carbon, hydrogen, and oxygen, the bulk of the plant nutrients are absorbed through the root system. Most of the essential nutrients are absorbed by plants as simple cations with the exception of phosphorus as phosphate, nitrogen as nitrate, boron as borate, sulfur as sulfate, and molybdenum as molybdate which are absorbed as anions. The soil solution seldom contains any of these ions in amounts adequate for normal plant growth. The exchangeable form of the basic elements is the principal immediate source from which the nutrients in the soil solution are renewed as they are taken up by plants. The exchangeable form also can be absorbed by plants directly through physical contact between the root hairs and colloid particles, forming an interface at which the hydrogen on the root can be directly exchanged for a cation on the colloid. An analogous system exists for the uptake of adsorbed anions by plants.

The clay minerals contain basic elements, *i.e.*, potassium and magnesium, in their internal structure, in addition to those in the exchange form. Illite is such a clay mineral rich in potassium. These forms are the reserve from which the more available exchange forms are renewed in the soil, and their abundance is a measure of the durability of chemical fertility. An analogous situation may exist with regard to phosphates.

Chemical soil tests for the plant-available forms extract and measure the adsorbed (exchangeable) bases and anions as well as the water-soluble nutrients, and they also dissolve the less soluble forms which contribute significantly, though to a lessser degree, to plant growth. Specifically, the objective of soil testing is the

measure of the amounts of exchangeable potassium, calcium and magnesium, the amounts of adsorbed and also easily acid-soluble phosphates, and the amounts of the ammonium and nitrate forms of nitrogen. Close correlations between these forms and plant response to added nutrients have been established by chemical soil tests and fertilizer experiments.

Almost all soils that have been cropped without fertilizer use for 100 years or more will have become deficient in one or more of the plant nutrients. Nitrogen is deficient in all of the agriculturally important soils of the world. Phosphorus and potassium deficiencies are very common. Zinc, manganese, iron, and copper deficiencies are commonly associated with sands and alkaline soils, while boron and molybdenum deficiencies are usually associated with the older, highly weathered soils. Calcium and magnesium deficiencies, which sometimes occur in the humid and semihumid temperate regions, are less of a problem since the soils are acid and the deficiencies are automatically corrected in the required liming program. (*See* also FERTILIZERS AND MANURES.)

7. Soil Variations with Climatic Environment.—Very young soils reflect the character of the parent material, but as weathering and its effects proceed, soils tend increasingly to show the effects of the climatic environment while the resemblance to the parent material grows gradually weaker or disappears.

Up to this point the discussion has pertained principally to soils formed under humid and subhumid temperate climates. With decreasing rainfall chemical weathering becomes less intense, soluble bases released by weathering are not leached from the soil and consequently the soil reaction becomes more alkaline, and the amount of organic matter present decreases. The colloid clays are dominantly of the 2:1 lattice type with montmorillonite the most common. The chemical fertility of such soils is usually variable and often quite low. The lack of moisture limits crop yields; therefore, such soils respond very favorably to irrigation and fertilization, particularly with nitrogen.

As one passes to subtropical and tropical regions of high temperature and high rainfall with intensified weathering and leaching, the silicate minerals are more highly weathered and bases and silica are leached away. The resulting soils are known as oxisols. Kaolinite is the dominant mineral, along with oxides of iron and aluminum, in the clay fraction. These soils are low in cation exchange capacity, low in present and potential fertility, and usually low in organic matter. Since bacterial activity is high throughout the year organic matter, in the upland soils, is so rapidly destroyed that its content is usually low in spite of its rapid production by vigorous growth of vegetation. In the deltaic regions, where most of the population resides, the soils are medium to high in organic matter. When properly fertilized and well managed, these soils have a tremendous capacity for food production.

In the cool to cold humid zones where forest vegetation is mainly coniferous the typical spodosol, or podsol, soil is formed. Decaying pine needles on the forest floor produce a highly acid leaching solution which not only removes the basic ions excessively but also dissolves iron and leaches it downward, where, with some of the aluminum and dissolved organic matter, it is redeposited in a characteristic tight subsoil, or B horizon layer. The surface layer immediately above this deposition is bleached white and ashy, and is high in silica. Kaolin, a 1:1 lattice mineral, dominates the soil clays. These soils are leached of their plant nutrients and are initially low in productivity. With good management, liming, and high fertilizer applications they become quite productive. (S. W. ME.)

8. Alkali and Saline Soils.—The term alkali has been rather loosely applied to all soils containing sufficient amounts of soluble salts to cause injury to plant life. When correctly used, the term alkali soils refers only to those soils that have a high pH value caused by sodium ions absorbed by the clay particles, and to those that sometimes have considerable amounts of the alkaline salt, sodium carbonate. Reaction between water and the sodium-saturated clay liberates sodium and hydroxyl ions into the solution, making it alkaline. These soils were formerly called black alkali in the U.S.

Saline soils are those that contain less absorbed or exchangeable sodium and that are high in nonalkaline salts such as sodium chloride and sodium sulfate, or those that contain enough salts such as sodium chloride to maintain flocculation and repress hydrolysis of the clay portion of the soil which is itself the insoluble salt of a strong base and a weak acid. The term white alkali frequently has been used for these soils. Alkali and saline soils are widely distributed in the drier areas of the world. On soils where salts may be a problem a great deal of foodstuff is produced. In regions of low rainfall, salts accumulate where drainage is poor, their origin being caused by insufficiency of percolating moisture to wash out and carry away in the drainage the soluble salts present in the parent soil material or those that are formed by its weathering. They are frequently formed in localized areas by seepage water from higher elevations. Soil-forming materials reclaimed from the sea, as on the coast of the Netherlands, are also saline. The salts that most frequently predominate are the carbonates, chlorides and sulfates of sodium, magnesium and calcium, or mixtures of two or more of these salts.

The tolerance of the various species and varieties of plants to soluble salts in soils is very different. Among the more tolerant agricultural plants are sugar beets, garden beets, milo, Bermuda grass, Rhodes grass, alfalfa, cotton, tomatoes, and barley. The less tolerant ones include beans, field peas, red clover, vetch, oats, and peaches. It has been found that the presence of high concentrations of salts in the soil solution (the liquid portion of the soil in which the plant roots obtain nutrients) limits the intake of water by plant roots. Yields are reduced approximately in proportion to the osmotic pressure of the nutrient solution. Although definite limits cannot be set for the amount of soluble salts required to reduce the yield of plants, less than 0.2% of salts in a soil is usually not harmful to salt-tolerant crops; in an alkali soil, lower percentages are harmful because of more toxic effects of the ions present. Relatively large amounts of salts may be added to the soil in irrigation water. In order to maintain the "salt balance" as much salt must be removed in drainage water from the soil as is introduced in the irrigation water.

Many irrigation projects have failed because of salinization caused by lack of proper drainage facilities or failure to use enough water to move excess salts down and out of the soil profile. Others have failed because of the high percentage of sodium in the irrigation water.

The chemical reactions by which a normal soil, the exchange complex of which (essentially the fine or clay fraction) is saturated with calcium and magnesium ions, is converted to an alkaline soil are as follows: Ca-saturated soil + $2NaCl \rightarrow$ Na-saturated soil + $CaCl_2$. When the sodium-saturated soil is leached by water little happens to it until the excess of soluble salts is removed. Further leaching brings about the reaction that is involved in the formation of the alkali soils of columnar structure called solonetz. This reaction is Na-saturated soil + H_2O = H soil + NaOH. Then, in the presence of carbon dioxide of the soil air the salt sodium carbonate is formed—$2NaOH + CO_2 + H_2O = Na_2CO_3 + 2H_2O$. Alkali soils are sticky, impervious to water, and unfavourable for agriculture. At the high pH values of alkali soils, dissolved organic matter coats the soil grains and gives a black colour. The reclamation of alkali soils is the reverse of the above processes. A soluble calcium salt such as gypsum is applied to the soil—2Na-saturated soil + $CaSO_4$ = Ca-saturated soil + Na_2SO_4. The sodium salt is then removed in the drainage water. Sulfur may be used in place of gypsum if the soil contains calcium carbonate. By biological action sulfur is converted to sulfuric acid, which reacts with the calcium carbonate to give gypsum. Good drainage is necessary to remove end products so that the reactions can proceed. Sodium-saturated colloids, in the absence of flocculating salts, "run together" or "freeze up." Sodium-saturated colloids in the presence of excess sodium salts may not be high in pH and are flocculated. *See* also IRRIGATION; LAND RECLAMATION. (L. T. A.)

VI. SOIL MICROBIOLOGY

The soil is not a dead, inert material; it is full of life, for a thimbleful of soil contains billions of living microorganisms. The

organic matter already present in the soil, as well as that constantly being added by higher plants, animals, and microorganisms, serves as a direct source of carbon and energy for the heterotrophic soil organisms; that is, microorganisms that derive their carbon and energy from organic materials. Indirectly, organic matter serves as food material for autotrophic bacteria—organisms that obtain their carbon supply from the carbon dioxide in the soil air and their energy from the oxidation of simple chemical substances.

Bacteria, fungi, actinomycetes (a filamentous form, including *Streptomyces*), algae, and protozoa—which have received the most attention from soil microbiologists—have a marked effect upon the productivity of soils. In general, highly productive soils contain an abundance of microorganisms, whereas less productive soils support a small population. While these facts do not prove that soils are productive because of their high microorganic population, or vice versa, there is much experimental evidence to show that microorganisms contribute to the productivity of soils through their varied activities.

1. **Decomposition of Organic Matter.**—One of the functions of soil microorganisms is the decomposition of organic matter added to the soil through the use of green manures—crops grown for soil-improvement purposes—animal manures, crop residues such as roots and stubble, and microorganisms. The sugars, starches, cellulose, and similar compounds in these organic materials serve as a source of energy for bacteria, fungi, and actinomycetes, and largely disappear into the soil air in the form of carbon dioxide. Other compounds more resistant to decay remain in the soil longer and are important contributors to the humus in the soil. Bacteria, fungi, and actinomycetes need nitrogen for their growth during the process of decomposition. Where the organic materials are low in nitrogen, microorganisms use the ammonium and nitrate nitrogen in the soil. Consequently, they may compete with grain crops for useful nitrogen where low-nitrogen organic materials are plowed under, resulting in a lower immediate yield than would be obtained in the absence of added organic material. Nitrate nitrogen is often lost in drainage water because of its high solubility. Such losses may be decreased by the use of low-nitrogen organic material on land having a high nitrate-supplying power during extended periods when no crops are grown.

Both farmers and city people, when preparing artificial manure, have taken advantage of the ability of microorganisms to decompose organic substances. In this process, leaves, lawn clippings, garden refuse, and similar organic materials are mixed with ammonium sulfate or nitrate, superphosphate, and crushed limestone. Microorganisms use the nitrogen in their growth processes; the limestone prevents the accumulation of acids; and the phosphate reinforces the manure. This process of making artificial manure is called composting. (*See* also FERTILIZERS AND MANURES.)

2. **Solution of Minerals.**—During the process of decomposition of organic materials, large amounts of carbon dioxide are produced. When dissolved in soil moisture, carbonic acid is formed and this weak acid may account for some solution of soil minerals. Some bacteria and fungi produce organic acids capable of dissolving minerals. Other microorganisms producing mineral acids are effective, to some extent at least, in releasing some of the nutrient elements from relatively insoluble soil minerals.

3. **Nitrogen-Fixing Bacteria.**—In the air above every acre of land there are about 69,000,000 lb. of elemental nitrogen. Curiously enough, most crop plants may be starving in this sea of nitrogen. Only when the nitrogen in the air is combined with other elements is it possible for crops to use this nitrogen in their growth processes. Similarly, most microorganisms need nitrogen that has been changed into organic or inorganic combinations.

Some bacteria are able to use the elemental form of nitrogen that exists in the air. Thus, they can grow in the absence of combined nitrogen and, at the same time, produce nitrogenous substances which may be used later by crop plants.

The legume-nodule bacteria (rhizobia) are examples of organisms that can use air nitrogen. They grow in association with leguminous crops and are responsible for the nodules or tubercles which are associated with most leguminous or pod-bearing crops, such as peas and beans.

Role of Legumes.—The nodule bacteria furnish nitrogen compounds for the leguminous plant, and the leguminous plant furnishes energy material for the bacteria. Such a relationship, in which dissimilar organisms live together for their mutual benefit, is known as symbiosis.

Because leguminous plants are able to use air nitrogen indirectly, they will grow well, if nodulated, on soils that do not contain enough useful nitrogen for maximum growth of a nonleguminous crop, such as corn. A plentiful supply of nutrient elements other than nitrogen is necessary, however.

The amount of nitrogen nodulated leguminous crops obtain from the air depends upon several soil conditions. Under favourable conditions, it is estimated that from 50 to 150 lb. of nitrogen per acre may be derived from the air. The amount of useful nitrogen in the soil, the kind of crops and the fertility of the soil (the relative amount of plant nutrients which are present) are factors affecting the amount of nitrogen that can be obtained from the soil.

The nodule bacteria vary also with respect to the amount of nitrogen they can take from the air. Some bacteria are highly efficient; other species have little capacity to obtain nitrogen from the air. Still others actually may be unable to secure air nitrogen and thus may be parasitic.

The supply of nitrogen in the soil may be replenished by growing nodulated leguminous crops if all or a part of the leguminous crops is returned to the soil. Obviously, the way the crop is handled determines the extent to which nitrogen-replenishment occurs.

No one kind of nodule bacteria will find a home on all leguminous crops. Of the approximately 10,000 species, only a relatively small number of leguminous crops and their nodule bacteria have been studied. These include about 20 kinds (strains) of nodule bacteria, recognized by their ability to produce nodules on different leguminous plants. There are some leguminous plants on which nodules never have been found.

Inoculation.—Because the success or failure of a crop may depend, in part, on the presence of nodule bacteria, farmers practise inoculation—the process of adding suitable nodule bacteria to the seed or soil—in order to be sure the leguminous crop will be nodulated. Where the leguminous crop is native to the area, inoculation is practised as a means of cheap insurance. Inoculation is very important for leguminous crops which are new to a locality.

Reliable commercial inoculants are sold by many seed stores. The information printed on the container gives the name of the crop or crops for which the inoculant should be used, as well as directions for its use. The cost per bushel of seed is comparatively low.

Other Bacteria.—In addition to the rhizobia, there are bacteria living independently in the soil; *i.e.*, they are free living and are capable of using air nitrogen in their growth processes. These bacteria are said to be nonsymbiotic. Some of the nonsymbiotic bacteria are aerobic, requiring free oxygen, while others are anaerobic, or able to live in the absence of free oxygen.

The *Azotobacter* are the most important among the aerobic nonsymbiotic nitrogen-fixing bacteria—those able to convert nitrogen from the air into nitrogen compounds. They are present in small numbers, rarely more than a few thousand in a spoonful of soil. However, the *Azotobacter* are found in most soils which are neither distinctly acid nor highly alkaline. Their importance in soil fertility is in doubt. Under highly favourable laboratory conditions, they fix appreciable amounts of nitrogen, but the evidence in natural soils is largely circumstantial. It has been estimated that the *Azotobacter* may fix from 15 to 40 lb. of nitrogen per acre per year.

Nitrogen fixation by free-living nitrogen-fixing bacteria may occur in poorly drained soils because several different microorganisms may live under anaerobic conditions and fix appreciable amounts of nitrogen. (*See* also NITROGEN, FIXATION OF.)

4. **Nitrogen Mineralization.**—Organic matter is the storehouse of nitrogen in the soil. The nitrogen fixed by both symbiotic and nonsymbiotic bacteria is present in the form of organic com-

pounds such as proteins and proteinlike substances. Before this nitrogen can be utilized by growing crops, it must be converted into inorganic forms. Some of the processes by which organic nitrogen is changed into inorganic forms are carried on by several kinds of microorganisms, but the final steps in the formation of nitrate nitrogen, the form most readily used by higher plants, are the result of the activity of highly specific organisms.

5. Ammonification.—Ammonification is the process by which ammonium is formed in soils from nitrogenous organic compounds. It results from the action of microorganisms. The majority of the soil bacteria, many fungi, and the actinomycetes decompose protein materials, utilizing the carbon for energy purposes and liberating ammonium produced in excess of the small quantities needed for growth of the microorganisms. Where large quantities of highly nitrogenous green manure crops are plowed under, correspondingly large amounts of ammonium are released. However, the ammonium that accumulates in most soils is transient, because it is utilized quickly by other microorganisms. Because of this quick utilization, it is not possible to measure directly the production of ammonium in soils, but accumulation at any specific time can be determined.

6. Nitrification.—Nitrification is the process of converting ammonium nitrogen to nitrate. When ammonium is oxidized, it is first changed to nitrite which, in turn, becomes nitrate upon further oxidation. Unlike ammonification, which is accomplished by numerous kinds of bacteria, fungi, and actinomycetes under both aerobic and anaerobic conditions, the change from ammonium to nitrite is achieved by the specialized bacterium, *Nitrosomonas*. Another specialized bacterium changes the nitrite into nitrate. Both of these bacteria are strictly autotrophic and aerobic. Their activity in the soil is restricted by acid conditions, consequently, they are most active in well-drained soils where acidity has been corrected by a liming program.

Nitrification varies directly with the temperature and moisture of the soil within certain well-defined limits. In a cold, wet spring, little nitrate is formed and the leaves of higher plants often are yellowish in colour, indicating a lack of nitrogen. As the temperature rises and the soil reaches an optimum moisture content, nitrification increases and the leaves of the plants become darker green, indicating a supply of useful nitrogen. Higher plants growing in low areas in fields are more apt to show nitrogen starvation in the early spring because of water and temperature relationships.

In any good soil-management program, the aim of the farmer is to make sure that conditions are favourable for nitrification at the time of year when nitrogen is used most heavily by the growing crop. Because nitrate nitrogen is subject to leaching, the removal of substances in water solution, it is undesirable to have nitrate accumulate when crops are not growing on the land. For example, ammonium is not leachable, but under favourable temperature conditions it is changed by nitrification into nitrate, which is leachable. Therefore, when farmers wish to plow under nitrogen fertilizers in the autumn, it is desirable to wait until the soil temperature is about 50° F or lower, to minimize the conversion of nonleachable ammonium into leachable nitrate.

7. Algae.—Soil algae occur in relatively small numbers in soils when compared with bacteria. Their numbers are in hundreds of thousands per gram of soil rather than in billions, as is the case with the bacteria. The algae contain chlorophyll—the green colouring matter in higher plants—and range in colour through greens, blue-greens, browns, and reds. The volume of individual cells is roughly 3 times that of bacteria and ⅓ that of protozoa. Typical water forms of algae are much larger.

The blue-green algae are able to fix nitrogen and may be agents contributing to the maintenance of crop yields on low-producing soils. Because algae are capable of getting their carbon from the carbon dioxide of the air and fixing their own nitrogen, they are credited with the formation of organic matter on volcanic ash and on burned-over areas.

8. Actinomycetes.—Actinomycetes are somewhat intermediate in size and shape between bacteria and fungi. In some respects they resemble both. They occur in smaller numbers than the bacteria, but are larger in size and frequently surpass the bacteria in weight per given unit. One of their main functions is the decomposition of organic materials. The characteristic odour of newly plowed soil is attributed to the activity of these organisms. Most actinomycetes are aerobic in nature and, like the fungi, are more abundant in dry than in wet soils. A few are parasitic on some plants and animals.

9. Fungi.—Fungi exist in soils in many different forms. Some may be seen only with a microscope, whereas others grow on the surface of the soil and have characteristic, visible shapes. Mushrooms are familiar examples.

The numbers and kinds of fungi present in soils depend upon moisture, temperature, the kind and amount of organic matter present, and the acid condition of the soil. Many fungi are active under conditions of acidity and temperature that many bacteria cannot tolerate. For this reason, certain groups, particularly the heat-tolerant organisms, are important in the decomposition of grain straw and strawy manure. Others are involved in the decomposition of nitrogenous organic materials with the liberation of ammonia.

Some fungi are capable of living in association with higher plants, particularly trees. They enter the roots of these trees, forming an association called mycorhiza (*q.v.*). They appear to benefit the host plant by increasing the capacity of the roots to absorb nutrients.

Microorganisms in general, but particularly fungi, are important agents in maintaining in soils a granular structure called aggregation. The threadlike network that they form tends to hold soil particles together in crumbs—a condition favourable for the absorption of water and for the exchange of air between the soil and the atmosphere.

Neither the enumeration of the kinds nor a description of the characteristics and activities of individual groups gives a complete picture of the importance of microorganisms in the soil. Except under very unusual and highly unfavourable conditions there is a continuous cycle of changes. Complex substances are broken down into simpler materials by some microorganisms, while the simple materials are combined to form more complex substances by other organisms. At any given time the status of the microbial population and its products are a result of the interaction of many groups of microorganisms. For instance, one-half of the phosphorus may be present in the organic fraction of the soil and the other half in the inorganic form. Microorganisms decompose a part of the soil's organic matter and change the phosphorus to an inorganic form that is useful for higher plants. At the same time, other microorganisms use the inorganic phosphorus as food and incorporate it in their bodies as organic phosphorus. This organic phosphorus is not available for the growth of higher plants until it is changed again into the inorganic form. Thus, in assessing the activity of the soil microorganisms, a series of checks and balances must be considered. (O. H. S.)

VII. SOIL FAUNA

Many members of the animal kingdom are present in both cultivated and uncultivated soils. Some of the animals are found within the soil for their entire life cycle, occasionally coming to the surface for short periods, while others spend only some definite portion of their life span in this subterranean environment. The animal population of the soil consequently consists of both permanent and transient inhabitants.

The soil fauna is composed of a variety of animals ranging from microscopic protozoa to vertebrates such as moles and ground squirrels. The major group of significance, however, is the invertebrate population. The simplest forms of underground animal life are the single-celled protozoa; this group, although part of the fauna, is discussed elsewhere with other soil microorganisms (see *Soil Microbiology* above).

The group of invertebrates discussed here includes a variety of distinct types, some ubiquitous, others occurring only infrequently and in small numbers. The major animals include the flatworms, the nematodes, certain mollusks—especially the slugs and snails—the earthworms, wood lice, millipedes, centipedes, a variety of

arachnids, and a broad spectrum of insects. In addition, a few vertebrate species make their homes in the soil.

On the basis of their feeding habits, the soil residents may be divided into categories, depending upon whether they feed on decaying organic matter (saprophagous), on living plants (phytophagous), on other animals (predaceous), or upon animal excretory products (coprophagous). Those that attack cultivated plants are a great menace to agriculture; but others play a positive role by aerating the soil and in many other ways.

The abundance of invertebrates in soil is determined in large part by the supply of available nutrients and by certain physical characteristics of the environment. For their optimum development, good aeration and drainage are necessary. Moisture conditions are significant, and the fauna tends to be present in smaller numbers when moisture is inadequate. Temperature is another important environmental influence, with little animal activity occurring at low temperatures. The addition of organic materials increases the population of animals, and each of the major groups is more abundant in manured than in unmanured soil.

The number of invertebrates in soil varies widely according to the environment, but in fertile, well-aerated soils the population ranges from several million to several hundred million per acre, exclusive of the protozoa. Most abundant are insects, nematodes (roundworms or eelworms), earthworms, and millipedes. Although there are vast numbers of smaller animals, these are so small that they make up only a fraction of the total weight of the fauna. The larger invertebrates, such as the earthworms and the myriapods (millipedes, centipedes, etc.), therefore, make up the bulk of the total weight, sometimes up to 90%. The fauna contributes as much as several hundred pounds of animal tissue per acre.

The common soil nematodes are nonsegmented worms barely visible to the unaided eye. They are endowed with the capacity for active movement when in a liquid film. Many different types maintain a subterranean habitat, and the number of species is immense (*see* ROUNDWORM). Some are saprophytic, using organic materials and even microorganisms as food; others prey upon members of the soil fauna including other nematodes, while a very important group can attack and parasitize growing plants. Some of the plant parasites prey upon one or a few plants, but others seem to be nonselective.

Mollusks such as snails and slugs spend part of their life in soil, and many mollusks will leave only infrequently. They prefer a moist environment and are common in shady areas at the soil surface as well as under rocks and surface debris. Although they will sometimes attack plants, they tend to utilize decaying materials. The mollusks are never abundant but may be present in numbers up to about 50,000 per ac.

Since the classical studies of Darwin, soil scientists have appreciated the significance of the earthworm (*q.v.*). These segmented worms spend their entire life cycle in the soil, emerging only occasionally. They are unfavourably affected by desiccation, low temperatures, and poor drainage, and they are generally absent when the soil reaction is highly acid. They show a distinct preference for well-aerated and well-drained localities. An available supply of organic matter is important, and the earthworm population is highest in soils receiving barnyard manure and in pasture and forest soils. Earthworms can feed upon organic litter but have certain distinct preferences in their nutritional habits. When present, animal manures are used. Not only is the number greater in organically rich soils, but the worms themselves are larger. Under optimum conditions, their numbers range up to several million per acre with a large bulk of the entire animal weight as earthworm tissue—up to a hundred or more pounds per acre. As they burrow, the worms excavate channels that increase aeration, improve drainage, loosen the soil structure, and provide paths for plant roots. The casts left by the earthworms tend to promote soil crumb aggregation as well. Darwin estimated that earthworms bring about 20 tons of material per acre to the surface each year—an amount sufficient to cover a field to a depth of one-fifth inch.

There are three major arachnid representatives: mites, spiders, and ticks. The arachnids are usually free-living in nature, although some can be parasites on plants or higher animals. Their burrowing habits probably improve the structure, aeration, and water movement in the soil.

Only two myriapods are encountered in appreciable numbers, the millipedes and the centipedes. Of the two, the millipedes are the more numerous, and soils rich in earthworms usually support the growth of many myriapods, with the population in each acre extending into the millions. Certain myriapods can utilize organic matter and living plants for their growth, while others are predaceous. This group requires adequate moisture in the soil.

Many genera and species of insects are present, either in the adult form or in the larval or pupal stage, and these are frequently the most abundant of the invertebrate animals in the soil. Among the more common insects are ants, springtails, termites, and beetle and fly larvae. Insects may feed upon subterranean portions of plants or may live saprophytically, while certain species lead a predatory existence. The plant parasites of this group pose a serious threat to agriculture.

Certain vertebrates burrow underground and spend part of their lives under the surface. Prairie, forest, and desert have their characteristic, burrowing species. Among the mammals are the moles, ground squirrels, mice, hares, shrews, woodchucks, badgers, and gophers; among the reptiles are snakes, lizards, and some turtles. The rodents in particular can be quite active in their movements through the soil.

The soil fauna has a great significance in soil structure and fertility and in plant pathology. The invertebrates are capable of causing a number of injuries to plants as well as being parasitic. There is evidence that the fauna may play an appreciable role in promoting the decay of organic matter, particularly in forest areas, although its significance is secondary to that of the microflora. As the animals burrow through the soil, they leave behind openings that improve aeration and drainage, although they may damage a cultivated field or wreak havoc on a well-tended lawn.

The fauna also can move a considerable quantity of plant residues as well as large numbers of microorganisms for considerable distances. The animal population serves to mix surface plant debris and forest litter with the underlying soil and also to move soil upward and leave it in the form of mound. In all respects, the soil fauna provides an interesting area of study for the soil scientist, the zoologist, and the plant pathologist. (M. AR.)

VIII. SOIL PRODUCTIVITY

Soil productivity is the capacity of a soil, in its natural climatic environment, to produce a specified plant or sequence of plants under a specified system of management. In their natural state, soils differ widely in productivity because of differences in their ages and in the materials from which they are formed. Most silt loam soils, for example, are naturally more productive than sandy soils, and mature soils are more productive than soils that are either young, geologically speaking, or very old. The productivity of a soil is also partly determined by the way it has been managed in cultivation. In discussing soil productivity, it is appropriate to do so in terms of its relation to management.

1. Soil Management.—Management can be defined as the combination of tillage, cropping, and soil-treatment practices used to produce a crop. The three practices compete with and complement each other. Desirable combinations of the three tend to eliminate the unfavourable effects on the soil of crop production, to reduce the adverse effects of climatic, insect, and disease hazards, and to push crop yields toward their potential. Under continuing poor management (poor tillage, undesirable sequence or choice of crops, and little or no soil treatment) the productivity tends to be lower and to vary widely. Under complete and balanced management, on the other hand, a soil can maintain or even improve its productivity, and consequently grow crops more economically. Good management increases the efficiency of the soil in meeting crop requirements and protects the soil from the deteriorating influences of crop removals, and from excessive drainage and erosion. (*See* also *Soil Erosion and Conservation* below.)

In practical agriculture there is a tendency to underuse and neglect cropping and soil-treatment practices, which leads to soil

exhaustion, greater susceptibility to hazards and declining productivity and lower economy of production. Field experiments with cropping systems and soil treatments dating from 1876 (Morrow Plots, University of Illinois) have shown a decline in yield of 60% by the second half of the 20th century for corn in continuous culture; the yield of corn declined at a slower rate when corn and oats were alternated; and the yield of corn remained constant when corn was rotated with oats and clover (*see* ROTATION OF CROPS). Under each cropping system, soil treatment (applications of manure, limestone, and phosphate; with a legume green-manure crop included in the corn-oats rotation) resulted in increased yields over the years.

Treatment practices are used to restore constituents that become deficient in the soil. The amount to apply depends upon the needs of the crop for soil-derived nutrient elements and the ability of the soil to fill these needs. Background information for planning these practices is obtained from chemical analyses of crop plants and from various tests for soil deficiencies. Chemical analyses of crop plants show that species differ widely in their requirements for nutrient elements and that the requirements for the different vegetative portions of a plant differ widely.

The data in Table II, giving the approximate quantities of five elements removed from the soil in moderate yields, show these relationships for some of the more commonly grown field crops. The legumes, such as alfalfa and soybeans, differ somewhat from the nonlegumes in that they can obtain from one-half to two-thirds of the nitrogen they need from the air if they are

TABLE II.—*Nutrient Elements in Field Crops*

Crop	Acre yields (cwt.)	Pounds in acre yields				
		Nitrogen (N)	Phosphorus (P)	Potassium (K)	Calcium (Ca)	Magnesium (Mg)
Corn—grain	28	50	8	10	1	4
—stover	30	25	4	32	1	4
Wheat—grain	15	33	6	6	1	2
—straw	25	14	2	20	5	2
Soybeans—grain	15	84	10	29	3	4
—straw	33	60	6	42	43	23
Potatoes	90	31	13	45	2	3
Cotton—lint	5	1	*	2	*	
—seed	10	38	6	10	2	
—stalks	20	28	3	16	11	
Alfalfa hay	60	158	14	103	117	24
Timothy hay	40	44	6	52	11	6

*Trace

grown in association with the proper nitrogen-fixing organisms (see *Soil Microbiology* above).

Various tests for soil deficiencies can be used to indicate whether a particular soil can supply the requirements of crop plants. These include: (1) comparing crop yields and production practices for similar soils in the same community; (2) studying information that can be obtained from agricultural experiment stations; (3) watching crops during the growing season for symptoms revealing inadequate supplies of nutrient elements; (4) using chemical soil tests to detect deficiencies in available nutrients; and (5) using plant tissue tests to learn whether the plant is obtaining sufficient nutrients at various stages of growth.

In humid regions, low nutrient-supplying powers can usually be improved by returning deficient constituents to the soil. The constituents most likely to be deficient and the materials most commonly used to restore them are as follows:

Organic Matter.—Many cropping systems bring about the destruction and loss of soil organic matter. Crop yields tend to decline in proportion to the reductions taking place. The unfavourable effects of these losses can be offset by the regular return of fresh supplies. In field crop husbandry, this is done with animal manure, legume and nonlegume green manures and crop residues.

Nitrogen.—Organic matter is the source of soil-supported nitrogen. In some regions, as in the North American corn belt, the soil resources, the climatic conditions, and the type of farming followed make it possible to provide a high level of nitrogen supply, with animal manures and the residues of legume and nonlegume crops used in association with each other. It is often necessary to supplement the farm resources of nitrogen with commercial supplies. Even in the corn belt supplementary nitrogen is often needed, particularly if crop sequences are used that emphasize nonlegume crops and if little or no manure is returned to the soil. In some regions or under some conditions, it may be necessary to rely almost entirely on commercial nitrogen. Providing adequate supplies of nitrogen is a major problem in crop production.

Lime.—Lime affects soil productivity by modifying soil properties and supplying the nutrient elements, calcium and magnesium. Deficiencies are revealed by the development of soil acidity, to which most legume crops are especially sensitive. The proper use of lime, in the form of pulverized limestone, chalk, marl, or hydrated or burned lime, occupies a place of first importance on many soils.

Phosphates.—Phosphorus is supplied from both mineral and organic soil constituents. Deficiencies may be reduced or temporarily eliminated on some soils by utilizing subsoil supplies with deep-rooted legumes. Effective correction, however, is dependent on the use of phosphatic fertilizers such as superphosphate, basic slag, and rock phosphate. The economy of these carriers, which vary in chemical properties and behaviour, depends upon costs, soil conditions, and the type of farming followed. The need for phosphatic fertilizers is widespread.

Potash.—Crops acquire potassium from various potash minerals. Some soils are not abundantly supplied with these minerals. Other soils are well supplied with some of the less available forms, but are unable to provide some crops with sufficient quantities of potassium for good yields. Deficiencies in potassium-supplying power can be reduced by returning crop residues and animal manure to the soil; effective correction, however, is usually dependent upon the proper use of the potash fertilizers such as muriate of potash.

Minor (Trace) Elements.—In addition to the nutrient elements supplied in the above mentioned materials, crops also depend upon the soil for small quantities of manganese, sulfur, zinc, copper, boron, iron, and other elements. Deficiencies may be corrected by the direct application of suitable carriers or by the use of fertilizers containing the elements needed.

In correcting the nutrient-supplying deficiencies of soils, two general systems of practices have arisen: (1) the use of treatment materials in such ways as to improve the capacity of the soil for producing crops; and (2) the use of treatment materials in such ways, forms and rates as will favour the immediate response of specific crops in particular soil environments. The first system is directed toward the soil and involves a very careful planning of the cropping system and the incorporation, when needed of liberal quantities of crop residues, animal manures, lime, phosphates, and potash into the soil. The second system is directed toward the crop and involves the use of readily available nutrients, singly or in mixtures to provide sufficient quantities of nitrogen, phosphorus, and potassium for effective growth and yield.

In the United States, the interest in this system has led to the large-scale manufacture of fertilizer mixtures in the form of grades for specific conditions. Either system is effective when properly used, but the most favourable long-time economy in crop production is more likely to be attained when the essential features of both systems are properly combined.

2. Soil Treatment as an Investment.—On most soils the use of some soil-treatment system will increase the yields of crops more than enough to cover the additional costs of the treatment. As would be expected, poorer soils benefit more from soil treatment than do soils that are naturally highly productive. Soil treatment does not entirely close the gap between soils of different productivity, but it can effectively reduce the amount of variation.

The results of applying treatments to soils of different productivity are shown in Table III, where the benefits are indicated as reductions in acreages needed to produce the same amounts of crops. These reductions emphasize the importance of management systems that increase productivity, for if the same produc-

TABLE III.—*Acres of Soils of Different Quality, With and Without Soil Treatment, Required to Produce the Same Total Yield as 100 Acres of Untreated Group I Soils**

Soil Groups†	Management Practices		
	Tillage and cropping only (ac.)	Tillage, cropping and treatment (ac.)	Reduction for treatment (ac.)
I	100	75	25
II	107	85	22
III	123	86	37
IV	138	90	48
V	171	103	68
VI	174	102	72
VII	380	167	213
VIII	500	137	363
IX	648	156	492
X	884	136	748

*Based on average corn-equivalent yields of all crops grown on 21 Illinois soil experiment fields over time periods averaging 45 years.

†The soil groups are listed in declining order of natural productivity, with Group I the most productive, and Group X, the least.

tion can result from less acreage, land is released that can be used to produce more crops or converted to other economic or social uses.

See FERTILIZERS AND MANURES; ROTATION OF CROPS.

(F. C. BR.)

IX. SOIL EROSION AND CONSERVATION

Soil erosion is usually described as the carrying away of soil by either wind or water. The Grand Canyon, the fiords of Norway, or the sinkholes and caves of the Karst region of the Balkan Peninsula are all examples of normal or geologic erosion, a slow process that occurs under natural conditions. Over long periods of time, usually thousands of years, it makes definite changes in the earth's surface. There is little man can do to change this process or the end result. For the mechanics of erosion *see* GEOLOGY. *See* also BREAKWATER; IRRIGATION; LAND RECLAMATION; RIVER; WIND EROSION AND DEPOSITION.

The erosion dealt with under this heading is less spectacular but much more vital to the welfare of mankind. It is the carrying away by wind or water of the soil from farms and ranches so necessary for the raising of food, fibre, and other agricultural products. Erosion of this type has been greatly speeded up by man's mismanagement of land, even to the point of destroying it for practical agricultural use within one generation.

1. Factors Affecting Erosion.—Chief among factors affecting water erosion is the amount of cover given the soil. Erosion is slow wherever soil is covered by trees or grass. At Zanesville, O., on comparable watersheds, woodland lost almost no soil during a nine-year period and pasture land very little, but cropland lost an inch of soil.

Closely allied to the matter of cover is the amount and intensity of the rainfall and its distribution during the season. Often a few hard rains which come at a time when crop cover is light will cause most of the soil loss. At La Crosse, Wis., four storms a year during a four-year period caused 95% of the total soil loss and 84% of the runoff from cornland.

Slope is another determining factor. The longer the slope the more soil and water are lost when cultivated. If the length of slope is doubled the soil loss is usually increased about 1.5 times. Similarly, steepness of slope affects the speed of runoff water and consequent soil loss. If the slope per cent is doubled the amount of loss is increased 2.5 times.

The nature of the soil affects the amount and seriousness of soil loss. The topsoil is generally the most valuable part of the soil profile because it contains more available plant food and organic matter. The surface soil can usually absorb water much faster than the subsoil. As the surface layer becomes thinner less water soaks into the soil and more runs off the surface, thus increasing the rate of soil loss. Losses from deep, open soils are less in amount and less serious than from shallow soils. The hazard increases for slowly permeable soils and those with tight, nearly impermeable subsoils.

The cropping system, kind and amount of tillage, and conservation measures used all influence the amount of erosion. Heavy cropping of a soil causes the organic matter to decompose rapidly. Tillage hastens the process. Reduction of organic matter causes a change in the structure of the soil. The granular condition or clumping of the soil particles into crumbs like grains in a popcorn ball is largely lost. The soil particles lose their tendency to stick together and are less able to resist the action of wind and water.

The main cause of wind erosion is lack of vegetative cover. This may be due to drought and high temperatures which reduce growth of vegetation; to tillage, and growing of cultivated crops that afford insufficient cover; and, in the case of pastures and ranges, to overgrazing. When soil is relatively bare, alternate freezing and thawing or wetting and drying loosens the surface soil and breaks down soil aggregates to granules that can easily be picked up by wind. The most critical seasons are late winter and early spring, when the wind usually blows strongest, the land has the least vegetation and soil is most susceptible to movement. Wind erosion is particularly serious in arid and semiarid regions.

2. Progressive Erosion Damage.—Soil erosion may be so gradual that serious damage is done before the land occupant is aware of what is happening. Each time a hard rain hits bare land, soil is torn loose and swept away. Each raindrop hits the bare earth like a tiny bomb, sending the wet soil in every direction. Not only is soil torn loose to be washed down the slope but available plant food is dissolved and carried away.

Soil damage occurs in successive stages. At first the principal loss is by sheet erosion; that is, each time it rains the runoff water will remove a thin layer of surface soil. As the surface soil becomes thinner small rills (miniature gullies) appear. On cropland these rills are erased by the next cultivation and the field looks much as before. After most of the surface soil is gone gullies may become the principal problem.

In gully erosion, channels are cut by water that has concentrated

BY COURTESY OF U.S. DEPARTMENT OF AGRICULTURE

FIG. 6.—SOIL EROSION

(Top left) A single drop of rain; (bottom left) the amount of soil it displaces as it hits the ground; (right) if unchecked, the combination of rain, runoff water, and wind quickly renders land unfit for cultivation or even for pasture

in the natural drainage ways or draws in a field. These channels may vary from a few inches to many feet in depth and up to many feet in width. As a result of gullying soil is lost, tile lines are washed out and fields are divided into two or more parts. Gully erosion is the most familiar type of erosion, doing serious injury to farm land. In the most advanced stages gullies may render land unfit for practical use for cultivated crops or even pasture. Reclamation of badly eroded land is expensive in any case, and full restoration is usually impossible except by natural processes of soil formation which require long periods of time. Sheet erosion, however, is actually more common and more serious than gully erosion because it takes place so gradually that the damage may not be recognized.

Soil damage by wind erosion is serious and extensive. Farmers in affected areas suffer crop damage or complete loss of crops. Serious as this may be, the greatest loss is that of the fine soil fractions (silt, clay, and organic matter). The wind exerts a sorting action on some soils much like a sieve. The finer particles, which contain the major part of the plant food elements, are gradually sorted out and carried away. Left behind is the coarser portion, which is less fertile and is often more erodible than the original soil.

Water and wind erosion result in much additional damage. Productive lowland soils often are covered with sand, gravel, stones, and low-fertility subsoil washed from slopes above. Water-borne silt fills stream channels, drainage ditches, and reservoirs, thus interfering with drainage and shortening the usefulness of lakes built for flood control, water supply, and recreation. Winds often cover productive land with infertile sand and bury railways, highways, fences, hedges, shelter belts, and even farm buildings.

3. Erosion as a Worldwide Problem.—Surveys made in different countries have shown that erosion has damaged or ruined for practical use hundreds of millions of acres of once-productive land all over the world. Induced erosion continues to be a major problem in every agricultural region in the world except in northwestern Europe. In Great Britain and other countries of northwest Europe, well-distributed rainfall, lack of torrential rains, and centuries of careful husbandry have preserved soil fertility remarkably well. During World War II, increase in the misuse of land changed this situation, however, and serious erosion in parts of the uplands of England, Scotland, and Wales resulted from up-and-down hill farming of potatoes and other row crops.

The erosion problem appears to be common throughout the world. This conclusion was borne out by the Food and Agriculture Organization of the United Nations (FAO) in a bulletin, *Soil Conservation, an International Study* (1948). After reviewing conditions in the United States and citing examples of serious and widespread erosion in China such as, "The Yellow River is the muddiest great river in the world," the report continued: "But these are only examples—the problems of soil conservation are world wide. They are particularly acute in India, where, as in China, there is tremendous pressure of population on natural resources even more meager; in the Mediterranean region of Europe and the Near East, where centuries of human occupation and the nature of the climate and land forms have favored rapid soil erosion; in Latin America, where many countries with dense populations have rapidly eroding lands; and in South Africa and Australia, where conditions and problems and the history of occupation are similar to those of parts of the United States of America." These conditions to a large extent still prevail.

In the United States a National Inventory of Soil and Water Conservation Needs, completed in 1961 under the direction of the U.S. Department of Agriculture, Soil Conservation Service, reported the amounts and usages of nonfederal rural land in the 48 mainland states as shown in Table IV.

The report revealed that nearly two-thirds of all nonfederal rural land needed conservation treatment of some kind and 8,323, or 65%, of the delineated small watersheds needed community-type projects for flood prevention and water management.

4. Conservation and Productivity.—In its simplest sense, soil conservation means the saving or preserving of the soil. This would be a relatively easy problem were it not that while saving

TABLE IV.—*Amounts and Usages of Nonfederal Rural Land in the U.S.*

Use	Acreage	Percent of tota.
Cropland . . . . Land currently tilled	447,399,000	31
Pasture and range. . . . . Land in grass or other long-term forage growth used primarily for grazing	484,716,000	33
Forest and woodland . . . . Land with at least 10% canopy of trees or other woody growth	449,651,000	31
Other land . . . . Farmsteads, farm roads, "idle" land, wildlife areas, etc.	66,271,000	5
Total	1,448,037,000	100

*Excluding Alaska and Hawaii.

the soil it is necessary to use it. Man must depend on it for most of his food and fibre. Soil conservation, therefore, involves saving the soil from erosion and improving its fertility and productivity while at the same time using it to grow needed products.

In addition to protecting and improving the land, soil conservation yields many side benefits. It helps lower the cost of farm production. This, in turn, increases farmers' profits and helps lower the cost of food and clothing to city dwellers. Soil conservation helps check drought damage in dry seasons and reduces flood crests in small and large streams in wet times. It reduces siltation of streams, reservoirs, and harbours, thus helping to insure a cleaner water supply for cities and towns, improve navigation and provide a better habitat for fish and other forms of wildlife. Soil and water conservation go hand in hand, for soil conservation is dependent on good water management. Conservation methods do away with the ugly scars of erosion, thus providing a more beautiful countryside in which to work or play.

5. The Conservation Movement.—Soil depletion and erosion began to take their toll early in the history of the American Colonies. Most of the early colonists came from areas where rains fall gently. Therefore they were not prepared to handle soil in the North American climate, where rain often falls with great intensity. They cut down the forests along the eastern coast, built their homes and planted corn and other crops in the cleared fields. More and more land was plowed to produce food for the increasing population and for production of tobacco and grain that could be traded abroad for needed commodities.

Then a few farmers in the older sections noticed a change in the soil. At first it had been dark, in some places almost black, but it gradually became lighter in colour. When a heavy rain came the loose soil mixed with the water and the muddy mixture flowed down the hills, carrying with it the richest portion of the soil—topsoil.

Destruction of soil by gullies and floods was noted by writers even before the Revolutionary War. By 1775 rivers that once ran clear were described as being black with mud. Early Massachusetts records indicate that most of the land near the coast was abandoned at least once before 1800.

Some of the leading men of that day were deeply concerned. Patrick Henry is credited with saying, "He is the greatest patriot who fills the most gullies." George Washington, who was an industrious farmer, was continually striving to conserve his land and check erosion. In his final message to Congress in 1796, he urged the creation of a board of agriculture, but it was not until Lincoln's administration in 1862 that the United States Department of Agriculture was created. Thomas Jefferson was constantly looking for ways to improve the productivity of his soil. While Jefferson was president his son-in-law, Thomas Mann Randolph introduced a system of horizontal plowing on his land that was the forerunner of contouring. The results delighted Jefferson. But people generally were not concerned about soil damage since there was so much virgin land. During the following century they moved west, clearing the forest for cropland, plowing the grasslands of the middle west and much of the Great Plains, and on to the Pacific coast.

Conservation sentiment was built up slowly. Early in the 20th century Pres. Theodore Roosevelt called the first conference of

governors to consider conservation of natural resources. Interest thus aroused led to the purchase of the first national forests and the establishment of the U.S. Forest Service.

Interest in soil conservation lagged during World War I and the agricultural depression of the 1920s. A soil surveyor in the Bureau of Soils, H. H. Bennett, was, during all this period, persistently calling attention to the serious erosion he had found in his survey work and trying to arouse interest in a corrective program. The movement took a big step forward in 1929 when Congress approved the establishment of soil conservation experiment stations to: (1) measure the rates of soil and water loss; (2) determine the extent and location of damage by erosion; and (3) develop methods for controlling erosion. Soon there were ten stations in operation gathering the information needed for an attack on the erosion problem.

In 1933 the Soil Erosion Service was set up in the Department of the Interior as one of the emergency agencies to help the country out of the economic depression of the early 1930s. Bennett was made chief of the first erosion control agency ever established by an important nation. In 1935, on the heels of the worst dust storm the U.S. had ever seen—dust from the Great Plains reached Washington, D.C., and drifted far out to sea—Congress passed Public Law 46, the first soil conservation act in the history of the United States or any other nation. The law established the Soil Conservation Service in the Department of Agriculture, replacing the Soil Erosion Service. The new agency was charged with carrying out, in cooperation with farmers, a program demonstrating good land use and erosion control. To facilitate the program, labour from the Civilian Conservation Corps (CCC) was made available. The work was carried out first on farms in project areas covering representative watersheds throughout the country, a watershed being that area of land which drains into a single stream, river or other drainageway. Later it was extended to farmer cooperators in CCC camp areas in each state. But all this was demonstration of limited scope. There was need for farmers and government to work together to apply the lessons to all farm and grazing lands.

Early in 1937 Pres. Franklin D. Roosevelt wrote to the governors of the states pointing out the need for state legislation to enable farmers to take necessary cooperative action in furthering the soil conservation program. With the letter was a proposed standard state soil conservation districts act that provided for the organization by farmers of soil conservation districts as governmental subdivisions of the state to carry on projects for erosion control.

The suggestion was well received. Arkansas became the first state to have a soil conservation law on its books. In Aug. 1937 the Brown Creek Soil Conservation District in North Carolina was organized as the first in the United States. By 1948 all states and territories had passed soil conservation district laws. In the second half of the 20th century approximately 3,000 districts had been formed to include the major part of the land in farms and ranches throughout the United States and its territories.

The U.S. Soil Conservation Service shifted its program to do most of its work in cooperation with these districts. The service assigns to each district one or more technicians to help farmers plan and apply conservation programs on their farms. The assistance includes: (1) a detailed soil survey of the farm or ranch; (2) a conservation plan drawn up by the farmer and technician working together; (3) application of practices called for in the plan; and (4) maintenance of the established practices. Any farmer may participate by agreeing with the district directors to conserve and use his land properly.

Many conservation practices can be applied by the farmer without assistance. But drainage, terraces, strip cropping, farm ponds, diversions, land smoothing, gully control structures, and the like require more technical skill than the farmer ordinarily has. The technical knowledge is supplied by technicians who work with the soil conservation district.

The U.S. Soil Conservation Service is extending its service to all users of land whether farm or nonfarm. The Food and Agriculture Act of 1962 assigned new tasks to the Soil Conservation Service in rural recreation, resource conservation and development projects, watershed development, and cropland conversion. Soil surveys are made for use in rural-urban planning and information on soil and water use is provided for urban fringe areas. Plans for watershed projects include recreation and wildlife development, water supply for municipal or other use as well as conservation of farmland and flood control.

The program of the soil conservation service has stressed work on a watershed basis. Early project areas were based on watersheds and the boundaries of many soil conservation districts are those of a watershed. In 1944 the service was authorized to apply special treatment for flood control on 11 watersheds in 12 states. In 1953 work was authorized and started on a pilot-watershed program covering 60 small watersheds in 34 states. In 1954 under the Watershed Protection and Flood Prevention Act (Public Law 566) the service was designated as the U.S. Department of Agriculture representative agency to cooperate with local organizations in small watersheds throughout the nation in work aimed at upstream watershed conservation and flood control. Response to this program was widespread in all states. In all of these watershed efforts the work is carried on in cooperation with local landowners.

Costs are shared by local organizations and the federal government. Amendments to Public Law 566 encouraged multiple use of watershed reservoirs. Recreation and water supply for municipal and industrial use are included in many of the small watershed projects.

Results following high intensity, flood producing storms in several well-developed watersheds in different parts of the country have been most encouraging. Upstream retarding dams used along with recommended soil and water conservation measures on the watershed lands have greatly reduced flooding, erosion and siltation.

In 1956 a Great Plains Conservation Program was authorized by Congress to help farmers and ranchers of the region develop long-term plans for their land, including soil and water conservation measures and the land-use adjustments needed to achieve a more stable agriculture. It provided technical aid and cost-sharing assistance over a contracted period of years to farmers and ranchers in designated counties for applying measures as part of an approved plan. The soil conservation service cooperates with local, county, state, and regional agencies and organizations to promote this program.

6. Methods of Conservation.—Appropriate use of the land depends upon land contour, type of soil, and climate. Land may be level, slightly or steeply sloping or even mountainous. Soil may be sandy or clayey, fertile or almost barren. Climate also varies. Therefore the farmer or rancher has a real problem in deciding how best to use his land, for there may be several kinds in his holdings. There are many examples of land that has been ruined or badly damaged because it was put to a use for which it was not fitted. To help a farmer or rancher with this problem, soil conservationists make a careful survey of the entire farm and put each area into one of eight broad classes of land-capability according to kind of soil, slope and degree of erosion. Each class is shown by a different colour or Roman numeral on the map of the holding. A farmer can refer to the map and quickly see whether a field can be safely cultivated or whether it should best be used for grazing, forestry or wildlife.

An important point in land management is to treat each acre according to its needs. After the best use for land is determined, the next problem is to treat each acre in such a way that it will produce as much as possible without injury from erosion. As stated above, good vegetative cover is the first defence against soil loss and runoff. Therefore, proper treatment of all classes of land will provide adequate cover.

For forest and woodland this involves protection from fire, managed cutting, and replanting if needed. For ranges and pastures approved practices include avoidance of overgrazing, resting to encourage growth of better pasture plants, renovating, fertilizing, and reseeding. The program varies widely from the pastures of humid regions to the ranges of arid and semiarid areas.

Cropland is used more intensively than other land on the farm.

More fertility is removed and cultivation encourages the breakdown of the organic matter which helps a soil resist erosion. The soil, when left bare, is open to the beating action of raindrops. This is especially true of row crops, such as corn, soybeans, cotton, and potatoes, which are cultivated for weed control. Special care must therefore be taken to maintain and improve fertility, add new organic matter, and protect the soil from erosion.

7. Minimum Tillage.—Working the soil less helps reduce the erosion loss from row crops. For years, in preparing a field for a crop of this type, it has been customary for farmers to work the soil into a fine seedbed by plowing and then going over the field several times with disk, harrow, or other tillage tool. Under the system called "minimum tillage" the aim is to use the least amount of tillage that will insure quick seed germination and satisfactory growth of the row crops. Minimum tillage saves one or more trips over a field, resulting in less expense for labour and for machine operation. In the minimum tillage system the seedbed is left with larger clods; more water sinks into the soil, less water flows down the slopes; consequently less soil is eroded from the field.

In the cropping system row crops are usually used in succession or rotation with small grains such as wheat, oats, barley or rye, and meadow crops—legumes or grasses or mixtures of the two. As protection against erosion, meadow crops are most effective, row crops the least and the small grains in between. How often each of the three classes of crops should be used in the rotation depends on the capability class of the land and the erosion-control practices employed (see *Soil Productivity* above). (*See* also ROTATION OF CROPS.)

8. Other Erosion Control Practices.—Even though good land use, a suitable crop rotation, and soil treatment are used, some additional measures are often needed on sloping fields and where wind erosion is a problem.

Grass waterways rank first in prevention of erosion and the control of runoff water from cultivated land. They are usually located in the natural drainage ways, but they may be constructed along field boundaries in some cases. Waterways are best made dished or saucer-shaped and rather shallow to keep water from concentrating to any great depth. Width should be sufficient to carry heavy rains without overflowing. A dense vegetative cover is required to keep the water from cutting a gully through the centre. Grasses that make a dense sod are preferred for this purpose.

Where a waterway discharges into an open ditch or stream at a lower level, some sort of structure is needed to keep the overfall from cutting a gully back into the waterway. These structures usually are constructed of concrete or masonry. Earth dams often are built across a large drainage way to prevent gullying and to form a pond for water supply, wildlife and recreation.

Contour tillage is one of the simpler practices that reduce soil losses about 50% from those occurring with up-and-down hill cultivation. In contouring, all furrows and rows follow around the slope and are level from end to end. The ridges left by farm machinery in preparing the soil and cultivating the crop act as small dams, holding back the water and giving it time to soak into the ground. Contour guide lines may easily be laid out by the use of a simple level fitted with sights. (*See* also CONTOUR FARMING.)

On longer slopes, contour strip-cropping has proved very effective. Instead of the entire slopes being planted in one crop, as in simple contouring, strips of row crops are planted between strips of close-growing crops such as small grains or meadow. If the contour ridges in the cultivated crop are overtopped in a hard storm, the close-growing strip below will slow down the water and catch much of the soil it carries. Strip-cropping has been found to reduce soil losses about 75%.

Terraces are widely used, especially on long slopes, and give the most effective control if they are properly laid out and constructed and are given good care. A terrace is a low, flat ridge of earth built across the slope, with a channel for runoff water just above the ridge. Usually terraces are built on a slight grade so that the water caught in the channel moves slowly toward the terrace outlet. Where several terraces are used on a long slope they have the effect of breaking it up into a series of short slopes that lose less soil than one long one. Terrace cultivation, of course, has long been practiced, for example in China, Japan, the Philippines, and other areas of Oceania and Southeast Asia; around the Mediterranean; and in the Andes of South America. In areas where soils are able to take in water readily and rainfall is relatively low, level terraces may be used. These terraces merely impound the excess water until it can have time to soak into the soil.

Diversion terraces, or diversions, are built in the same way as field terraces, but they are somewhat larger. They are often used to protect hillside fields from water falling on land above them, or to take water away from a gully and thus assist in controlling it. The practices already discussed are the major ones used in the control of erosion by water. Additional practices in use on farms and ranches include land contouring or leveling for irrigation, irrigation water management, diversion construction, and pond construction. Watershed projects include floodwater retarding and diversion structures, grade stabilization structures, silt and debris basins, and stream channel improvement.

Wind erosion control requires a somewhat different technique, which involves the following practices:

BY COURTESY OF U.S. DEPARTMENT OF AGRICULTURE

FIG. 7.—EROSION CONTROL PRACTICES

(Left) Contour strip-cropping with alternate strips of close-growing grass-legume and corn, the row crop; (above) narrow base terraces, each occupying as much space as three rows of corn; (below) grass waterway planted with thick grass

1. Protecting the soil surface with cover, usually vegetation. Tillage methods, called mulch tillage or stubble mulching, in which crop residues are kept on or in the surface of the soil, have proved effective.

2. Keeping the surface of the soil rough to slow down the wind and trap drifting soil.

3. Managing the soil so as to produce soil aggregates large enough to resist wind action.

4. Using barriers, such as crop strips, ridges, or field windbreaks to trap the drifting soil and keep it from spreading.

See also DRY FARMING.

9. Worldwide Conservation.—By the middle of the 20th century, conservation had gained support throughout the world. The conservation movement in the United States has been widely studied and the programs of many nations have been patterned after it. The United States has sent many technicians abroad to study erosion problems and suggest courses of action. This work is carried forward through programs of international cooperation and assistance. An extensive training program for technicians from various lands was established through the U.S. Soil Conservation Service. Several hundred specialists from most of the countries of the world have been trained in this manner.

A publication of the Soil Conservation Service (*Misc. Pub. 908*, 1962) contains a world map indicating the countries carrying on soil and water conservation works and programs. Fifteen countries shown as conducting special programs, nationwide in scope, are: United States, Mexico, Colombia, Venezuela, Spain, Crete, Kenya, Northern Rhodesia (now Zambia), Tunisia, South Africa, Jordan, India, Pakistan, Australia, and Philippines. Twenty-nine other countries have soil and water conservation incorporated in agriculture, public works, or other national programs. Nine additional countries have local soil and water conservation programs or projects, provincial or state, or subsidiary to research or educational programs.

See also references under "Soil" in the Index. (E. D. W.)

BIBLIOGRAPHY.—*General:* U.S. Department of Agriculture, "Soil," *1957 Yearbook of Agriculture* (1957); E. L. Worthen and S. R. Aldrich, *Farm Soils*, 5th ed. (1956). *Soil Classification:* For detailed information about criteria and systems of soil classification, see *Soil Science*, 67:77–191 (1949) and U.S. Department of Agriculture, "Soil Survey Manual," *Agricultural Handbook No. 18* (1951). For discussions of the character, principles and uses of the new system of the United States, see *Soil Science*, 96:1–74 (1963). G. R. Clarke, *The Study of the Soil in the Field*, 4th ed. (1957); *Soil Surveys for Land Development*, FAO (1953); N. H. Taylor and L. T. Pohlen, "Soil Survey Methods," *N.Z. Soil Bur. Bull.*, 25 (1962). *Soil Testing and Analysis:* F. E. Bear *et al.*, *Diagnostic Techniques for Soils and Crops* (1948); Organization for European Economic Cooperation, *The Organisation and Rationalisation of Soil Analysis*, European Productivity Agency, project no. 156 (1956); M. L. Jackson, *Soil Chemical Analysis* (1958). *Physical and Chemical Properties:* L. D. Baver, *Soil Physics*, 3rd ed. (1956); H. O. Buckman and N. C. Brady, *The Nature and Properties of Soils*, 6th ed. (1960); C. E. Millar, L. M. Turk, and H. D. Foth, *Fundamentals of Soil Science*, 3rd ed. (1958); L. M. Thompson, *Soils and Soil Fertility*, 2nd ed. (1957); F. E. Bear (ed.), *Chemistry of the Soil*, 2nd ed., "American Chemical Society Monograph Series," no. 160 (1964). *Microbiology and Fauna: See* Francis E. Clark, "Living Organisms in the Soil," in U.S. Department of Agriculture, "Soil," *1957 Yearbook of Agriculture* (1957); S. A. Waksman, *Soil Microbiology* (1952); D. K. M. Kevan (ed.), *Soil Zoology* (1955). *Productivity:* E. J. Russell, *Soil Conditions and Plant Growth*, 9th ed. (1961); U.S. Department of Agriculture, "Soils and Men," *1938 Yearbook of Agriculture* (1938), "Climate and Man," Part iii, *1941 Yearbook of Agriculture* (1941). *Erosion and Conservation:* J. H. Stallings, *Soil Conservation* (1957); H. H. Bennett, *Elements of Soil Conservation*, 2nd ed. (1955); E. D. Walker and A. B. Foster, *This Is Our Soil* (1949); U.S. Department of Agriculture, "Soils and Men," *1938 Yearbook of Agriculture* (1938), "Soil," *1957 Yearbook of Agriculture* (1957), "Soil and Water Conservation in Many Countries," *Misc. Publs.* 908, "Soil and Water Conservation Needs a National Inventory," *Misc. Publs* 971 (1965), "Soil Erosion: the Work of Uncontrolled Water," *Agriculture Inf. Bull. 260*, "Multiple-Purpose Watershed Projects Under Public Law 566," PA-575, "Agricultural Land Resources—Capabilities, Uses, Conservation Needs," *Agriculture Inf. Bull.*

SOIL ANALYSIS: *see* SOIL: *Soil Testing and Analysis.*

SOIL MECHANICS is the branch of engineering science dealing with the behaviour of soil when subjected to stresses or to the action of percolating water. The subject furnishes a rational basis for foundation engineering and earth-dam design and provides new tools for geological investigation.

Soil is a natural aggregate of mineral grains, with or without organic constituents, that can be separated by gentle mechanical means such as agitation in water. It may consist of the products of rock decomposition found essentially in their original position after weathering, or of materials transported and deposited by glaciers, streams or wind. Soil differs from most structural materials in that it consists of three phases: solid mineral matter, water, and air or other gas. Many soils comprise two-phase systems, of mineral matter and water, and are then said to be saturated.

Principle of Effective Stress.—Although soils constitute possibly the oldest construction materials known to man, the arts of foundation and earth-dam engineering remained essentially empirical until about the third decade of the 20th century, largely because the implications of the interaction between the soil grains and water had not been correctly evaluated. In 1923 Karl Terzaghi published the first clear statement of the principle of effective stress, a concept that proved to be the key to an understanding of the engineering behaviour of soils. According to this principle, the intensity of stress p acting normal to any section through a mass of soil is equal to the sum of two parts, the intergranular or effective stress p' which is transmitted from grain to grain at points of contact, and the neutral or pore stress u which is transmitted through the water in the pore space; that is,

$$p = p' + u \tag{1}$$

Moreover, the engineering behaviour of the mass of soil is exclusively a function of the intergranular stress and is unaffected by the neutral stress.

The significance of the principle of effective stress is illustrated in connection with the strength of saturated cohesionless granular materials such as sand or gravel. As early as 1773, C. A. Coulomb expressed the resistance s possessed by such a material against failure by sliding as

$$s = p \tan \phi \tag{2}$$

where p was the pressure per unit area normal to the section along which sliding took place and ϕ the angle of repose of the material. The quantity ϕ, now more precisely designated as the angle of internal friction, is a soil property depending somewhat on the kind and shape of the mineral grains and primarily on the looseness or denseness of the arrangement of the grains. It may range between about 30° and 44°. Since slopes subjected to percolating water often proved less stable than those on dry sand, it was commonly assumed that the values of ϕ for wet sand were at least 6° smaller than those for dry sand. Yet, direct shear tests by Terzaghi indicated essentially identical values in either state. The discrepancy was resolved by replacing p by p' in equation (2), whence

$$s = p' \tan \phi = (p - u) \tan \phi \tag{3}$$

In this form, Coulomb's equation is valid.

Curiously enough, Coulomb himself made no measurements to determine ϕ and assumed a value of 45° in his application of equation (2). Indeed, until the principle of effective stress was understood, meaningful tests of the physical properties of soils could not be performed.

Seepage Pressure.—If water flows through a deposit of sand, the viscous drag tends to move the grains in the direction of the flow and produces a force, known as a seepage pressure, between the grains. The seepage pressure is an intergranular stress. If upward flowing water produces a seepage pressure just equal and opposite to the submerged weight of the sand above a given horizontal section, the intergranular pressure becomes zero and, according to equation (3), the strength of the sand becomes zero. The sand is said to be in a "quick" condition and is incapable of supporting a load on its surface.

Where flowing water emerges from a cohesionless deposit the seepage pressure has a component directed outward and away from the surface and tends to dislodge and remove some of the smaller particles. If the process continues, a tunnel-shaped cavity

or "pipe" may develop. The formation of such pipes beneath or within dams has led to several catastrophic failures. It may be prevented by blanketing the surface where the seepage emerges with coarser materials that permit escape of the water but prevent erosion of the fines. If the seepage pressure has a great enough upward component, it may be necessary to add weight to the top of the filter to counterbalance the upward forces.

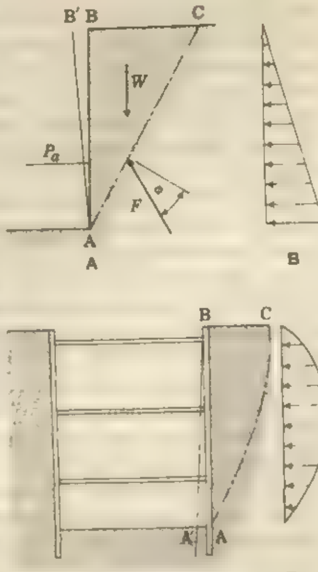


FIG. 1.—SHEARING RESISTANCE
(A) Forces assumed by Coulomb in calculating earth pressure against retaining wall; (B) distribution of pressure against wall; (C) braced cut; (D) distribution of earth pressure against sheeting

Shearing Resistance.—The shearing resistance of soil is a primary factor in several problems of engineering importance. It governs the pressure against retaining walls, bulkheads and the timbering of braced cuts. It also determines the bearing capacity of footings and piles and the stability of slopes, embankments and dams.

One of the first problems to receive attention in soil mechanics was the calculation of the pressure of earth against retaining walls. Coulomb solved the problem by considering the equilibrium of a triangular wedge of earth behind the wall (fig. 1[A]). He assumed the surface of sliding AC to be plane. The resultant force F on the surface of sliding was assigned the inclination ϕ to the normal to the surface, implying fully developed shearing resistance along AC. By allowing the inclination of AC to vary, Coulomb found the plane on which slip was most likely to occur, and the corresponding value of earth pressure P_a, designated as the active earth pressure. He assumed this pressure to act horizontally at the lower third point of the wall.

Coulomb's theory proved, with slight modifications, to be satisfactory for ordinary retaining walls and is widely used today. However, it failed to account for the pressures that practical men observed against the bracing of open cuts. Experience indicated that the load in the upper braces was substantially greater than that given by the theory, whereas the bracing at the bottom was very lightly loaded.

Theorists generally ignored these observations, and practical men generally discounted the theories. The state of contradiction was not clarified until in 1920 Terzaghi recognized the great importance of the manner in which the support may yield. In connection with a retaining wall, the supporting surface AB yields into the position AB′ and permits a uniform degree of lateral expansion in the entire sliding wedge ABC; the resulting lateral pressure increases linearly with depth (fig. 1[B]). On the other hand, in a braced cut, lateral expansion of the upper part of the wedge is restricted as soon as the uppermost brace is set; the deformation of the sheeting is likely to resemble BA′ in fig. 1(C). The large expansion at the bottom of the wedge is associated with a reduction of pressure at the bottom of the cut and an increase at higher levels. The corresponding pressure distribution is roughly parabolic (fig. 1[D]). Moreover, the surface of sliding is no longer plane but is curved. The general wedge theory of earth pressure (Terzaghi, 1939) took account of these factors and provided the means for calculating the loads in temporary bracing systems. The theory was confirmed by full-scale measurements on the bracing of the Berlin subway. Subsequently, it was found applicable to clay soils and confirmed by numerous full-scale field measurements, including a series on the Chicago subway.

Slope Stability.—Closely related to problems in earth pressure are those dealing with the stability of slopes. Before about 1920 it was assumed that every soil possessed an angle of repose and that slopes would be stable if established at angles slightly smaller than the appropriate angles of repose tabulated in various handbooks. Occasional catastrophic slides, particularly in clay soils, indicated from time to time the inadequacy of the procedure. In particular, a series of slides on the Swedish state railways near the beginning of the 20th century, culminating in one with a loss of 41 lives, led to the appointment of a Royal Swedish Geotechnical commission to study the degree of safety of existing slopes and to suggest remedial and precautionary measures. The report of the commission, published in 1922, contained the most complete analysis of landslides up to that time. It made the first extensive use of what is now called the Swedish method of stability analysis. The method is based on the empirical observation that the surface on which a slide occurs is almost always an approximately circular arc. The principle of the method is shown in fig. 2. The mass ABC tends to rotate about O under the influence of the weight W_1. The rotation is resisted by the moment of W_2 and by the moment of the shearing resistances s acting along the circular arc AC. The value of s required to prevent a failure (factor of safety = 1) is

$$s = \frac{W_1 l_1 - W_2 l_2}{R\widehat{AC}} \tag{4}$$

The factor of safety of the slope may be determined by comparing s from equation (4) with the value of shearing strength determined by laboratory or field tests. The position of the surface of sliding for which the factor of safety is a minimum is found by successive trials.

Most currently used procedures for stability analysis are derived from the Swedish method. Since slides frequently occur in clayey soils, information is needed regarding the shearing strength of clays. Coulomb expressed the relation between shearing strength s and normal pressure on surface of sliding as

$$s = c + p \tan \phi \tag{5}$$

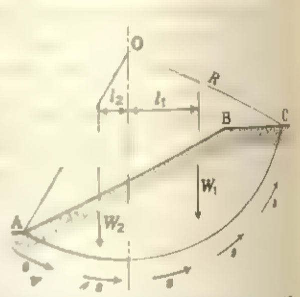


FIG. 2.—FORCES TENDING TO CAUSE AND PREVENT FAILURE OF A SLOPE ALONG A CYLINDRICAL SURFACE

where c was called the cohesion, or the shearing strength at zero normal pressure, and the other symbols have the same meaning as in equation (2). However, even with the replacement of p by p', equation (1), Coulomb's equation was found to be inadequate because the value of s was found to be a function of the loading history of the clay. Research has not yet fully clarified this aspect of the subject. For saturated clays that do not experience change in water content under stress, however, the shearing strength with respect to total stresses is independent of the pressure. That is, such clays behave as frictionless materials and their strength is

$$s = c = q_u/2 \tag{6}$$

where q_u is the ultimate strength of a cylindrical or prismatic specimen tested axially in a simple compression test. Many problems of practical importance can be treated in this manner.

Settling of Structures.—The settling of structures located above deep beds of soft clay long proved baffling to foundation engineers, particularly since such settlings often continued to increase many years after completion of the structures. Inasmuch as the damaging effects of the settling often did not develop until the settling had reached an advanced state, buildings frequently appeared to crack and deteriorate spontaneously. Several engineers had suspected that the settling was the result of slow squeezing of the water from the pores of the clay, but no satisfactory quantitative explanation was available until Terzaghi proposed the theory of consolidation in 1923. According to this theory, clay consists of mineral particles arranged in a structure that is at once relatively compressible and fairly impermeable. The voids between the mineral particles may constitute an appreciable fraction of the total volume, but the openings between

the voids are very small and restrict the flow of water. The voids are considered to be filled with water, which is inherently less compressible than the mineral skeleton. Hence, a load when first applied to the clay soil is carried almost exclusively by the water. The compression of the clay structure, and hence the settling of the ground surface, take place, in accordance with the principle of effective stress, only as the stress is transferred from the water to the clay structure.

Terzaghi proposed an instructive analogue to the process of consolidation of clay. In its simplest form, it consists of a water-filled cylindrical container closed by a piston supported by a spring. The spring represents the clay skeleton. The piston represents the boundary between two void spaces in the clay. It is pierced by a hole AB which permits escape of water at a slow rate corresponding to the low permeability of clay. If a load P is suddenly placed on the piston, no shortening of the spring can occur at the first instant because no water has yet escaped from the cylinder. Since the spring must shorten to carry additional load, the load P must produce an additional pressure P/A in the water, and the pressure gauge will indicate this value. There is now set up across the opening AB a pressure differential equal to P/A, which produces a flow of water from the chamber. As the water flows out the spring correspondingly shortens and carries load in proportion to its shortening. With the spring carrying part of the load, the pressure in the water is less than P/A. The pressure differential across AB is reduced, the rate of flow decreases and the rate of settling decreases. The process continues at a decreasing rate until, after a very great time, the excess pressure in the water approaches zero. The rate of settling of the piston corresponds to the rate of transfer of the total stress P/A from the water to the spring. By analogy, the rate of settling of a clay stratum is governed by the rate of transfer of the total pressure caused by the construction from porewater pressure to intergranular or effective pressure. The rate of transfer, in turn, is a function of the compressibility of the soil structure, the permeability of the clay and the viscosity of the water.

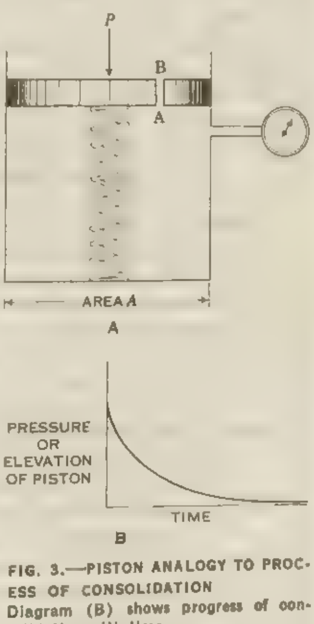


FIG. 3.—PISTON ANALOGY TO PROCESS OF CONSOLIDATION
Diagram (B) shows progress of consolidation with time

The mechanism of consolidation provides an explanation for several hitherto unexplained phenomena, including the settling of the ground surface associated with pumping from underlying aquifers. The reduction in pore pressure associated with pumping induces consolidation of any clay layers above or below the aquifer. Large settlings due to this phenomenon have occurred at many places, including Mexico City and the Santa Clara valley in California.

The accuracy with which settlings can be predicted on the basis of soil mechanics varies to a considerable extent with the geology and character of the deposits involved. Under the most favourable circumstances, computed and observed settlings agree satisfactorily until the settling reaches about 60% of the computed final value. Thereafter the real settlings are likely to exceed the computed ones by an amount that is usually, but not always, a small fraction of the total settling. The additional settling partakes of the character of a slow creep; the discrepancy is greatest for highly organic soils.

Sampling and Testing.—The structure of most compressible soils is sensitive to disturbance. If the samples from which values of compressibility and other physical properties are determined have experienced disturbance, the predictions based on the test values may be seriously in error. To avoid this difficulty, drilling and sampling techniques have been devised to permit procurement of specimens as undisturbed as possible. The most refined samplers consist of thin-walled tubes that are pushed steadily, without driving, into the ground, with an internal piston that remains at a fixed elevation such that the length of sample recovered is equal to the distance of penetration. Techniques have been developed for investigating the properties of the soils *in situ;* research is actively being conducted in this field. Nevertheless, errors due to disturbance remain significant; their evaluation requires full-scale field observations.

The greatest source of error in the application of soil mechanics lies in the heterogeneous character of most natural soil deposits. The products of natural processes are almost always complex and cannot economically be investigated in sufficient detail to determine soil properties completely. In some instances, only average values can be estimated, together with probable variations from the average. Nevertheless, useful predictions of settlement and other aspects of engineering behaviour can be derived from such estimates and the principles of soil mechanics.

See Karl Terzaghi, "Origin and Functions of Soil Mechanics," *Trans. Amer. Soc. Civ. Engrs.*, vol. CT, pp. 666–696, which contains extensive bibliography (1953); Karl Terzaghi and Ralph B. Peck, *Soil Mechanics in Engineering Practice* (1948). (R. B. P.)

SOISSONS, COMTES DE, nobles and princes important in French history. The Soissonnais, or country round Soissons, became an hereditary countship in the early period of feudalism. Having been held successively by the Houses of Vermandois, Eu, Nesle, Hainaut, Châtillon, and Coucy, it was shared, early in the 15th century, by those of Bar and Orléans. The Orléans rights to it were granted by King Louis XII to his daughter Claude, whose son, King Henry II, finally reunited them with the French crown, but the Bar rights passed by marriage (1435) to the House of Luxembourg. Marie de Luxembourg, comtesse de Soissons, who died in 1547 as the widow of François de Bourbon, comte de Vendôme (1470–95; *see* BOURBON: Table III), left the title to her grandson Jean. When he was killed in the Battle of Saint-Quentin (1557), his rights passed to his brother Louis I, prince de Condé (d. 1569).

Louis de Condé left the title of comte de Soissons to his youngest son, Charles (1566–1612). This comte fought for Henry of Navarre (later Henry IV of France) against Henry III of France at Coutras (1587), for Henry III against the Holy League in Brittany (1589), and finally for Henry IV against the League in Normandy (1590–92) and against Savoy (1600). His desire to marry Henry IV's sister Catherine was frustrated. Governor of Dauphiné from 1601, he protested when Marie (*q.v.*) de Médicis was named regent in 1610; but he was appeased with the governorship of Normandy. He died at Blandy-en-Brie on Nov. 1, 1612, leaving a son and three daughters by his marriage (1601) to Anne de Montafié.

As the princes de Condé (*q.v.*) came to be called simply Monsieur le Prince, so Charles was called simply Monsieur le Comte. His son Louis (1604–41) inherited this designation with the title comte de Soissons. After taking the side of Marie de Médicis in 1620, he served Louis XIII against the Huguenots in 1622. Involved in intrigues against the cardinal de Richelieu (*q.v.*), he is alleged to have plotted to assassinate him at Amiens in 1636, after a campaign against the Spaniards in Picardy (*see* THIRTY YEARS' WAR). In 1637 he fled to Sedan, just across France's eastern frontier; other malcontents joined him; and in 1641 he published a manifesto against Richelieu and invaded France with a Habsburg army. He defeated the marshal de Châtillon (Gaspard III de Coligny) at La Marfée on July 6, 1641, but was killed by a mysterious shot at the moment of his victory.

The rebel's surviving sister, Marie, had married Thomas of Savoy, prince of Carignano (*see* SAVOY, HOUSE OF) in 1625. Their youngest son, Eugène Maurice de Savoie-Carignan (1633–73), who married Cardinal Mazarin's niece Olimpia (Olympe) Mancini in 1657, assumed the title comte de Soissons and served in Louis XIV's armies. The comtesse, after a brief liaison with Louis XIV, had to leave France suddenly in 1680 to avoid interrogation over the Affair of the Poisons (*q.v.*); the allegation that

she poisoned the queen of Spain, Maria Luisa of Orléans, in 1689, is discredited. Her eldest son Louis Thomas (1658–1702), comte de Soissons from 1673, followed his brother, the famous Prince Eugene (*q.v.*) of Savoy, into the Austrian service in 1695. His grandson Eugène Jean François (1714–34), duke of Troppau and, from 1729, titular comte de Soissons, died childless.

In 1888 King Humbert I of Italy granted the title of conti di Villafranca-Soissons to the issue of Prince Eugenio Emanuele of Savoy-Carignano (1816–88; a member of a junior branch of the senior line of Thomas of Carignano's descendants) by his morganatic marriage (1863) to Felicita Crosio.

SOISSONS, a city of France, capital of an *arrondissement* in the *département* of Aisne, Picardy region, 60 mi. (97 km.) NE of Paris by road. Pop. (1962) 22,890. The city, surrounded by wooded hills, stands on the left bank of the Aisne, the suburb of Saint-Vaast lying on the right bank. The Cathedral of Saint-Gervais and Saint-Protais, partly ruined in World War I, was begun in the second half of the 12th century and finished toward the end of the 13th. It is 328 ft. (100 m.) long and 87 ft. (26.5 m.) wide, and the vaulting of the triple nave is 100 ft. (30.5 m.) above the pavement. The single tower dates from the middle of the 13th century and is an imitation of those of Notre Dame of Paris at a height of 216.5 ft. (66 m.). The south transept, the oldest and most graceful portion of the building, terminates in an apse. The facade of the north transept dates from the end of the 13th century. The apse and choir retain some fine 13th-century glass.

Considerable remains exist of the magnificent Abbey of Saint-Jean-des-Vignes, where Thomas Becket resided for a time. These include the ruins of two cloisters (the larger dating from the 13th century), the refectory, and above all the imposing facade of the church (restored). Above the three portals (13th century) runs a gallery, over which is a large window; the two unequal towers (the higher being 256 ft. [78 m.]) of the 15th and early 16th centuries are surmounted by beautiful stone spires, which command the town. The former abbey church of Saint-Léger, which belongs to the 13th century, now houses a museum.

The wealthiest of all the abbeys in Soissons, and one of the most important of all France during the first two dynasties, was that of Saint-Médard, in the Saint-Vaast district, founded about 560 by Clotaire I, beside the villa of Syagrius, which had become the palace of the Frankish kings. St. Médard, apostle of Vermandois, and kings Clotaire and Sigebert were buried in the monastery. It was there that Childeric III, the last Merovingian, was deposed, Pepin (Pippin) III the Short was crowned by the papal legate, and Louis the Pious was kept in captivity in 833. In 1530 Saint-Médard was visited by a procession of 300,000 pilgrims. But the religious wars ruined the abbey and, although it was restored by the Benedictines in 1637, it never recovered its former splendour. Now only the crypt, dating from around 840, remains.

Soissons is an important road junction and on the Paris-Laon railway. Among the industrial establishments are iron and copper foundries and factories, producing boilers, agricultural implements and other iron goods, rubber goods, cardboard and packaging materials, glass, enamel, and sugar. There is a large trade in grain for the provisioning of Paris.

History.—Soissons is generally identified with the oppidum of Gallia Belgica, called Noviodunum by Caesar. Noviodunum was the capital of the Suessiones, who occupied 12 towns, and whose king, Divitiacus, one of the most powerful in Gaul, had extended his authority even beyond the sea among the Britons. In 58 B.C. Galba, king of the Suessiones, separated from the confederation of the Belgians and submitted to the Romans. At the beginning of the empire Noviodunum took the name of Augusta Suessionum, and afterward that of Suessiona, and became the second capital of Gallia Belgica, of which Reims was the metropolis. The town was before long surrounded with a wall and defended by a citadel, and it became the starting point of several military roads.

Christianity was introduced by St. Crispin and St. Crispinian, who were martyred in *c.* 286 but whose work was continued by St. Sinitius, the first bishop of Soissons. After the barbarians had crossed the Rhine and the Meuse Soissons became the metropolis of the Roman possessions in northern Gaul, and on the defeat of Syagrius by Clovis in 486, the Franks seized the town. It was at Soissons that Clovis married Clotilde, and, though he later settled in Paris, Soissons was the capital of his son Clotaire I, and afterward of Chilperic I, king of Neustria. In the time of Chilperic's son, Clotaire II, the kingdom of Soissons was incorporated with that of Paris. Louis the Pious did penance there after being deposed by the assembly at Compiègne.

Under Charles the Fat (886) the Normans failed to win the town, but laid waste Saint-Médard and the neighbourhood. In 923 Charles the Simple was defeated outside the walls by the supporters of Rudolph of Burgundy, and Hugh the Great besieged and partly burned the town in 948. Under the first Capets Soissons was held by hereditary counts. (*See* SOISSONS, COMTES DE.)

The communal charter of the town dates from 1131. At a synod held at Soissons in 1121 the teachings of Abelard were condemned and he was forced to retract them. In 1155, at an assembly of prelates and barons held at Soissons, Louis VII issued a famous decree forbidding all private wars for a space of ten years; and in 1325 Charles the Fair replaced the mayor of Soissons by a royal provost dependent on the bailiwick of Vermandois.

The town suffered severely during the Hundred Years' War, in 1414, when it was held by the Burgundians, it was captured and sacked by the Armagnacs under the dauphin. The Treaty of Arras (1435) brought it again under the royal authority. It was sacked by Charles V in 1544 and in 1565 by the Huguenots.

In 1814 Soissons was captured and recaptured by the Allies and the French. In 1815, after Waterloo, it was a temporary rallying point for the vanquished. In 1870 it capitulated to the Germans after a bombardment of three days.

During World War I Soissons was for most of the time just behind the Franco-British lines, but the Germans overran it in 1918 in their thrust for Paris (May 27); it was retaken in the Franco-British offensive of July 18, 1918. In World War II the town sustained some damage, and the bridges were blown up and the railway station completely destroyed by the explosion of a U.S. munitions train.

SOKOLLU (SOKOLLI) **MOHAMMED PASHA** (1505–1579), one of the most celebrated of the Ottoman grand viziers, was born at the village of Sokol in Bosnia. Recruited into the Ottoman service through the *devshirme* (child-tribute) levied in the Balkan lands, he was educated at the Porte during the earlier years of the reign of Sultan Suleiman the Magnificent (1520–66). He became *kapudan pasha* (high admiral of the Ottoman fleet) in 1546 and then *beylerbeyi* (governor-general) of Rumeli. During his tenure of this latter office he conquered Temesvár in Hungary (1552). He also took part in the third campaign of Sultan Suleiman against Persia in 1554–55, being raised thereafter to the rank of vizier. Mohammed Pasha commanded the forces of Selim during the conflict (1559–61) between Selim and Bayazid, the sons of Sultan Suleiman, over the succession to the throne and in 1562 received in marriage the hand of Esmi-Khan, a daughter of Selim. He was made grand vizier in June 1565 and retained this office until his death in October 1579. During the 14 years and more of his grand vizierate he became in effect the real ruler of the empire, although he had to face powerful influences which opposed his personal domination and the policies that he advocated. The main event during his tenure of the grand vizierate was the war against Venice (1570–73) which led to the Ottoman conquest of Cyprus. Mohammed Pasha was in general inclined to maintain peaceful relations with foreign powers. Reluctant to enter into the war with Venice, he also opposed the decision to begin a new conflict (1578–90) against Persia. After the death of Sultan Selim II in 1574, however, his influence at the Porte became much diminished. Mohammed Pasha was assassinated in Istanbul on Oct. 11, 1579. (V. J. P.)

SOKOLOW, NAHUM (1861–1936), Jewish writer and Zionist leader, was president of the World Zionist Organization and of the Jewish Agency (1931–35) and chairman of the World Zionist Executive (1922–31). Descendant of an ancient rabbinical family in Poland, he was born at Wyszogrod on Feb. 3, 1861. He displayed remarkable gifts as a writer and scholar at an early age and was well-known by his contributions to the Jewish

press in Hebrew and other languages. At 24 he became assistant editor of the Hebrew scientific weekly *Ha-Zefirah* in Warsaw and later its editor, transforming it into a modern daily newspaper with wide circulation. He also edited in Warsaw the literary and historical periodicals *Ha-Asif* and *Sefer Ha-Shanah* (1885–1902).

In 1897 Sokolow joined the Zionist Organization. In 1906 he became its secretary general and editor of its official organs *Die Welt* and the Hebrew weekly *Ha-Olam.* He traveled all over the world propagating Zionist ideas. After the outbreak of World War I he went to England, becoming later a naturalized British subject. He took a prominent part in Anglo-French negotiations leading to the Balfour declaration of Nov. 2, 1917 (*see* ZIONISM). Sokolow secured similar declarations in favour of a Jewish National Home from France, Italy, Poland, South Africa, and other countries. In May 1917 Sokolow was received in the Vatican by Pope Benedict XV, who expressed his sympathy for the Zionist cause.

At the Paris Peace Conference of 1919 Sokolow led the Zionist delegation and later was instrumental in obtaining minority rights for the Jews in some East European countries. He died in London on May 17, 1936. In 1960 his remains were transferred to Israel.

His numerous writings include *The History of Zionism, 1600–1918,* 2 vol. (1919); *Hibbath Zion* ("Love of Zion," 1934); *Baruch Spinoza u-Zemano* ("Spinoza and his Time," 1928–29); *Yischim* (biographical essays), 3 vol. (1911–35).

See *Sefer Sokolow,* in Hebrew (1960); S. Kling, *Nachum Sokolow* (1960). (C. So.)

SOKOTO, a town and provincial capital of Northern Nigeria, Africa, lies on the south bank of the Sokoto River, about 621 mi. (1,000 km.) NNE of Lagos by road, at approximately 500 ft. (152 m.) above sea level. Situated in a well-populated area, Sokoto is the most northerly of the larger towns of Nigeria. Pop. (1963) 89,817. To the north of the river the land develops into rolling sand dunes that merge into the desert across the border in the Republic of Niger. West of Sokoto the Rima River joins the Sokoto and creates a fertile area around the town.

In the 1820s Sokoto was adopted as his capital by Sultan Bello, one of the sons of Shehu Usman Dan Fodio, after the successful *jihad* (holy war) which had placed the Fulani dynasties in control of Northern Nigeria. Sokoto was recognized by the Fulani amirs as the capital of the empire and the sultans of Sokoto have since continued to be regarded as the *sarkin mussulmi* or leaders of all Muslims in the western Sudan. The tomb of Dan Fodio and other holy shrines have made the town a place of pilgrimage.

Sokoto is a typical Hausa-Fulani town of mud dwellings, though the more prosperous merchants build houses faced with concrete. The sultan's palace is in the centre of the town, close by the mosque of the Shehu. Houses are generally grouped in compounds with a retaining wall within the main wall of the town. The Sokoto River is subject to periodic flooding which made it necessary to construct an approach road to the town over a raised causeway. Sokoto lies virtually at the northwestern end of Nigeria's internal system of communications. There is an airstrip which has connections with Kaduna to the east and a tarmac road links Sokoto with Kano, via Gusau, where a branch of the Nigerian Railway passes en route to Kaura Namoda, the railhead. There are road links southward to Argungu and Birnin Kebbi and thence to Yelwa on the Niger. There is also a road to the north into the desert where it joins up with the ancient caravan routes.

Sokoto is an important centre of commerce for an area extending far into the western Sudan. The famous red goats of Sokoto supply the leather of the incomparable moroccan leather work while Sokoto itself has a flourishing tanning and dyeing leather industry. There is also hand weaving and pottery, and large numbers of people are engaged in fishing on the Sokoto River and in agricultural pursuits. Caravans from the north bring in salt and potash from the desert, the potash being required for the tanning of leather. To the west of Sokoto successful experiments are being carried out for the cultivation of rice. Sokoto provides a meeting place and commercial exchange centre for the desert tribes and for pilgrims, who come from as far away as Ilorin and Katagum, and in exchange for the salt and potash take away the palm oil, kola nuts, soap, and cloth produced further south.

(W. H. I.)

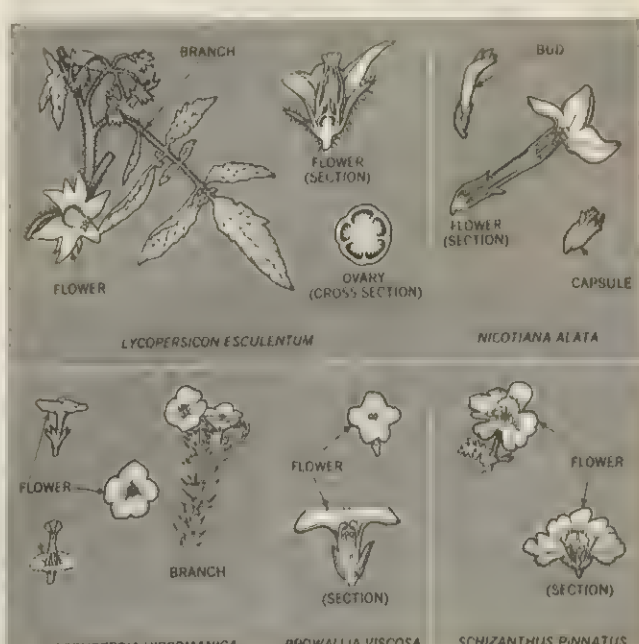
SOL, the sun, a name given to two distinct deities at Rome. The original Sol, or Sol Indiges, had a shrine on the Quirinal and an annual sacrifice on Aug. 9. The worship appears to be native. The Roman poets, however, equate him with the Greek Helios.

The worship of Sol assumes an entirely different character with the later importation of various sun cults (*e.g.,* Mithraism; *q.v.*) from the east (Syria). Elagabalus (*q.v.*) built a temple to him as Sol Invictus on the Palatine and attempted to make his worship the principal religion at Rome. Aurelian reestablished the worship and erected a magnificent temple to Sol in the Campus Agrippae. The worship of Sol as special protector of the emperors and the empire remained the chief imperial cult until Christianity replaced it. (R. B. LD.)

SOLANACEAE, the nightshade family of plants, of considerable economic importance, includes a number of species that produce food and drug narcotics or are used as fumitories or ornamentals. Prominent among these are the potato and eggplant (*Solanum*), tomato (*Lycopersicon*), garden pepper (*Capsicum*), tobacco (*Nicotiana*), belladonna (*Atropa*), thorn apple (*Datura*), and henbane (*Hyoscyamus*). Some of the genera that contain ornamental species are *Petunia, Lycium, Solanum, Nicotiana, Datura, Salpiglossis, Browallia, Brunfelsia, Cestrum, Schizanthus, Solandra, Streptosolen,* and *Nierembergia.*

This family of dicotyledons belongs to the order Tubiflorae. It includes about 90 genera with approximately 2,800 species, which, though found throughout the world except in polar regions, are most abundant and widely distributed in the tropical regions of Latin America, where about 40 genera are endemic. Very few are to be found in temperate regions, and about 50 species are found in North America, north of Mexico. Only four species in three genera are known from Britain. The genus *Solanum,* which embraces more than two-thirds of all the species in the family, includes all the wild potatoes, which are distributed primarily at high elevations in the mountains from Utah and Colorado southward through Mexico and Central and South America. Some of these wild potatoes, especially *Solanum demissum* of Mexico, have been widely used in the improvement of the domesticated potato.

Alkaloids, especially the glucosidal alkaloid solanine, are the narcotic-poisonous principles prevalent in many of the species. These have given the family its sombre vernacular name "nightshade." The alkaloids are of increasing interest and importance to the



FROM L. BAILEY, "MANUAL OF CULTIVATED PLANTS," 1949; THE MACMILLAN CO.

REPRESENTATIVES OF THE FAMILY SOLANACEAE

chemical and pharmaceutical industries. Many species in the Solanaceae have been used by primitive people in various parts of the world in their folklore and therapeutic practices.

Included in the family are annual herbs, as *Solanum nigrum*, a common weed in waste places, and perennial herbs, as *Atropa belladonna*, the deadly nightshade; shrubs, as *Lycium halimifolium*, an old-fashioned garden plant; small trees, as *Datura sanguinea*, a South American ornamental with large reddish flowers; lianas, as *Solanum juglandifolium*, a showy rampant vine of South America; and prostrate creepers, as some Latin-American species of *Solanum*. The alternate leaves (or sometimes opposite near the inflorescence) are simple and entire, as in most species of *Lycium*, or are variously dissected, as represented by the extreme condition found in various species of *Lycopersicon* (tomato). The flowers are usually borne in extra-axillary cymes.

The hermaphroditic, generally regular, flowers have their parts in fives—five sepals, five petals, five stamens, all in alternating whorls. There are two carpels, which are generally placed obliquely. The corolla is regular and rotate, as in *Solanum tuberosum*, or bell shaped, as in *Datura* and *Atropa*, or somewhat irregular as in *Hyoscyamus;* in the tribe Salpiglossideae, which forms a link with the closely allied family Scrophulariaceae (*q.v.*), it is zygomorphic, forming, as in *Schizanthus*, a two-lipped flower. The stamens are inserted on the corolla tube and alternate with its lobes; in zygomorphic flowers only two or four fertile stamens are present. The flowers are generally conspicuous, and honey is secreted on the disk at the base of the ovary or at the bottom of the corolla tube between the stamens. The ovary is usually bilocular but in *Capsicum* becomes unilocular above, while in some cases an ingrowth of a secondary septum makes it four celled, as in *Datura*, or irregularly three to five celled, as in *Nicandra*. The anatropous ovules are generally numerous on swollen axile placentae. The style is simple and bears a bilobed or capitate stigma. The fruit is a many-seeded berry, as in *Solanum*, or capsule, as in *Datura*, where it splits lengthwise, and *Hyoscyamus*, where it opens by a transverse lid forming a pyxidium. The embryo is bent or straight and embedded in endosperm. The persistent calyx may serve to protect the fruit or aid in its distribution, as in the bladdery structure enveloping the fruit of *Physalis*.

Some additional species of less importance than those already considered are *Physalis alkekengi* (winter cherry) and *P. peruviana* (cape gooseberry); *Lycopersicon pimpinellifolium* (currant tomato) and *Cyphomandra betacea* (tree tomato), both of South America; *Datura metel* of India and *D. meteloides* of southwestern United States, large-flowered annuals that are grown in gardens in warm-temperate countries, as are the treelike shrubs *D. suaveolens* of Brazil and *D. arborea* of the central Andes; *Solanum muricatum* (pepino), *S. pseudocapsicum* (Jerusalem cherry), and *S. integrifolium* (scarlet eggplant), also cultivated for their fruits and aesthetic appearance. (D. S. Cl.)

SOLAR ENERGY, UTILIZATION OF. Energy is radiated from the sun at the rate of about 43,000 kw. per square metre of solar surface. Solar radiation occupies only the small portion of the electromagnetic spectrum between 0.22 and 3.3 microns (μ). The invisibly short but potent ultraviolet waves below 0.4μ carry 9% of the incoming solar energy, while the invisible infrared heat waves beyond 0.7μ are responsible for 50% of the total. The "light" waves between these limits convey the remaining 41%, including the maximum intensity radiation at 0.48μ in the green portion of the visible spectrum. Both the amount and the spectral distribution of the energy radiated from the sun indicate that the effective temperature of its surface is close to 6,000° C (10,800° F).

Solar radiation falling on a surface normal to the sun's rays at the outer limit of the earth's atmosphere has an average intensity of 2 (±2%) cal. per minute per square centimetre (1.41 kw. per square metre; 445 BTU per hour per square foot). This quantity, known as the solar constant, has apparently remained virtually unchanged, except for periodic minor variations, for several billion years. Only about 46% of the incoming solar energy actually reaches the earth's surface, since, on the average, 35% is reflected back into space by clouds and the remaining 19% is absorbed by the atmosphere. The intensity of the sunshine reaching any particular area on the earth varies from zero at sunrise through a noon maximum that may be as high as 300 to 350 BTU per hour per square foot on a cloudless day. Local weather bureau data must be consulted to obtain specific values for any particular locality (see SUN; SUNSHINE).

The total amount of solar energy reaching the earth is about 7×10^{17} kw-hr. per year, more than 30,000 times as much as that used in all man-made devices. This prodigious inflow of energy warms the earth and produces, through photosynthesis, all of the food, fuel, and oxygen upon which life depends.

Though man has sought for many centuries to make direct use of the sun's radiant energy, his success has been meagre, primarily because sunshine is intermittent in availability, variable in direction, and relatively low in intensity. Apparatus intended for use with solar energy must in general be large in area and must be provided with some means for storing energy for use when the sun is not shining. Despite these formidable difficulties, significant advances have been made since 1750 in using solar energy to generate heat, to bring about chemical reactions, and, most recently, to produce usable amounts of electricity.

Early Uses.—Man learned long ago that heat is produced when the sun's rays are absorbed by a blackened surface and that the temperature thus produced could be greatly increased by using reflection or refraction to concentrate a large area of sunshine onto a small target. The concave silver mirrors used by the Incas and the convex quartz lens found in the ruins of Nineveh were probably used to light sacred fires by means of concentrated sunbeams. Archimedes is said to have devised a battery of mirrors to defend Syracuse in 212 B.C. by burning the sails of an invading fleet "at the distance of a bowshot." In 1747 G. L. L. Buffon set up a group of 140 flat mirrors in a Paris garden and ignited a stack of wood placed about 200 ft. from his reflectors, thus demonstrating that the feat attributed to Archimedes might indeed have been accomplished.

A. L. Lavoisier, who experimented with solar energy as early as 1774, was probably the originator of the device known today as the solar furnace. Enclosing specimens of various substances in transparent quartz vessels, he placed them at the focal point of a 52-in. diameter lens and used concentrated solar radiation to heat them in partial vacuum and in controlled atmospheres of oxygen and other gases. He first called attention to one of the principal virtues of the solar furnace when he wrote, "the fire of ordinary furnaces seems less pure than that of the sun."

Modern Applications.—*Heat Production.*—Because of its unique ability to heat materials by intense radiation alone for long periods of time without contamination, the solar furnace has come into prominence as a tool for high-temperature research. Scientists have found that parabolic reflectors taken from antiaircraft searchlights can concentrate solar radiation so effectively that temperatures as high as 3,500° C (6,332° F) can be attained. Equatorial or altazimuth mountings are used to enable the concentrator to follow the apparent motion of the sun across the sky; sleeves or shutters are used to regulate the amount of radiation impinging on the target.

In most of the large solar furnaces the inconvenience caused by target movement is eliminated by using a flat mirror called a heliostat to track the sun across the sky and reflect its rays into a fixed paraboloidal concentrator. The largest furnaces of this type are those at Mont-Louis, France; Sendai, Japan; and Natick, Mass. These furnaces use concentrators that are more than 30 ft. in diameter, made of many small curved glass mirrors arranged approximately in a paraboloidal configuration.

Solar steam generators, built by many pioneers during the 19th century, used arrangements of movable mirrors to concentrate large amounts of solar radiation upon blackened pipes through which water was circulated and turned to steam. Both steam and hot-air engines were operated in this way with some degree of success as early as 1870, and ice was produced in Paris in 1878 in an ammonia-absorption refrigerator operated by a solar boiler. The use of solar energy for pumping irrigation water was tried in Arizona and California, employing a number of solar pumping

stations using conical concentrators 30 ft. in diameter. The largest of all solar-power installations was erected in 1912 on the bank of the Nile at Meadi, near Cairo, using a total of 14,000 sq.ft. of concentrating surface in the form of seven parabolic troughs, each 205 ft. in length. The 100-hp. steam engine connected to this great solar boiler actually produced between 50 and 60 hp. continuously during one five-hour test, but the system was not economically competitive with other pumping apparatus and was abandoned during World War I.

All of these early installations suffered from the same deficiencies—irregularity of operation and excessive cost—and none survived. Interest in solar power revived with the advent of artificial earth satellites, since these applications avoid such earthly problems as clouds, atmospheric absorption, and low-cost competitive energy sources. Serious consideration is again being given to vapour cycles which would receive the full complement of solar radiation, 1.4 kw. per square metre, whenever their collectors turn toward the sun and reject the unused energy by radiation into space.

When heat is needed at moderate temperatures for such purposes as distillation of salt water, growing and drying of agricultural products, cooking, and heating of buildings and domestic hot-water supplies, solar energy can be used with considerable success, thanks to the "greenhouse" effect. Common window glass and many plastic films are able to transmit 90% of the solar radiation that falls on them, but they are relatively opaque to the long waves (8 to 40μ in wave length) emitted by any surface heated to 100°–200° C (212°–392° F). Thus a simple blackened wooden box, covered with several sheets of glass or plastic film, acts as an effective trap for solar energy when it is turned toward the sun. If its sides and bottom are insulated, the internal temperature can be raised as high as 150° C (302° F) by the unaided rays of a bright sun. M. K. Ghosh in India (1945) and M. Telkes in the United States (1955) improved this simple solar stove by adding reflective wings to direct more sunshine into well-insulated ovens, achieving temperatures up to 200° C (392° F).

Production of salt by evaporation of sea water was probably man's first planned use of solar energy. Conversely, the production of potable water from brackish wells or sea water was the first successful large-scale application. A solar still erected in Chile in 1872 used 51,000 sq.ft. of collecting surface and produced some 6,000 gal. of water per day during a period of 40 years. These stills usually employ a shallow blackened box covered with a single layer of glass or plastic film, through which solar radiation enters and evaporates some of the water. The vapour is condensed on the lower surface of the transparent cover and trickles down into channels that lead the condensate into appropriate containers. Since the output of such solar stills is relatively low (approximately 1 gal. per day for each 10 sq.ft. of collection area), they offer little promise for producing irrigation water. They can, however, supply the drinking water requirements for limited numbers of people, and they are being used for this purpose on several of the arid Greek islands, notably at Simi.

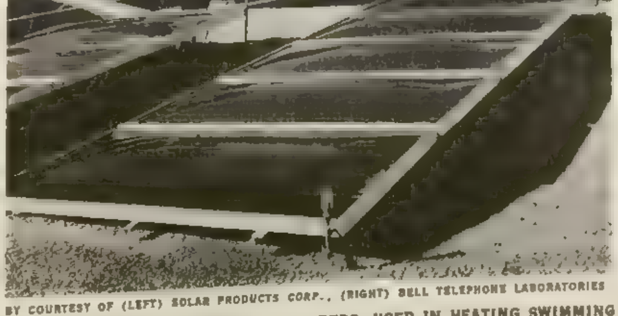
Solar water heaters have been in relatively wide use in Florida and the southwestern U.S. for many years. Most of these consist of flat sheets of blackened metal provided with tubes through which water circulates by natural or forced convection whenever the sun is shining. Wastage of the absorbed solar energy is minimized by covering the upper surface of the metal with a glazing of glass or plastic film, separated from the sheet by a small air space. The lower surface is insulated with rock wool, vegetable fibres, or a reflective coating that reduces radiation. Heaters of this type do not have to follow the sun's motion, since they operate effectively if they are mounted in a fixed position facing the south (in the Northern Hemisphere) and tilted so that their angle with the horizontal is equal to the local latitude plus 20°. Temperatures up to 60° C (140° F) can readily be reached on sunny winter days, and the domestic hot-water requirements of a typical family living in regions south of latitude 35° N can be met during most of the year by a heater having as little as 40 sq.ft. of surface.

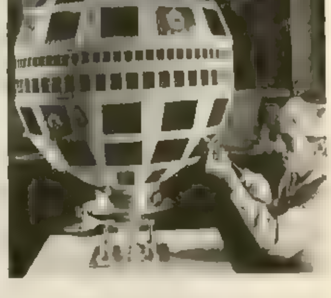
Similar but much larger collectors can supply most of the heat needed in homes of moderate size in favourable localities. Storage of heat for use at night or during cloudy periods is accomplished by using large insulated tanks to store water heated during daylight hours. This system is particularly effective when radiant heating is used, with the warmed water flowing through tubes in floors and ceilings. Air can also be heated in suitably designed collectors and used in conventional warm-air systems. Heat storage can then be provided by using beds of gravel or containers filled with chemicals (*e.g.*, Glauber's salt) that absorb heat at a suitable temperature when they melt and give off the stored heat when they solidify again.

Solar Power Systems.—Direct conversion of solar radiation into electricity can be accomplished on a small scale in several ways. The thermoelectric effect, first discovered in 1821 and greatly improved during the 1950s by both Russian and U.S. physicists, can be utilized by placing one set of thermocouple junctions in the focal region of a solar furnace while the other set is kept cool. The efficiency of conversion of solar to electric energy was extremely low, less than 1%, when only metallic thermocouple junctions were available. An important advance came with the development of semiconductors, which transmit electricity well but are poor conductors of heat; these give energy conversion efficiencies reportedly as high as 10%. The most efficient and convenient way to convert solar radiation into electricity is through the use of the silicon photovoltaic cells, which were developed in the U.S. in 1954 by the Bell Telephone Laboratories. The Bell solar cell is made of thin wafers of ultrapure silicon to which traces of arsenic and boron are added. In bright sunlight it can produce direct current at 0.6 v. and approximately 0.030 amp. per sq.cm. of exposed cell area, with a conversion efficiency as high as 16%.

Terrestrial solar-power systems have achieved little success because of irregularity of operation and excessive cost. In space, the situation is entirely different, because the supply of solar energy is certain and predictable, and there are no competitive sources of unending energy. The power requirements of transistorized radio and television receivers and transmitters are well within the capability of a moderate area of solar batteries, and their cost, while far too high for earthly applications, is moderate in comparison with their usefulness.

Solar radiation was first used as a source of electricity in space by the U.S. satellite Vanguard I, which was put into orbit on March 17, 1958, carrying 108 silicon solar cells in six groups of 18 cells each. Radiation encountered in the Van Allen belt caused the cells to deteriorate, but in the mid-1960s they continued to





BY COURTESY OF (LEFT) SOLAR PRODUCTS CORP., (RIGHT) BELL TELEPHONE LABORATORIES

(LEFT) MODULAR FLAT-PLATE ABSORBERS, USED IN HEATING SWIMMING POOLS; (RIGHT) TECHNICIAN ATTACHES SOLAR BATTERIES TO THE SURFACE OF A PROTOTYPE OF THE TELSTAR COMMUNICATIONS SATELLITE

transmit intelligible signals. Subsequent satellites and space vehicles have used much larger numbers of cells arranged in series-parallel circuits so that they could charge sealed storage batteries and thus provide relatively large amounts of power for short periods of transmission. (*See* SPACE EXPLORATION.)

Weather and communications satellites (Tiros, Relay, Telstar) have their outer surface literally covered with thousands of silicon cells, some of which are always facing the sun and generating the power needed to charge their batteries and to operate their radio and television equipment. Space vehicles intended to transmit signals over vast distances employ more than 20,000 cells mounted on panels that can be folded into the launch rocket and then deployed when the desired course has been attained. These vehicles (Ranger, Mariner) are established at such attitudes that their solar panels face the sun continuously.

Future Uses.—To meet the larger power requirements of second-generation space vehicles, thermodynamic power systems that are being developed will use liquid metals as their working fluids, with metallic hydrides for high-temperature heat storage. Very large paraboloidal concentrators have been devised to fit within the narrow confines of a rocket's nose cone and later to unfold or inflate in space. Thermionic converters are also being developed for space applications, since they are apparently relatively immune to radiation damage and can operate at the very high temperatures attained in the focal region of solar concentrators.

Biological systems also under development use sunlit algae to convert exhaled carbon dioxide and water into oxygen and protein-rich carbohydrates. These systems can ultimately be used to sustain astronauts on protracted excursions into space.

The major potential areas of usefulness for terrestrial applications of solar energy lie between 30° N and 30° S latitude, where sunshine is plentiful but conventional energy sources are scarce and fossil fuels expensive. Small solar-powered pumps and refrigerators, still to be developed, would help the billions of people living in these regions to raise their standards of living. Where heat is needed rather than mechanical power or electricity, solar devices are already technologically available, but fuel-burning equipment is still less expensive in most parts of the world.

BIBLIOGRAPHY.—Association for Applied Solar Energy, *Proceedings of the World Symposium on Applied Solar Energy, 1955* (1956); F. Daniels, *Direct Use of the Sun's Energy* (1964); A. M. Zarem and D. D. Erway (eds.), *Introduction to the Utilization of Solar Energy* (1963). (J. I. Y.)

SOLARI (SOLARIO), **ANDREA** (ANDREA DEL GOBBO) (active 1495–d. 1524), Italian painter of the Milanese school and one of the most important followers of Leonardo da Vinci, was probably born at Milan. He received his early training from his brother Cristofano, a distinguished sculptor and architect, who was employed extensively on work at the cathedral at Milan and at the Certosa di Pavia. He probably accompanied his brother to Venice, where he seems to have been strongly influenced by Antonello da Messina, who was then active in the city. The fine portrait of a man with a pink (National gallery, London) displays Antonello's plastic conception of form and was probably painted about 1492. Solari's earliest dated work is a "Holy Family and St. Jerome" (Brera, Milan), with a fine landscape background, executed for S. Pietro at Murano in 1495. The Leonardesque type of the Madonna proves that Solari after his return from Venice became strongly influenced by the great Florentine artist, who was then carrying everything before him. To this period belong a small "Crucifixion" (1503, Louvre, Paris); the portrait of Charles of Amboise (Louvre); the portrait of Longono (1505, National gallery, London); "The Annunciation" (1506, Fitzwilliam museum, Cambridge); the beautiful "Virgin on a Green Cushion" (Louvre), for which a sensitive drawing of the Virgin's head is in the Ambrosiana at Milan; and the "Head of the Baptist in a Silver Charger" (1507, Louvre). In 1507 Solari went to France with letters of introduction to the cardinal of Amboise and was employed for two years on frescoes in the chapel of his castle of Gaillon in Normandy (demolished during the French Revolution). It has been suggested that Solari may have visited Flanders before returning to his native country, and this may account for the Flemish character of his later work. In 1515 he painted the "Flight Into Egypt" (Poldi-Pezzoli, Milan), with its harmonious and detailed landscape background. To this period belong the "Procession to Calvary" (Borghese gallery, Rome); the portrait of the chancellor Domenico Morone (Palazzo Scotti, Milan); and the "Woman Playing a Guitar" (Hertz collection, Rome). Solari's last work was an altarpiece representing "The Assumption of the Virgin," left unfinished at his death and completed by Bernardino Campi about 1576. Solari died between Aug. 18 and Oct. 7, 1524.

See K. Badt, *Andrea Solario* (1914); W. Suida in *Art Quarterly*, viii, pp. 16 ff. (1945). (I. A. R.; X.)

SOLAR SYSTEM consists of the sun and the complex assemblage of bodies revolving around it in closed orbits under the dominating influence of its gravitational attraction. For a detailed discussion of the various members of the solar system *see* SUN; PLANETS; COMET; METEOR; METEORITES; MOON; and the planets under their individual names (*see* below). Only a brief survey of the solar system as a whole is given here.

The solar system is usually defined as extending to the orbit of the outermost known planet, Pluto, 40 astronomical units from the sun. (The astronomical unit is equal to the semi-major axis of the earth's orbit, 93,000,000 mi. in length.) If the solar system is considered to extend to the aphelia of the comets with nearly parabolic orbits, however, its extent is much larger—approximately 100,000 astronomical units. More than 99% of the mass of the solar system is in the sun. The rest of the solar system consists of empty space traversed at wide intervals by bodies of relatively insignificant mass. These bodies may be grouped into two solar families: (1) the 9 major planets, with 31 satellites attending 6 of them, and the 1,600 known asteroids; and (2) about 1,000 comets, a horde of meteors and meteorites, and a thin cloud of interplanetary dust. These are the known members of the solar system. It has been estimated that there are 40,000 asteroids bright enough to be photographed with a 100-in. telescope. A long search with a large telescope would probably reveal more satellites, especially of the giant planets. New comets are discovered at the rate of half a dozen per year on the average. But it is very doubtful if any unknown distant major planets brighter than the 16th magnitude exist.

The Sun.—The sun is the only star close enough so that its surface can be viewed from the earth. All other stars are so distant that they appear merely as points of light in the largest telescopes. In a sense we may be in actual contact with the sun at times, for there is evidence that the earth is occasionally enveloped in particles emitted by the solar surface. Compared with the average star, the sun is slightly more luminous, slightly larger and more massive, and slightly less dense. It is 109 times the diameter of the earth and 333,000 times as massive. The temperature at the surface of the sun is 6,000° K and is believed to increase to perhaps 14,000,000° K at the centre. Although the density of material near the centre of the sun is eight times the density of lead, it is gaseous throughout.

The sun radiates at the rate of 9×10^{25} cal. per sec., emitting enough energy to melt a 3,000-ft. layer of ice surrounding it in 90 min. This energy is derived from thermonuclear processes involving the transformation of hydrogen into helium. Although the sun loses 4,000,000 tons of hydrogen per second in this process, it is believed that the sun will continue to radiate at essentially its present rate for hundreds of millions of years.

The bright surface of the sun that is seen from the earth is called the photosphere. Above the photosphere is a region called the chromosphere. During a total solar eclipse, when the light of the photosphere is completely covered by the moon, the chromosphere appears as a brilliant scarlet ring surrounding the sun. Often ruby-coloured clouds or prominences may also be seen at the edge of the sun. The most spectacular sight is the halo of pearly white light or corona that flashes out around the sun during the total phase of an eclipse. The spectrum of the corona shows bright lines that have been identified with lines emitted by atoms of iron, nickel, and calcium ionized from 9 to 15 times.

The high level of ionization in the corona indicates a temperature of about 1,000,000° K. With special instruments it is possible

to make certain observations of the prominences, chromosphere, and even the corona without waiting for a total eclipse.

The most conspicuous markings visible directly on the disk of the sun with only a small telescope are the groups of dark spots (sunspots). They range in size from mere specks a few hundred miles in diameter to areas 30 times the equatorial cross section of the earth. These large spot groups can be seen by the unaided eye when suitably protected by dark glasses. (A telescope should never be pointed at the sun without using proper filters.) The spots appear dark only by contrast with the hotter photosphere around them, being at a temperature of about 4,500° K. The number of spot groups increases and decreases at irregular intervals that average 11 yr. in length.

The Major Planets.—The major planets in order of increasing distance from the sun are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, and Pluto. These may be divided into the terrestrial planets, Mercury, Venus, Earth, and Mars, nearest the sun; and the giant planets, Jupiter, Saturn, Uranus, and Neptune, beyond them. Pluto is believed to be about 0.45 the size of the earth but is beyond the orbit of Neptune. There are reasons for believing that Pluto may once have been a satellite of Neptune, which escaped from its gravitational control. The terrestrial planets are of about the same size and mass as the earth and of high density. The giant planets have from 15 to 318 times the volume of the earth but have only about 20% the density of the terrestrial planets. The high density of the terrestrial planets is attributed to their large metallic content. The giant planets are believed to consist mostly of the lightest elements, hydrogen and helium, which are probably found deep in the interior of the planets in a solid or possibly even metallic state.

The Asteroids or Minor Planets.—The orbits of more than 1,600 asteroids have been computed since the discovery of the first one in 1801. The size of these minor planets ranges from about 500 mi. in diameter to one mile or less. Their total mass is estimated at only 1/3,000th that of the earth. Almost all of the asteroids revolve between the orbits of Mars and Jupiter. But a few exceptional ones have elongated orbits that extend from Mercury to Mars, in the case of Icarus, and from Mars to Saturn, in the case of Hidalgo. Some of the exceptional asteroids, such as Hermes and Eros, have come within a few million miles of the earth.

Natural Satellites.—A satellite is defined as a body which revolves around a planet. It is thus distinguished from a planet by its motion rather than by any physical property. Mercury, Venus, and Pluto have no known satellites, the Earth has 1, Mars has 2, Jupiter has 12, Saturn has 9, Uranus has 5, and Neptune has 2. Some satellites are only a few miles in diameter, as in the case of the two tiny moons of Mars and the outer satellites of Jupiter. A few satellites are about the size of Mercury and Mars, as in the case of Ganymede and Callisto of Jupiter, Titan of Saturn, and Triton of Neptune, which are each about 3,500 mi. in diameter.

Comets.—Comets are bodies of great size but small mass. They are popularly pictured as bright stars flashing erratically across the sky followed by long luminous tails. But very few comets become bright enough to be seen without a telescope, and their motions are not erratic but, as with the planets, in accordance with the law of gravitation. Apparent slight deviations from gravitational motion observed in the case of a few comets like Encke's may be due to the action of internal forces within the comet's nucleus.

In contrast to the nearly circular orbits in which the planets and most of the asteroids move, most comets revolve in greatly elongated orbits. Cometary orbits may also be inclined at any angle to the ecliptic. The comets of short period, less than 100 yr., move in orbits of moderate inclination. Halley's comet is the only periodic comet that moves in a retrograde direction. A few comets revolve in nearly circular orbits.

Meteors and Meteorites.—A meteor or "shooting star" refers both to the flashes of light occasionally seen darting across the night sky and to the objects that produce the light. Meteors may be no larger than a grain of sand or a pea. The glow is produced by collisions between the atoms of the meteor and the molecules of the air. Such meteors are consumed in the atmosphere at altitudes of about 60 mi. Some meteors, having a mass of a few ounces to many tons, survive their rush through the atmosphere and reach the earth. They are called meteorites. Micrometeorites a few microns in diameter sift through the atmosphere and reach the earth as dust particles.

When the earth encounters a swarm of meteors moving in an orbit around the sun, there occurs a meteor "shower," during which the number of meteors counted per hour may amount to several dozen or more. Very seldom does a display resembling a real shower occur. But there have been a few times when meteors appeared to stream from the sky as thickly as snowflakes. The orbital elements of some meteor swarms are so similar to those of certain comets that there can be no doubt that they are identical. The meteoric material evidently originates from debris left behind by the disintegration of the comets.

Interplanetary Dust.—Under favourable conditions a wedge-shaped glow may sometimes be seen along the ecliptic extending upward from the western horizon after sunset or from the eastern horizon before sunrise. This is the zodiacal light which apparently originates from a swarm of dust particles surrounding the sun in the plane of the earth's orbit or ecliptic and extends out to the orbit of the earth and beyond. The total quantity of dust must be very small, as it does not noticeably interfere with the motion of the inner planets.

Another phenomenon attributed to dust or meteorites is the gegenschein or "counterglow," which may sometimes be seen as a faint hazy patch of light in the night sky always directly opposite the sun. It is variously attributed to light reflected from meteorites or dust trapped by dynamic or electrical forces opposite the earth from the sun.

Regularities in the Solar System.—The nine major planets revolve around the sun in the direction customarily referred to as direct. They move in elliptic orbits that depart only slightly from circles. These orbits lie nearly in the plane of the earth's orbit, which is used as a reference plane. Except for the orbit of Pluto, which has a tilt of 17°, no other planet has a tilt to this plane greater than 7°. The sun also rotates on its axis in the same direction that the planets revolve around it. The earth's moon, the 2 satellites of Mars, 8 of the 12 satellites of Jupiter, and 8 of the 9 satellites of Saturn revolve around their planets in this same direction. But four of the small satellites of Jupiter and the outermost satellite of Saturn revolve in a retrograde direction. These tiny bodies may be captured asteroids. However, this explanation is not so plausible in the case of a large satellite of Neptune that also revolves in the retrograde direction.

All of the 1,600 asteroids revolve around the sun in the same direction as the planets and in about the same plane. Most of the asteroids also move in nearly circular orbits. As already noted, there are a few exceptional asteroids that move in elongated orbits like most of the comets. But there are also several comets that move in nearly circular orbits like the majority of the asteroids.

Theories of the Origin of the Solar System.—The orderly arrangement and motion of the planets and asteroids are impressive evidence that they form a true solar system. It seems impossible that such a system could have resulted from chance encounters between these hundreds of bodies. Another significant factor is the extreme isolation of the solar system. The nearest star is 60,000 times as far away as the outermost planet, Pluto. The evidence is thus overwhelming that the members of the solar system had a common origin. Theories of the origin of the solar system have been of two general types: (1) origin by an orderly process of evolution; and (2) origin by catastrophe.

The classic example of the first type is the nebular hypothesis advanced by the French mathematician Pierre Laplace in 1796. He postulated a vast discus-shaped mass of cold gas in slow rotation extending beyond the orbit of the farthest planet. As this nebula contracted under the mutual gravitation of its parts, its rate of rotation necessarily increased to conserve angular momentum. The angular momentum of a body rotating around a centre is the product of its mass, velocity, and the radius connecting it

with the centre. It can be proven that the angular momentum of a closed system cannot change. If, for example, the radius decreases the velocity increases proportionately so that the product remains the same. Eventually the rate of rotation increased so much that the centrifugal force at the periphery of the nebula exceeded gravitation, causing a ring of material to separate from the main mass. As contraction continued, other rings broke off at successively smaller distances from the centre.

These rings were not of uniform width all the way around. The densest portion gradually drew material to it and condensed into a planet. The satellites were formed by condensation from the contracting planets. The comets and meteors were presumably formed from material left over between the planets. Today the sun is all that remains of the primeval nebular mass.

Laplace's nebular hypothesis was one of the most successful theories in the history of science. For a century it was practically unchallenged. About 1900, however, it encountered criticism which caused it to be discarded. The chief objection to the nebular hypothesis is the peculiar distribution of angular momentum found in the solar system. The planets, which possess less than 1% of the mass of the solar system, by some means have been able to acquire 98% of its angular momentum. Scientists cannot conceive of a natural physical process during which the angular momentum of a system could have been so unequally distributed. Also it is more likely that matter would have separated from the nebula particle by particle rather than in the form of a ring.

One way to get a large amount of angular momentum into the planets is to put it there. This could be done by having the necessary energy forcibly injected from outside the system. Under this theory of catastrophe it is assumed that the sun suffered a collision or near-collision with a passing star. As the star approached, great tides were raised on the sun. If the star came close enough, there was not merely a tidal bulge but an ejection of matter from the sun in the form of a long filament. The end of the filament tended to follow the star into space, causing it to move in a curved direction around the sun. Some of the material also fell back on the sun and started it rotating in the same direction as the revolution of the filament around it. According to this theory, the filament broke into separate parts which ultimately condensed into the planets. Again the asteroids, comets, and meteorites consist of debris left over between the planets.

This theory sounds fairly plausible, but calculations indicate that the ejected filament, instead of revolving around the sun, would probably have followed along behind the passing star into outer space. To meet this objection the theory has been modified by assuming that the sun was originally a double star and that the collision was with the sun's companion. The colliding stars could have gone off in different directions leaving the central part of the filament torn from them nearly motionless near the sun. But highly heated material dragged from the interior of a star would be more likely to explode, forming a thin atmosphere around the sun, than to condense into planets. However, the most serious objection to any encounter theory is the great distance between the stars. The probability that two stars will collide is practically nil. In an attempt to circumvent this difficulty it has been assumed that the filament was formed when a companion of the sun exploded and became a nova.

Modern theories of the origin of the solar system generally reject catastrophic processes and prefer as their starting point a primeval mass of cold gas, as in the old nebular hypothesis. There would be considerable turbulence in such a large mass, with currents forming and dying down and forming again. Eventually this mass would break up into gas clouds or protoplanets. The material at this time would have been cold, but eventually, with contraction, the central mass would have become hot enough to radiate. This radiation would have caused the planets to lose much of their mass into space by evaporation. In this way the difficulty in connection with angular momentum is avoided.

It should be emphasized that no theory of the origin of the solar system has as yet won general acceptance. All involve highly improbable assumptions. But the difficulty is in trying to find a theory with any degree of probability at all.

Is the Solar System Unique?—With our largest telescopes we cannot hope to detect a planet even as large as Jupiter revolving around the nearest stars. Hence we have no direct knowledge of any stellar planetary systems. But we know that stellar associations are common in the Milky Way. Probably more than half of the "stars" are actually binaries or systems of higher multiplicity. We know of stars with companions only 0.016 as massive as the sun, or 17 times the mass of Jupiter. It seems reasonable to suppose, therefore, that many stars are accompanied by bodies of planetary dimensions.

See also references under "Solar System" in the Index.

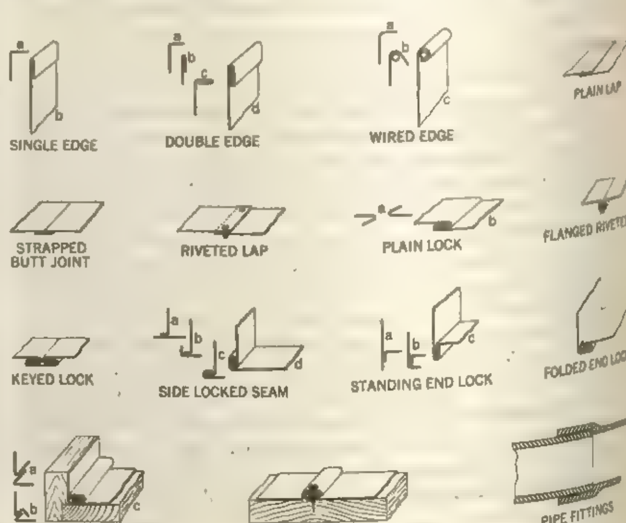
(R. S. Rn.)

SOLDERING. Soldering and brazing are processes for joining metals by the application of heat. A common characteristic of both processes is the use of a filler metal that melts and wets the surfaces of the joint at temperatures below the melting points of the metals being joined. The distinguishing difference between the processes is the strength of the joint and the temperature required for making it.

The soldering process relates only to joints made below 800° F (about 430° C) with solders melting well below this temperature (formerly called the soft solders). The major requirement is sealing and solidifying the assembly and the minor requirement is high joint strength. Fabrication at low temperatures is often the easiest and most convenient method of joining to avoid destruction of heat-sensitive materials surrounding the joint.

The brazing process relates to joints made with brazing filler metals (hard solders) at temperatures in the range of 1,100° to 2,150° F (595° to 1,175° C). Brazed joints are stronger than soldered joints. Brazing is sometimes classified as a welding process, although the latter usually implies fusion of the metals being joined.

Solder Alloys.—Solders of the tin-lead alloy system constitute the largest portion of all solders in use. Alloys containing from 5 to 70% tin, with the balance lead, are available in a melting range of 361° to 594° F (183° to 312° C). They are used for joining copper, brass, nickel-silver, iron, tinplate, and other coated metals. For special purposes, tin-antimony, tin-antimony-lead tin-zinc, tin-silver, and lead-tin-silver solders are used. Special low melting alloys, melting below 300° F contain bismuth cadmium, tin, and lead (*see* also FUSIBLE ALLOYS). The composition melting characteristics, and applications of the commonly used solders are given in Table I. The melting characteristics of the various alloys are indicated in terms of their solidus, or temperature below which the alloy is completely solid, and their liquidus above which it is completely molten (*see* METALLOGRAPHY: *Metal and Alloy Structure*).



BY COURTESY OF THE AMERICAN WELDING SOCIETY

FIG. 1.—TYPICAL SOLDER JOINT DESIGNS

TABLE I

PROPERTIES OF SOLDER ALLOYS

Nominal composition (%)			Temperature ° F (° C) Solidus	Liquidus	Uses
Tin	Lead				
100	0		450 (232)	450 (232)	nontoxic solder
63	37		361 (183)	361 (183)	eutectic (having lowest melting point for tin-lead), for electronic uses
60	40		361 (183)	374 (190)	electronics and general purpose
50	50		361 (183)	421 (216)	most popular general purpose
40	60		361 (183)	460 (238)	sheet metal, radio and TV
35	65		361 (183)	477 (247)	plumber's solder, wiping (filling) lead joints
20	80		361 (183)	531 (277)	radiator solders
10	90		514 (268)	570 (299)	coating metals
Tin	Lead	Antimony			Note: Not recommended for use on galvanized iron, brass
40	58	2	363 (184)	517 (269)	general purpose
35	63.2	1.8	364 (185)	504 (262)	wiping solder
95	0	5	452 (233)	464 (240)	nontoxic high strength solder
Tin	Lead	Silver			
1	97.5	1.5	588 (309)	588 (309)	high strength at moderately elevated temperatures
0	97.5	2.5	570 (299)	579 (303)	
Tin		Silver			
96.5		3.5	430 (221)	430 (221)	for fine instrument work
Tin		Zinc			
91		9	390 (199)	390 (199)	aluminum solders
70		30	390 (199)	518 (270)	

form with rosin or zinc chloride (acid) cores. Paste solders consist of powdered solder mixed with flux. This is a convenient method of applying a controlled amount of flux.

Brazing Filler Metals.—The choice of brazing filler metal depends on the base metals and service requirements of the brazed part. The composition, melting characteristics, and applications for brazing filler metals are given in Table II. Brazing filler metals are classified into six main groups.

Brazing Process.—Preparation of the surfaces by mechanical or chemical cleaning is important. The extensive use of silver-based brazing alloys melting at temperatures below 1,200° F (650° C) required development of fluxes that are fluid and active at 1,100° F. Combinations of borates, fluoborates, fluorides, chlorides, borax, and boric acid are commonly used as fluxes.

Brazing filler metals are free flowing when heated to the proper temperature. Proper attention must be given to design and alignment. In most applications (fig. 2) joint clearances should be in the order of 0.002 to 0.005 in. (0.005 to 0.013 cm.) except

Soldering Process.—Typical designs for soldered joints are illustrated in fig. 1. The basic steps in making a joint are: (1) The joining surfaces are thoroughly cleaned by mechanical or chemical means. (2) A soldering flux is applied to remove oxide films, promote wetting, and prevent reoxidation of the surfaces during heating. (3) The joint is aligned to provide a clearance of 0.001 to 0.005 in. (0.003 to 0.013 cm.) and to promote capillary flow. (4) Heat is applied with a soldering iron or torch. Induction, resistance, or oven heating, and dip soldering are other methods of applying heat. (5) Solder is hand-fed or preplaced in the joint as a preform. (6) The joint is cooled without movement. (7) Corrosive flux residues are removed.

In a technique developed for assembling microminiature electronic components, a small globule of solder is melted on the tip of a soldering iron; when the joint is touched to the globule, solder flows up and around it, heating it only to the melting point of the solder, or substantially below the temperatures required for conventional hand or dip soldering.

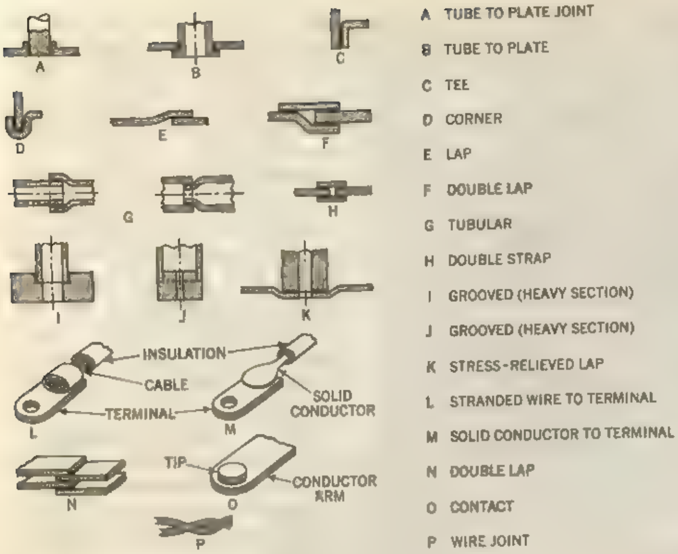
Fluxes.—Soldering fluxes are available commercially for any soldering application. A quick acting corrosive flux is a water solution of three parts zinc chloride and one part ammonium chloride. Hydrochloric acid or stannous chloride, or both, are added to zinc chloride fluxes for soldering stainless steels. An alcoholic solution of rosin is noncorrosive but slower acting. Rosin fluxes are used for soldering electronic equipment, easily solderable metals, and metals that have been precoated with a solderable coating. The lead-tin alloys can be supplied in tubular

TABLE II

BRAZING FILLER METALS

Classification*	Composition			Temperature ° F (° C) Solidus	Liquidus	Uses
	Al	Si	Cu			For aluminum
Aluminum-Silicon 1	95	5	0	1,070 (577)	1,165 (629)	furnace and dip brazing [illegible]
2	92.5	7.5	0	1,070 (577)	1,1[illegible] (613)	cladding [illegible] sheet
3	86	10	4	970 (521)	1,085 (585)	general purpose for controlled flow
4	88	12	0	1,070 (577)	1,080 (582)	general purpose for good corrosion resistance
5	90	10	0	1,070 (577)	1,095 (590)	
	P	Ag	Cu			For copper and copper alloys
Copper-Phosphorus 1	5	0	95	1,310 (710)	1,650 (899)	resistance brazing applications
2	7	0	93	1,310 (710)	[illegible]	very [illegible] for close fits
3	6	5	89	[illegible]	[illegible]	for [illegible] fits
4	7	6	87	[illegible]	[illegible]	very [illegible] for close fits
5	5	15	80	1,190 (643)	1,475 (802)	for loose fits
	Ag Cu	Zn	Cd			For most metals except aluminum and magnesium
Silver 1	45 15	16	24	1,125 (607)	1,145 (618)	application requiring low temperatures
2	35 26	21	18	1,125 (607)	1,295 (702)	general purpose to work brass, [illegible] and food [illegible]
5	45 30	25	0	1,250 (677)	1,370 (743)	
8	72 28	0	0	1,435 (779)	1,435 (779)	for use [illegible] controlled atmospheres no flux
	Al	Zn	Mg			For joining magnesium
Magnesium 1	9	2	89	830 (443)	1,110 (599)	torch or dip brazing. Contains [illegible]
2	12	5	83	770 (410)	1,050 (566)	torch and dip brazing
	Ni B	Si	Fe			For joining heat resisting alloys
Nickel-3	91 3	4.5	1.5	1,800 (982)	1,900 (1,038)	for brazing highly stressed parts
	Au	Cu	Ni			For step brazing applications
Gold 1	37.5	62.5	0	1,815 (991)	1,860 (1,016)	[illegible] electron tubes
2	80	20	0	1,635 (891)	1,635 (891)	heat resisting [illegible]
3	35	62	3	1,785 (974)	1,885 (1,029)	heat resisting iron and nickel irons

*Based on the American Welding Society Classification



BY COURTESY OF THE AMERICAN WELDING SOCIETY

FIG. 2.—TYPICAL JOINTS FOR BRAZING

in the case of aluminum where clearances of 0.006 to 0.025 in. (0.015 to 0.064 cm.) are used. Capillary forces play an important part in distributing the alloy. The alloys may be preplaced or fed into the joint by hand.

Heating for brazing is done with torches, inductive heating, electrical resistance, molten salts, and baths of molten metal. The wide use of these processes has led to the development of special furnaces and automatic equipment with special attention being given to accurate control of the temperature and regulation of the atmosphere. Jigs and fixtures are necessary for dip brazing.

Brazing fluxes are highly corrosive and the residue that remains in the joint area must be removed. Some flux residues are soluble in hot water. In stubborn cases it may be necessary to use warm 10% sulfuric acid followed by hot water.

The soldering and brazing manuals of the American Welding Society are comprehensive sources of information on soldering and brazing. (R. M. MacI.)

SOLENOIDS, ATMOSPHERIC: *see* Baroclinic.

SOLENT, THE, a strait of the English Channel, between the mainland (the coast of Hampshire, Eng.) and the coast of the Isle of Wight. It is the submerged valley of an eastward-flowing river of which the present Frome was the headstream and the Itchen and Test, two tributaries. Spithead, the drowned eastern portion of the same valley, is on the whole wider than the western part, and therefore affords a safer approach to Southampton for very large craft. From Southampton Water to the Needles is 15 mi., (24 km.) and from Southampton Water to Spithead about half the distance; the breadth is from 1¼ to 4 mi. (3 to 6 km.). Opposite the Needles there springs from the mainland a great pebbly bank, nearly 2 mi. in length, on the end of which stands Hurst Castle, dating from the time of Henry VIII. At the mouth of Southampton Water another pebbly bar, projecting from the western shore, also bears a Tudor fortress, Calshot Castle. The low coast of the mainland is broken by the estuaries of the Beaulieu River and the Lym; the coast of Wight, which rises more steeply, is cut by the Medina, Newton, and Yar estuaries. The Solent is frequently the scene of yacht races, especially from Cowes, and also of naval reviews, off Spithead.

SOLER, ANTONIO (1729–1783), the most important Spanish instrumental and church music composer of the later 18th century. Born at Olot de Porrera, Gerona, on Dec. 3, 1729, he was educated in the choir school of Montserrat and at an early age was appointed *maestro de capilla* at Lérida Cathedral. In 1752 he joined the order of St. Jerome (Hieronymites), and in the same year became organist at the Escorial monastery, near Madrid. Later he was a pupil of Domenico Scarlatti, then in residence at the Spanish court, and he himself taught members of the royal family. Scarlatti's influence may be perceived in the lively keyboard technique, the form, and the often unexpected harmonic progressions of his numerous harpsichord sonatas. Besides these he wrote much church music, in which he indulged a taste for intricate canons; incidental music for plays by Calderón and others; six quintets for organ and strings; and six concertos for two organs. He was also the author of a theoretical treatise, much discussed by his contemporaries, entitled *Llave de la modulación* (Madrid, 1762), and experimented in microtones, inventing for the purpose a keyboard instrument called the *afinador*. He died at the Escorial, on Dec. 20, 1783.

See F. M. Carroll, *An Introduction to Antonio Soler* (1960). (L. Sa.)

SOLESMES, French village on the Sarthe River in the *département* of Sarthe, 29 mi. (47 km.) WSW of Le Mans by road, is well known for its abbey, the mother house of the French Benedictines. Pop. (1962) 489. Founded at Solesmes in 1010, the monastery was reformed by the Maurists (1664), suppressed at the Revolution (1791), and refounded by P. L. P. Guéranger (*q.v.*), under whom it became an abbey (1837) and the centre of the reform of plainsong (*q.v.*). Expelled from France in 1901, the monks built Quarr Abbey, Isle of Wight, England. They returned to Solesmes in 1922. The church (11th–16th centuries) contains fine Renaissance sculptures. *See* also Benedictines: *The Reformation and After.* (P. J. G.)

SOLI, an ancient seaport of Asia Minor, situated at the western end of the Cilician plain, about 5 mi. (8 km.) W of modern Mersin, the older parts of which were largely built from the ruins of Soli. It was founded by colonists from Rhodes and was so prosperous when taken by Alexander the Great (333 B.C.) that he was able to exact from it a fine of 200 talents for its attachment to Persia. It was destroyed and depopulated by Tigranes of Armenia but revived by Pompey, who settled the defeated Cilician pirates there and renamed it Pompeiopolis. The bad Greek spoken there gave origin to the term solecism (Greek *soloikismos*). The remains of the artificial harbour, slightly awash, can be traced in the form of two parallel moles with curving ends. The most striking remains in the city are the relics of a long colonnade which flanked its main highway, leading to the harbour. (Wm. C. B.)

SOLI (Greek Soloi), an ancient Greek city of Cyprus, just west of modern Karavostasi on Morphou Bay. It was believed to have been founded after the Trojan War by the Attic hero Acamas. According to another tradition it took its name from the Athenian lawgiver Solon, who was supposed to have visited Cyprus. Soli is probably the town Sillu, whose king Irisu was an ally of Esarhaddon of Assyria in 672 B.C. In Hellenic times Soli had little political importance, though it stood a siege from the Persians *c.* 497 B.C.; its copper mines, however, were famous, and a neighbouring monastery is dedicated to "Our Lady of the Slag Heaps" (Panagia Skouriotissa). Excavated monuments include a theatre, a temple of Aphrodite and Isis, and a 5th-century palace situated 5 mi. W of the town at Vouni.

See E. Gjerstad and A. Westholm in *The Swedish Cyprus Expedition*, vol. iii, pp. 76–582 (1937).

SOLICITOR: *see* Legal Profession.

SOLIDS, GEOMETRIC, are portions of space that correspond to physical solids (*e.g.*, spheres, cubes, pyramids). If a surface separates three-dimensional space into two regions, one of which is finite, the other is, of course, infinite. The finite region is called the interior of the surface. The surface plus its interior is called a solid; the surface itself is called the boundary of the solid (*see* Surfaces). From the standpoint of topology, such surfaces, being closed and orientable, fall into a sequence of families such that any two surfaces belonging to the same family are homeomorphic; *i.e.*, each can be distorted or stretched into the other by a continuous point-to-point transformation. (For example, a cube of clay can be distorted by squeezing to form a sphere.) Each family is characterized by a nonnegative integer called the genus. A sphere has genus zero, and a surface of genus k is homeomorphic to a sphere with k handles. (Physically, a handle would be represented by a half-doughnut with both cut ends pasted to the spherical surface.) The case $k = 1$ is a ring or torus (*see* Topology, Algebraic; Topology, General).

A solid is said to be convex if it contains the whole of the segment joining each pair of its points. The boundary of a convex solid is called a convex surface. Every convex surface has genus zero; but a sphere (*q.v.*) can easily be distorted to lose its convexity without changing its genus. (For instance, a ball of clay can be molded to make a bowl, in which case many segments that connect pairs of points pass through the concave depression rather than through the solid clay itself.)

Any solid has a convex hull: the unique smallest convex solid that contains it. Similarly, any solid has a circumsphere: the unique smallest sphere that contains it; the radii of circumspheres for some geometric solids are given in the Table.

Let C denote the volume of a solid, and A the surface area of its boundary. Then the ratio A^3/C^2 has the value $36\pi(= 113.09 \ldots)$ for a sphere (of any radius) and a greater value for any other shape; *e.g.*, it is 216 for a cube (*q.v.*). Thus, of all solids with a given volume, the sphere has the smallest surface; and of all solids with a given surface, the sphere has the greatest volume. (*See* also MENSURATION.)

Any solid can be approximated as closely as desired by a polyhedral solid, whose boundary is a polyhedron.

Polyhedrons.—A polyhedron is a finite set of plane polygons so arranged that every side of each polygon (*q.v.*) belongs to just one other, and no two adjacent polygons lie in the same plane (with the restriction that no subset has the same property). The polygons and their sides and vertices are called faces, edges, and vertices of the polyhedron; their numbers are conveniently denoted by F, E, and V. For instance, the cube has $F = 6$, $E = 12$, and $V = 8$. Every convex polyhedron satisfies Euler's formula

$$V - E + F = 2$$

Conversely, every Eulerian polyhedron (*i.e.*, every polyhedron for which $V - E + F = 2$) is homeomorphic to a convex polyhedron. More generally, a polyhedron of genus k satisfies

$$V - E + F = 2 - 2k$$

Fig. 1 shows one for which $V = F = 16$ and $E = 32$, so that $k = 1$. The simplest polyhedron is the tetrahedron (or triangular pyramid) for which $V = F = 4$ and $E = 6$ (*see* PYRAMID).

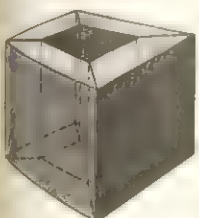


FIG. 1.—POLYHEDRON OF GENUS ONE

Since at least three edges meet at a vertex, $V \leqq 2E/3$; similarly $F \leqq 2E/3$. There is no polyhedron with exactly 7 edges; for $E = 7$ would make both V and F less than 14/3 (that is, $V = F = 4$) and this would leave a tetrahedron (for which $E = 6$, not 7).

Concern here is with polyhedrons that can bound solids, and no mention has been made of nonorientable (or one-sided) polyhedrons. For such a polyhedron, the Euler-Poincaré characteristic $V - E + F$ may take any one of the values $1, 0, -1, -2, \ldots$.

Uniform Polyhedrons.—A polyhedron is said to be uniform if all its faces are regular polygons while all its vertices are alike. The latter requirement means that, for any edge AB, it is possible to transform A into B by a rotation or reflection that transforms the whole polyhedron into itself. A uniform polyhedron is said to be regular if its faces are all alike. The five regular solids (fig. 2) are all mentioned in Plato's *Theaetetus* and are described in detail in the last books of Euclid's *Elements*.

Each uniform polyhedron can be characterized by indicating the number of edges for each of the faces that come together at a vertex (in their proper cyclic order). In this notation the cube is 4.4.4 (since three 4-edged faces meet at each vertex), the uniform n-gonal prism (*q.v.*) is $4.4.n$, and the uniform n-gonal antiprism (or prismoid) is $3.3.3.n$. The only convex uniform polyhedrons are the five Platonic solids, the prisms and antiprisms, and the 13 so-called Archimedean solids. These are listed in the Table with the names that Johannes Kepler gave them in 1619. In the F column, the faces of different kinds are given separately; *e.g.*, in the case of the icosidodecahedron 3.5.3.5, the entry $20 + 12$ means that the faces consist of 20 triangles and 12 pentagons.

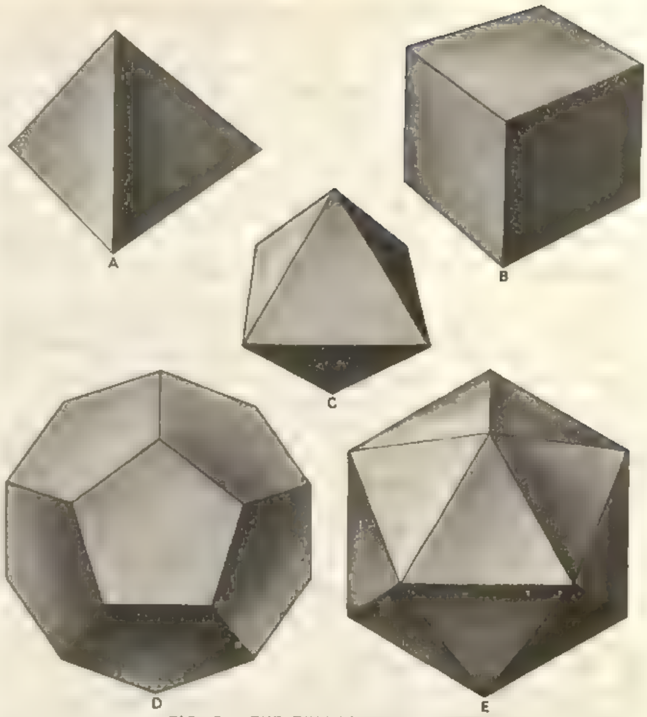


FIG. 2.—THE FIVE PLATONIC SOLIDS

(A) Tetrahedron; (B) cube (hexahedron); (C) octahedron; (D) dodecahedron; (E) icosahedron

The truncated cube 3.8.8 is derived from a solid cube by cutting off the 8 corners to leave 8 triangles, while the 6 squares become 6 octagons. Truncating more deeply, so that the squares are reduced to smaller squares (with one vertex on each original edge), gives the cuboctahedron, so called because, instead of being derived from a cube, it could just as easily have been derived from an octahedron. The cuboctahedron itself can be truncated; but its 12 corners yield 12 rectangles. Kepler's truncated cuboctahedron is derived by distorting these rectangles into squares. (For this reason, its name was changed in earlier editions of *Encyclopædia Britannica* to great rhombicuboctahedron.) Truncating more deeply, and then applying the same kind of distortion, produces the rhombicuboctahedron (or small rhombicuboctahedron). The snub cube (Kepler's *cubus simus*, which could more logically have been named snub cuboctahedron) is derived from the truncated cuboctahedron by first selecting half the vertices so that every edge joins a selected vertex to a rejected vertex, and then performing another slight distortion. The 24 selected vertices belong to a solid bounded by 24 scalene triangles (one for each rejected vertex), 8 equilateral triangles (one inscribed in each hexagon), and 6 squares (one inscribed in each octagon); thus the entry under F in the Table reads $(24 + 8) + 6$ for the snub cube. The distortion is needed to make the scalene triangles equilateral.

The remaining Archimedean polyhedrons (including the snub dodecahedron, which could more logically have been named snub icosidodecahedron) can be derived in a corresponding manner from the other Platonic solids.

Any maker of models (*see* MATHEMATICAL MODELS) can observe that the amount $\delta°$ by which the sum of the face angles at a vertex falls short of 360° is smaller for a complicated solid like the truncated icosidodecahedron than for a simple one like the tetrahedron. René Descartes (*q.v.*) showed that if δ is the same at every vertex, the number of vertexes is $720/\delta$. For instance, in the case of 4.6.10, the face angles at a vertex are 90°, 120°, 144°; therefore $\delta = 6$ and $V = 120$.

In many ways polyhedrons play the same role in space as polygons in the plane. But the theory of dissection provides a remarkable contrast. Any polygonal region can be dissected into a finite number of triangular pieces that can be reassembled to form any other polygon having the same area. But it was discovered by

Table of Uniform Polyhedrons

	Name	Face characteristics at each vertex	V	E	F	Circumradius for edge 2 (or circumdiameter for edge 1)
Platonic	Tetrahedron	3.3.3	4	6	4	$\sqrt{3/2}$
	Cube (Hexahedron)	4.4.4	8	12	6	$\sqrt{3}$
	Octahedron	3.3.3.3	6	12	8	$\sqrt{2}$
	Dodecahedron	5.5.5	20	30	12	$(\sqrt{5}+1)\sqrt{3}/2$
	Icosahedron	3.3.3.3.3	12	30	20	$\sqrt{(5+\sqrt{5})/2}$
	Prism	4.4.n	$2n$	$3n$	$n+2$	$\sqrt{1+\operatorname{cosec}^2\pi/n}$
	Antiprism	3.3.3.n	$2n$	$4n$	$2n+2$	$\sqrt{1+\frac{1}{4}\operatorname{cosec}^2\pi/2n}$
Archimedean	Cuboctahedron	3.4.3.4	12	24	$8+6$	2
	Icosidodecahedron	3.5.3.5	30	60	$20+12$	$\sqrt{5}+1$
	Rhombicuboctahedron	3.4.4.4	24	48	$8+(12+6)$	$\sqrt{5+2\sqrt{2}}$
	Rhombicosidodecahedron	3.4.5.4	60	120	$20+30+12$	$\sqrt{11+4\sqrt{5}}$
	Truncated tetrahedron	3.6.6	12	18	$4+4$	$\sqrt{11/2}$
	Trucated cube	3,8,8	24	36	$8+6$	$\sqrt{7+4\sqrt{2}}$
	Truncated octahedron	4.6.6	24	36	$6+8$	$\sqrt{10}$
	Truncated dodecahedron	3.10.10	60	90	$20+12$	$\sqrt{(37+15\sqrt{5})/2}$
	Truncated icosahedron	5.6.6	60	90	$12+20$	$\sqrt{(29+9\sqrt{5})/2}$
	Truncated cuboctahedron	4.6.8	48	72	$12+8+6$	$\sqrt{13+6\sqrt{2}}$
	Truncated icosidodecahedron	4.6.10	120	180	$30+20+12$	$\sqrt{31+12\sqrt{5}}$
	Snub cube	3.3.3.3.4	24	60	$(24+8)+6$	2.6874267 . . .
	Snub dodecahedron	3.3.3.3.5	60	150	$(60+20)+12$	4.3116747 . . .

Max W. Dehn that there is no finite number of tetrahedral pieces into which a solid cube can be dissected to make a solid regular tetrahedron.

Reciprocal Polyhedrons.—The clumsy symbol 3.3.3.3.3 for the icosahedron may be replaced by the Schläfli symbol $\{3,5\}$, indicating that five 3-edged faces meet at each vertex. In this notation the remaining Platonic solids are characterized as follows: tetrahedron $\{3,3\}$, octahedron $\{3,4\}$, cube $\{4,3\}$, and dodecahedron $\{5,3\}$. Properties of any Platonic solid may be expressed in terms of the general Schläfli symbol $\{p,q\}$; for instance, the total number of edges E for the solid is given by

$$\frac{1}{E} = \frac{1}{p} + \frac{1}{q} - \frac{1}{2}$$

Thus for the cube ($p = 4$; $q = 3$) the value of $\frac{1}{E}$ is found to be $\frac{1}{4}+\frac{1}{3}-\frac{1}{2}=\frac{1}{12}$; showing a total of 12 edges for that solid.

Two polyhedrons are said to be reciprocal (*see* DUALITY) if their vertices, edges, and faces can be paired to make any two vertices that terminate an edge of one polyhedron correspond to two faces that share an edge of the other, and vice versa. In particular, $\{p,q\}$ and $\{q,p\}$ are reciprocal; the centres of the faces of polyhedron $\{p,q\}$ are the vertices of an inscribed $\{q,p\}$. Then the sphere inscribed in the former is the circumsphere of the latter.

More symmetrically, a smaller $\{p,q\}$ and a larger $\{q,p\}$ may be combined so that every edge of each perpendicularly bisects the corresponding edge of the other. Then both solids have the same midsphere; *i.e.*, the same sphere touches all their edges. Such a combination of two reciprocal tetrahedrons $\{3,3\}$ occurs in nature as a crystal twin. Kepler called it *stella octangula;* the 12 edges of the two tetrahedrons are the diagonals of the 6 faces of a cube, which is the convex hull of the twin (*see* also DIAMOND: *Crystallography*). Similarly, a cube $\{4,3\}$ and an octahedron $\{3,4\}$ with crossing edges form a combination whose convex hull (the rhombic dodecahedron) is the reciprocal of the cuboctahedron. The analogous icosahedron-dodecahedron combination yields the rhombic triacontahedron, which is the reciprocal of the icosidodecahedron.

The cube, hexagonal prism, truncated octahedron, and rhombic dodecahedron all have an interesting space-filling property: each can be repeated by translations to form a kind of honeycomb to fill any desired amount of space without overlapping and without interstices. Another solid with the same property is Fedorov's elongated dodecahedron, which is obtained by holding a rhombic dodecahedron so that a vertex where four faces meet is at the top and stretching it so that the four vertical faces become four hexagons, meeting in pairs along four new vertical edges.

FIG. 3.—A SELF-RECIPROCAL POLYHEDRON

Fig. 3 shows a polyhedron with four triangles and four trapezoids as faces. It may be regarded as a modified regular octahedron that has had two opposite vertices replaced by two new edges, both parallel to (but shorter than) two of the original edges; thus $V = F = 8$ and $E = 14$. This polyhedron shares with the pyramids the property of being self-reciprocal (*i.e.*, similar to its reciprocal, though differently placed). It is the simplest centrally symmetrical polyhedron of this kind; there are infinitely many others.

See also CRYSTALLOGRAPHY.

BIBLIOGRAPHY.—V. Eberhard, *Zur Morphologie der Polyeder* (1891); M. Brückner, *Vielecke und Vielflache* (1900); E. Steinitz, "Polyeder und Raumeinteilungen," *Encyklopädie der math. Wissenschaften* (1922); O. Veblen and J. W. Young, *Projective Geometry* (1910, 1918); H. G. Forder, *The Foundations of Euclidean Geometry* (1927); W. W. R. Ball, *Mathematical Recreations and Essays*, rev. by H. S. M. Coxeter (1960); L. Fejes Tóth, *Regular Figures* (1964); H. S. M. Coxeter, *Regular Polytopes*, 2nd ed. (1963); L. A. Liusternik, *Convex Figures and Polyhedra* (1963). (H. S. M. C.)

SOLID STATE PHYSICS deals with the physical properties, behaviour, and internal structure of solid materials. Although fundamental in relating the characteristics of solids to basic physical laws, the subject has spawned important technological developments. The topic is so broad that many aspects of solid state physics are treated separately, as in PIEZOELECTRICITY; THERMIONICS; ELECTRICITY, CONDUCTION OF.

A solid has been defined as any body that tends to return to its original shape after forces applied to it have been removed. This definition fails to draw a distinction between solids that undergo slow deformation (creep) when they are loaded, and thixotropic liquids that behave in a gelatinous fashion under small stresses and do not flow under small loads. The reason is that a sharp boundary between what are thought of as ideal solids and ideal liquids does not seem to exist in nature.

This article is organized as follows:

I. HISTORICAL DEVELOPMENT

In the 16th century a new interest in natural crystals was part of the Renaissance growth of science in Europe. Systematic measurement with the relatively crude tools of the day showed that crystals often are anisotropic; *e.g.*, their mechanical or optical properties differ from one crystal axis to another. By 1830 crystalline materials had been divided into 32 classes on the basis of symmetry; some are completely anisotropic (*i.e.*, have no two directions with the same properties); another class has 48 symmetry elements corresponding to a cube. Angles between crystal faces were found to be accurately the same for each mineral regardless of the source or the chance form of a given specimen (*see* CRYSTALLOGRAPHY). This macroscopic type of study reached its peak late in the 19th century; although it still attracts great interest as a source of data concerning the uses to be made of crystals, it is no longer the central field of research.

Lattice Theory.—Late in the 18th century interest arose in the origin of the remarkable regularity and symmetry of crystals. With the development of atomic theory it seemed clear that the macroscopic properties reflect a highly ordered arrangement of atoms or molecules. A. Bravais, at mid-19th century, proposed that the molecules in crystals are arranged in regular three-dimensional lattices; that is, he supposed all crystals to have translational symmetry in three directions. He showed that this hypothesis would account for the 32 symmetry types found in nature. By 1890 W. Barlow, E. S. Federov, and A. Schoenflies had demonstrated that 230 types of lattice symmetry (*i.e.*, 230 space groups) are possible. Little was known about the arrangements of atoms in lattices although ingenious proposals were made for the simpler crystals such as rock salt.

Lattice theory came into its own in 1912 when M. T. F. von Laue (*q.v.*) and colleagues demonstrated that lattice spacing in typical crystals permits them to diffract X rays much the same as a finely ruled grating will diffract ordinary light. This discovery permitted the measurement of X-ray wavelengths (*see* SPECTROSCOPY, X-RAY) and the structural study of many crystals. Within ten years the structures of many simple inorganic materials were known; during the 1920s analysis was extended to more complex inorganic minerals, and after 1930 to organic crystals. X-ray techniques also revealed imperfections in lattice structure through deviations from ideal diffraction. X-rays were later supplemented by electron and neutron (*qq.v.*) diffraction.

Electronic Properties of Ideal Crystals.—Soon after discovery of the electron (late 19th century) great interest arose in the role electrons play in matter. P. K. L. Drude and H. A. Lorentz (early 20th century) presented an incomplete theory that the valence (*q.v.*) electrons in many metals behave as a gas that is free to move under electric and magnetic fields. Similarly, E. Madelung and M. Born developed a particularly useful theory that ionic crystals are composed of regular arrays of the ions found in aqueous solutions.

By 1930, through the use of quantum statistics, A. Sommerfeld had essentially modified the Drude-Lorentz theory of metals, successfully applying it to problems in electrical conductivity. Studies of the behaviour of electrons in a potential field possessing lattice symmetry led to the development of the band theory of solids (discussed below) and provided clarification of the striking differences among the ideal crystal types. Semiconductivity was explained semiquantitatively by A. H. Wilson and N. F. Mott, while L. D. Landau and R. Peierls turned to the interaction of electrons in solids with magnetic fields. In 1932 E. P Wigner and F. Seitz discovered a practical method for calculating wave functions for simple solids; it was later extended by J. C. Slater, C. Herring, and others. By 1940 a fairly detailed semiquantitative picture of the behaviour of electrons in solids was available.

After World War II more sophisticated calculations with high-speed computers and new experimental data on single crystals in electromagnetic fields yielded detailed information about electron energy levels in solid materials. By the 1960s the electronic structure of simpler metals and of a few valence semiconductors was known quantitatively, and that of many other inorganic solids was known semiquantitatively.

Imperfection-Determined Properties.—Clarification of those electronic properties that are determined by imperfections in ideal crystal structure developed slowly until the 1940s. Just prior to World War I, K. Baedeker noted that what are now known as impurity semiconductors owe their electrical properties to imperfections (*i.e.*, foreign atoms). In 1927 J. Frenkel demonstrated that the electrolytic conductivity of ionic crystals is associated with mobile charged lattice imperfections; proposed models for such imperfections were amplified in 1930 by W. Schottky and C. Wagner.

By 1934 E. Orowan and G. I. Taylor had proposed models of those imperfections (called dislocations) that accounted for the plastic properties of ductile crystals. Thus the importance of imperfections in crystals was appreciated by 1940; however, reduction to relatively broad principles came only after World War II.

Later Developments.—High-purity silicon and germanium (*qq.v.*) became available after World War II and, along with increasing understanding, led to high-quality semiconductors with reproducible and controllable properties. As an outgrowth, the point-contact transistor was invented by J. Bardeen and W. H. Brattain (1948), followed by the development of junction rectifiers, the junction transistor, and in 1954 by a solar photovoltaic cell of relatively high conversion efficiency. In 1957 L. Esaki found an anomalous current-voltage relationship in narrow P-N junctions that led to the tunnel diode.

Solid-state advances revolutionized the electronics industry, providing a host of low-impedance devices (*see* ELECTRICITY: *Induced Electromotive Forces*) and allowing substantial miniaturization of electronic components. In the 1960s interest centred on devices made from thin films (produced by evaporation, chemical deposition, or epitaxial crystal growth), spawning a field called microelectronics.

For many years magnetic iron oxide (magnetite: Fe_3O_4), a ferromagnetic insulator, was a scientific curiosity. During World War II J. L. Snoek developed magnetic oxides (ferrites) with the same crystal structure as magnetite; these have found many high-frequency applications because their eddy-current losses are small. The use of ferrites in radio-frequency inductors and transformers also contributed to the miniaturization of these electronic components.

H. Kamerlingh Onnes' discovery of superconductivity (*q.v.*) in 1911 promised the production of powerful, sustained magnetic fields by currents flowing in wires at temperatures approaching absolute zero. Unfortunately, intense magnetic fields tend to destroy superconductivity. In 1961 certain alloys were found to carry substantial current densities in fields approaching 100,000 gauss without losing superconductivity. Solenoids constructed of these materials offer convenient sources of intense magnetic field, and despite the need to maintain them at low temperature, they are much more economical to operate than are sources using conventional materials. Most applications of powerful magnetic fields were expected to benefit; *e.g.*, in the development of new communication devices (*see* LOW-TEMPERATURE PHYSICS).

II. PHYSICAL PROPERTIES OF SOLIDS

1. Types of Solid.—Crystalline solids have a very high degree of order in atomic arrangement; practically all metals and minerals are of this type. Noncrystalline solids (including glass, plastics, and gels) are those in which the atoms and molecules are not arranged in a definite lattice pattern.

Four distinct types of crystalline solid may be recognized: metals and alloys; salts (ionic crystals); valence crystals; molecular crystals. Metals and alloys are distinguished by their high electrical and thermal conductivity and metallic lustre. The conductivity of metals and alloys is known to arise from migration of free electrons. Salts are commonly formed from ionic solutions and may be viewed as aggregates of charged ions. They commonly have ionic conductivity that increases with temperature. Valence crystals are hard, often brittle, materials such as diamond, silicon, germanium, and silicon carbide. The more abrasive types have high melting points. Like salts, they tend to form in accordance with the rules of valence (*q.v.*) chemistry. In simpler monatomic

types (such as diamond) each atom is surrounded by a number of atoms equal to its valence. Unlike salts, they are not ionic conductors in the pure state and cannot be formed by aqueous precipitation. Molecular crystals are solids such as dry ice (carbon dioxide), the solid forms of the rare gases, and crystals of many organic compounds; these have relatively high vapour pressure at ordinary temperatures (*see* VAPORIZATION). Their sublimation products, such as CO_2 gas in the case of dry ice, are usually stable molecules.

Many alloys, salts, valence crystals, and molecular crystals that are good electrical insulators at low temperatures become conductors at elevated temperatures, conductivity increasing rapidly with temperature. Such materials are termed semiconductors. Their electronic conductivity is usually low compared with that of such metals as copper, silver, or gold, in which conductivity decreases with increasing temperature. In general, the class cuts across the four types of crystalline solid. Some (*e.g.*, silicon, germanium, and tellurium) are semiconducting even when highly purified; others are semiconductors only when appropriate foreign additions are present, or when their composition deviates from ideal combining proportions. The two types are commonly called intrinsic and impurity semiconductors. Cuprous oxide and zinc oxide are typical impurity semiconductors. Actually, the conductivity of most intrinsic semiconductors can be altered substantially by suitable additions of impurities. Semiconducting solids attracted early interest (*c.* 1920) because of their ability to act as electrical rectifiers.

Most crystalline solids have the properties of two ideal types. For example, silicates have properties that resemble those of both salts and valence crystals; many alloys (such as Mg_3Sb_2) have compositions that resemble those of ionic crystals. Most noncrystalline organic materials have properties common to molecular and valence crystals. They frequently are composed of large molecules, between which there is a relatively weak cross-linkage and an occasional strong valence bond. Most types of glass lie between the ideal valence and ionic types. It is noteworthy that ideal metals (and alloys) and ionic substances always form crystalline solids.

2. Atomic Structure of Solids.—Differences among solids may be ascribed to differences in the forces between their atoms and to differences in the way the atoms are arranged. It is a general principle that atoms attract one another when far apart and repel one another at small separations (*see* fig. 1). If the energy of interaction is arbitrarily chosen to be zero at infinite separation, it is found first to decrease as the atoms approach one another, and then to increase very rapidly as they come very close together. The separation at which the minimum occurs is that for which the forces of attraction and repulsion are at equilibrium and represents the separation of atoms in the diatomic molecule. The depth of the minimum and the position at which it occurs depend on the chemical species of the atoms. Thus, the minimum is shallow for two atoms of the rare-gas type and very deep for atoms that form strong chemical bonds. The depth of the minimum is a direct measure of the strength of the bond between atoms.

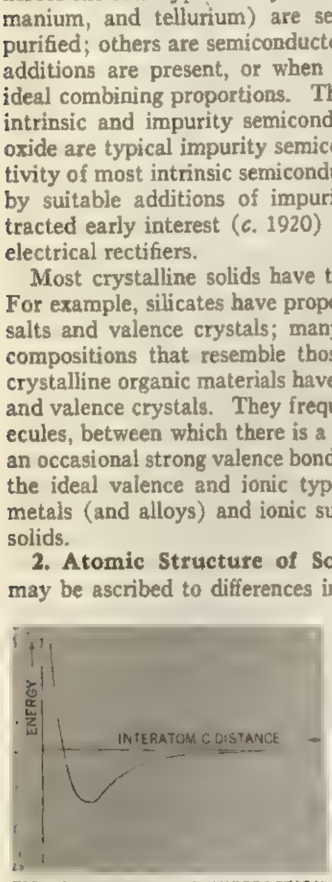


FIG. 1.—ENERGY OF INTERACTION OF TWO ATOMS AS A FUNCTION OF INTERATOMIC SPACING

The energy of interaction between any pair of atoms usually changes in the presence of other atoms. As a result, it is not feasible in general to determine the cohesive forces in a solid by treating it as composed of isolated pairs of atoms. But it is still useful for qualitative work to regard each atom as a definite sphere of fixed diameter that attracts neighbouring atoms, the sign of the forces reversing on contact. Many empirically observed properties of solids can be accounted for in terms of this simple rigid sphere concept (*see* ATOM: *Size and Mass of Atoms*). Assuming spherical structure, the distance between the centres of adjacent atoms in a sample of pure elemental substance is equal to the atomic diameter; similar considerations hold for ionized (charged) atoms. The following table gives estimates of typical atom and ion diameters as determined from X-ray data.

Atom and Ion Diameters in Å (10^{-8} cm.)

H	0.74	F	1.36	F^-	2.66
C	1.54	Cl	1.94	Cl^-	3.62
Fe	2.48	Br	2.26	Br^-	3.92
Zn	2.66	Ga	2.66	Ga^{3+}	1.34
Ti	2.90	Na	3.72	Na^+	1.96
Li	3.00	K	4.54	K^+	2.66
Mg	3.19	Rb	4.86	Rb^+	2.98

The distribution of atoms in solids as determined with X rays is basically as follows:

Within each individual crystal or grain in crystalline solids, the atoms are arranged in a three-dimensional lattice. Fig. 2A shows the face-centred cubic lattice characteristic of many metals such as copper and nickel; fig. 2B shows the structure of sodium chloride, characteristic of most alkali halides.

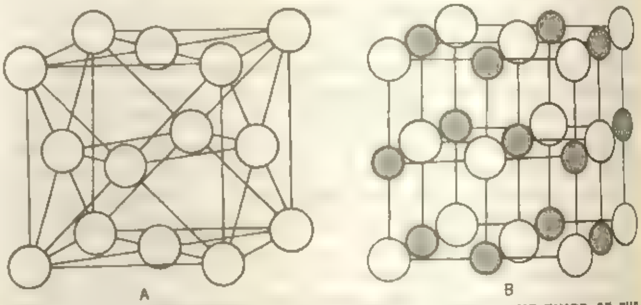

A B

FIG. 2.—(A) FACE-CENTRED CUBIC LATTICE. ABOUT ONE-THIRD OF THE METALS HAVE THIS STRUCTURE. (B) SODIUM CHLORIDE STRUCTURE; SODIUM IONS ARE SHOWN SHADED, CHLORINE IONS AS OPEN CIRCLES. THIS ARRANGEMENT IS TYPICAL OF IONIC CRYSTALS

Noncrystalline solids have a less orderly atomic arrangement. In the extreme case, which occurs in glass (and in liquids), the immediate neighbours of any atom are symmetrically disposed about it; however, relative to a given atom, this order does not extend beyond the first few sets of neighbours. Crystalline solids sometimes are characterized as having both long-range and short-range order; noncrystalline solids as having only short-range order. Fig. 3 shows the structure observed in boric oxide glass. Similarly, each silicon atom in silica glass is surrounded tetrahedrally by four oxygen atoms and each oxygen atom has two silicon atoms as nearest neighbours (*see* GLASS MANUFACTURE). Fibrous materials may have parallel long-chain molecules linked to produce an appreciable amount of long-range order relative to the fibre-axis; thus rubber and similar plastic substances (*see* PLASTICS) frequently show semicrystalline structure under tension.

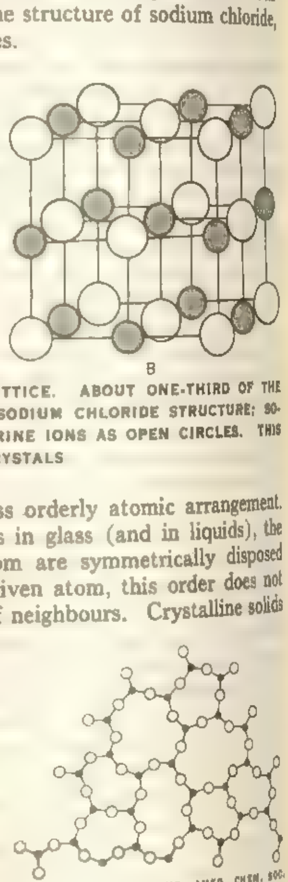

W. H. ZACHARIASEN, JOUR. AMER. CHEM. SOC.

FIG. 3.—TWO-DIMENSIONAL NETWORK OF BORIC OXIDE GLASS. THE STRUCTURE LACKS PERIODIC (LONG-RANGE) SYMMETRY

3. Chemical Composition.—Most solids with low electrical conductivity (*e.g.*, saturated hydrocarbons; oxides and chlorides of the simpler metals) conform to the usual rules of combining proportions summarized in elementary valence chemistry. Specific examples include alkali halides, magnesium oxide, and paraffins. This obedience to simple valence theory is reflected in the definite and simple way their atoms are arranged. In crystalline materials each type of atom forms its own lattice; the complete lattice may be regarded as a combination of these lattices. Similarly, in insulating noncrystalline solids, the short-range order shows regularity and simplicity, as in the case of silica glass.

Metals, in combining with each other, do not obey the usual rules of valence. They tend to combine over wide ranges of composition; *e.g.*, copper will dissolve up to 30% of zinc without

altering its face-centred cubic lattice structure. They commonly form compounds that bear little relation to the usual rules of combining proportions. Thus copper and zinc form compounds of approximate composition CuZn, Cu_5Zn_8 that differ in lattice structure (*see* INTERMETALLIC COMPOUNDS).

W. Hume-Rothery and G. Hägg, respectively, attempted to systematize rules of combining proportions of metals for substitutional and interstitial alloys. In substitutional alloys, atoms of one type are replaced by another as composition is varied. For example, copper atoms are replaced by zinc as the amount of zinc in brass is increased. Interstitial alloys differ in that one type of atom is sufficiently small to fit into the interstices of the lattice formed by the other. Typical interstitial alloys are formed by dissolving carbon in iron or hydrogen in palladium (*see* ALLOYS).

Hume-Rothery noted that substitutional alloys form for wide ranges of composition as long as the atomic diameters of the constituent elements do not differ by more than 15%. When the difference in diameter is more than this, few substitutional phases form and these have very narrow solubility ranges. He found a close relationship between the crystal structure of a substitutional alloy and the ratio of the number of valence electrons within it to the number of atoms. For example, the body-centred cubic lattice, characteristic of beta brass (CuZn), forms whenever the ratio of valence electrons to atoms is 1.5. In beta brass, each zinc atom contributes two valence electrons and each copper atom contributes one, so that the ratio is 3:2 = 1.5. The corresponding phase in the bronze (tin-copper) system has the approximate composition Cu_5Sn.

Hägg found that interstitial alloys form readily if the metals are very close in the electromotive series (*see* BATTERY) and if the diameter of one atom is less than 0.59 times that of the other. The only systems of alloys that satisfy these conditions are those that combine the so-called metalloid atoms (hydrogen, boron, carbon, and nitrogen) with the transition metals, such as those of the iron, platinum, and palladium groups. The metalloid atoms fit the interstices of the metal-atom lattices. Hägg also found a close correlation between composition and structure among the interstitial alloys.

Two elements will not alloy well if the ratio of their diameters falls in the range 0.59–0.85. Within this range the smaller atom is too large to fit into the lattice interstices of the other, and too small to form an ideal substitutional alloy. The range is a kind of no-man's land for alloy formation. The ratio is close to 0.59 for the iron-carbon system; were it somewhat larger we might have been denied the valuable alloy system of steel.

The relatively feeble electronic conductivity of intrinsic semiconductors indicates that they lie between ideal metallic solids and ideal insulators. Practically all the impurity semiconductors also fall between insulators and metals. Either they have appreciable solubility for foreign atoms having appropriate size relations, or they are compounds that tend to form over an appreciable range of composition. Ferrous oxide (FeO), a typical semiconductor, tends to deviate from ideal combining proportions by about 1%. Similar deviations are observed in zinc oxide; lead sulfide, selenide, and telluride; cuprous oxide; and even in alkali halides transformed into semiconductors by heating with alkali metal vapour. In such cases the crystal lattice exhibits imperfections on an atomic scale that are distributed throughout the specimen. Thus in iron oxide some lattice sites that would be occupied by iron ions in the normal crystal are vacant. In zinc oxide the deviation from ideal composition is associated with the presence of interstitial zinc atoms in the lattice of the oxide. The electronic conductivity of any of these semiconductors increases with the degree of deviation from ideal composition.

4. **Principles Determining Physical Form.**—The laws of thermodynamics (*q.v.*) indicate that the stable state of a system at constant pressure is determined by the condition that the free energy

$$F = E + PV - TS \tag{1}$$

be a minimum. Here E is the energy of the system, P is external pressure, V is volume, T is temperature, and S is the entropy. Variations in PV that normally occur when solids are maintained at atmospheric pressure are negligible, so that this term may be discarded. It usually is adequate for practical purposes to use

$$F = E - TS \tag{2}$$

At absolute zero temperature the TS term vanishes and the equilibrium state is determined by the condition that E be a minimum. This is just the condition for equilibrium in a mechanical system. Thus the TS term in equation (2) acts as a correction for the influence of temperature fluctuations in altering the equilibrium value from what it would be at absolute zero. As the temperature of a solid rises, energy is absorbed and E increases. The entropy S, which measures the thermally induced disorder of the system, also increases. Thus the terms on the right side of equation (2) compete as temperature rises, E tending to increase and TS to decrease F. At low temperatures atoms in solid systems tend to be in highly ordered arrangements for which E is small; at high temperatures disordered states for which S is large are favoured. The tendency toward disorder as temperature rises may be manifested in one or more of several ways.

The atoms are set into oscillation about positions of equilibrium that theoretically obtain at absolute zero. This type of disordering is universal for solids at any given temperature above absolute zero.

The system may change from one lattice structure to another in which the atoms have greater freedom of movement. This is termed an allotropic transformation and is very common. For example, iron changes from the body-centred cubic to the face-centred cubic lattice at 910° C (*see* IRON AND STEEL INDUSTRY: *Steel Metallurgy*). Sulfur exhibits several allotropic transformations at successively higher temperatures. (*See* also PHASE EQUILIBRIA; SULFUR.)

Pairs of atoms of different types may change places to disorder the lattice. This interchange is most commonly observed in alloys, in which the net energy required for interchange is smaller than for materials in which the constituents differ widely (*e.g.*, salts). For example, at absolute zero temperature the stable form of beta brass (CuZn) is a body-centred cubic structure in which one type of atom occupies the body-centred positions and the other occupies the position at cube corners. As temperature rises, atoms begin to change places until, at 480° C, it is equally probable that a given position will be occupied by either type of atom. The basic lattice remains unchanged in this type of transformation.

A similar kind of disorder may arise through rotation (libration) of individual molecules or radicals in molecular crystals of salts such as ammonium chloride.

Disordering may result from occasional migration of atoms to interstitial positions in the lattice or to the surface, leaving the lattice with vacant sites. The number of disordered atoms, or lattice defects, per unit volume increases rapidly with temperature, obeying the same law as does the density of vapour above a liquid. Diffusion (*q.v.*) and ionic conductivity in solids are intimately related to lattice defects of this type. Diffusion occurs in substitutional alloys because vacancies produced in the lattice by migration of atoms to the surface can move from one site to another. Their motion permits a shuffling of atoms within the lattice. Similarly, ionic conductivity of alkali halides is related to migration of vacancies for halogen and alkali-metal ions.

It is evident that changes from solid to liquid or gaseous state represent even more extreme types of disordering; nevertheless, melting and sublimation obey the same basic principles.

Motion of free electrons is responsible for the high thermal and electrical conductivity of metals. The free electrons may contribute to disorder by an increase in their rate of motion. Ferromagnetism (*q.v.*) originates in an ordered coupling of a fraction of the valence electrons. This order is disrupted at elevated temperatures and the ferromagnetism vanishes, *e.g.*, it disappears at 780° C in iron.

The onset of disorder increases the internal energy of the system and requires the absorption of a corresponding amount of thermal energy. This absorbed energy is evident as the specific heat necessary to raise the temperature, or the latent heat of phase transi-

tion. It frequently is possible to identify the nature of the disordering process by its contribution to the absorbed thermal energy. For example, change in specific heat associated with an increase in amplitude of oscillation of the atoms varies as the third power of the temperature near absolute zero and reaches a limiting value at high temperatures (*see* CALORIMETRY).

Allotropic changes like melting are accompanied by the appearance of a latent heat at the temperature where the phase change occurs. The observation of this latent heat is commonly used to detect allotropic changes (*see* LATENT HEAT).

Onset of order-disorder in alloys and of molecular rotation in molecular solids is usually distinguished by a large increase in specific heat over the temperature range near that at which disordering (or rotation) becomes complete. The transition may also be marked by the appearance of a latent heat.

Disorder associated with electron changes is also marked by thermal effects. For example, free electrons in metals make a contribution to the specific heat that increases linearly with temperature; it is particularly large in the transition metals such as iron and nickel. Similarly, disappearance of ferromagnetism in iron is accompanied by a peak in specific heat of the type observed during the disordering of alloys.

5. Elasticity.—Recall that each atom in a solid is normally at equilibrium with its neighbours. The atoms take up new positions when external forces are applied, *i.e.*, when the solid becomes strained. As long as the applied forces are sufficiently small, the atoms return to their original positions when the forces are relieved, and the strain is said to be elastic. The material is strained beyond its elastic limit when the process is not reversible. The range of elasticity depends much on the nature of the applied forces, and on the material and history. Most homogeneous materials are highly elastic under hydrostatic pressure. In general, a tensile stress, or a unidirectional compressive stress, or a shearing stress is necessary for a deviation from elasticity.

P. W. Bridgman (*q.v.*) employed hydrostatic pressures as high as 100,000 atmospheres to produce large elastic strains. Changes in crystal structure may be produced by hydrostatic pressure and occasionally persist after the pressure is removed. Pressure approaching 100,000 atmospheres and temperature in excess of 2300° K were used by others to convert graphite into diamond. Bridgman discovered a variety of forms of ice at high pressures, most of which sink in water (*see* ICE).

For small pressures the relation between stress and strain is linear for small stresses; however, quadratic and higher terms appear in the stress-strain relation at elevated pressures. If no phase change is induced by the pressure, ever-increasing pressure is required for the same increment of strain. This decrease in compressibility as the pressure increases results from an increase in repulsive forces between atoms with decreasing atomic separation (fig. 1). The linear force-strain relation for small forces arises because interatomic forces vary linearly for small displacements.

Although hydrostatic pressure can produce a large elastic deformation, the range of elastic strain is much more limited for shearing stresses or tensile stresses. In such cases the material either flows plastically or breaks before the absolute magnitude of the stress components becomes as large as those in Bridgman's experiments. Work is done on a solid strained under applied forces. The reversible component of this work represents a change in the free energy of the solid; the elastic properties may readily be described in terms of changes in free energy that accompany atomic displacements associated with strains. In most solids this change in free energy is principally a function of change in E, the internal energy of the material; that is, the TS terms in equation (2) merely determine the relatively small amount by which the elastic constants at a finite temperature differ from those at absolute zero. There are, however, notable exceptions; for example, the elastic properties of rubber are principally related to how entropy changes as the material is strained. The elastic constants of materials of this type are usually strongly dependent on temperature.

6. Plasticity.—Most noncrystalline solids (*e.g.*, glass) become brittle at low temperatures; their elastic range is determined by the stresses at which they break. Many crystalline solids, however (notably metals), exhibit surprising ductility even near absolute zero. This type of plastic flow in large single crystals was found to obey several laws.

Crystals deform plastically by displacement of one part of the crystal relative to another along particular crystallographic planes (*see* CRYSTALS, DISLOCATION OF). These slip planes (or glide planes) are the octahedral planes of closest packing in face-centred cubic crystals and are the basal planes (those normal to the hexagonal axis) in hexagonal crystals. There are three different types of slip plane in body-centred cubic metals. Displacement along a given plane may even be visible to the naked eye or under a low-powered microscope. The crystal shows lamellae that are displaced like cards when a deck is sheared. X-ray evidence indicates plastic flow even within the lamellae.

In general, slip will occur in a given set of parallel planes only if the resolved shearing stress exceeds a critical value that depends on the previous history of the specimen and on its purity. This critical shearing stress σ_c increases as the amount of plastic strain increases, a phenomenon referred to as work-hardening; thus solids become harder as they are deformed by drawing or rolling. Critical shearing stress may also be raised by soluble impurities; in this way metals such as copper, silver, and gold can be made sufficiently hard for coinage. Values of σ_c for pure metals annealed at elevated temperatures range between 1 and 100 kg./cm.^{2}; values as high as 1,000 kg./cm.^{2} are common for work-hardened or alloyed metals.

There is a more or less well-defined temperature above which the effects of work-hardening may be removed by annealing in a reasonable time, perhaps half an hour. This resoftening temperature for metals such as lead, zinc, and tin is near or below room temperature; however, it is well above room temperature for many useful metals. Resoftening probably occurs at all temperatures above absolute zero, but is negligibly slow below the resoftening temperature.

Critical shearing stress is slightly dependent on temperature below the resoftening point; *e.g.*, the value for very pure zinc was found to increase from 75 to 140 kg./cm.^{2} in going from room temperature to 20° K (−253° C). This means that metals retain much of their ductility at low temperatures; *e.g.*, zinc bends under its own weight near absolute zero.

The high ductility of metals appears related to the presence of imperfections. If two parts of an ideally perfect crystal were to be displaced relative to one another along a slip plane as rigid units, the stresses required would be about $\frac{1}{10}$ of Young's modulus (*see* ELASTICITY; MATERIALS, STRENGTH OF); that is, about 500,000 kg./cm.^{2}. This is at least 1,000 times larger than the values normally observed in pure single crystals. Hairlike microscopic crystals about 0.0001 in. in diameter sometimes have this theoretical strength. Such crystals probably lack the imperfections responsible for the ductility of most materials.

7. Fracture.—From the above discussion the rupture stress of an ideal crystal (*i.e.*, the stress required to break chemical bonds) is seen to be about $\frac{1}{10}$ Young's modulus for the material. The fact that materials usually break under much lower stresses probably arises from magnification of stress in the vicinity of small (even submicroscopic) cracks. If the magnified stress is large enough the cracks will propagate, weakening the material, shifting the applied load to a smaller effective cross-sectional area of crystal, and activating a new set of microcracks. Such a chain of events leads to fracture, the applied stress required being called the fracture stress.

Recall also that critical shearing stress for plastic flow (*i.e.*, the stress needed to move dislocations through the crystal) is also much lower than ideal rupture stress. Whether the material fractures abruptly or yields plastically depends on which of the two (fracture stress or critical shearing stress) is lower. If fracture stress is lower the material is brittle; lower critical shearing stress produces ductile behaviour. Presumably glass is brittle at low temperatures because its nonperiodic structure does not allow easy movement of dislocations. Ionic crystals usually show brittle fracture, but the fracture stress seems strongly dependent on surface condition. Small surface flaws apparently have associated micro-

cracks that produce substantial stress magnification. Alkali halide crystals can be bent in a ductile manner when immersed in water, and ductile behaviour has been reported for crystal specimens plated with a thin layer of a soft metal.

At normal temperatures pure metals deform in a ductile manner because critical shearing stress is relatively low, while fracture stress is much higher. Onset of plastic flow relieves stress concentrations, preventing local stresses from greatly exceeding average stress level. Extreme work-hardening will raise critical shearing stress above fracture stress, and the metal can be made to break in a brittle manner.

Yield stress in metals increases more rapidly than does fracture stress as temperature drops; thus a normally ductile metal can become brittle at lower temperatures, as in the brittle failure of steel.

8. Electrical Conductivity.—The electrical resistivity of solids covers a wide range. At normal temperatures the resistivity of a typical metal is 10^{-6} ohm-cm., while that of fused quartz (a good insulator) is about 10^{20} ohm-cm. (*see* INSULATING MATERIAL [ELECTRICAL].) Alloys, semiconductors, and less efficient insulators fall between these extremes. As temperatures approach absolute zero the resistivity of a pure metal falls markedly; depending on residual impurity content, this decrease may be as much as 100,000 times.

Solids can conduct electricity if they contain mobile ions. The probability that a given ion will become sufficiently free (as a result of thermal fluctuations) to jump from one equilibrium position to another determines conductivity in this case. Ionic conductivity is not very large, however, and solids that conduct only in this way are usually among the less efficient insulators; they also conduct much better at high temperatures.

If a solid contains free electrons, its total conductivity will heavily overshadow any ionic conductivity present; thus good electrical conductors are characterized by electronic conductivity. One of the greatest achievements in applying quantum mechanics to solids was in understanding electronic conduction in metals and semiconductors. It is known that energy levels for electrons in solids fall into quasi-continuous bands separated from each other by forbidden regions called energy gaps. A metal is characterized by a partially filled band of electrons or by a group of partially filled overlapping bands. In such a situation electrons are present adjacent to unoccupied levels, and can be excited into the empty levels by absorbing energy from the electric field. The upshot is that a substantial fraction of the electrons in a metal behave as if they were free.

If the number of electrons in a solid is just sufficient to fill one of the quasi-continuous bands of energy levels, and if this band is separated from other bands by substantial energy gaps, the material is an insulator. However, a number of circumstances perturb this ideal situation: If an energy gap is not too large, some thermally excited electrons may cross it, leaving vacant levels in the nominally filled band, and putting some electrons into a practically empty band; this represents an intrinsic semiconductor. Through inclusion of suitable impurities the electron-atom ratio in the solid can be increased to force some electrons into a vacant band, or decreased to partially empty the original band; this describes an extrinsic semiconductor. The number of electrons that participate in conduction in a semiconductor is quite small on a per-atom basis. In an extrinsic semiconductor the number varies from 10^{-5} to 10^{-8} electrons per atom; in intrinsic semiconductors it is even smaller.

Through quantum mechanics it is indicated that metals can have infinite conductivity if all atoms are completely at rest and the lattice is perfect. Conductivity is limited only because free electrons accelerated by the applied field are deflected by lattice imperfections. Most important of these imperfections in a pure metal at ordinary temperatures are those arising from thermal vibrations of the lattice. Foreign atoms also make an important contribution to the resistivity of impure metals and alloys, and are the principal source of residual resistance in pure metal crystals at very low temperatures.

The conductivity of most electronic conductors is modified in a magnetic field. The effect is particularly pronounced in pure materials at temperatures near absolute zero. In most cases conductivity is reduced because conduction electrons are continually deflected by the magnetic field, increasing the length of their overall trajectories. The electrical conductivity of isotropic solids or cubic crystals may be regarded as a scalar quantity until a magnetic field is applied. The current then no longer need be in the direction of the applied electric field; in other words, the scalar conductivity is replaced by a tensor conductivity (*see* ANALYSIS: *Tensorial Functions*). Electrical properties of metals and semiconductors have been studied in magnetic field strengths up to 200,000 gauss. Such experiments have yielded considerable information about electron energy levels in single crystals.

Electrical conduction in a temperature gradient is subject to thermoelectric effects, and a circuit of two dissimilar materials can produce electrical energy from heat. The origin of such effects is understood in terms of electronic band structure, and measured thermoelectric coefficients give information about electron energy levels in the solid under study.

Some metals and alloys become superconducting at very low temperatures, abruptly losing all resistivity near absolute zero. Superconductivity (*q.v.*) was first observed in metallic mercury, and many other superconductors have been found, each with a distinct superconducting transition temperature. The highest transition temperature known in the 1960s was that of niobium tin (Nb_3Sn) at 18.05° K. J. E. Kunzler and co-workers (1961) discovered that some hard superconductors (*e.g.*, Nb_3Sn or niobium-zirconium alloys) could carry large current densities in magnetic fields approaching 100,000 gauss. This discovery opened the door to high-intensity magnets that can operate without electrical power dissipation; in 1965 a magnet cooled with liquid helium was reported to develop 132,000 gauss using only six automobile batteries as a power source.

Onset of superconductivity is accompanied by a marked change in magnetic properties: the superconductor tends to expel the magnetic field, becoming an ideal diamagnetic (*see* DIAMAGNETISM). Most macroscopic aspects were correlated by F. London's and H. London's phenomenological theory (1935). Observing that superconducting metals are those with relatively high resistivity at normal temperatures, H. Fröhlich (1950) proposed that superconductivity arises when conduction electrons strongly interact with lattice vibrational waves. This approach was developed by Bardeen and others into a successful microscopic theory (1957).

9. Thermal Conductivity.—The high thermal conductivity of metals (as does their electrical conductivity) depends on a high density of essentially free electrons that transport thermal energy from hotter to cooler regions. When thermal conduction via free electrons is the dominant process, thermal and electrical conductivity tend to be proportional, independently of the detailed electronic level structure of the solid. This proportionality, the coefficient of which can be calculated theoretically, has been known (for more than 100 years) to hold for many metals at normal temperatures; it is known as the Wiedemann-Franz law.

Another important mechanism of thermal transport through solids is by lattice vibrations (frequently called conduction by phonons). Atoms in a crystal, interacting with forces that obey Hooke's law in first approximation, can be made to vibrate; *e.g.*, a sound wave with fixed frequency and direction of motion may be passed through the crystal, or vibrations can be stimulated by heating. When the crystal is heated at one end vibrational energy will increase there, and lattice vibrational waves will propagate to the cooler end. These waves carry more energy from the hot region than is transported back by waves from cooler parts. Such vibrational waves have corpuscular properties; their energies are limited to discrete values, much as light is quantized (*see* LIGHT: *History*). The excitation of a vibrational wave proceeds in steps, each excitation unit being called a phonon. Sometimes it is convenient to regard the phonons as a gas of particles that conduct heat as do molecules in gas, with the essential difference that phonons can be "created and destroyed" through energy transfer. Excess phonons arise in hotter regions of a crystal, propagate to cooler regions, and disappear.

Insulating crystals conduct heat by phonon transport. In semiconducting solids conduction by phonons is of comparable magnitude to electronic thermal conductivity.

10. Magnetic Properties.—All solids are influenced by a magnetic field, but in many cases very slightly. The five types of magnetic behaviour exhibited by solids are diamagnetism, paramagnetism, ferromagnetism, antiferromagnetism, and ferrimagnetism. The first two arise from orbital and spin motion of individual electrons, usually with a small effect on the magnetic field except at very low temperatures. The last three arise through cooperative interaction of a large fraction of the electrons in a solid, the magnetic moments of the atoms becoming aligned or ordered (*see* MAGNETISM).

Ferromagnetism is exhibited by the transition metals and by some of their alloys. Ideally ferromagnetism is depicted by complete alignment of the magnetic moments on the individual atoms (even in the absence of an applied magnetic field); as the solid is heated it reaches its Curie temperature, at which point the arrangement becomes so disordered that ferromagnetism disappears. Antiferromagnetism is represented by an ordering in which equal magnetic moments alternate, pointing in opposite directions to yield a net moment of zero. Ferrimagnetism is the most general type of magnetic ordering in solids; it is seen in ferrites, which are composed of more than one type of atom. While the alignment of one atomic species may be opposite to that of another, their unequal moments lead to a resultant magnetic moment other than zero. Magnetite (Fe_3O_4) is a mineral ferrite known since ancient times, but only since World War II has its magnetic structure (and those of related magnetic oxides) been understood. Although not as strongly magnetic as the metals iron, nickel, and cobalt, ferrites have the advantage of being electrical insulators, and serve in high-frequency applications where eddy-current losses in conducting materials otherwise would be very large.

Quantum theory has been used to account for differences among diamagnetic, paramagnetic, and ferromagnetic solids. Most prominent among these developments are a theory of paramagnetism in simple metals (W. Pauli); a theory of ferromagnetism (W. K. Heisenberg, F. Bloch, and others); a treatment of paramagnetic salts (J. H. Van Vleck); and an investigation of strongly diamagnetic metals such as antimony and bismuth (N. F. Mott and H. Jones).

At very low temperatures the magnetic susceptibility of single crystals of pure metals shows the oscillatory De Haas-Van Alphen effect when plotted against reciprocal magnetic field. The oscillatory behaviour results from a quantum effect that bunches the electron energy levels in the presence of a magnetic field. Measurements of the effect yield detailed information about electron energy levels in the solid.

After World War II, changes in absorption of high-frequency radio waves by solids in stationary magnetic fields received considerable attention. The energies of atomic nuclei and electrons depend upon their orientation in the magnetic field if they have resultant magnetic moments (*i.e.*, if they behave as tiny magnets). Nuclei of odd mass number and unpaired valence electrons usually have such magnetic moments. Electromagnetic waves of appropriate frequency can lose energy to the solid in changing the orientation of nuclei or electrons (*see* NUCLEAR MAGNETIC RESONANCE; ELECTRON PARAMAGNETIC RESONANCE). Studies of magnetic resonance have furnished much information concerning properties of solids and other systems.

In the presence of a magnetic field the conduction electrons of a metal or semiconductor move in orbits that are large compared to atomic dimensions. If the material is in a stationary magnetic field irradiated with high-frequency radio waves, the electrons can absorb energy when wave frequency coincides with their orbital frequency; this is called cyclotron resonance. Resonance studies of this type give information about the effective mass of electrons in solids.

11. Optical Properties.—The optical properties of a solid describe its behaviour under radiation from the infrared, visible, or ultraviolet parts of the electromagnetic spectrum. With overlapping band structure containing many vacant energy levels, metals strongly absorb infrared and visible radiation. On the other hand, because of the energy gap, a semiconductor does not absorb long-wavelength photons appreciably, showing strong absorption only when the energy of the photon is enough to lift an electron across the energy gap.

The colours of some metals and alloys, the transparency of alkali metals to ultraviolet, and the detailed shape of the absorption spectrum in semiconductors have been explained in terms of quantum mechanics (electron energy levels). In some high-resistivity semiconductors absorption of photons with enough energy to lift electrons across the gap can reduce resistivity significantly; such materials (*e.g.*, *see* SELENIUM) are called photoconductors. The photovoltaic effect depends on absorption of photons in a junction region between two extrinsic semiconductors (one conducts electricity by excess electrons and the other conducts by a deficiency). Devices utilizing the photovoltaic effect can convert solar energy into electricity with more than 10% efficiency (*see* PHOTOELECTRICITY; ENERGY CONVERSION, DIRECT).

III. THEORETICAL APPROACHES

The most important developments in the theory of solids after 1930 arose from the application of quantum mechanics to the behaviour of valence electrons. The approach closely resembles its application to molecular systems in that the coordinates of atomic nuclei are treated as parameters when the wave functions for the electronic states are determined. The procedure is permissible because the electrons move so much faster than the heavier nuclei that they may be regarded as being in a stationary state for each position of the nuclei (even when the latter are in motion, as in thermal vibration). This method of separating electronic from nuclear motion was introduced by Born and J. R. Oppenheimer and is accurate for most practical problems.

The complete wave function for the electrons in a solid is too complicated to determine precisely, and simplifications are usually made in solving problems.

1. One-Electron Approximation.—The electrons in the inner shells of an atom (which are not appreciably affected by changes in chemical binding) are treated as though rigidly attached to the nucleus; their influence on the valence electrons is assumed to be that they merely contribute a static potential field centred on the nucleus.

In first approximation each valence electron is assumed to possess its own wave function. This approximation (known as the Hartree-Fock method, or one-electron approximation) assumes that interaction between valence electrons again derives from a static potential field, but in this case a field calculated self-consistently. Although the results it yields are rough, and at best semiquantitative, it permits the motion of each electron to be described independently; that is, each electron can be assigned its own constants of motion. There are two ways in which the one-electron method is commonly used: atomic approximation and free-electron approximation.

In atomic approximation it is assumed that each electron is attached to a given nucleus or group of nuclei, as if it were in an isolated atom or molecule. Thus, the solid has the aspect of an aggregate of atoms or molecules, and is conceived as an insulator since the electrons are bound to definite nuclei and cannot move about the solid. This approximation is particularly useful in describing the normal state of a rare-gas solid (such as solid argon), a molecular solid (such as solid methane or any molecular solid with a low boiling point), or an ionic crystal (such as one of the alkali halides), where the electrons are bound to ions rather than neutral atoms or molecules.

In free-electron approximation it is assumed that each valence electron is free to move in the lattice so that its wave function has finite amplitude through the specimen. In the simplest case this function has the same form as that for a free electron; however, the approximation is not restricted to this simple situation. At first sight a solid described with free-electron functions seems to be a metallic conductor, since the electrons are free to wander everywhere. Actually, the motion of the electrons is restricted in certain cases so that they cannot conduct a current. The ground

state of any insulator may be described with equal accuracy with either free-electron or atomic approximation; thus, free-electron approximation is of more general use in treating the normal states of solids.

It may be asked if atomic approximation should ever be used; however, excited states described by the two one-electron approaches differ notably. Any neutral atom or molecule is characterized by excited electronic states in which electrons are bound to nuclei. Similarly, a solid described by atomic approximation may exhibit states in which excited electrons are bound to definite atoms or molecules so that the solid is not conducting. On the other hand, an excited state of a solid described with free-electron approximation is always conducting, even though the ground state represents an insulator. Thus, atomic approximation can describe excited states that cannot be treated appropriately by the free-electron approach. Nonconducting excited states have been observed in alkali halide crystals; in these cases atomic approximation is the more appropriate.

To summarize, free-electron approximation can always be used for metals and the ground state of insulators; atomic approximation is frequently better for the excited states of insulators; it also has sufficient flexibility to describe conducting excited states of insulators (analogues of ionized states of atoms and molecules).

Neither one-electron approximation furnishes an exact solution of the quantum mechanical equations of motion. In spite of this defect, one-electron methods provide useful qualitative pictures of the behaviour of electrons in solids and illuminate many experimental findings. The difficult problem of modifying the approximations to give more nearly exact solutions has been solved in a few simple cases. Better solutions were obtained with alkali halides, for which atomic approximation is almost exact, and with alkali metals, for which the free-electron approximation is almost correct.

Atomic Approximation.—In applying atomic approximation to insulating solids most of the concepts gained from studying isolated atoms and molecules may be used; for example, the idea of electron shells. Also used is the notion that no two electrons can move in a given orbit unless they have opposite spin; this essentially is the Pauli exclusion principle (*see* NUCLEUS: *Further Properties of Nuclei;* PERIODIC LAW; SPECTROSCOPY). This principle can also be expressed by saying that no two electrons can have the same one-electron wave function unless they have opposite spin; in this form it can be applied to the case in which free-electron approximation is employed.

Recall that a solid described by atomic approximation is necessarily an insulator because the valence electrons are not free to roam about the specimen. In addition, it may remain an insulator when the electrons are excited to higher states of motion provided the orbits of the excited electrons are still bound to the atoms. The excited electrons will become free if given sufficient energy for the same reason that atoms and molecules become ionized when excited sufficiently: only a finite energy is required to remove any electron from an atomic system. If the specimen is excited by light certain ranges of frequency may be absorbed that do not make the crystal conducting. This is the case in which there are nonconducting excited states. On the other hand, as the energy of the light quanta increases a point will be reached at which the electrons are freed and the specimen becomes photoconducting. Most of the common insulating solids behave according to this model. That is, they do not become conducting when irradiated with the longest wavelengths which are absorbed by atoms of the bulk material.

The insulating solid may contain impurity atoms, a stoichiometric excess of one of the constituents, or imperfections. In such case, the electrons attached to the foreign atoms, or those in the vicinity of the imperfections, will not be bound in the same way as the electrons of the bulk solid. As a result they will have their own characteristic excited states and ionization potentials. If these excited levels lie closer to the ground state than those of the bulk solid, the longest wavelengths of light absorbed by the specimen will be characteristic of the foreign atoms or imperfections. Hence an impure crystal may become photoconducting as a result of its impurities. Most photoconducting solids used in practice are of this variety. If electrons can be freed from the atoms of an insulator by thermal fluctuations at ordinary temperatures the solid evidently is an intrinsic semiconductor. Similarly, if an otherwise ideal insulator contains foreign atoms that may be ionized thermally at normal temperatures the material is an impurity semiconductor.

Free-Electron Approximation.—Electrons moving in identical orbits on different atoms are considered to have the same energy in atomic approximation. On the other hand, it is generally assumed that the electronic states have different energies in free-electron approximation. This result is a consequence of the Pauli exclusion principle which states that no two electrons in the same system can have the same set of quantum numbers.

The energy levels of individual electrons may be grouped into nearly continuous bands. Separation of neighbouring levels in a band decreases with increasing size of specimen and is so small in all practical cases that it cannot be resolved by direct measurement.

The different bands of levels may be separated or may overlap, depending on the material (fig. 4). In general, the larger the fluctuations in internal electric field, the greater the likelihood that the bands will be separated. It is more likely that the bands will be separated in a monatomic solid containing quadrivalent atoms, such as carbon, than in one containing alkali metal atoms, since the electrostatic field strength will be larger and will fluctuate more strongly in the first case. Similarly, it is more likely that the bands will be separated in a salt containing strongly charged ions than in a monatomic solid in which each is neutral.

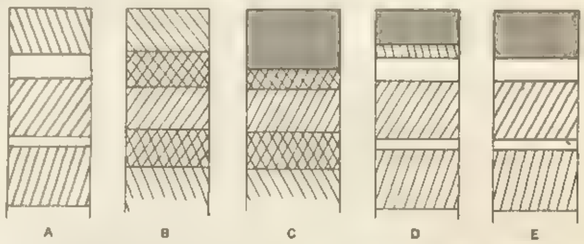


FIG. 4.—ENERGY LEVELS AS OCCUPIED BY ELECTRONS IN SOLIDS. HATCHED AREAS DESIGNATE QUASI-CONTINUOUS BANDS; SHADED AREAS, OCCUPIED LEVELS

(A) Quasi-continuous bands do not overlap; gaps represent forbidden regions of energy. (B) Bands overlap. (C) A metal; occupied levels border unoccupied levels infinitely closely. (D) A theoretical metal unknown in practice; uppermost occupied band only partly filled. (E) An insulator; occupied region extends to top of highest occupied band

Although the spacing between levels in a band is very small, the spacing itself is of practical importance because of the Pauli exclusion principle. A specimen of ordinary size contains about 10^{23} atoms and hence at least that number of valence electrons. Only two electrons at most can have the same wave function and these must have opposite spin. Although the energy levels are degenerate (that is, more than one wave function is associated with each), the degeneracy is small so that only a few electrons may occupy any given level. As a result the electrons are distributed over a large number of levels, or over a wide range of the spectrum. In the normal state of the solid the occupied levels will be those that lie as low as possible. Thus, all levels below a given point in the spectrum will be occupied as densely as possible. It is evident that the width of the occupied region depends upon the spacing of levels; hence, the importance of this spacing.

Under normal conditions the electrons in the densely occupied portion of the spectrum are so paired that there is no net flow of current even though each electron is moving through the solid very rapidly. In other words, for each electron moving in one direction there is another moving in the opposite direction at equal speed. If the bands overlap, it follows that the uppermost occupied level of the energy spectrum is always very near to the lowest unoccupied level (fig. 4C). Hence, when this type of solid is placed in even a very weak electrostatic field some electrons are excited into unoccupied levels, statistical balance of electron motion is upset, and current flows; this is characteristic of a metal.

Thus any solid in which the top of the occupied region is very close to unoccupied levels is a metal. This evidently will always obtain when the bands overlap, and even if the bands do not overlap provided the top of the occupied region does not extend to the upper edge of the highest band occupied (fig. 4D).

If the bands do not overlap and if the occupied range extends to the top of a band (fig. 4E), the solid will not be a metal. In this case the statistical balance among electrons can be disturbed only if the applied electrostatic field is so large (usually about 10^6 v./cm.) that electrons can be stimulated from the highest occupied band to the lowest unoccupied one. Thus, in free-electron approximation, an insulator is described as shown in fig. 4E in which the occupied region is separated from the unoccupied one by a large margin.

If the spacing between the highest filled band and the lowest empty band is sufficiently small, as in fig. 4E, electrons may be thermally excited to the upper band. A solid of this type will be a semiconductor even when pure and of ideal stoichiometric composition; that is, it will be an intrinsic semiconductor. The width of the gap may be determined from the dependence of electronic conductivity on temperature in such cases.

The crystal gains conductivity on electron transfer between bands because it now contains an unimpeded electron in a nearly empty band and the electrons in the filled band are no longer completely impeded, since one electron is absent. Such a nearly filled band is said to contain a hole, analogous to a bubble or void in an otherwise filled container of liquid. The hole acts as a positively charged particle, since a negative electron has left an otherwise neutral system. It has been shown that a hole behaves as a positive charge in electric and magnetic fields.

Considerable success has been achieved in determining the arrangements of bands and the manner in which they are occupied for a large number of simple solids. Reasonably accurate diagrams of the bands are available for many relatively simple metals and salts and for valence crystals such as diamond, silicon, and germanium.

Van Vleck and H. Brooks developed a method for relating the behaviour of the electrons in simple solids to the behaviour in the free atom without solving the equations of quantum mechanics directly. This method is successful in treating the cohesive properties of a number of simple solids, particularly metals. In a sense the procedure relates the properties of the solid to easily measured properties of free atoms without the complexities of solving the basic equations of the theory.

2. Cohesion.—The cohesion of a solid may be expressed in terms of the energy required to remove an atom from the surface to infinity; that is, the heat of sublimation, which can be computed once the energies of the electrons are known. The one-electron approximations usually fail to give accurate values of cohesive energy. More refined calculations made for simple solids (such as alkali metals and alkali halides) do, however, show reliable agreement with experimental data on bonding. The one-electron methods, on the other hand, have been adequate to estimate relative cohesive energies of different solids. Jones made a satisfactory interpretation of the Hume-Rothery rules with the use of band approximation, thus accounting for the importance of the electron-atom ratio in substitutional alloys. The same procedure was used to explain the comparatively great stability of metals with partly filled d and f shells.

IV. EXPERIMENTAL APPROACHES

Information about electron energy levels has been obtained by allowing solids to interact with electromagnetic fields. In metals such experiments provide information about the Fermi surface (the boundary between occupied and unoccupied levels). The experiments allow measurement of energy gaps in semiconductors, and yield information about electrons in the nearly empty bands or about holes.

A crystal is three-dimensional, and so is its energy-band structure; *i.e.*, detailed grouping of energy levels and the width of an energy gap depend on direction of motion of the electron in the crystal. Electron motion is wavelike, and may be described by a wave vector $\mathbf{k}$; thus electron energy is a function of the magnitude and direction of $\mathbf{k}$. In mathematical language energy $E = E(\mathbf{k})$. An electron energy level corresponding to the wave vector $\mathbf{k}$ with components $(\mathbf{k}_x, \mathbf{k}_y, \mathbf{k}_z)$ may be represented as a point in the three-dimensional wave-vector space. This space plays a fundamental role in the theory of energy bands; the Fermi surface is the boundary between occupied and unoccupied levels in the wave-vector space.

If the electrons were perfectly free (as they might be in an electron gas) the wave vector $\mathbf{k}$ would be simply proportional to the momentum $\mathbf{p}$ of the electron. In other words the De Broglie relationship $\mathbf{p} = \frac{h}{2\pi}\mathbf{k}$, or $|\mathbf{p}| = \frac{h}{\lambda}$ would hold, where h is Planck's constant and λ is the electron wavelength. The energy of an electron would thus be proportional to $\mathbf{k}^2$. For the lowest energy states in the gas to be populated, the Fermi surface would become spherical. In an actual solid the relationship between wave vector and momentum is more complicated; the Fermi surface is not spherical but its shape does reflect the basic symmetry of the crystal. Furthermore the energy gaps in the level structure depend on the direction of $\mathbf{k}$. The locus of energy gaps in $\mathbf{k}$ space is called the Brillouin zone boundary.

Fig. 5 shows the Fermi surface of copper as deduced from a theoretical model and experiment. The boundary of the first Brillouin zone is also shown. It is seen that the Fermi surface actually comes into contact with the Brillouin zone boundary at a number of points. If the potential field in the crystal were much weaker, the Fermi surface would approximate a sphere completely enclosed by the Brillouin zone.

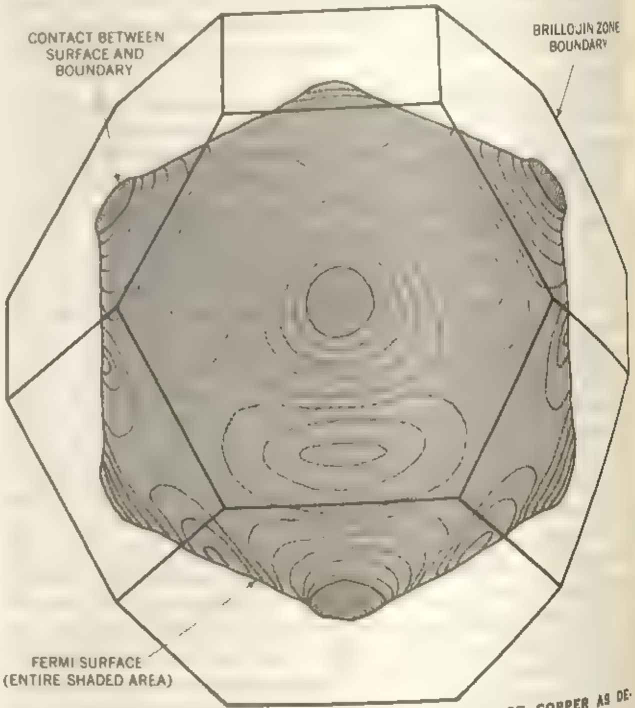


FIG. 5.—FERMI SURFACE AND FIRST BRILLOUIN ZONE OF COPPER AS DEDUCED FROM THEORY AND EXPERIMENT

1. Metals.—Various features of the Fermi surface for a given metal can be measured in separate types of experiment, and sometimes pieced together to get a good overall picture. These studies generally use a static magnetic field, and require single crystals of high purity; data are gathered at very low temperatures so that collision time for electrons is long compared with the time per orbit in the magnetic field.

De Haas-Van Alphen Effect.—The De Haas-Van Alphen effect involves oscillation in magnetic susceptibility observed in most pure metals at low temperatures when the magnetic field is increased or decreased. The oscillation has a constant period when the susceptibility is plotted against reciprocal magnetic field. Detailed calculations show that the period of oscillation measures

the extremal cross section of the Fermi surface in a plane perpendicular to the magnetic field direction. If the Fermi surface is complicated, several extremal cross sections may be present simultaneously, and the different periods corresponding to these cross sections beat together in the resulting oscillation. It is possible to separate the individual periods by means of Fourier analysis.

Cyclotron Resonance.—A static magnetic field moves the electrons in orbits about the magnetic field lines. If a high-frequency electromagnetic field is applied simultaneously, and if the frequency of this field coincides with the frequency of orbital motion, the electrons can absorb energy from the electromagnetic field. This is cyclotron resonance. The measured frequency of maximum absorption for a known static field is used to compute the effective mass of electrons at the Fermi surface.

Anomalous Skin Effect.—At high frequencies the electric current carried by a metal is confined to a thin surface layer of thickness called the skin depth. If the electron's mean free path for making a collision is large compared to the skin depth the conduction is called anomalous. The measured surface impedance of the metal in the anomalous range can be translated into a measurement of total free area of Fermi surface (*i.e.*, Fermi surface area not in contact with a Brillouin zone boundary).

Magnetoacoustic Absorption.—Magnetoacoustic absorption is anomalous attenuation of acoustic waves as they propagate through metal under a static magnetic field. When acoustic-wave frequency is appropriate to the frequency of electronic motion in the magnetic orbits, the electrons in the solid can absorb energy from the wave. In this type of experiment several values of magnetic field that lead to strong attenuation of waves of a fixed frequency are measured and can be used to determine dimensions of the Fermi surface.

Magnetoresistance.—Recall that the electrical conductivity of a material is modified by the presence of a magnetic field. If the limiting behaviour of the resistivity in strong magnetic fields can be determined, it can shed light on the topology of the Fermi surface. Specifically, if the orbit in wave-vector space of an electron at the Fermi surface touches the Brillouin zone boundary at a number of different points, the asymptotic behaviour of the resistance will differ from the case where the orbit does not touch the boundary. The quite complicated theory of magnetoresistance in metals was first worked out in detail by the Soviet scientists I. M. Lifshitz and V. G. Peschanskii.

2. Semiconductors.—Many experimental methods for metals (particularly cyclotron resonance) are used for semiconductors; measurements are not made of the Fermi surface, but of properties of conduction electrons and holes.

Piezoresistance.—The change in electrical resistance of a material when it is stressed is piezoresistance. In a single crystal the stress can be applied in different ways and the electric field can be applied in various crystallographic directions; appropriate combinations of these yield an array of piezoresistance coefficients for the material. In semiconductors the piezoresistance coefficients are usually much larger than are those of metals. Their measurement (first performed by C. S. Smith in 1954 for the semiconductors germanium and silicon) allows inferences about the location of excess electrons and holes in wave-vector space. Because of their large piezoresistance, small germanium wafers are extensively used as strain gauges.

Optical Absorption.—A semiconductor cannot absorb electromagnetic radiation unless the photon energy is sufficient to lift an electron across the energy gap. Information on minimum energy gap, the density of energy levels on each side of the gap, and other gaps between filled and vacant conduction bands can be obtained from measurements of the change in optical absorption as frequency of incident electromagnetic radiation is varied. Such measurements can be obtained from experiments in which a light beam is transmitted through or reflected from thin films of the material. Measurements of optical absorption have yielded information about the energy-band structure of valence semiconductors (*e.g.*, germanium, silicon, indium antimonide, and tellurium).

V. CRYSTAL IMPERFECTIONS

It has been noted that imperfections help determine some of the most important properties of crystals, including the electronic conductivity of semiconductors, the electrolytic conductivity of salts, and the plastic properties of all ductile materials. Before 1940 it might have been supposed that new imperfections would be introduced almost indefinitely to describe the properties of solids. However, by the 1960s it appeared that the complexity arises from a few types of imperfection that play an intricate role within the crystal. The principal imperfections seem to be the following.

1. Lattice Vibrations (Phonons).—Phonons have been identified as excitation units of lattice vibration propagating through and disordering a crystal. A random group of vibrational waves (such as might arise through heating the solid) may interfere with one another constructively and destructively. Thus phonons impart a large amount of energy to a single atom when constructive interference is appropriate. Such an event corresponds to a large thermal fluctuation, which may be enough to eject an atom from a normal position in the lattice or effect a similar change.

2. Free Electrons and Holes.—As mentioned, the ideal insulating solid is one in which an uppermost filled band is separated from an empty one by an appreciable energy gap. Such an insulator becomes conducting if free electrons are excited into the empty band to leave free holes in the filled band. The conductivity depends on the number of such electrons and holes and the ease with which each moves; in an otherwise perfect insulator they can be looked upon as a type of imperfection.

3. Excitons.—A hole and an electron in a given insulator have opposite charges and, as a result, attract one another in accordance with Coulomb's law (*see* ELECTRICITY: *Electrostatics*). Just as an electron and a proton in a hydrogen atom have bound states in which they rotate about a common centre, an electron and a hole can move as a closely coupled uncharged unit. Called an exciton, this unit can transport energy from one part of the crystal to another. The exciton can be viewed as being generated by putting the insulating crystal in an excited electronic state, similar to the discrete states of an atom which lie below the ionization limit. The exciton is relatively difficult to detect directly, being uncharged and not deflected easily by electric or magnetic fields. Excitons have been detected, however, in salts such as alkali halides and cuprous oxide. They probably are important in photochemical processes, particularly when energy associated with incident light mediates a chemical reaction at the surface of a crystal. In such cases the energy may be transferred by excitons from the interior (where the light was absorbed) to the surface.

4. Vacancies and Interstitial Atoms.—Thermal fluctuations may drive atoms from normal lattice sites into interstitial positions, producing vacant sites. Such imperfections can lead to even more disorder; *e.g.*, one of the atoms in a normal site neighbouring a vacancy may jump into the vacancy. Or an interstitial atom may replace a neighbouring normal atom by pushing it into another interstitial place. Such mechanisms give freedom of motion to the typical atoms of a lattice. The migration of thermally induced vacancies and interstitial atoms is responsible for diffusion and electrolytic conductivity in most solids.

Vacancies and interstitial atoms can be produced in pairs by bombarding a lattice with fast particles such as protons, neutrons, and electrons produced by high-voltage accelerators. The defects may also be produced by radioactive disintegration of atoms within the lattice, reflecting so-called radiation damage. Studies of such damage and its recovery illuminate the behaviour of interstitial atoms and vacancies in typical solids.

5. Foreign Atoms.—Atoms with loosely bound electrons or holes may transform an insulator into an impurity semiconductor. Similarly, appropriate foreign ions in ionic crystals may be accompanied by vacant lattice sites or interstitial atoms that become free at normal temperatures to increase electrolytic conductivity and diffusion. For instance, divalent ions such as Ca^{2+} replace two $Na+$ ions when dissolved in NaCl. The calcium ion occupies one site but the other is vacant and may become mobile.

Similarly, S^{2-} can be dissolved in silver chloride by substituting for Cl^-. The ion is accompanied by an Ag^+ ion in an interstitial site.

Extra electrons associated with foreign atoms in insulating solids usually absorb and emit their own characteristic frequencies of light. Thus such foreign atoms may absorb or emit radiation and make the crystal fluorescent (*see* LUMINESCENCE).

6. Dislocations.—Dislocations are the most intricate of the imperfections in the lattice; they were introduced earlier in this article to explain plastic flow in ductile solids. Those called edge dislocations also may act as relatively easy sources or sinks for vacancies or interstitial atoms. Such imperfections can be produced by thermal fluctuations at dislocations in the interior with the same ease as at the crystal surface. What is termed a screw dislocation can terminate at a ridge on the surface, and is an ideal catalyst for the growth of the crystal from solution or vapour. The regions where dislocations break through to the surface can be detected by etching the surface with appropriate reagents. The point of emergence becomes the centre of an etch pit.

7. Interaction of Imperfections.—Lattice waves (phonons) may interact with electrons and holes and affect the current flowing through a crystal. Similarly, appropriate impurity atoms may capture free electrons or holes produced by light or other excitation and prevent them from contributing to the current. When edge dislocations are forced to move (as in plastic flow) they apparently leave a trail of vacant lattice sites that can enhance atomic migration in crystals. Irregular dislocations, particularly those present after plastic flow, can capture free electrons and holes.

Negative ion vacancies in salts (which are positively charged) may capture free electrons and produce what can be regarded as new imperfections. These neutral units are termed *F*-centres. They alter the colour and many other properties of the specimens containing them. They may be introduced in controlled amounts by heating an alkali halide crystal (such as sodium chloride) in alkali metal vapour. An excess of metal is absorbed by the crystal but actually appears in such a form that some of the halogen ions are replaced by electrons. The vacant lattice sites necessary for *F*-centres can be produced in a number of ways; *e.g.*, by plastic flow or by irradiation with ultraviolet light or X rays.

When a photographic film is exposed to light, a latent image that can be developed chemically is formed in the grains of silver halide. The process appears to involve interaction of a large number of imperfections, and was described theoretically from the atomic point of view by R. W. Gurney and Mott in 1938. They postulated that incident light produces free electrons and holes that wander about the crystal. The electrons become trapped at imperfections, probably dislocations or possibly foreign atoms. The trapped electrons attract interstitial silver ions produced thermally and which are mobile at room temperature. Clusters of silver atoms formed by this union of electrons and interstitial ions constitute the latent image. Such specks of silver may catalyze the transformation of the entire grain into free silver when a developer is present. The deposition of silver atoms at dislocations in silver bromide has been demonstrated. It has also been shown that *F*-centres in the alkali halides will precipitate along dislocations if the crystals are held at a temperature favouring the formation of colloids.

VI. SOLID STATE ELECTRONICS

Applications of solid state physics are perhaps best exemplified in electronics. Semiconductor devices have had a heavy impact on the electronics industry; *e.g.*, by supplying miniature substitutes for conventional vacuum tubes. Such magnetic materials as ferrite-core inductances have replaced air-core inductances in many high-frequency applications with subsequent miniaturization. Permanent magnet materials have been developed; one of these, the elongated single-domain particle, had its origin in a theoretical study of magnetic domain structure.

1. Semiconductors and Rectifiers.—Recall that a semiconductor is a solid in which the occupied range of electron energy levels is separated from the unoccupied region by an energy gap sufficiently small that electrons can be thermally excited across it. The occupied range is called the valence band and the unoccupied region beyond the gap is called the conduction band. Excitation to create a free electron in the conduction band and a free hole in the valence band is usually termed thermal generation of a hole-electron pair. The electron and hole can collide and recombine.

The number of free charge carriers (hence the conductivity) can be greatly increased by including small amounts of suitable impurities when the semiconductor crystal is formed. These impurities can be donor atoms which furnish extra electrons to the energy level structure of the parent semiconductor, or acceptor atoms which remove electrons. For example, in the common semiconductors germanium and silicon the four valence electrons go into the valence band, and semiconduction in the pure material occurs only through thermal generation of hole-electron pairs. Pentavalent atoms (*e.g.*, phosphorus, arsenic, and antimony) are donors; when such an atom substitutes for germanium (or silicon) in the crystal, it contributes four electrons to the valence band and one to the conduction band. Trivalent boron, aluminum, and gallium are acceptors since each substituted atom contributes only three electrons to the valence band.

Extrinsic semiconductors (those in which conductivity is controlled by impurities) do not contain equal numbers of free electrons and holes; both types of charge carrier are present but the concentration of the minority carrier is usually negligible compared to that of the other. A semiconductor that contains donor atoms is called an N-type (or excess) semiconductor since conduction is primarily by free electrons in the conduction band. A semiconductor containing acceptor atoms is called P type, and conduction is primarily by holes. Most industrial and research applications are for extrinsic semiconductors.

Joining two semiconductors with different properties (or a metal and semiconductor) often produces a rectifying junction in which current is conducted most readily in one direction. One of these was the "cat's whisker" detector used in the early days of radio reception to rectify the feeble radio-frequency signal; it consisted of a metal wire making a point contact to a semiconductor (usually lead sulfide). Other early devices with larger current-handling capacity included copper oxide and selenium rectifiers. Known as barrier-layer rectifiers, they are more complicated in operation than the junction rectifier described below.

A P region and an N region in the same crystal produce what is called a P-N junction which functions as a good rectifier; such rectifiers made of silicon have largely replaced other types (except for high-voltage applications where vacuum diodes are superior; *see* ELECTRON TUBE).

Rectification in a P-N junction can be understood by reference to fig. 6. Two dissimilar substances brought into contact tend to come into equilibrium by exchanging heat and electrons. Equilibrium in terms of electrons is expressed by equality of the electrochemical potential (also called the Fermi level) in the two materials. In a semiconductor the Fermi level lies in the energy gap close to the valence band in a P type, and close to the conduction band in an N type. Fig. 6A represents energy-level structure at equilibrium; a few electrons in the conduction band can diffuse into the P region to combine with holes. Similarly a few holes can diffuse into the N region to combine with free electrons. both diffusion processes contribute current in the same direction. This is balanced by a thermally generated reverse current of electrons in the P region and holes in the N region (which move readily across the junction) and the net current is zero.

If the Fermi level in the N region increases by voltage V, as in fig. 6B, diffusion of electrons and holes over their respective barriers is increased; if the Fermi level drops diffusion is inhibited. In both cases thermal generation of electrons in the P region and holes in the N region is unaffected. The net current I through the junction is given by

$$I = I_o[\exp(qV/kT) - 1]$$

which leads to rectification when the voltage difference V across the junction is greater than kT/q. Here T is the absolute temperature, k is Boltzmann's constant, q is the electronic charge, and exp is the exponential function.

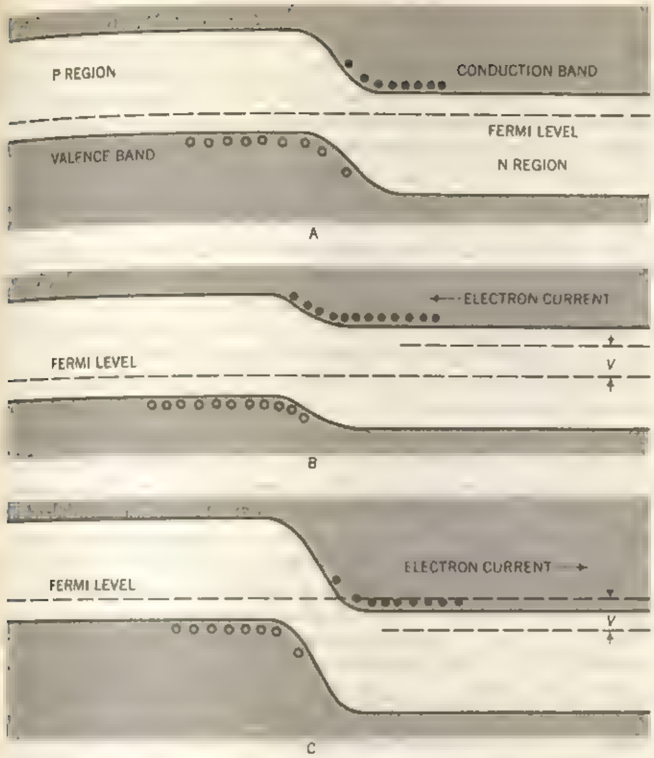


FIG. 6.—ELECTRON ENERGY LEVELS AT A P-N JUNCTION. BLACK DOTS REPRESENT ELECTRONS; CIRCLES REPRESENT HOLES
(A) Equilibrium; (B) forward voltage difference V applied across junction; (C) reverse V applied

To make practical devices some of the metallic contacts to the semiconductor must be nonrectifying. Metals with small work functions generally make nonrectifying (ohmic) contacts to N-type semiconductors while those with large work functions make rectifying contacts; the opposite is the case for P types. The work function of a metal is the minimum energy (usually expressed in electron volts) required to remove an electron from the surface of the metal.

Semiconductor crystals with more than one P-N junction are known as junction transistors and are discussed under TRANSISTOR.

2. Tunnel Diodes.—The tunnel diode is a two-terminal semiconducting device consisting of a single P-N junction, but its current-voltage characteristic is quite different from that of a conventional junction rectifier. The difference comes about through the inclusion of suitable impurities; the P and N regions are made highly conducting, their conductivity being about 1,000 times greater than that of the P-N junctions discussed above. This moves the Fermi level into the valence band on the P side, and into the conduction band on the N side, and under equilibrium conditions vacant levels on one side of the junction stand opposite filled levels on the other. Furthermore, the junction is quite narrow.

Under these conditions electrons can tunnel directly through the junction from occupied levels to vacant levels of the same energy on the other side. The phenomenon cannot be understood in terms of classical physics; it is a quantum mechanical effect intimately associated with the wavelike properties of the electron.

The forward characteristic of the tunnel diode is shown in fig. 7. As the Fermi level of the N region is raised relative to that of the P region through application of external voltage, more vacant energy levels on the P side are brought into coincidence with the filled conduction levels of the N region, and the tunnel current increases. If the forward bias is increased beyond a certain value (approximately 0.05 v. in germanium) the levels on opposite sides of the junction move out of coincidence, and the tunnel current drops. This decrease in current as voltage increases is generally called a negative-resistance or negative-conductance characteristic; in other words, an incremental increase in voltage produces an incremental decrease in current through the device. As forward bias is further increased, the normal rectifier characteristic takes over.

Because of its negative-conductance region, the tunnel diode can be used in electronic circuits to produce amplification or oscillation. The device has replaced transistors in many applications, offering the advantages of extremely high-frequency operation, small size, low noise, low cost of manufacture, and high reliability.

3. Photoconductors and Xerography.—Materials that lose electrical resistance when subject to illumination are photoconductors. The phenomenon is most striking when the unilluminated material has low conductance, since the mobile carriers created through absorption of light quanta can then produce a drastic reduction in resistance. A good photoconductor is usually an intrinsic semiconductor; each photon (light quantum) absorbed by the material excites an electron across the energy gap to create an electron-hole pair.

One of the more interesting applications is that of xerography, a photocopy process. If the photoconductor has a large enough resistance in the dark, it can store charge for appreciable periods of time; *e.g.*, amorphous selenium with a dark resistivity in excess of 10^{11} ohm-cm. can store charge for seconds. If a thin layer of photoconducting material is charged electrostatically and then illuminated with a focused or projected image, the bright regions become electrically conducting and discharge. After the illumination is removed, the photoconductor can be dusted with a frictionally charged powder that adheres to the remaining surface charge. The photoconductor thus carries a powder reproduction of the original image which can be transferred by contact to paper, and fixed by heat to provide an inexpensive photocopy. Since some of the powder tends to remain on the paper, a number of copies of the same image can be made. Commercially available xerographic equipment can reproduce information with a resolution in excess of 10 lines per millimetre at a speed of about 1,200 linear feet per minute.

4. Solar Batteries.—In the vicinity of a P-N junction the electron and hole in a pair are separated by the strong electrostatic field; electrons move into the N region and holes into the P region. Hole-electron pairs are thermally generated in small numbers at all times and contribute to the reverse current through the junc-

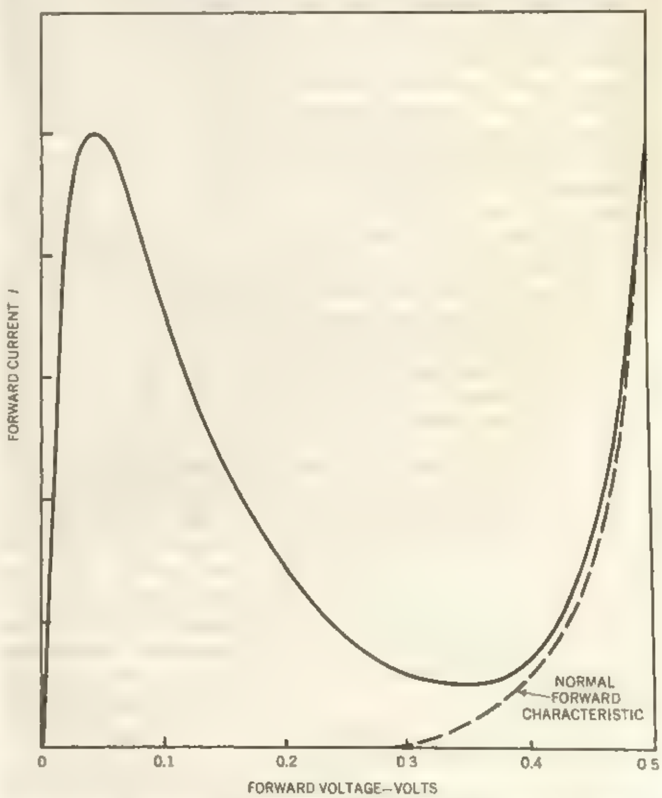


FIG. 7.—CURRENT-VOLTAGE CHARACTERISTIC OF A TUNNEL DIODE

tion; but if many pairs are created by absorption of light, the equilibrium of the junction is disturbed and the P region develops a positive potential relative to the N region. In other words, the P-N junction becomes a battery when illuminated.

The circuit may be completed by joining the P and N regions through an external load. The illuminated junction generates current continuously, and power is dissipated in the load. Photovoltaic conversion efficiency in such a device principally depends on width of energy gap in the semiconductor relative to photon energy in the incident radiation. Silicon cells have a theoretical conversion efficiency of about 20% for solar radiation; commercial cells with efficiencies greater than 12% have been constructed.

Solar batteries were found to be quite reliable in early U.S. space satellites (*see* SPACE EXPLORATION: *Typical Spacecraft*). They are also used as low-power generators for remote earth-based locations where climatic conditions permit their use.

5. Thermoelectric Materials.—Important applied thermoelectric effects are conversion of heat energy into electricity (thermoelectric power generation), and the use of electrical energy to maintain temperature below the ambient level (Peltier cooling or refrigeration). The efficiencies of both processes depend on the temperature interval and on certain material parameters. Theoretical analysis has shown that the important combination of material parameters is $\alpha^2/\rho\kappa$, where α is the Seebeck coefficient, ρ is electrical resistivity, and κ is thermal conductivity. The Seebeck coefficient is defined as the thermoelectric voltage developed by the material for a 1° C temperature difference (*see* THERMOELECTRICITY).

The best thermoelectric materials appear to be extrinsic semiconductors with carrier concentrations in the range of 10^{19}/cc. However, there are important differences among semiconductors, and the development of useful materials required substantial research. Modest conversion efficiencies are possible with the semiconductors: bismuth telluride or lead telluride. In the 1960s much effort was being devoted to study of high-temperature semiconducting oxides.

See also references under "Solid State Physics" in the Index.

BIBLIOGRAPHY.—W. Hume-Rothery, *The Metallic State* (1931), *Electrons, Atoms, Metals and Alloys* (1963); J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities* (1932); N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (1936); C. F. Elam, *The Distortion of Metal Crystals* (1935); E. Schmid and W. Boas, *Kristallplastizität* (1935); H. Fröhlich, *Theorie der Metalle* (1937); J. C. Slater, *Introduction to Chemical Physics* (1939); R. Becker and W. Döring, *Ferromagnetismus* (1939); N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals*, 2nd ed. (1948); F. Seitz, *The Modern Theory of Solids* (1940), *The Physics of Metals* (1943); C. S. Barrett, *Structure of Metals*, 2nd ed. (1952); R. Smoluchowski *et al.*, *Phase Transformations in Solids* (1951); W. Shockley, *Electrons and Holes in Semiconductors* (1950); W. Shockley *et al.*, *Imperfections in Nearly Perfect Crystals* (1951); W. T. Read, *Dislocations in Crystals* (1953); A. H. Cottrell, *Dislocations and Plastic Flow in Crystals* (1953), *Theory of Crystal Dislocations* (1963); R. Gomer and C. S. Smith (eds.), *Structure and Properties of Solid Surfaces* (1953); A. H. Wilson, *Theory of Metals*, 2nd ed. (1953); Physical Society of London, *Report of the Bristol Conference on Defects in Solids* (1955); C. Kittel, *Introduction to Solid State Physics*, 2nd ed. (1956); W. A. Harrison and M. B. Webb, *The Fermi Surface* (1960); H. M. Rosenberg, *Low Temperature Solid State Physics* (1963); D. Fox *et al.* (eds.), *The Physics and Chemistry of the Organic Solid State*, 2 vol. (1963–65). F. Seitz and D. Trumbull (eds.), *Solid State Physics*, 16 vol. (1955–1964), provides an extensive guide to the literature of the field. (F. Sz.; J. R. Rz.)

SOLIHULL, a county borough in Warwickshire, England, 8 mi. (13 km.) SE of Birmingham on the road to Warwick, and on the main Birmingham–London Railway. Pop. (1961) 95,977. Solihull grew rapidly after the 1920s from an urban district in 1932 (pop. 25,373) to a municipal borough in 1954 (72,000). It became a county borough in 1964, and includes Shirley, Olton, Sheldon, Elmdon, Knowle, and Dorridge. The parish church of St. Alphege, with its 13th-century chancel, is the most important old building. Others include Packwood House (15th century), owned by the National Trust, with a unique yew garden. Solihull public school was founded in 1560. Though mainly a residential area, Solihull has automobile and other light industries. (W. M. ML.)

SOLIKAMSK, a town in Perm *oblast* in the Russian Soviet Federated Socialist Republic of the U.S.S.R., stands on the small Usolka River, 7 mi. above its confluence with the Kama. Pop. (1959) 82,874. The town, probably the first Russian settlement in the Urals, was founded in the early 15th century as a salt-mining centre and it has remained so to the present. Its deposits of potassium salts are among the world's largest. The important chemical industry based on them produces fertilizers, chlorides, bromide, sulfides, and industrial alcohol. The magnesium works are the largest in the U.S.S.R. A railway and highway join the town to Berezniki on the Kama.

In 1959 the satellite town of Borovsk was merged with Solikamsk. (R. A. F.)

SOLIMAN: *see* SULEIMAN I; SULEIMAN II.

SOLINGEN, a town of West Germany in the *Land* (state) of North Rhine-Westphalia, Federal Republic of Germany. One of Europe's oldest centres for the manufacture of cutlery, it is equaled in importance only by Thiers, France, and Sheffield, Eng. Pop. (1961) 169,930. The town lies on the Wupper River, 17 mi. (27 km.) E of Düsseldorf, and the centre, severely damaged by bombing during World War II, has been rebuilt with wide streets and parks. The craft of making swords was established in Solingen in the Middle Ages and was still carried on in the mid-20th century. Cutlery, widely exported and chiefly produced in small- to medium-sized factories, accounted in 1965 for about 25% of the town's industry. Metallurgical industry predominates (machine tools, tubes, foundries); the remainder ranges from chemicals to sweets. Much of Solingen's immediate surroundings are forest and farmland. Chartered in 1374, the town was enlarged in 1929 to include Ohligs, Wald, Gräfrath, and Höhscheid. A museum at Gräfrath illustrates the development of cutting implements and cutlery from prehistoric to modern times. Solingen is connected to the Cologne-Ruhr Autobahn and the Cologne-Wuppertal railway, which pass west of the city centre.

SOLIPSISM (Latin *solus*, "alone," *ipse*, "self"), in philosophy, a term originally applied to moral egoism (*q.v.*) but now used in an epistemological sense to denote the extreme form of subjective idealism which denies that the human mind has any valid ground for believing in the existence of anything but itself. F. H. Bradley, in *Appearance and Reality*, characterized the solipsistic view as follows: "I cannot transcend experience, and experience is *my* experience. From this it follows that nothing beyond myself exists; for what is experience is its (the self's) states." Presented as a solution of the problem of explaining our knowledge of the external world (*see* KNOWLEDGE, THEORY OF), it is generally regarded as a *reductio ad absurdum*. *See* further IDEALISM; and compare PHENOMENALISM.

SOLÍS, JUAN DÍAZ DE (1470?–1516), Spanish navigator and one of the early explorers to enter the Río de la Plata estuary. As a pilot major of Spain he led an expedition which left Sanlúcar de Barrameda, Spain, on Oct. 8, 1515, in three ships. After touching the coast of modern Brazil around Cape São Roque, he sailed down the east coast of South America and in January 1516 reached the Río de la Plata (River Plate), which he called Mar Dulce. He entered the estuary and named the first island he found Martín García after one of his sailors who had died. Sailing partly up the Uruguay River, he landed with eight men on the left bank, in modern Uruguay, inhabited at the time by the Charrúa Indians, who attacked his party while they were taking possession of the land for Spain. All but one were killed and eaten by the Indians in sight of the remaining crewmen on the ships. Francisco del Puerto, the sole survivor, was made prisoner, and later gave valuable information about the area to Sebastian Cabot when he arrived in the region in 1526.

The ships sailed back to Spain, but one was wrecked off the island of Santa Catarina near the south coast of Brazil, and 11 of the crew saved themselves by swimming ashore. Survivors of this group were rescued by Sebastian Cabot and told him of great wealth in the interior of South America. Cabot, as a result, sailed south, entered the Río de la Plata and ascended the Paraná River looking for the kingdom of silver. After 1948 Argentine historians accepted the contention of their colleague (Roberto Levillier) that Amerigo Vespucci and not Solís discovered the Río de la Plata, as early as 1502. (F. L. HN.)

SOLITAIRE, the name given to certain American birds of the thrush family (Turdidae) and to an extinct flightless bird. The present-day solitaires (*Myadestes* and *Cichlopsis* species) are noteworthy songsters. The best known is Townsend's solitaire (*M. townsendi*), found in the Rockies from Alaska to northern Mexico; it is a slim gray bird with a white eye-ring, white sides of the tail, and a buffy wing patch.

The extinct solitaire (*Pezophaps solitaria*), which formerly lived on the island of Rodrigues, was exterminated, along with the closely related dodos, by man and introduced animals (*see* also Dodo).

SOLITAIRE is the preferred name in the United States for all games of patience, or card games that may be played by one person. There are more solitaires than all other card games together, and solitaire is the most widely played of all forms of cards.

Most games are played with one or two 52-card packs, the cards ranking king (high), queen, jack, 10, 9, 8, 7, 6, 5, 4, 3, 2, ace. Sometimes the sequence is continuous, the ace ranking next to the king. Many solitaire games are known by a variety of names; even the game probably played by most players, Klondike, is more often miscalled Canfield, properly the name of a quite different game. The following terms describe features common to most games:

1. *The layout.*—The array of cards as first dealt on the table, comprising all cards that do not remain in the hand.
2. *The tableau.*—A distinctive arrangement into which the cards that may be "worked" are first dealt.
3. *The foundations.*—Usually all the cards of a designated rank, upon which other cards are built in sequence.
4. *The stock.*—A special pile of cards that is part of the layout in some games.
5. *The hand.*—Any remainder of the pack not used in the layout.
6. *The talon.*—A waste or discard pile into which are placed those cards turned up from the hand that cannot be immediately built upon the tableau or foundations.
7. *Available Cards.*—Those cards of the tableau, stock, etc., that are subject to transfer to other parts of the layout. In most games, certain cards of the tableau are not available until uncovered by the transfer of cards above them.
8. *A space.*—An empty place in the tableau created when all the cards of one pile have been transferred elsewhere.
9. *Rows.*—Lines of cards across the table horizontally.
10. *Columns.*—Lines of cards extending vertically away from the player.

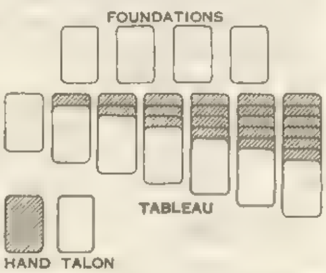


FIG. 1.—LAYOUT OF CARDS FOR KLONDIKE

A few of the popular forms of solitaire are here described.

Klondike.—One pack is used and a tableau dealt of seven columns in a row, with one card in the column at the far left, two in the next, three in the third, etc., the column at the far right containing seven cards. Only the topmost card in each column is turned face up. The foundations are aces, which are put in a row above the tableau as soon as they become available. Each foundation must be built up in the same suit in sequence to the king (*i.e.*, ace, deuce, trey, etc.). A card placed in a foundation may not thereafter be moved.

On the uppermost card of any tableau pile may be placed a card of next lower rank and opposite colour. When transferring cards within the tableau, all the face-up cards on any tableau pile must be moved as a unit onto a card of next higher rank and opposite colour to the highest card of the unit. If the highest card of the unit is a king, it may be moved only into a space, and a space may be filled only by a king (or a unit built upon a king). The lowest card of a unit may be removed and played in proper sequence upon a foundation pile.

Whenever a transfer leaves a face-down card uncovered, this card is turned up and becomes available.

The hand is turned one card at a time and is run through once only. A card not added to the foundation or the tableau is placed face up on the talon. Top card of the talon is always available.

Game is won if all cards can be built upon the foundations.

Double Klondike.—Many solitaires can be played by two, but Klondike is by far the favourite for this purpose and is the game referred to as double solitaire. Each player deals his own layout with his own deck. Neither may build on the other's tableau, but foundation piles are common property. Turn to play alternates, ending when a player lays a card on his waste pile. Play continues until one player has built all his cards on the foundations or until neither can make a further play. It is not compulsory to play a card to the centre; strategy may dictate withholding a play that would help the opponent get rid of a great number of cards.

In a blocked game, winner is determined by counting the cards each has played to the foundations.

Canfield.—From a single pack, a stock of 13 cards is counted out and the pile turned face up, with only the top card visible. The next four cards are dealt face up in a row to the right of the stock, forming the tableau. The next card is placed face up above the tableau to start the foundation row. Other cards of the same rank as this starter are placed in the foundation row as soon as available. Foundation piles are built in continuous ascending sequence in the same suit. (Thus, with ♡Q turned as the starter, other queens are played to the centre as soon as available. After the king has been added, the next card in sequence is the ace, then the 2, 3, etc.)

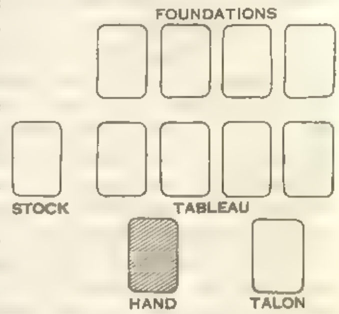


FIG. 2.—CANFIELD

Tableau columns are built in downward sequence and alternate colour, black on red, red on black. The king plays on the ace unless either rank is the foundation starter. Piles may be combined when the touching cards are of correct sequence and colour. Customarily, the entire column must be moved as a unit but some play that the available card may be individually shifted; *i.e.*, a column of ♡ 8, ♤ 7, ◇ 6 may be placed as a unit upon the ♤ 9 or ♧ 9, or the ◇ 6 may be shifted separately onto ♧ 7 if available, after which ♤ 7 may be placed on ◇ 8.

The hand is turned up in packets of three cards at a time, each packet being placed face up on the talon. The top card only of the talon or stock is available; the card immediately below becomes available when the top card is played. A space must be filled by the top card of the stock, but when the stock is exhausted the player may use any available card of the talon.

After the hand is run through once, the talon is turned over and, without shuffling, is run through again, continuing as often as plays can be made. Game is won if all cards are played onto the foundations.

Others.—The wide variety of solitaire games offers simple layout and count-out pastimes suitable for children, such as clock and accordion, as well as games that are a challenge to skill, such as calculation. Few who know how to play cards of any kind do not know at least one kind of solitaire.

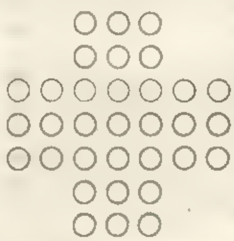


FIG. 3.—ARRANGEMENT OF HOLES FOR VERSION OF PEGBOARD GAME

Pegboard Game.—Solitaire also is the name of a pegboard game or solo puzzle, similar to a game the American Indians played with arrows stuck into the ground. In one version, the board is pierced with 33 holes arranged in a pattern of three vertical and three horizontal rows intersecting to form a cross. Pegs are set in all but one hole. One peg at a time is jumped over another and into a vacant hole, the peg thus jumped being removed from the board as in checkers. Jumps must be horizontal or vertical, never diagonal; the object of the basic puzzle is to remove from the board all but a single peg.

To make the game even more difficult, one variation calls for the last peg to be left in the centre hole.

See Morehead and G. Mott-Smith, *Complete Book of Solitaire and Patience Games* (1949). (R. L. Fy.)

SOLNA, an industrial suburb 3 mi. N.W. of Stockholm, Swed. Pop. (1960) 50,864. It is an ancient settlement and antedates Stockholm, having historic buildings including an 11th-century church and old palaces and manors. Ulriksdal palace on Lake Edsviken is the royal residence for part of the year; the Haga palace on Lake Brunnsviken is the crown prince's residence; and the Karlberg palace is now the Royal Military school.

There are also a Nobel institute for scientific research, an institute for the blind and the Karolinska hospital with a medical school. Sports facilities include a racecourse and football stadium. Its expanding industrial development includes printing works and electronic industries. It is also a centre of the Swedish film industry. (L. E. MA.)

SOLO (BENGAWAN), a river (*Kali*) in East Java, Indonesia, and the longest river (335 mi.) of Java. Its headwaters are on the slope of the Lawu volcano and the southern limestone range (Gunungkidul). Flowing through several small basins the river enters the densely populated Surakarta plain near Wonogiri. Several tributaries join it in this volcano-bordered plain. North of Surakarta (or Solo) the river gradually bends eastward round the central limestone range (Gunungkendeng). Its longest tributary, the Madiun, joins the Solo at Ngawi, where it begins its 20 mi. breakthrough of the Gunungkendeng and emerges into the region called the Solo valley. There the river has little gradient and meanders through a lowland 10 mi. wide and 80 mi. long between the Gunungkendeng and the northern limestone ridges in the district of Bodjonegoro.

A project to use the Solo river for large-scale irrigation was never completed. The delta is very marshy and used for fish ponds. In order to reduce silting of the strait between Madura and Java (the northern approach to the port of Surabaja) the main mouth of the Solo was diverted to the north, away from the strait.

In the dry season much of the river bed is dry; in the wet season (November–April) the average volume (in the Solo valley) is 440 cu.m., the maximum 2,800 cu.m.

The Solo is navigable for river *prahus* (fishing boats) from Ngawi downstream. Sea-going *prahus* in the wet season can reach Tjepu. *See* JAVA: *Archaeology* for an account of hominid remains in the region of the Solo river. (J. O. M. B.)

SOLOGNE, a district of central France contained within the great northward arc of the Loire river toward Orléans. It lies mainly in the *départements* of Loiret and Loir-et-Cher, with a southeastward extension into Cher, and is a flat area (*c.* 200 sq.mi.), covered with thick deposits of debris brought down from the crystalline rocks of the Massif Central.

The resultant soils, extremely deficient in lime, are very infertile, and the clays hold up water in slight hollows so that the countryside is studded with innumerable meres and ponds. Sluggish westward-flowing streams, the Sauldre, Beuvron and Cosson, traverse it toward the Loire.

For centuries Sologne remained a backward area, where poor crops of rye and buckwheat were grown on widely scattered farms, and most of the country was left in a wild state of woodland, marsh and water, devoted to the chase and owned in large estates surrounding châteaux, such as Chambord.

Reclamation for agriculture, undertaken after the middle of the 19th century, has made some headway, with drainage works and use of fertilizers, but there are large game preserves and buckwheat is still cultivated for pheasants. Generally a *polyculture* (mixed farming) is practised and poultry and dairy cattle are kept. The ponds, stocked with fish, and the woodlands play their part in the local economy.

The population is still sparse, and is served by market towns such as Gien, Orléans and Blois in the Loire valley, and Romorantin and Vierzon on the southern borders. (AR. E. S.)

SOLOGUB, FEDOR (pseudonym of FËDOR KUZMICH TETERNIKOV) (1863–1927), Russian Symbolist poet, novelist, and playwright, was born at St. Petersburg on Feb. 17 (Old Style; March 1, New Style), 1863. His father, a tailor, died when Sologub was quite young, and his mother's employers paid for his education. He became a teacher, and, finally, a superintendent of elementary schools, retiring in 1907. His first volumes of stories and poems and his first novel, *Tyazhelye sny* ("Bad Dreams") were published in 1896. As a poet, he became a leading representative of Russian Symbolism, combining classical limpidity of form with modernist, even "decadent," content. His work reflects a dualistic, Manichean view of life. His best and most successful novel is *Melkii bes* (1907; "The Petty Demon"; first trans. as *The Little Demon*, 1916). Also interesting and characteristic is the ambitious trilogy *Tvorimaya legenda* (1908–12; partially translated as *The Created Legend*, 1916). Sologub also wrote plays, and many highly original stories, some of which are translated in *The Old House and Other Tales, The Sweet-scented Name, and Other Fairy Tales, Fables, and Stories* (both 1915), and *Little Tales* (1917). He took a sharply negative attitude toward the Bolshevik Revolution of 1917, and, after his wife's tragic suicide in 1922, withdrew into himself. Much of the poetry of his last years remains unpublished.

Sologub died in Leningrad on Dec. 5, 1927.

BIBLIOGRAPHY.—*Sobranie sochinenii*, 19 vol. (1913–14); R. Poggioli, *The Poets of Russia: 1890–1930* (1960); J. Holthusen, *F. Sologubs Roman-Trilogie* (1960); G. Donchin, *The Influence of French Symbolism on Russian Poetry* (1958). (G. St.)

SOLOLÁ, a department in the central highlands of Guatemala and location of the beautiful Lake Atitlán. Area 410 sq.mi., of which 53 sq.mi. are covered by the water of the lake; pop. (1964) 108,815. The lake and town of Atitlán (*q.v.*), a favourite tourist area, occupies a valley dammed by volcanic ash at an altitude of about 4,900 ft. On the borders of the lake are three cone-shaped volcanoes, Atitlán, Tolimán, and San Pedro.

The Maya Indians, who make up the greater part of the population, grow maize, wheat, and beans. On the southern shore of the lake there are plantations of coffee. Much land in the department is used for the feeding of beef cattle and hogs. The lake is also used for fishing. The department is served by the Inter-American Highway which passes through Sololá on the way from Guatemala City to Quezaltenango.

Sololá, the departmental capital (pop. [1964] 21,382 [mun.]), is near the northern shores of the lake, about 45 mi. WNW of Guatemala City. It is 6,900 ft. above sea level. It is a colourful market centre for the Indian farmers of the highlands who come to buy or sell cotton and woolen textiles and foods. (P. E. J.)

SOLOMON (d. *c.* 934 [922] B.C.), the second king of the united Judah-Israel and the wise man par excellence in the Bible and in folklore. He succeeded his father, David (*q.v.*), *c.* 974 (or, according to those who favour the later chronology, *c.* 962) B.C. and reigned for 40 years. According to the lists of I Chron 3:5 and 14:4 he was the fourth son born to David in Jerusalem. Solomon ("peace," "prosperity") was the name his parents gave him, but Nathan the prophet called him Jedidiah ("beloved of the Lord"). He became king through a carefully planned plot by Nathan, swiftly carried out by Nathan and Bathsheba. When Adonijah, David's oldest surviving son, attempted to seize the throne in David's old age, Nathan acted with all speed, persuaded the aged king that he had promised the succession to Bathsheba's son, and led him himself to institute Solomon. In course of time Solomon astutely managed to eliminate Adonijah and his supporters without incurring blood guilt.

The picture of Solomon as painted by the compiler of I Kings 1–11 is previewed by the story of the young king's visit to the shrine at Gibeon (3:4–15), when Solomon asked not for riches and honour but for wisdom to be a good judge and ruler, with the result that he was granted all three. The rest of the account of Solomon's reign from 3:16 to the end of ch. 10 is designed to show the fulfillment of this promise, and it is done most thoroughly. All the untoward events of Solomon's reign are relegated to the last chapter (11), with the result that they are placed in decidedly low relief. The account of Solomon's reign in II Chron. 1–9 gives an even more impressive picture of peace and prosperity since the splendour of the Temple buildings is emphasized and all reference to rebellions and reverses is entirely omitted.

The King and His Successes.—The first ten chapters of I Kings undoubtedly convey the intended impression that Solomon's

reign was prosperous and peaceful and that he himself was wisest of the wise, richest of the rich. He married "the daughter of Pharaoh" and thus formed an alliance with Egypt. This marriage brought him the further gain of Gezer, a Canaanite fortress which the Hebrews had never been able to capture. The pharaoh had taken and reduced it, and he gave it to Solomon by way of dowry. Solomon rebuilt and fortified the city in company with numerous others which were used as barracks, arsenals, and storehouses. Even more advantageous was the extension of the treaty which David had made with Hiram, ruler of the Phoenician city of Tyre. In partnership with him, Solomon maintained a fleet of oceangoing ships, some possibly trading in the Mediterranean but others certainly sailing from Ezion-geber, a port on the Gulf of Aqaba, close by the modern Israeli port of Elath. These ships made three-year voyages and brought back "gold, silver, ivory, apes, and peacocks." The word *tukkiyyim* ("peacocks") is generally held to be the Malabar word, and the word for "apes" (*qophim*) is still in use on the Malabar coast, so presumably these eastern voyages extended at least as far as southwestern India. Perhaps that was the site of Ophir, with which Hiram traded, unless it was Ceylon, or perhaps even beyond to the Spice Islands (the Moluccas in Indonesia). It is evident that the treaties with Tyre and Egypt brought Israel for the first time fully into the current of Oriental commerce. Solomon's court was justly famed for its wealth and magnificence. The climax of the story comes with the visit of the queen of Sheba, rich and beautiful, who was astounded at what she saw. "There was no more spirit in her" (it took her breath away; II Chron. 9:4).

The long reign of Solomon, his wealth, and the temporary weakness of all neighbouring kingdoms enabled him to indulge his passion for building. He spent 13 years in constructing a splendid royal dwelling and with it a complex of buildings, including, above all, the Temple. This Temple was actually an appanage to the palace, built after the Phoenician style, and many of the symbols in it were of foreign origin: the two pillars before the entrance, the brazen "sea" and the palm and pomegranate decorations. Later days made the Temple the very centre and inspiration of Israel's life and devotion, and for this reason the description of it occupies a considerable place in the records. This is right and proper, because it was the writer's justification for extolling the deeds and splendour of Solomon the magnificent, Solomon the wise, Solomon the builder of the Temple. (*See* also TEMPLE, JEWISH.)

The King and His Difficulties.—In I Kings 11 the other side of the picture appears. The writer kept it to the last so that it appears to be only in Solomon's last years that things went wrong. Apparently Solomon's building of temples was not confined to that on Mt. Zion. He built at least one heathen temple on the Mount of Olives, east of and opposite to the Temple mount. He made provision for the worship of the many deities of his many wives and concubines. This is said to have taken place when he was old, but it must have been much earlier. Many untoward things happened comparatively early in his reign. After the death of Joab (11:21; this was early in the reign) the Edomite prince Hadad, who as a child escaped the ruthless massacre of the Edomites, returned and ruled in Edom. The Aramaean Rezon established himself in Damascus. These two events robbed Solomon of his command of the overland trade routes. He had to cede territory in Galilee to pay his debts to Hiram of Tyre (9:12–14).

There were troubles within the kingdom. In order to obtain supplies for his grandiose schemes, Solomon divided the country outside Judah into 12 districts, independent of the old tribal divisions. One month's supply was exacted from each district in turn. It is declared (9:20–22) that the Israelites were not treated as "bond servants" and that the forced labour needed for the building projects was exacted from the Canaanites, but it is plain that the labour levy was exacted from all Israel (*i.e.*, the northern tribes) and that all except the men of Judah were forced to work in the Lebanon for one month in three. Apparently the north was subject to the south; it was far from being an equal partnership. Jeroboam son of Nebat was in charge of "all the forced labor of the house of Joseph" (11:28). He raised a revolt against Solomon and had to seek refuge in Egypt, but he returned at Solomon's death to reign over the northern tribes, who refused any longer to suffer the exactions of a Judaean ruler.

However splendid the court may have been, the common people must have been far from happy. The king's lavish expenditure was a primary cause of the discontent which led to the breakup of the kingdom under his successor. His expenditure was far too great for his resources. He lost territory he could not afford to lose and with it considerable revenue. He oppressed the northern tribes both in labour and in taxes. Doubtless he uttered many wise sayings, enough to make him the father of wisdom, so that all subsequent wise sayings were ascribed to him. But he was not a wise ruler; he dissipated all that David had gathered, threw away all that David had gained.

Solomon as a Wise Man.—Solomon's shrewdness as judge is illustrated by the story of the two harlots, one of whom killed her baby by lying on it and took instead the live baby of the other (3:16–28). The motif of the story, whereby the king uses the power of mother love, by suggesting that the surviving baby should be divided between them, to show to whom the living baby belongs, is a common element in folk stories in many lands. It may well be that this particular story is an adaptation from folklore, especially since there is an Arabian story of the 11-year-old Solomon giving a wiser judgment than his father in a case of damage by trespass. Such a borrowing is also likely in the famous story of the queen of Sheba. In Arabic literature she is held to be Bilkis (or Maqeda), the daughter of the jinn, more beautiful than the houris of paradise. There are many legends concerning her and Solomon, including the ancient tradition that her son by Solomon was the ancestor of the present-day royal house of Ethiopia.

Jewish and Muslim writings, notably the Talmud and the Koran, contain many stories concerning Solomon. According to Jewish tradition, he had a magic carpet 60 mi. long and 60 mi. wide, and could set off at dawn, breakfast in Damascus, and sup in Media. During one journey he was reproved for his pride by the queen of the ants. Solomon had a magic ring, engraved with the name of God. It contained four jewels, given him by four angels in the shapes of a whale, an eagle, a lion, and a serpent. These four jewels gave Solomon power over various parts of nature. That given him by the eagle-angel gave him dominion over animals and birds, and the whole ring gave him power over spirits and animals, wind and water. It is also said that the seal of Solomon was a six-pointed star (shield of David; *see* MAGEN DAVID), though Western legends make it the five-pointed "pentacle," the "Druid's foot" (*Drudenfuss*) which protected Faust's threshold so that Mephistopheles could not cross it. Solomon's throne had mechanical figures of animals and birds on each step, from the ox on the first step, which held out its foreleg for the mounting king to lean upon, to the eagles on the sixth step, which raised him and placed him on his throne. Muslim tradition says that once he lost his magic ring, that the rebel angel Sakhr obtained it and ruled as king for 40 days while Solomon wandered about ragged and destitute; but Sakhr lost the ring in the sea, and Solomon recovered it from the inside of a fish that had swallowed it. Once Solomon made a pilgrimage to Mecca on his magic carpet, sheltered from the sun by a canopy of birds. The hoopoe only was missing, and it was he, to save himself punishment, who told Solomon about the beauty of Bilkis. There is a story that Solomon employed the jinn to complete the Temple; he died before they finished, but his death was concealed from them by God for a year because he died standing at prayer and his staff supported him, dead but upright, till the work was finished, and the jinn did not know.

Solomon's reputation for wisdom caused the authors of various works of wisdom literature to claim his name. Among these works are three biblical books: Proverbs, Ecclesiastes, and the Song of Songs or Song of Solomon; one book of the Apocrypha, the Wisdom of Solomon, or Book of Wisdom; and two pseudepigraphal Jewish books, the Psalms of Solomon and the Odes of Solomon.

See also JEWS: *Earliest Times to A.D. 135: David and Solomon*; KINGS, BOOKS OF THE; and references under "Solomon" in the Index.

BIBLIOGRAPHY.—For history of Solomon *see* bibliography to JEWS

and KINGS, BOOKS OF THE. For Muslim and other stories of Solomon see E. A. Wallis Budge, *The Queen of Sheba* (1922); L. Ginzberg, *The Legends of the Jews*, vol. iv (1913). (N. H. S.)

SOLOMON, ODES OF, is a collection of 42 short pseudepigraphical hymns resembling the canonical Psalms in their general nature. They are characterized by a spirit of adoration; only exceptionally do they touch some theological theme, and then the religious lyric tends to veil it, wrapping it in mystical speech. But not everything fades away from our sight: the centre of salvation is not the cross but God's self-humiliation (as in Phil. ii); virginity is a prerequisite for perfection, and the congregation is envisaged as a mystical union.

Until 1908, when a Syriac manuscript containing both the Odes and the Psalms of Solomon was discovered, they were known only from fragments: a citation of one ode (xix, 6) in Lactantius' *Divinae Institutiones* (early 4th century A.D.); and five odes (i, v, vi, xxii, xxv) in the *Pistis Sophia*, a 3rd–4th-century Gnostic work in Coptic. One ode (xi) in Greek has subsequently come to light in a 3rd-century papyrus.

The question of provenance has been much discussed. The idea that the odes stem from Gnosticism is untenable; dualism is lacking and there is nothing else in them that indicates a Gnostic background. Nor are there any reasons for postulating a Montanist origin. These texts reflect the oriental Christianity of the 2nd century and are not unorthodox. Greek has been regarded as their original language, but their content—an amalgam of covenant consciousness, baptismal imagery and Encratism—indicates that their home must be sought in Mesopotamia. There seems to be an echo of the history of Edessa, the home of Syriac literature, and the style and rhythm savour of a Syriac original, which shines through even in the Greek version. Syriac provenance also would explain some affinities with the Thanksgiving Psalms found at Qumran (*see* DEAD SEA SCROLLS).

BIBLIOGRAPHY.—*Texts: Pistis Sophia*, Coptic text ed. by C. Schmidt (1925), Eng. trans. by G. Horner (1924); R. Harris and A. Mingana, *The Odes and Psalms of Solomon*, 2 vol., with facsimile of the manuscript and Eng. trans. (1916–20); *Papyrus Bodmer XI*, ed. by M. Testuz (1959).

Literature: G. Kittel, *Die Oden Salomos überarbeitet oder einheitlich* (1914); J. de Zwaan, "The Edessene Origin," in *Quantulacumque* (1937); L. G. Rylands, *The Beginning of Gnostic Christianity* (1940); A. Vööbus, *Celibacy, a Requirement for Admission to Baptism in the Early Syrian Church* (1951), and "Neues Licht zur Frage der originalen Sprache der Oden Salomos," in *Le Muséon* (1962). (AR. VO.)

SOLOMON, PSALMS OF, a collection of 18 pseudepigraphical psalms extant in Greek and Syriac versions. The ascription to Solomon, which probably goes back to the editor who first put them into their present form, makes them analogous to the canonical psalter ascribed to David. The poet (or poets) who composed the Psalms of Solomon seems not to have thought of ascribing them to Solomon, for they contain not even an allusive reference to him. It makes little difference whether more than one author is concerned, for the psalms all come from the same period and milieu and share the same religious and political views. The most that can be said is that the three great political psalms (ii, viii and xvii) were written by one and the same author, a man of poetic vigour and imagination.

The Hebrew original of the Psalms of Solomon is not extant, but the Syriac translation as well as the Greek seems to have been made directly from it. This is indicated by a number of passages in which the Greek is incomprehensible because the Hebrew it translated must have been corrupt, whereas the Syriac reveals the correct reading in the Hebrew. (The reverse situation also occurs.) Only occasionally did the Syriac translator correct himself by reference to the Greek version.

The Psalms of Solomon were clearly composed in Palestine, and some certainly originated in Jerusalem. This accounts for the important role which Jerusalem plays in them. The three political psalms can be dated fairly exactly because they make a clear reference to Pompey the Great. Thus most of xvii was composed shortly before Pompey entered Jerusalem in 63 B.C., though verses 11–14 are a later addition which refer back to this event and date from after 61 B.C. In ii, 26 ff. there is a reference to the death of Pompey in 48 B.C. As all the psalms belong to the same period, the collection can be dated to between 80 and 40 B.C.

The importance of the Psalms of Solomon lies chiefly in their vivid and contemporary presentation of the political and religious thought of the "pious of Israel" (*i.e.*, the Pharisees) in the last century B.C. The author of the psalms was certainly a Pharisee. Psalm xvii is particularly interesting for its detailed description of the eschatological messianic king and the political and religious expectations of his reign. This is the conception of the Messiah still current among the Jews during Jesus' lifetime; as the Gospels show, it lies behind the reaction of the common people to Jesus' teaching, the opposition of the Pharisees, high priest and scribes, and the hopes and expectations of the disciples.

BIBLIOGRAPHY.—Greek version ed. by O. von Gebhardt in *Texte und Untersuchungen*, vol. xiii, part 2 (1895), by H. E. Ryle and M. R. James with Eng. trans. (1891), by J. Viteau with French trans. (1911). Syriac version ed. by R. Harris and A. Mingana with Eng. trans., 2 vol. (1916–20). *See* also E. Schürer, *Geschichte des jüdischen Volkes im Zeitalter Jesu Christi*, 4th ed., vol. iii, pp. 205–212 (1909); O. Eissfeldt, *Einleitung in das Alten Testament*, 2nd ed., pp. 754–758 (1956); K. G. Kuhn, *Die älteste Textgestalt der Psalmen Salomos . . . mit einer Bearbeitung und Übersetzung der Psalmen Salomos 13–17* (1937). For the Syriac version as a translation of the Greek *see* J. Begrich, "Der Text der Psalmen Salomos," *Zeitschrift für die Neutestamentliche Wissenschaft* 38:131–164 (1939). (K. G. K.)

SOLOMON, SONG OF (or SONG OF SONGS; in the Douai version of the Bible, CANTICLE OF CANTICLES, from *Canticum Canticorum* in the Vulgate; derived from the title in the Hebrew Bible, SHIR HASHIRIM), a book of the Old Testament, one of the Megilloth, or "Rolls," the others being Ruth, Lamentations, Ecclesiastes and Esther. The Megilloth are read in the synagogue at certain festivals, the Song of Solomon being used at Passover, which celebrates the Exodus from Egypt. The reasons for this are not entirely clear, but the fact that spring is referred to in ii, 11 ff. and vi, 11 (Passover being also celebrated in spring), and the allegorical interpretation of the book as referring to God's love for Israel and the events of the Exodus, especially the Covenant at Mt. Sinai, should be considered in this connection.

Contents.—There is no coherent story in the book. It consists of a series of love poems in which a man and a woman speak alternately. A number of songs describe systematically the beauty and excellence of the beloved (called by a modern Arabic term *wasf*, "description"): iv, 1–7; v, 10–16; vi, 4–7; vii, 1–6. There is reference to the bride's seeking for her beloved: i, 7 ff.; iii, 1–5; v, 6–8. There is a description of a festive procession: iii, 6–11. And there are dialogues between the two partners: *e.g.*, i, 9–ii, 7, etc.

Author.—The Hebrew title of the book mentions King Solomon as its author, but several considerations make this improbable: Solomon is referred to in the third person; the language of the book seems to reflect a much later epoch (*e.g.*, there are Persian and perhaps even Greek loanwords); and the city of Tirzah is referred to (vi, 4) in such a context as to make it probable that it is the capital of the northern kingdom of Israel (9th century B.C.). This, alongside the fact that the language contains some archaic elements, seems to suggest that an older text has been revised at a later epoch. The author of the book remains unknown.

Interpretation.—The Song of Solomon has been interpreted in many different ways of which four, described below, are important.

Allegorical.—The allegorical interpretation takes the book as an allegory of God's love for Israel or of Christ's love for the church. This type is represented by the Jewish Targum (the free Aramaic rendering given in the synagogue when Hebrew was no longer understood), which makes it refer to the lawgiving on Mt. Sinai, the history of Israel and the coming of the Messiah. In the Christian Church this interpretation is maintained by the early Fathers, by the Eastern Churches and by many modern Roman Catholic scholars. In medieval mysticism the Song of Songs was applied to the love between Christ and the human soul, and Bernard of Clairvaux gave a famous exposition of it to that effect. The earliest traces of an allegorical interpretation seem to be Rev. iii, 20 (which may reflect S. of Sol. v, 2) and II Esd. v, 23–26 (1st century A.D.).

Dramatic.—The dramatic interpretation is based on the fact that a considerable part of the book is clearly in dialogue form. According to this view, the heroine of the book is a peasant maiden in Solomon's harem, who longs for her shepherd lover and utters her feelings to the ladies of the court; they lead her on to describe her lover and to tell how she came to be carried off by Solomon. Finally her constancy secures her release, and the curtain falls on the sentiment of true affection in viii, 6, 7. This theory is attractive in certain respects, but in view of the absence of drama in the Semitic literatures it is not very probable. Moreover, any progress of action from first to last is difficult to prove. Most advocates of the theory find it necessary to rearrange the sequence of the sections considerably in order to get a coherent story, which seriously weakens their case.

Literal.—The so-called literal interpretation considers the book to be a collection of secular love songs without any religious implications. This theory was introduced in modern research by J. G. Wetzstein, who during the 19th century was Prussian consul at Damascus and whose observations of the wedding customs of Syrian peasants led him to believe that the biblical book is substantially a collection of songs originally sung at such festivities. During the first seven days after the wedding the man and his young wife play the part of king and queen. A threshing board is erected on the threshing floor as a throne, and after the royal couple have taken their seats a dance in their honour begins. The accompanying song to a great extent consists of *wasf, i.e.*, a description of the physical perfections and the ornaments of both. Wetzstein's contribution, which was originally published in the *Zeitschrift für Ethnologie* in 1873, was reprinted by F. J. Delitzsch in an appendix to his *Commentary*, and his theory was accepted by K. Budde (1898) and by many other scholars.

The songs edited by Wetzstein, and folk songs from Palestine later published, indeed show a striking similarity in language and imagery with the Song of Songs—*e.g.*, apart from the *wasf*, the description of the bride as a garden (iv, 12; vi, 2, 3) and many other details. This theory would also account for the reference to Solomon in the Song of Songs: it would be another way of saying that the bridegroom is playing the role of a king. However, the very fact that the young couple in Syria are celebrated as king and queen calls for an explanation.

Another difficulty arises from the fact that almost throughout the Song of Solomon it is the bride and not the bridegroom who takes the initiative, which would be inconceivable in an ordinary Syrian or Palestinian wedding. This has given rise to a fourth theory.

Cultic-Mythological.—The cultic-mythological interpretation is based on the fact that the poems of the Song of Songs are strongly reminiscent of the songs connected with the Sumerian rite of the sacred marriage. This ceremony, an element of the fertility cult of ancient Mesopotamia, most frequently implies the consummation of a ritual marriage between the king and a goddess, represented perhaps by a temple prostitute or priestess, the king playing the role of the fertility god Tammuz. Description songs (*wasf*), the goddess's search for the god and several other typical motifs and expressions occur here for the first time in history, much earlier than the Song of Songs and the modern Arabic folk songs.

The cultic interpretation was inaugurated by Erbt in 1906 and further elaborated by a number of scholars. It implies that rites similar to the Sumerian ones were practised in ancient Canaan—this is to some extent substantiated by the Ras Shamra texts, dating from the 14th century B.C.—and that these practices were in some way or other taken over by Israel. It is known at least that this was the case during the reign of King Manasseh (II Kings xxi). But it is difficult to explain how a collection of more or less pagan cult songs was introduced into the Hebrew canon of Holy Scriptures. One suggested explanation assumes that these songs were used for centuries in popular celebrations in springtime until their original meaning was forgotten, and then were adopted because of the allegorical interpretation which was given to them partly because their performance happened to coincide with the celebration of the Exodus at Passover.

This theory has the advantage over all the others: it accounts for the dialogue form; it is "literal" in that it understands the text as it is without looking for hidden implications; and it even provides a certain basis for an allegorical interpretation, if it may be assumed that the sacred marriage forms the background of the prophetic description of the relationship between Yahweh and Israel as a marriage (Jer. ii, 2; Isa. liv, 4 ff.; lxii, 4 ff.).

The cultic interpretation is accepted by an increasing number of writers, especially since the discovery and publication of many new parallels in Sumerian texts. The main reason for rejecting it is the difficulty of explaining the book's presence in the canon on these premises. The only alternative is the interpretation as secular love songs, the similarity with Sumerian cult poems being accounted for by the assumption that they are dependent on a common tradition of love songs. Some Egyptian collections of love songs, which also offer some parallels to the Song of Solomon, seem to be of a more secular character, although many of them are thought to have originated in the cult of the love goddess Hathor.

BIBLIOGRAPHY.—A good survey of the different interpretations is given by H. H. Rowley in his collection of essays *The Servant of the Lord* (1952). For commentaries, *see* T. J. Meek in *The Interpreter's Bible* (1956); H. Ringgren in *Das Alte Testament deutsch* (1958). For folk songs, *see* St. H. Stephan, *Modern Palestinian Parallels to the Song of Songs* (1923; also in *Journal of the Palestine-Oriental Society*, vol. ii, 1923); S. Linder and H. Ringgren, *Palästinische Volksgesänge*, 1–2 (1952–55). (H. Ri.)

SOLOMON ISLANDS, a chain of islands in the Melanesian area of the southwestern Pacific, which extends in a northwest-southeast direction for almost 900 mi. between latitudes 5° and 11° S., and longitudes 154° 40′ and 162° 30′ E. The group includes the Santa Cruz Islands, about 300 mi. to the east of the southernmost Solomons, and the Lord Howe (Ontong Java) Islands.

The northernmost islands, Buka and Bougainville, are administered by Australia as part of the trust territory of New Guinea (*q.v.*); the remaining islands—including Guadalcanal, San Cristóbal, Malaita, Santa Isabel, Choiseul, New Georgia, Santa Cruz, Lord Howe, Reef and Duff groups and Mitre Island—form a British protectorate. The area of the Australian-administered islands is about 4,100 sq.mi. and that of the British protectorate about 11,500 sq.mi.

The group contains a number of low coral atolls, but the larger islands are all volcanic and rugged. Heights of more than 8,000 ft. are reached on Bougainville and Guadalcanal. Geologically the islands form a continuation of the volcanic arc to the east of New Ireland.

Bougainville (*q.v.*) is the largest of the group. South of Bougainville the islands are arranged in two parallel chains, the western including Vella Lavella, Kolombangara, New Georgia and Guadalcanal, and the eastern, Choiseul Island, Santa Isabel Island and Malaita. The two chains converge on the southernmost island, San Cristóbal.

The climate is equatorial, with constant heat, high humidity and no dry season. Rainfall on most of the islands exceeds 100 in. a year. The vegetation is a heavy forest and, as in most island groups, the animal life is restricted.

The islands were discovered in 1568 by the Spaniard Alvaro de Mendaña, who tried to establish a colony in 1595. The expedition, which failed miserably, is the subject of a novel by Robert Graves (see *Bibliography*). Only toward the end of the 18th century was European knowledge of the group completed, by the voyages (1766–92) of Philip Carteret, Count Louis Antoine de Bougainville, John Shortland, Antoine Raymond Joseph de Bruni d'Entrecasteaux and others. The British protectorate was established over the southern islands in 1893, after a century of European contact with generally unfortunate results.

A German protectorate over the northern Solomons was declared about 1885, but in 1899 all except Buka and Bougainville were transferred to Great Britain in return for recognition of German claims in western Samoa. The German islands were occupied by Australian troops in 1914, and after 1920 the two northern islands were administered by Australia. In 1942 the Japanese advance

into the Pacific swept the Solomons where its further progress was checked. The fighting in the following three years was among the bitterest in the Pacific, particularly the long struggle on Guadalcanal (*see* WORLD WAR II: *The War in the Pacific*).

The total population of the Solomons is about 154,000, of whom 124,076 (1959 census) live in the British protectorate. The most populous island is Malaita, with (1957 est.) 47,000 inhabitants. The larger islands are almost entirely populated by Melanesians, but there has been much racial admixture and there is a great variety of physical types and linguistic groups. Ten separate languages are spoken on Bougainville. A few Polynesians are found on the atolls.

There are only two towns, Kieta on the east coast of Bougainville, and Honiara, administrative capital of the protectorate, on the north coast of Guadalcanal.

In 1960 a new constitution was brought into operation for the protectorate. The legislative council, presided over by the high commissioner for the western Pacific, has an official majority. The executive council is divided equally between official and unofficial members.

Primary education is mostly provided at subsidized mission schools, and there are a secondary boarding school and a training college for teachers and carpenters. There is a central hospital at Honiara, besides a leprosarium at Tetere, and district, rural and mission hospitals.

Copra is the mainstay of the economy of the Solomons, and is produced by both native growers and European estates, the latter mainly on Guadalcanal, the Russell Islands, San Cristóbal Island and Santa Isabel Island. Great damage was done to the coconut plantations during World War II, and even by 1960 only about two-thirds of the prewar area of 64,000 ac. were productive. No new estate plantings of coconuts had taken place, although there had been a small increase in native copra production from Santa Isabel Island and Malaita. There had, however, been a marked improvement in the quality of copra produced, following the granting of a premium for the better grades by the copra board, which is the sole purchaser of copra in the islands.

The administration had endeavoured to introduce cacao as an alternative cash crop; it was taken up by some native growers on Malaita, and a little is produced by Europeans in the western Solomons.

Only two other commercial activities have more than a local importance: the Kieta area of Bougainville produces a little gold, and the islands of Guadalcanal and Vanikoro support a lumbering industry that makes a small contribution to the export trade. About four-fifths of the exports by value, however, are accounted for by copra.

See PACIFIC ISLANDS; *see* also references under "Solomon Islands" in the Index.

BIBLIOGRAPHY.—Colonial Office, *Reports on the British Solomon Islands* (occasionally); Commonwealth of Australia, *Reports on the Trust Territory of New Guinea* (occasionally); R. Graves, *The Islands of Unwisdom* (1949); S. E. Morison, *U.S. Naval Operations in World War II* (1950). (D. W. F.)

SOLOMON'S-SEAL, a perennial herb of the genus *Polygonatum,* of the lily family (Liliaceae). There are about 30 species native to north temperate regions. The plant springs from a fleshy, creeping rootstock on which the scars of the previous year's shoots leave curious seallike marks—hence the common name.

Two species occur in eastern North America and three are found in Great Britain.

The single erect or arching stem, one to eight feet high, bears ovate or lance-shaped, sharp-pointed, sessile leaves in the axils of which appear narrow greenish, whitish or pinkish bell-shaped flowers, on slender, drooping stalks, followed by globular, pulpy, usually bluish berries.

In the United States various species of *Smilacina,* a closely allied genus, are called false Solomon's-seal.

SOLOMOS, DIONYSIOS, COUNT (1798–1857), was the first poet of modern Greece to show the capabilities of its spoken language (the *dimotiki*) when inspired by wide culture and a first-rate lyrical fancy. Born in Zante, he was educated in Italy (1808–18), at Venice, Cremona and Padua, after which he returned to Zante. His earliest poems were written in Italian but in 1822 he finally determined to write in the spoken tongue of Greece. His noble *Hymn to Liberty,* in 158 four-line stanzas, was composed in May 1823, and his poem on the death of Lord Byron in 1824–25. The unfinished *Lambros,* a romantic poem of the revolutionary times written in ottava rima (*q.v.*), was begun in 1826. To this period (1823–28) belong also some shorter lyrical pieces and some satires, of which the most notable is *I Gynaika tis Zakynthos* ("The Woman of Zante"), a long sketch written mostly in poetical prose. In 1828 he migrated to Corfu, where he remained until his death. His lyrical exuberance was curbed by a growing preoccupation with German theories of dramatic form and by an inhibiting dissatisfaction with the as yet meagre resources of his chosen linguistic medium. These impediments, together with a disastrous family quarrel (1833–38), explain why his major poems of this period remain fragmentary. Nonetheless, *O Kritikos* ("The Cretan," 1833), the second and third sketches of *Oi Eleftheroi Poliorkimenoi* ("The Free Besieged") of Missolonghi (before and after 1844), and *O Porphyras* ("The Shark," 1849) exhibit, even in their fragments, a sense of rhythm, a "curious felicity" and a melody of cadence not found in his *juvenilia.*

BIBLIOGRAPHY.—Dionysios Solomos' collected works were ed. by L. Politis, 2 vol. (1948–55). *See* also L. Politis, *O Solomos sta grammata tou* (1956). For biographical sketches *see* R. Jenkins, *Dionysius Solomós* (1940); P. Michalopoulos, *Dionysios Solomos* (in Greek) (1931); M. Sigouros, *O Solomos* (in Greek) (1958). (R. J. H. J.)

SOLON (active about the beginning of the 6th century B.C.), an Athenian statesman and the earliest Greek lawmaker about whose achievements anything is known in detail. He was born about 640, son of Execestides, a nobleman or eupatrid (*see* EUPATRIDAE).

Political Background.—Solon grew up in a period of bitter political strife, the symptoms of which were the attempt of Cylon, supported by his father-in-law, Theagenes, tyrant of Megara, to become tyrant of Athens (in 636 or 632), its savage suppression by the Alcmaeonid Megacles, involving (so his opponents averred) an act of gross impiety, the breach in the eupatrids' monopoly of knowledge of the law by the publication of Draco's code *c.* 621 and the vengeance, backed by Delphi, taken on the Alcmaeonids and their followers for impiety, probably toward the end of the century (*see* ALCMAEONIDAE). This political strife arose from the interplay of two factors, rivalry among a number of powerful aristocratic families and economic stress among the common people. Under this stress many had fallen into debt-slavery and many had lost effective ownership of their ancestral plots of land. There seem to have been some aristocrats, among them Solon, who were prepared to make concessions to the oppressed lower class partly by attacking the more or less complete monopoly exercised by the eupatrids over the control of the state, partly by attempting to relieve the economic burdens of the poor. Draco's publication of the law was a step in the former direction. A bolder step toward solving both sides of the problem was taken when Solon was appointed archon with legislative powers in 594/593 and was given the special task of reconciling the contending factions. This role of standing as a kind of umpire between two extreme parties rests on the unimpeachable evidence of Solon's own poems, as does the fact that, like all moderate reformers, he was criticized by both right and left, by the right more severely. The detail of his reforms, for which *see* below, rests on less secure evidence.

Solon's Participation in Athenian Foreign Policy.—Solon seems to have played a part in external affairs also during this period. His name is connected with two of the three main enterprises which engaged Athenian arms at this time. In a prolonged struggle with her neighbour, Megara, the Athenians wrested from her the possession of the island of Salamis off the coast of Attica. According to Plutarch in his *Life of Solon* it was the esteem Solon won in this war that led to his appointment as archon and, if that is right, the campaign must be dated *c.* 600–595. Later, according to some accounts, Salamis was lost again and was recovered by Pisistratus, acting with the support of Solon, *c.* 565. Some historians, both ancient and modern, have held that only this second war, under the generalship of Pisistratus, actually occurred. Solon

himself refers in his poems to the war for Salamis, but says nothing to fix the date decisively.

At some time in the 590s Thessaly, Sicyon and Athens combined successfully against Cirrha, in what was known as the First Sacred War, for the control of Delphi and its oracle. Solon was said to have advocated Athenian participation in this enterprise, but the general in command was Alcmaeon.

It is fairly certain that Athens engaged in a war for Sigeum in the Troad at the end of the 7th century. Solon's name is not connected with this war. Nevertheless it is highly probable that the war was to some extent concerned with trade through the Hellespont and this would fit with the view that Solon saw an extension of foreign trade as a partial solution to Athens' economic difficulties.

Absence from Athens.—Tradition said that after his year of reform Solon voluntarily absented himself from Athens for a long period of travel, partly for its own sake, partly for trade. He is supposed to have visited Egypt, the court of Croesus in Lydia, and Cyprus, where he stayed with the prince Philocyprus and gave his name to the city called Soli. The poems contain a hint of a visit to Egypt and a specific mention of a visit to Soli in Cyprus, though neither reference gives any clue as to the date of the visit. The visit to Croesus (*q.v.*), though the most famous, and vouched for and described in detail by Herodotus, must be rejected as mythical on chronological grounds.

Tyranny of Pisistratus.—The last recorded chapter in Solon's life was his impassioned resistance to the setting up of Pisistratus as tyrant of Athens *c.* 561. Dramatic colour was added to the story by the alleged facts that the mothers of the two men were related and that Pisistratus had once been Solon's lover. The poems voice some sharp warnings against the dangers of tyranny, but there is no certainty that these fragments are to be dated to the end of Solon's life. He cannot have survived long after the setting up of the tyranny, if indeed he lived so long.

The story of Solon's life was much obscured in antiquity by his inclusion in the list of the Seven Sages with their many apocryphal encounters.

Evidence of Reforms.—For the details of Solon's reforms we depend largely on two main sources, Aristotle's *Constitution of Athens* (*q.v.*) and Plutarch's *Life,* and they in turn must have drawn heavily on the writings of the Atthidographers, local historians of Attica who wrote between the end of the 5th and the middle of the 2nd century B.C. Modern scholars have differed widely as to the reliance to be placed on this evidence, one school holding that virtually nothing more was known to these writers than is available to us today and that most of what was attributed to Solon was the result of arguments, often faulty, from survivals, while to others it has seemed probable that these writers had access to much that has now been lost, whether it was oral tradition or, more important, the remains of the original wooden tablets (*axones* or *kurbeis, see* below) on which Solon published his laws, and that it is therefore possible to recover in some detail the shape of his reforms. The latter view has been adopted in the following account, which treats the reforms under two main heads (1) social and economic and (2) constitutional.

Social and Economic Reforms.—Solon's boldest step, and one which he himself puts in the forefront of his reforms, was the so-called "shaking-off of burdens" (*seisachtheia*). Much of the land of Attica was occupied by smallholders who had fallen into debt and had to pay their creditors a sixth part of the yearly produce (hence they were called "sixth parters" or *hectemoroi*). The fact that a man's land was thus encumbered was marked by the setting up of a mortgage pillar (*horos*) upon it. By the harsh law of debt anyone who defaulted on these payments could be enslaved. Solon describes how he freed the earth by plucking up these pillars (canceling, we must suppose, the debts to which they witnessed) and released those who had been enslaved, even buying back those who had been sold abroad. He prohibited enslavement for debt in the future, thus putting Athens in the van of progress in this respect—even at the end of the 5th century there were several Greek states where enslavement for debt still prevailed. He tried to ensure that debtor slaves actually secured their freedom, and that in future children, for example, were not illegally pledged into slavery, by providing that in certain cases anyone who wished (*ho boulomenos*) could proceed judicially against a wrongdoer and not, as heretofore, only the wronged party. To help those who still could not hope to secure a living from the land he encouraged the teaching of crafts, offered citizenship to craftsmen who wished to settle permanently in Attica and banned the export of all produce of the land with the significant exception of olive oil. He was also alleged to have altered the coinage to fit the Euboic-Corinthian instead of the Aeginetan standard so as to facilitate Athenian trade with the west. It is doubtful however whether Athenian coinage goes back as far as Solon and it may be that all he did here was to alter the system of weights and measures from the Aeginetan to the Euboic-Corinthian. In the early days of coinage the effect of this would have been almost equally important. All this suggests that Solon had an idea, however rudimentary, that where a population can no longer be fed on home produce the necessary imports of food can be paid for only by exports. Other laws aimed at strengthening the individual family and the state on the one hand vis-à-vis the powerful clan on the other. Thus he allowed a man who had no male issue to adopt a son and leave his property to him by will; hitherto such property had devolved on the kin in a prescribed order. He made sumptuary restrictions on funerals and instituted or confirmed a state cult of the dead (*genesia*) peculiar to Athens, over against the clan cults. Finally several rules analogous to Roman praedial servitudes are attributed to Solon; to take a single example, a law prescribed the distance various kinds of tree must be planted from the neighbour's boundary.

Constitutional Reforms.—Here Solon's principal step was to break the eupatrid monopoly of the chief magistracies, especially the archonships. To this end he used a classification of the free population by the produce of their land. There were four classes, *pentacosiomedimnoi* whose land produced more than 500 *medimni* of grain (1 *medimnus* = $1\frac{1}{2}$ bushel) or 500 *metretae* of wine or oil (1 *metretes* = $8\frac{1}{2}$ gal.), *hippeis* (horsemen) with more than 300 *medimni, zeugitai* (probably hoplites or infantrymen) with more than 200 and *thetes* (day labourers) all the rest. Anyone who belonged to the highest class (or perhaps to one of the two highest classes) might now become an *archon* (*q.v.*) whether he was an eupatrid or not. Minor offices were open to members of the third class. The *thetes* could hold no office, but they were assured membership of the popular assembly (*ecclesia; q.v.*). This assembly probably gained in power, though less no doubt than the oppressed class would have liked, and less than was later asserted by those who tried to push back all important democratic institutions to Solon. It now played a part in the election of the archons, who had hitherto probably been chosen by the aristocratic council of the Areopagus (*q.v.*). Each of the four tribes of which it was composed chose 10 men and out of these 40 the archons were selected by lot. This was less democratic than would have been the right to elect the archons direct. Solon allowed appeals in certain cases (it is not possible to say precisely which) from the judicial decisions of magistrates to the assembly, called when exercising this function *heliaia*. Aristotle, in his *Constitution of Athens*, followed by Plutarch, tells us that Solon instituted a council (*boule; q.v.*) of 400 whose function would have been to prepare business for the assembly. We can say nothing of the composition or method of election of this council except that it would not have included *thetes* and that it would have been open to noneupatrids. This Solonian council of 400 has been doubted, perhaps without sufficient ground. If accepted, it implies that Solon contemplated fairly regular meetings of the assembly and that he was not prepared either to allow these meetings a completely free rein or to entrust control of them to the Areopagus, which, being composed of former archons, was bound for many years to be dominated by eupatrids. Finally Solon attempted to bring home to each citizen his responsibility for the stability of the state by the ordinance that, when civil strife broke out, no one should remain neutral. Some such constitutional provisions as these would square well with Solon's own claim that he had "stood casting a strong shield round both parties and had allowed neither to triumph unjustly"

and with the stress that he lays in several passages on the need for a reform in the administration of justice.

Solon's laws, not of course neatly docketed into social, economic and constitutional but probably arranged under the heads of the various magistrates to whose spheres of duty they severally related, were published on wooden prisms (*axones* or *kurbeis*) which could be revolved to show their three faces. A substantial proportion of these archaic objects probably survived at least to the end of the 5th century B.C. and remnants of them down to Plutarch's day in the 1st century A.D. There were at least 16 of them since we have a quotation from the 16th. The law of homicide remained substantially as Draco had published it, but Solon made drastic changes in the rest of Draco's code. Thus 4th-century writers and speakers tended to regard the law of Athens as being basically Draco's law of homicide together with the laws of Solon, though they were well aware that in the course of two centuries many amendments and additions had been made to them.

In neither the economic nor the constitutional sphere was Solon wholly successful. Thus, though we hear nothing further of the *hectemoroi*, a discontented class of smallholders and city dwellers remained to form part of the backing for Pisistratus' bid for tyranny. And, although he had made noneupatrids eligible for office, he could not ensure that they actually be elected: struggles over elections to the archonships remained bitter and, after an illegal usurpation by one Damasias in 582–580, a compromise had to be adopted—we do not know for how long it lasted—by which the office was shared roughly equally between eupatrids and noneupatrids.

Solon was also the earliest Athenian poet, using verse in the elegiac, iambic and trochaic metres as a vehicle both for his political and social programs and for his views on morals and religion. His poems express, with a measured passion, respect for the individual and loyalty to the state as reciprocally enhancing aims in the life of the community. A much more skeptical view of the evidence for Solon's reforms is taken by, for example, C. Hignett (*see* bibliography), who also holds that the legislation should not be placed in Solon's archonship but was passed toward the year 570. *See* also references under "Solon" in the Index.

BIBLIOGRAPHY.—G. Busolt and H. Swoboda, *Griechische Staatskunde* (1920–26); K. Freeman, *Work and Life of Solon* (1926); W. J. Woodhouse, *Solon the Liberator* (1938); C. Hignett, *A History of the Athenian Constitution* (1952); A. Andrewes, *The Greek Tyrants* (1956); V. L. Ehrenberg, *The Greek State* (1960); C. M. Bowra, *Early Greek Elegists* (1938). Fragments of Solon's poems in E. Diehl, *Anthologia Lyrica Graeca* (1949) in the Teubner series and with Eng. trans. by J. M. Edmonds, *Elegy and Iambus* (1951) in the "Loeb Series." (A. R. W. H.)

SOLOR and **ALOR ISLANDS,** two closely associated island groups of the Nusa Tenggara Islands, Indonesia, stretching 150 mi. (240 km.) eastward from Flores to Timor. They consist of five large islands, Solor (pop. [1930] 35,000), Adonara (25,000), Lomblen (32,000), Pantar (8,000), and Alor (51,000, the largest), together with a number of smaller ones. The population in the 1960s was not much higher, because of the slow natural growth and emigration.

Alor, divided from Timor by Ombai Strait, 17 mi. (27 km.) wide, is very mountainous (Mt. Kolana 6,033 ft. [1,839 m.] and Mt. Muna 4,724 ft. [1,440 m.], both old volcanoes), and is much broken up by steep ravines, with only one plateau and some small coastal plains. It is largely covered with low trees and alang grass. The coast is mostly rocky, with few indentations, but on the west coast, Kalabahi Bay, 10 mi. (16 km.) long and nearly a mile wide, divides a northwestern promontory from the rest of the island, with which it is connected by a very low alluvial isthmus about 2½ mi. wide. Pantar is high (Mt. Delaki 4,478 ft. [1,365 m.]), with a rugged coast; Lomblen has several active volcanoes, Ili Labalekang (5,394 ft. [1,644 m.]) and Ili Lewotolo (4,757 ft. [1,450 m.]); Adonara has a volcano, Ili Boleng (5,443 ft. [1,659 m.]), and a plateau, ringed with hills, reaching 3,000 ft. (914 m.); Solor is hilly also. The islands have a pronounced seasonal rainfall, October to April being the wet season. All the islands are administratively attached to the districts of the Nusa Tenggara Province (Lesser Sundas).

Racially the people are Oceanic Negroids, short statured with dark brown to black skin and wavy to kinky hair. Many languages are spoken, all belonging to the Malayo-Polynesian family. Apart from the Muslims, who are often immigrants, and some Christian communities along the coasts, the local inhabitants are animists. The Dutch introduced a political hierarchy, giving local rule to the coastal rajas, but these rajas have only nominal control over the interior tribes. Social organization there centres around kinship and village groups. The villages are small, usually on defensive mountain sites, and consist of several communal houses, each occupied by a lineage group. House styles differ, but the rectangular stilt house with high pyramidal roof predominates. The high points of village life are weddings, burial feasts, and housebuilding ceremonies, each one accompanied by much dancing. Each occasion also involves exchange of the local currency, consisting of pigs, gongs, and *mokos* (the highly prized metal kettle drums). The women cultivate the fields, on which dry rice, maize (corn), beans, and various tubers are grown. Pigs and chickens are generally raised, but carabao and horses are few, mostly raised on the western islands, and often owned by the rajas. Lomblen and Adonara export copra, and Lomblen also has a fishing industry.

The Portuguese, who had settlements on the Islands in the 16th century, were dislodged by the Dutch in the 17th and 18th centuries, but the Portuguese claims were not given up until a treaty between the Netherlands and Portugal was ratified in 1859. In 1949 the islands became part of Indonesia.

See C. A. Du Bois, *People of Alor* (1944). (J. O. M. B.)

SOLOTHURN (French SOLEURE), an ancient little town of Switzerland and capital of the canton of Solothurn, lies on both banks of the Aare River, 25 mi. (40 km.) S of Basel and 18 mi. (29 km.) NNE of Bern. Pop. (1960) 18,394, German-speaking, with a small Protestant majority. The chief building is the 18th-century Cathedral of SS. Ursus and Victor, which stands on the site of a far older edifice. Since 1828 it has been the cathedral church of the bishop of Basel. Other notable buildings are the Jesuit Church, the Clock Tower (13th century), and the 15th-century town hall. The early 17th-century arsenal contains the finest collection of armour and old weapons in Switzerland, and the modern museum houses a collection of fossils from the Jura Mountains and a Madonna by Hans Holbein the Younger. There are some fine 16th-century fountains in the town, which in its older parts with the massive bulwarks still keeps much of its medieval aspect. In the modern suburbs and neighbouring villages there is some industrial activity, which includes watchmaking, engineering, and the manufacture of electric motors, electrical and technical equipment, and screws. The position of Solothurn at the foot of the Jura and close to the navigable portion of the Aare has always made it a meeting point of various routes; seven railway lines branch off there. A mile north of the town is the Hermitage of St. Verena, in a striking rock gorge, above which rises the Weissenstein ridge.

A 16th-century rhyme claims for the town the fame of being the oldest place in Celtis save Trier. Certainly its name, Salodurum, is found in Roman inscriptions, and the remains of the Roman castrum still exist. Its position as commanding the approach to the Rhine from the southwest has led to its being more than once strongly fortified. The medieval town grew round the castrum and the house of secular canons, founded in the 10th century in honour of St. Ursus and St. Victor by Queen Bertha, wife of Rudolph II, king of Burgundy. The provost and canons had many rights over the town, but in 1218, having shaken off the jurisdiction of the canons, who were then taken under its protection, Solothurn became a free imperial city. In 1295 it allied itself with Bern, through which it was drawn into association with the Swiss Confederation. An attempt to surprise it in 1382, made by the Habsburgs, was foiled, and resulted in Solothurn's sharing in the Sempach War. It was included in the Sempach ordinance of 1393 and in the great treaty of 1394 by which the Habsburgs renounced their claims to all territories within the Confederation. In 1411 Solothurn sought in vain to be admitted into the Confederation, a privilege only granted to it in 1481 at the diet of Stans.

In the 15th century also, by purchase or conquest, the town acquired the main part of the territories forming the present canton. In 1529 the majority of the communes went over to the reformed faith and assisted Huldreich Zwingli at the Battle of Kappel (1531), but in 1533 the old faith regained its sway, and in 1586 Solothurn was a member of the Golden, or Borromean, League. Solothurn was the residence of the French ambassador from 1530 to 1792. From the 16th century it had an aristocratic form of government, but this was finally broken down in 1831, and in 1832 Solothurn joined the League to guarantee the maintenance of the new cantonal constitutions. Though distinctly a Roman Catholic centre, it did not join the Sonderbund, and it voted in favour of the federal constitutions of 1848 and 1874. (*See* also SWITZERLAND: *History*.)

SOLOTHURN CANTON has an area of 305.5 sq.mi. (791 sq.km.), of which about 97% of the land is productive. Pop. (1960) 200,816, almost entirely German-speaking, with three-fifths Roman Catholics and two-fifths Protestants. The canton has an irregular shape because it consists of territories acquired at different dates by its capital town, from which it took its name. It includes the foothills of the Jura Mountains (*q.v.*) and consists of a plain along the Aare Valley projecting into Bern Canton, with the Bucheggberg in the southwest between the Aare and Emme (a right-bank tributary which joins the Aare near the town of Solothurn). Farther downstream, a part of the Aare Valley belongs to Bern, but Solothurn resumes possession farther on, and there the Aare receives the Wigger on the right bank at Aarburg and the Dünnern from the Jura on the left bank at Olten. The Hasenmatt (4,751 ft.; 1,448 m.) is the highest point of the Weissenstein section of the Jura, across which the canton stretches in a broad belt to the north where, at various points, it shares with Bern the Birs River, a tributary which joins the Rhine near Basel. Beyond the Birs are two separated districts that belong to the canton: Leimental, including the well-known pilgrimage centre Mariastein, and Kleinlützel. The northern part of the canton and its detached fragments touch the canton of Bern and the half cantons of Basel-Stadt and Basel-Land. The districts of the canton form part of the diocese of Basel, with the town of Solothurn as the site of the bishop's palace. Besides Solothurn, the other chief towns are Olten (20,044) and Grenchen (18,000), both on the Aare River.

Until the mid-19th century the cantonal activities were mainly agricultural and pastoral, and although these are still important, the population is largely engaged in the manufacture of a variety of products, particularly around Solothurn, Olten, and Grenchen. The chief manufactures are watches, jewelry, shoes, cotton goods, paper, cellulose, cement, auto parts, iron and steel products, and electrical equipment for communications. The canton has excellent rail and road connections. Olten is a railway junction where the direct lines from St. Gotthard Pass via Lucerne, and from Geneva, Zürich, and Basel all unite.

Solothurn Canton is divided into ten administrative districts with 132 communes. The cantonal constitution dates from 1887, but was substantially revised in 1895. The *Kantonsrat* (legislative assembly, consisting of 144 deputies) is elected on the principle of proportional representation. The *Regierungsrat* (executive) consists of five members. Both groups hold office for four years, but any 4,000 citizens can demand a popular vote on *Abberufung*, or recall, to decide whether the existing members shall continue or not. Also by the obligatory referendum (since 1869 all laws and financial resolutions passed by the *Kantonsrat* must be approved by a popular vote) and by the right of "initiative" 2,000 electors can compel the legislative assembly to consider any legislative proposal. Since 1856, the demand of 3,000 electors has been sufficient to necessitate a popular vote as to the advisability of any change in the constitution. The representatives to the federal *Ständerat* (Council of States) and to the federal *Nationalrat* (National Council) are also chosen by popular vote. (E. A. BH.)

SOLOVETS ISLANDS (SOLOVETSKIYE OSTROVA), a group of islands of the U.S.S.R. located in the White Sea off the end of the Onega Peninsula (Onezhskiy Poluostrov). Administratively they are part of Archangel Oblast'. The islands are formed of morainic deposits and cover 134 sq.mi. (347 sq.km.). The largest, Solovetskiy Island, has an area of 100 sq.mi. (259 sq.km.). On it is located the former monastery, founded in 1436, which was one of the largest, richest, and most famous in north Russia. Its wealth derived largely from salt working. After the reform of the Orthodox Church, Old Believers of the monastery underwent a siege from 1667 to 1676. (R. A. F.)

SOLOVIEV, SERGEI MIKHAILOVICH (1820–1879), author of the first comprehensive, scholarly account of Russian history, whose historiographic approach influenced later scholars, was born on May 5 (old style; 17, new style), 1820, in Moscow. After studying at the university there, he traveled abroad, returning to become professor of Russian history (1847–79). In his monumental *Istoria Rossii s drevneishikh vremën* ("History of Russia from Ancient Times" [to 1774]; 29 vol., 1851–79; 15 vol., 1959–65), influenced by Hegelian thought, he viewed Russian society as having undergone a process of organic evolution, developing from the primitive clan (*rod*) to the patriarchal state, and finally to enlightened absolutism. By stressing environmental factors, he set the achievements of individual rulers in historical perspective. Russia, he maintained, was inherently part of Europe, despite the cultural backwardness which it was the function of the state power, assisted by the "best" elements of society, to overcome. His mild liberalism was combined with Great-Russian patriotism and loyalty to the Orthodox Church, and he supported Alexander II's expansionist Balkan policy as well as his domestic reforms. He died in Moscow on Oct. 4 (O.S.; 16, N.S.), 1879.

See K. D. Grothusen, *Die Historische Rechtsschule Russlands* (1962). (J. L. H. K.)

SOLOVIEV, VLADIMIR SERGEEVICH (1853–1900), Russian philosopher and theologian whose work foreshadowed the 20th-century movement for the reconciliation of the Christian churches with one another and with scientific thought. He was born in Moscow on Jan. 28 (new style; 16, old style), 1853, the second son of the historian S. M. Soloviev (*q.v.*). Soon outgrowing the fashionable materialism that he had adopted as a child, he studied science, philosophy, and theology before graduating as doctor of philosophy at Moscow University. His thesis, *The Crisis of Western Philosophy—Against the Positivists* (1874), won him a teaching post; but the liberal character of his Christianity soon brought him into conflict with the reactionary forces in the university, and in May 1875 he was encouraged to go abroad for 15 months. After visiting London, Cairo, and Paris, he returned to Moscow, but was suspended from teaching again in 1877 and then transferred to the Ministry of Public Instruction in St. Petersburg. An appointment at the University of St. Petersburg (1880) was terminated after he had appealed for clemency toward the murderers of the emperor Alexander II (1881). Thereafter he lived as an independent writer. He died at Uzkoye, near Moscow, on Aug. 13 (N.S.; July 31, O.S.), 1900.

Soloviev was a man of vast and conscientious learning, familiar with patristic and gnostic literature on the one hand and with modern philosophy on the other. Among his most significant writings are *Philosophical Principles of Integral Knowledge* (1877); *Lectures on Godmanhood* (1878–81; Eng. trans. 1944); *Critique of Abstract Principles* (1880); three discourses *In Memory of Dostoevski* (1881–83); *Russia and the Universal Church* (in French, 1889; Russ. trans. 1913; Eng. trans. 1948); *The Meaning of Love* (1892–94; Eng. trans. 1946); *The Justification of the Good* (two editions, 1897 and 1898; Eng. trans. 1915); and *Three Conversations on War, Progress, and the End of Universal History* (1899; Eng. trans. 1915). He also published poems, influential on the Russian Symbolists. "Godmanhood" (*Bogochelovechestvo*) meant the raising of the individual man to identification with Christ, the historically incarnate God, by means of "free theurgy" (the cooperation of workers, artists, and mystics), "free theocracy" (the Christian organization of society), and "free theosophy" (the integration of religion, philosophy, and science). Opposed to the intolerant and exclusively "Russian" religion of the Slavophils (*q.v.*), Soloviev strongly advocated an understanding between the Russian Orthodox Church and Roman Catholicism and blamed the Byzantines for their part in the original schism between East and West. Nationalism he deplored as generative

of hatred among mankind, but he had a patriotic belief in Russia's destiny in a Christian world. The second collected edition of his works, in ten volumes, appeared in 1911–14.

BIBLIOGRAPHY.—E. L. Radlov, *Vladimir Soloviev* (1913), in Russian; M. d'Herbigny, *Un Newman russe: Vladimir Soloviev*, 6th ed. (1934); N. M. Zernov, *Three Russian Prophets* (1944); E. Munzer, *Solovyev, Prophet of Russian-Western Unity* (1956).

SOLSTICE, in astronomy either of the two points at which the sun reaches its greatest declination north or south (Lat. *solstitium,* from *sol,* "sun," and *sistere,* "to stand still"). Each solstice is upon the ecliptic (the projection of the plane of the earth's orbit on the celestial sphere) midway between the equinoxes, and therefore 90° from each. The term is also applied to the time at which the sun reaches the point thus defined (about June 21 and Dec. 21). Compare EQUINOX.

SOLUNTUM (Greek SOLOEIS or SOLOUS; Phoenician KFRA; modern Italian SOLUNTO), an ancient town on the north coast of Sicily, 10 mi. (16 km.) E of Palermo and about 2 mi. NE of Bagheria. Lying on the southeast side of Mt. Catalfano (1,227 ft.), in a naturally strong position 600 ft. (180 m.) above sea level, it commands a fine view. Soluntum was one of the three chief Phoenician towns in the island and remained a Carthaginian possession until the First Punic War, when after the fall of Panormus (Palermo; 254 B.C.) it went over to the Romans. Excavations brought to light considerable remains of the Hellenistic and Roman periods; the older town was probably nearby.

SOLUTIONS are homogeneous mixtures of two or more substances in relative amounts that can be varied continuously up to what is called the limit of solubility. Solutions in water of common salt, sugar, alcohol, and carbon dioxide gas are familiar examples. The history of chemistry is witness to the extraordinary importance of solubility. The apparently mysterious nature of solution and recrystallization invited the speculations of ancient philosophers. Medieval alchemists were almost as interested in finding a universal solvent (alkahest) as they were in gold and in eternal life. Modern chemists take advantage of differences in solubility to separate and purify materials, and for methods of chemical analysis. Materials for laboratory vessels and chemical manufacturing equipment are selected to resist the solvent action of their contents; many liquids are valuable chiefly as solvents. Most chemical reactions occur in solution and are influenced by the solubilities of the reagents.

Life processes depend largely on solutions. Oxygen from the lungs goes into solution in the blood plasma, unites chemically with the hemoglobin in the red blood cells, and is released to the body tissues (*see* BLOOD). Products of digestion also are carried in solution to the different parts of the body (*see* DIGESTION). The dependence of life on the freedom of complex molecules in solution to move and rearrange is strikingly shown by the cessation of metabolic processes on freezing. Nitrogen is soluble in water, blood, and fatty tissues. At sea level, the body of a man of average weight contains about one litre (1.15 g.) of dissolved nitrogen. Living at an altitude of 6,000 ft. (where air pressure is about 0.8 atmosphere) he has 0.8 as much (0.92 g.). The pressure of the air breathed by a deep-sea diver increases by one atmosphere for about every 34 ft. of depth; thus at 100 ft. below the surface the amount of nitrogen dissolved in his body builds up toward roughly 4×1.15 g. As he ascends this extra nitrogen leaves solution and may form bubbles that produce intense pain, block circulation in his blood vessels, and even result in his death. Sometimes called the bends, this condition may be relieved if the victim is put into a pressure chamber and compressed until the bubbles disappear. The pressure is then reduced only as fast as the excess nitrogen can escape through his lungs. Helium is less soluble than nitrogen in body fluids; and helium-oxygen mixtures sometimes are used in place of compressed air in especially deep dives (*see* DIVING APPARATUS; CAISSON DISEASE).

Common salt (sodium chloride) can be dissolved to a limit of approximately 26 weight percent (*i.e.,* about 26 g. of salt per 100 g. of solution); at this concentration the solution is saturated since any added salt remains undissolved. The solubility of common salt increases only slightly with temperature. Sugar dissolves in water up to concentrations of 64.2% at the ice point (0° C; 32° F), 69.2% at body temperature (37° C; 98.6° F), and 82.3% at the boiling point (100° C; 212° F). Alcohol and water mix in all proportions. Carbon dioxide CO_2 gas at one atmosphere pressure dissolves in water up to a limit of 0.145% at 25° C (77° F)—less as temperature increases, more at higher pressures. In a bottle of carbonated water (*see* SOFT DRINKS) at a pressure of two atmospheres, twice as much could be dissolved. It is convenient to remember that one litre of CO_2 dissolves in one litre of water under one atmosphere pressure at 20° C (68° F). Carbon dioxide is 13 times as soluble in benzene as it is in water. The gases of the air all are soluble in water; fishes extract the dissolved oxygen through their gills.

A solid substance dissolved in a liquid is easily detected as a residue by evaporating the solvent. Its presence may also be detected if, when part of the liquid is removed by evaporation or freezing, any property such as density, colour, refractive index, vapour pressure, or freezing temperature changes.

Although the term solution usually suggests a liquid solvent, it applies equally to mixtures of gases. Air is basically a solution of nitrogen, oxygen, argon, water, carbon dioxide, and minute amounts of hydrogen, neon, krypton, xenon, and helium (*see* AIR). Liquid air has a blue colour from its oxygen. If allowed to boil away, the nitrogen escapes more rapidly than the oxygen; the blue colour gradually deepens, and the boiling point gradually rises (*see* OXYGEN: *Preparation*).

There are also solid solutions. For example, ordinary alum $KAl(SO_4)_2 \cdot 12H_2O$ (colourless), and $KCr(SO_4)_2 \cdot 12H_2O$ chrome alum (violet) have crystals so nearly alike in form that they crystallize together from water solution. The crystals vary in colour depending on the relative concentrations of the two alums in the solution (*see* ALUM).

The adjective soluble broadly indicates that a substance goes into solution (whether or not the process involves a chemical change). The colour of iodine (*q.v.*) dissolved in carbon tetrachloride is practically identical with that of iodine vapour (*i.e.,* violet); evidently the iodine molecules are scarcely altered by contact with those of the solvent. However, a solution of iodine in ether is brown, indicating the formation of a chemical complex. The iodine can be recovered by letting the ether evaporate. But when copper dissolves in a dilute solution of nitric acid in water, nitric oxide NO gas is evolved, the copper loses electrons and forms the blue, hydrated copper ion: probably $Cu(H_2O)_4^{2+}$. If the liquid is evaporated, a solid, deep-blue salt remains: cupric nitrate $Cu(NO_3)_2 \cdot 6H_2O$.

Solvent and Solute.—For a solution of a solid or a gas in a liquid, the liquid is customarily designated as the solvent, the substance added is called the solute. If both components are liquids, the distinction loses significance; the one present in smaller concentration is likely to be called the solute.

Concentration.—The concentration of any component in a solution may be expressed in units of weight, or volume, or in moles (gram molecules; *see* MOLECULE). These may be mixed; *e.g.,* weight percent, mole fraction, mole percent, moles per litre, moles per kilogram. The last two are usually used for solutions in water.

Electrolytes and Nonelectrolytes.—Crystals of some salts contain lattices of ions; *i.e.,* atoms or groups with alternating positive and negative charges (*see* ELECTRICITY: *Ions*). To dissolve such a crystal the attractions of the oppositely charged ions (largely responsible for cohesion in the crystal) must be overcome by electric charges in the solvent. These may be furnished by the ions of a fused salt, or by electric dipoles in the molecules of the solvent, as indicated by a high dielectric constant (*see* DIPOLE MOMENTS). Such solvents include water, methyl alcohol, dioxane, liquid ammonia, and hydrogen fluoride. The ions of the solute, surrounded by dipolar molecules of solvent, are detached from each other and free to migrate to charged electrodes. Such a solution can conduct electricity, and the solute is called an electrolyte (*see* ELECTROCHEMISTRY).

The potential energy of attraction between simple, nonpolar molecules (nonelectrolytes) is of very short range; it decreases

approximately as the 7th power of the distance between them. According to this, the energy of attraction between a pair of molecules separated by two molecular diameters is only $\frac{1}{64}$ of what it is at one diameter. For electrolytes the energy of attraction and repulsion of charged ions (like charged spheres) drops only as the first power of the distance; thus their solutions have very different properties from those of nonelectrolytes. The best solvents for one are poorest for the other; although solutions of electrolytes conduct electricity, those of nonelectrolytes do not.

Metallic Solutions.—Metals conduct electricity by having free electrons that can wander through the metallic solid or liquid. Many different metals are miscible in the liquid state, sometimes forming recognizable compounds; some are sufficiently alike to form solid solutions (*see* ALLOYS). A few metals dissolve slightly in their own molten salts. Mercury dissolves at 25° C without chemical change in liquid active phosphorus P_4 to a concentration of 0.03%.

Theories.—The earliest general theory of solutions was based on the concept of osmosis (*q.v.*). If a solution is separated from the pure solvent by a membrane that is permeable to the solvent but not to the solute, the solution will tend to become more dilute by absorbing solvent through the membrane. This can be stopped by establishing on the solution a specific excess of pressure (called the osmotic pressure). J. H. van't Hoff (1886) showed that if the solute is so dilute that its partial vapour pressure above the solution obeys Henry's Law (*i.e.*, is proportional to its concentration in the solution) then osmotic pressure varies with concentration and temperature approximately as it would if the solute were a gas occupying the same volume. This relation led to equations for determining molecular weights of solutes in dilute solutions through effects on the freezing point, boiling point, or vapour pressure of the solvent.

Van't Hoff specified that the gas law as applied to osmotic pressure has limited application; some of his contemporaries assumed that it held also in concentrated solutions. But, for example, when it was used in an attempt to distinguish solvent water from water of hydration in a concentrated solution of calcium chloride, the alleged water of hydration so calculated exceeded all the water in the solution. It became evident that the difference of water pressure in balance on two sides of a semipermeable membrane depends not only on the concentration of solute molecules but also on their attraction for water molecules.

The modern theory of nonpolar solutions is based on the statistical behaviour of a mixture of molecules in chaotic thermal disorder. If the attractions among molecules of like and unlike species are all the same, then the escaping tendency of any species depends simply on its fraction in the mixture (Raoult's Law). If they are unlike and there is a valid theory of how the attraction between the unlike species depends on the attractions in the pure liquids then (if the distribution is still in a state of maximum chaos) there is a basis for calculating the relative amounts of the three sorts of attraction in the mixture. This affords a foundation for a theory of solution applicable for all fractions of the components.

There is abundant evidence that pure liquids and mixtures of nonpolar, reasonably symmetrical molecules are kept by thermal motions in a state of maximum disorder, altogether different from the long-range order existing in crystals. For example, phosphorus, composed of rotating, tetrahedral molecules P_4, melts at 44.1° C, but the liquid can be kept at room temperature (about 20° C) for long periods, and droplets in vacuo have been cooled to −70° C. If the liquid contained any trace of crystallike structures, it surely could not remain liquid under such enormous supercooling.

SOLUTIONS OF NONELECTROLYTES

The properties of nonionic solutions are best described in terms of departures from those of a hypothetical model in which the intermolecular forces of attraction and repulsion are all the same. Such a model is called an ideal solution. There is no change in total volume or temperature when it is formed by mixing its pure liquid components. The escaping tendency of individual molecules is the same in pure liquid and in solution, and the vapour pressure of one species therefore simply depends upon the ratio of the molecules of that species to the total number in the solution. If the solution is composed of n_1 molecules of species 1 and n_2 of species 2, the respective mole fractions of the two species are $x_1 = n_1/(n_1 + n_2)$ and $x_2 = n_2/(n_1 + n_2)$ and $x_1 + x_2 = 1$. If the vapour pressures of the two pure components are p_1^0 and p_2^0, their partial vapour pressures over the solution are $p_1 = p_1^0 x_1$ and $p_2 = p_2^0 x_2$. Total vapour pressure is $p = p_1 + p_2 = p_1^0 x_1 + p_2^0 x_2$ (*see* VAPORIZATION).

This relation was discovered in 1886 by François Marie Raoult (*q.v.*). Its simple, statistical significance was apparently not appreciated, and it was not until much later that it was recognized as an appropriate ideal solution law. That function was temporarily ascribed to van't Hoff's Law, according to which the osmotic pressure of a dissolved substance is equal to the pressure it would exert if in the form of a gas in the same volume. Raoult's Law is obeyed by many mixtures with only moderate deviations throughout the range of composition. It is possible, furthermore, to predict which components are likely to show fair agreement with it. The van't Hoff Law, on the other hand, holds reasonably well only at high dilution, because it takes no account of the specific character of the solvent; this is treated simply as supplying volume for a quasi-gas solute.

Chemical Interactions.—There are several sources of departure from ideality. One is an excess attraction between unlike molecules, such as would be expected in a mixture of phenol C_6H_5OH (weakly acidic) with aniline $C_6H_5NH_2$ (weakly basic). This reduces what otherwise would be the escaping tendency of each component, and the vapour pressures are less than ideal. It occurs also in mixtures of chloroform $CHCl_3$, (weakly acidic) with acetone $(CH_3)_2CO$ (weakly basic), as illustrated in fig. 1. The dotted lines correspond to Raoult's Law; this interaction is to be regarded as specific, selective, and hence chemical. Note that in spite of the interaction the vapour pressure of each component approaches the ideal line as the mole fraction of the other approaches zero.

The terms acid and base are used in the way proposed by G. N. Lewis, a base being a substance with electrons that fit into vacant orbitals of an acid. Bases and acids are in this sense electron donors and electron acceptors (*see* ACIDS AND BASES). Basic character is closely related to low ionization potential. Typical bases include pyridine, amines, acetone, ether, and aliphatic sulfides and iodides. Base strength of aromatics increases in this order: benzene, toluene, *p*-xylene, *o*-xylene, *m*-xylene, mesitylene, naphthalene, phenanthrene. Typical acids are $SnCl_4$, SO_2, BCl_3, and BF_3. The halogens are acids with strengths that increase as follows: Cl_2, Br_2, I_2, IBr, BrCl, and ICl.

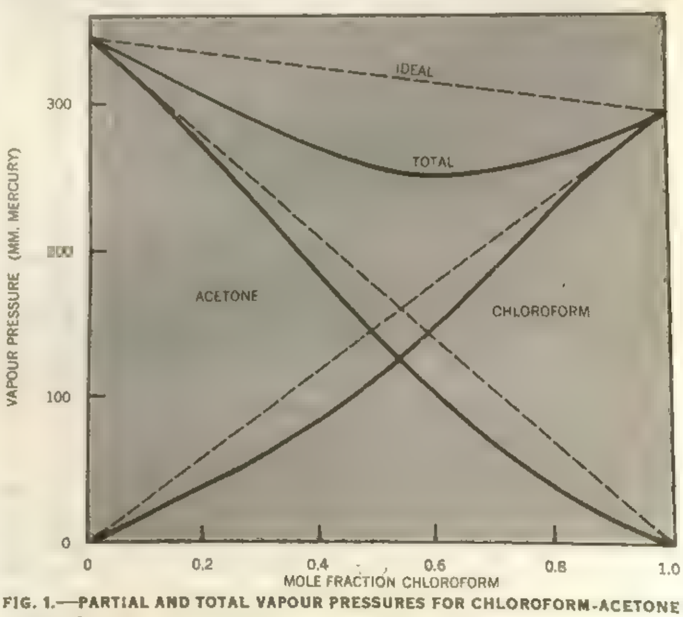


FIG. 1.—PARTIAL AND TOTAL VAPOUR PRESSURES FOR CHLOROFORM-ACETONE AT 35.17° C

Association.—In a mixture of a polar component (its molecules contain electric dipoles) with a nonpolar species, the nonpolar species tends to be squeezed out of the mixture, resulting in vapour pressures in excess of ideal (especially if the polar species is associated through hydrogen bonds, as are water, liquid ammonia, liquid hydrogen fluoride, the alcohols, and the carboxylic acids).

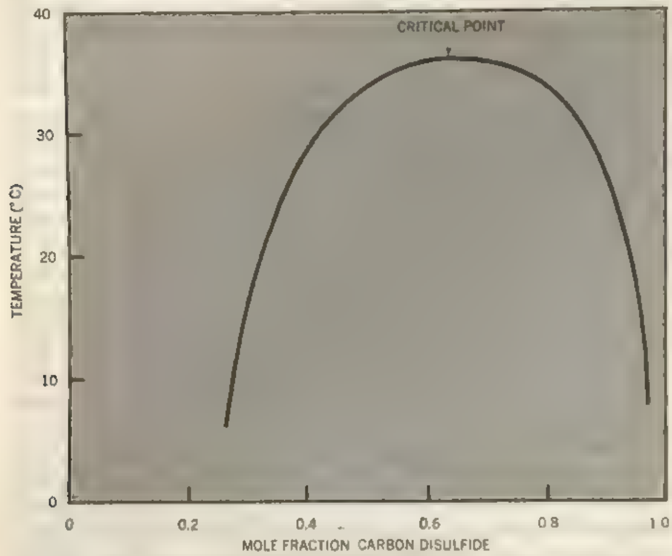


FIG. 2.—SOLUBILITY OF CARBON DISULFIDE WITH METHYL ALCOHOL

This effect often proceeds to the extent of forming another liquid phase, as is the case with mixtures of water with such nonpolar liquids as the hydrocarbons, carbon tetrachloride, and carbon disulfide. Methyl alcohol, CH_3OH, less strongly hydrogen-bonded than water, forms two liquid layers at room temperatures with nonpolar carbon disulfide CS_2. Their mutual solubility increases with temperature by reason of increasing molecular motion until the two phases coalesce at the critical temperature 36.5° C as shown in fig. 2.

Regular Solutions.—A third factor contributing to departure from ideality is inequality in forces of a general, nonselective nature (London forces), such as are responsible for the attractions between the symmetrical, monatomic molecules of the rare gases. They arise from mutual perturbations of the rapidly moving electrons in molecules that are close to each other. The theory of this kind of interaction was developed by F. London (1930) on the basis of quantum mechanics. The attractive potential between a pair of molecules depends on the number and what may be called looseness of their electrons, as indicated by optical polarizability and zero-point energy or, alternatively, low ionization potential (*see* PHYSICS: GENERAL SURVEY: *Atomic and Nuclear Structure*). The potential is of very short range, is additive (*i.e.*, not saturating or selective), and it is not affected by temperature.

According to London's analysis, energy of attraction ϵ_{12} between two unlike molecules is approximately the geometric mean of the potential between the like species ϵ_{11} and ϵ_{22} (*i.e.*, $\epsilon_{12} = \sqrt{\epsilon_{11}\epsilon_{22}}$). This is less than the arithmetic mean of ϵ_{11} and ϵ_{22}; therefore, when two liquids with different intermolecular attractions are mixed, expansion and absorption of heat result, and vapour pressures rise higher than Raoult's Law predicts.

Assuming that thermal agitation suffices to mix molecules with maximum randomness, J. H. Hildebrand and S. E. Wood integrated all the pair potentials in a mole of liquid to give the energy of vaporization, and derived a formula for deviation from Raoult's Law in terms of the energy of vaporization:

$$RT \ln (f_2/f_2^0 x_2) = V_2\phi_1^2(\delta_2 - \delta_1)^2 \qquad (1)$$

Here R is the constant of the ideal gas laws: 1.985 calories per mole per degree absolute temperature T; f_2^0 is the escaping tendency (or fugacity) of pure liquid component *2* and is p_2^0 corrected for the nonideal gas behaviour of the vapour; f_2 is p_2 similarly corrected. V_2 is the volume per mole of component *2*; ϕ_1 is the volume fraction of component 1. The δ values, designated by Hildebrand and R. L. Scott as solubility parameters, are theoretically the square roots of the respective energies of vaporization per mole of the pure liquid components. Practically, the parameters are better determined directly from solubility data. Deviations from Raoult's Law increase mainly with the square of the difference between the solubility parameters of the components. Values for representative liquids are given as Table I.

TABLE I.—*Solubility Parameters δ and Molal Volumes V at 25° C*

Liquid	Formula	V(cc/mole)	δ(cal/cc.mole)^{½}
Perfluoroheptane	C_7F_{16}	225	6.0
2,2,4-Trimethyl pentane	i-C_8H_{18}	166	6.9
n-Heptane	C_7H_{16}	147	7.4
Silicon tetrachloride	$SiCl_4$	115	7.6
Cyclohexane	c-C_6H_{12}	109	8.2
Carbon tetrachloride	CCl_4	97	8.6
Chloroform	$CHCl_3$	81	9.0
Benzene	C_6H_6	89	9.2
Carbon disulfide	CS_2	60	10.0
Bromoform	$CHBr_3$	88	10.5
Bromine	Br_2	52	11.5
Phosphorus (liquid)	P_4	70	14.4

With increasing unlikeness, as measured by $(\delta_2 - \delta_1)^2$, the deviations from ideality increase, as illustrated in fig. 3. The ordinate is the ratio $f_2/f_2^0 = a_2$, the activity; in its pure liquid $a_2 = 1$. Two nonpolar, nonreacting substances with equal parameters can be expected to give an approximately ideal solution, with activities corresponding to line A. Moderate differences in solubility parameters but equal molal volumes give activities of type B. Greater differences yield curve C, indicating a critical point. Still greater differences give curves of type D, with two compositions (at the ends of the horizontal line) with the same activity. This corresponds to two liquid layers. Equation (1) shows that the lower the temperature the greater the deviation from ideal for the same value of the difference $(\delta_2 - \delta_1)^2$. These curves all approach the ideal line A at the upper end, but deviate increasingly toward the lower end. When only a small proportion of component *2* is present, a molecule of component 1 is surrounded mainly by those of the same species, and hence its escaping tendency is not greatly altered. It is altered most at the low end, where the molecules of one kind tend to be surrounded by those of the other.

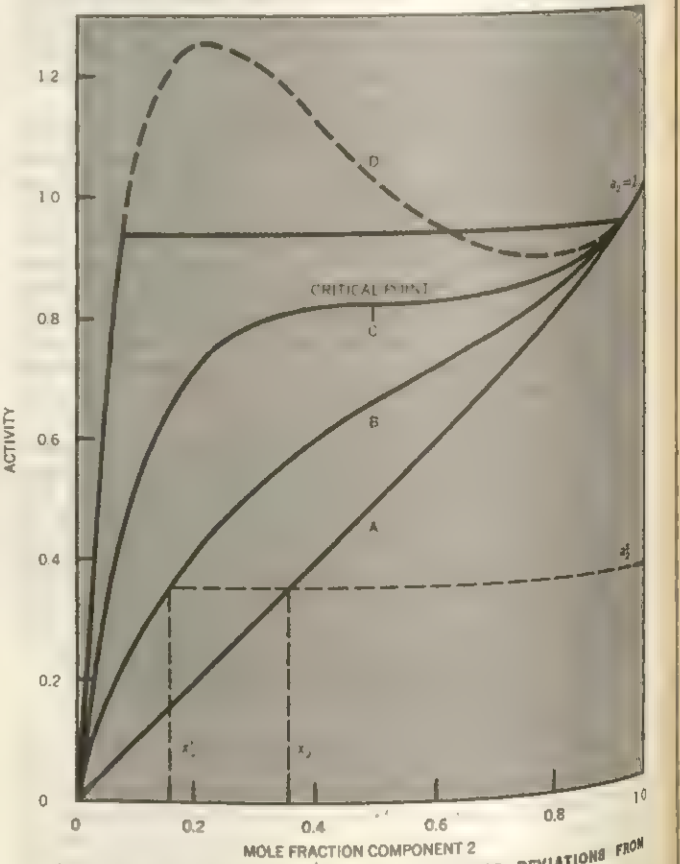


FIG. 3.—ACTIVITY *v.* MOLE FRACTION FOR VARYING DEVIATIONS FROM RAOULT'S LAW (*see* TEXT)

Liquid-Liquid Solubility.—The unlikeness that opposes complete mutual miscibility of liquids in the case of carbon disulfide and methanol (fig. 2) consists largely of hydrogen bonding; *i.e.*, of methanol molecules with each other. Fig. 4 shows curves for the solubility of perfluoroheptane C_7F_{16} which has a very low solubility parameter (6.0) with five other liquids. From Table I it may be seen that the critical temperatures required for complete mixing rise in order of increasing unlikeness as measured by the difference between solubility parameters.

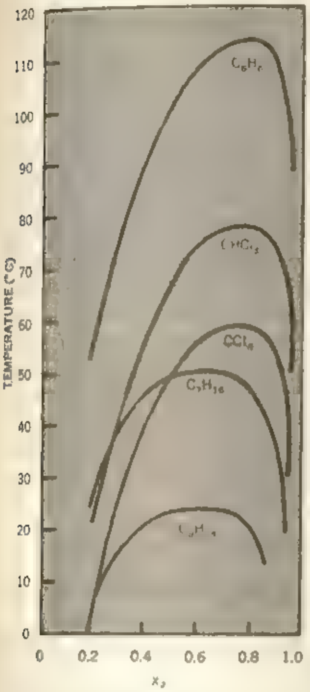


FIG. 4.—SOLUBILITY OF PERFLUOROHEPTANE C_7F_{16} WITH OTHER LIQUIDS

Solubility of Solids.—The fugacity of a crystalline solid is lower than that of its supercooled liquid, since the solid is the stable form below the melting point. The activity ratio f_2^s/f_2^0 is less than 1, and depends on the molal heat of fusion ΔH^F and the distance below the melting point T_m as given by

$$\log_{10}\frac{f_2^s}{f_2^0} = \frac{-\Delta H^F}{2.3R}\,\frac{T_m - T}{T_m T} \quad (2)$$

If the solid dissolves to form an ideal solution, then $f_2^s/f_2^0 = x_2$; and x_2 can be calculated by aid of this equation. It is the value of x_2 in fig. 3 where a_2^s cuts line A.

In a solvent in which the deviation from ideality corresponds to curve B, the solubility would be x_2'; *i.e.*, where the activity along curve B reaches a_2^s. It would be still less if the system followed curve C, and so on. In a solvent with which the solute combines chemically (as in a system like that shown in fig. 1) the solubility would be greater than the value predicted by $\delta_2 - \delta_1$, possibly even greater than ideal.

The way in which the solubility of a solid varies with solvent and temperature is illustrated in fig. 5 for solutions of iodine. Plots of log x_2 against log T have the advantage of giving virtually straight lines except near the melting point. The solid lines represent solutions in which there is practically no chemical interaction, as indicated by their violet colour, identical with that of iodine vapour. The dotted lines represent solutions of different colours: red in benzene, brown in ether, yellow in water, and so on. In the violet solutions only London forces are acting; in the other cases, the specific chemical interaction alters their positions and slopes. The lines for the violet group present a regularity not exhibited by the others. They were designated by Hildebrand as regular solutions, characterized by maximum randomness in the mixing of the molecules, one of the assumptions made in deriving equation (1). This equation accounts reasonably well for the positions in this family of lines, despite the enormous range in solubility.

The other solutions present a more complex problem, involving selective chemical interactions, reducing the randomness, enhancing the solubility over what it would be if the solution were regular, and diminishing the rate of increase of solubility with temperature. The colours of the complexes not only differ from that of iodine in the visible spectrum, but (as discovered by H. A. Benesi and Hildebrand, 1947) show strong absorption bands in the ultraviolet region.

Iodine has a high solubility parameter (14.1) and its solubility falls off rapidly with decrease in the parameters of violet solvents, as illustrated by the following data. These give the solubility of iodine at 25° C in mole percent ($100x_2$) in solvents of widely different solubility parameters δ_1. (Because the parameter for iodine is higher than those for the solvents, the order is the reverse of that in fig. 4.)

	Ideal	CS_2	$CHCl_3$	CCl_4	$SiCl_4$	C_7F_{16}
$100x_2$	25.8	5.46	2.28	1.15	0.50	0.019
δ_1	14.1	10.0	9.0	8.6	7.6	6.0

Solubilities calculated through equation (1) agree reasonably well with these experimental values. Solubilities in the solvents that give nonviolet solutions are higher than they would be in the absence of complexing. Ether $(C_2H_5)_2O$ (solubility parameter 7.5), for example, would in the absence of complexing be only about as good a solvent as heptane C_7H_{16} (*see* fig. 5).

The effect of difference in melting point on the solubility of otherwise similar solid substances is illustrated by the isomers of dinitrobenzene. Their solubilities in benzene are given in Table II together with their melting points and densities. These three

TABLE II.—*Melting Points, Densities, and Solubilities of Dinitrobenzenes in Benzene at 25° C*

Dinitrobenzene isomer	Melting point, °C	Density	Solubility, mole fraction
meta	37 6	1.575	0.376
ortho	116 0	1.590	0 175
para	170 0	1 625	0.031

isomers, when dissolved in benzene, lower its vapour pressure equally. The lowering is less than ideal, because of the polar nature of the two nitro groups, but the effect is practically the same for each isomer; it is the number of nitro groups (not their positions or the overall dipole moment) that is responsible.

The different solubilities of the solid forms of the dinitrobenzenes depend almost entirely on the different crystal-lattice energies. The molecules of the para form are most symmetrical, pack most tightly in their crystal, and are least soluble (*see* SOLID STATE PHYSICS).

Gas Solubility in Liquids.—The solubility of a gas that can be liquefied under pressure can be treated by the method used for the vapour pressure of liquid mixtures. Liquid chlorine, for example, has a vapour pressure of 3.66 atmospheres at 0° C. Its ideal

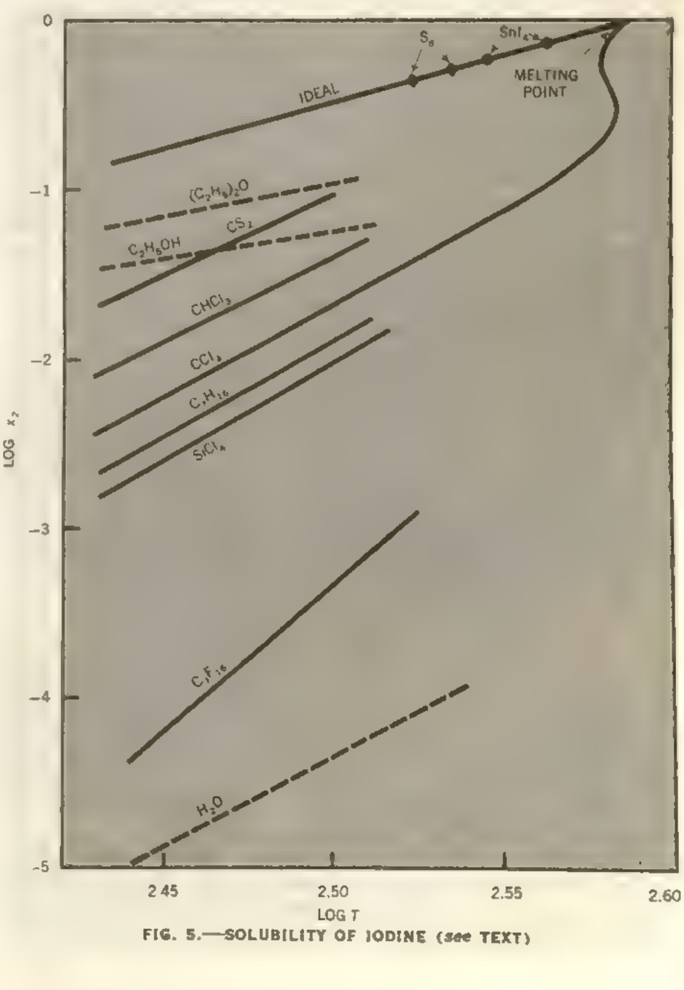


FIG. 5.—SOLUBILITY OF IODINE (*see* TEXT)

solubility at 1 atmosphere and 0° C is therefore $x_2 = 1/3.66 = 0.273$. Its measured solubility in carbon tetrachloride is nearly the same (0.298). The agreement is even closer if the vapour pressure is corrected to fugacity, which gives $f_2/f_2^0 = 0.286$. Its solubility in perfluoroheptane, which has a solubility parameter much smaller than that of chlorine, is considerably less than ideal (0.164).

The higher the pressure required to condense a gas (*i.e.*, the lower its boiling point) the lower its solubility (in the absence of chemical interaction). Also, the low-boiling gases, with their low attractive forces, dissolve to a greater extent as the solubility parameter of the solvent decreases. These relations are illustrated in Table III.

TABLE III.—*Solubility of Gases*

Solvent	δ	Solubility ($10^4 x_2$) at 1 atm. and 25° C					
Gas		He	H_2	N_2	O_2	Ar	Kr
n-C_7F_{16}	6.0	—	14.0	38.7	55.3	—	—
n-C_7H_{16}	7.5	2.5	6.9	—	—	25.0	67.6
c-C_6H_{12}	8.2	—	—	7.6	—	14.9	46.7
CCl_4	8.6	—	3.2	6.4	12.0	13.4	—
C_6H_6	9.2	0.78	2.6	4.5	8.2	8.8	27.3
CS_2	10.2	—	1.6	2.2	4.4	4.9	—
H_2O	—	0.07	0.14	0.12	0.23	0.25	0.45
Boiling point ° K		4.2°	20°	77°	90°	87°	121°

These gases have difficulty in penetrating the strong hydrogen bonds in water. But ammonia, which is moderately hydrogen-bonding and also strongly basic, dissolves readily in ethyl alcohol, and still more easily in water. The figures for x_2 at 1 atmosphere and 0° C are typical: toluene 0.26, ether 0.79, ideal 23.8, ethyl alcohol 39.8, water 48.1.

Molecular Weights of Solutes.—For the more quantitative aspects of solution theory the point of departure remains the ideal solution. Allowance for deviations from ideality can be made fairly well; also, the activity of the solvent approaches ideality as the mole fraction of the solute approaches zero.

For example, molecular weights of substances in dilute solutions (even if they are not ideal) can be determined by measuring the effect of moderate amounts of solute on the activity of the solvent (*see* MOLECULE). The more nearly the solution conforms to Raoult's Law the more precise the result; it is advantageous, therefore, to choose an efficient solvent, or at least to extrapolate a series of measurements to zero mole fraction of solute. The solute will lower the vapour pressure of the solvent, increase its boiling point (nonvolatile solute), depress its freezing point (pure crystals of solvent), and affect the osmotic pressure (*i.e.*, the hydrostatic pressure that must be applied to the solution to keep it in balance with pure solvent from which it is separated by a membrane to which the solvent, but not the solute, is permeable).

Vapour Pressure.—Let a solution be composed of N_1 moles of solvent and N_2 moles of solute. If molal weight is M, and direct weight is w, Raoult's Law may be written $(p_1^0 - p_1)/p_1 = N_2/N_1 = w_2M_1/M_2w_1$. Taking known w_2, w_1 and M_1, and measuring the drop in vapour pressure produced by a nonvolatile solute, M_2 (the molal weight of the solute) can be calculated.

Boiling Point.—The drop in vapour pressure that would be produced by adding solute at constant temperature can be offset by raising the temperature enough to restore the pressure. The following simple mathematical identities can be written.

$$\left(\frac{\partial x_2}{\partial T}\right)_p = -\left(\frac{\partial x_1}{\partial T}\right)_p = \left(\frac{\partial x_1}{\partial p_1}\right)_T \left(\frac{\partial p_1}{\partial T}\right)_{x_1}$$

For an ideal solution $(\partial x_1/\partial p_1)_T = 1/p_1^0$ and in general $(\partial p/\partial T)_v = p\Delta H^{vap}/RT^2$. Since there is no heat of mixing $\Delta H^{vap} = \Delta H_1^{vap}$, and at the boiling point $(\partial x_2/\partial T)_p = \Delta H_1^{vap}/RT_b^2$. For small values of solute mole fraction x_2, write $x_2 = \Delta T_b \Delta H_1^{vap}/RT_b^2$, from which the molal weight M_2 of the solute can be calculated from the measured rise ΔT_b in boiling point.

Freezing Point.—Equation (2) for ideal solubility may be used to find the molal weight of a solute from the effect of a known weight of solute in depressing the freezing point of a known amount of solvent (mole fraction x_1); thus

$$-\ln x_1 = \frac{\Delta H_1^F}{R}\frac{T_m - T}{T_m T} = \ln(1 - x_2) = x_2 + \frac{x_2^2}{2} + \frac{x_2^3}{3} + \dots$$

When the solute mole fraction is small, higher powers may be neglected; $T_m - T = \Delta T$ the freezing-point depression, and $T_mT \cong T_m^2$; therefore $x_2 = \Delta H_1^F \Delta T/RT_m^2$. By determining x_2 from a measurement of ΔT for solute weight w_2 in N_1 moles of solvent, the molal weight M_2 of the solute can be calculated. To obtain precise values of N_2 a series of points for x_2 against ΔT should be taken and extrapolated to $w_2 = 0$.

Osmotic Pressure.—A mole of solvent can be transferred to a solution in either of two reversible ways. If the solvent is transferred by distillation the change in free energy F is $\Delta F = -RT \ln (f_1/f_1^0)$. If it is forced out at pressure P through a semipermeable membrane $\Delta F = P\bar{V}_1$. V_1 is the partial molal volume of the solvent, V_1 its molal volume. The two are virtually identical when the solute is dilute, or when the solution is ideal. Equating these $P\bar{V}_1 = -RT \ln (f_1/f_1^0)$; if the solution is ideal $f_1/f_1^0 = N_1/(N_1 + N_2)$ and $\bar{V}_1 = V_1$, the latter because there is no volume change on mixing the components of an ideal solution. Therefore

$$PV_1 = RT \ln \frac{N_1 + N_2}{N_1} = RT \ln \left(1 + \frac{N_2}{N_1}\right)$$

Now

$$\ln\left(1 + \frac{N_2}{N_1}\right) = \frac{N_2}{N_1}\left[1 - \frac{1}{2}\frac{N_2}{N_1} + \frac{1}{3}\left(\frac{N_2}{N_1}\right)^2 - \cdots\right]$$

When N_2/N_1 is small the higher powers can be neglected, and $PV_2 \cong \dfrac{N_2}{N_1} RT$ or $PV = RT$, where V is the volume of solvent containing 1 mole of solute.

The identity in form of the equations for osmotic pressure and ideal gas pressure led to the unwarranted assumption that osmotic pressure is produced through bombardment of the membrane by solute molecules. However, this would assume that any other two energetically equivalent processes must take place by the same mechanism. Osmotic pressure is simply the hydrostatic pressure required to offset the reduction in escaping tendency of the solvent when solute is added.

Experimental measurement of osmotic pressure with most solutions is very difficult, because a suitable semipermeable membrane is usually difficult, if not impossible, to find. However, it is practical for determining molecular weights of high polymers (*see* POLYMERIZATION), for which membranes are easily devised. With these solutes, high concentrations by weight represent very small mole fractions, and exert but minute effects on the vapour pressure or freezing point of a solvent. For example, a solvent vapour pressure of 100 mm. of mercury would be reduced only to 99.9 mm. by 0.001 mole fraction of a solute. The osmotic pressure of the same solution, if the molal volume of the solvent is 100 cc., would be an easily measured 186 mm.

The Gibbs-Duhem Relation.—There are two energetically equivalent ways to remove one component from a mixture containing N_1 and N_2 moles of the respective components. One way is to distill out, say, dN_1 moles of component 1. The other is to remove solution containing dN_1 moles of component 1 with $(N_2/N_1)dN_2$ moles of component 2, and then distill the latter back into the solution. Equating the partial molal free energies $\bar{F}_1$ and $\bar{F}_2$ of these two steps gives $N_1d\bar{F}_1 + N_2d\bar{F}_2 = 0$. This may also be written in the following especially useful way.

$$\left(\frac{\partial \ln f_1}{\partial \ln x_1}\right)_T = \left(\frac{\partial \ln f_2}{\partial \ln x_2}\right)_T$$

Or activities a may be substituted for fugacities f since $a_1 = f_1/f_1^0$, and so on.

This relation was derived by J. W. Gibbs, and later independently by P. Duhem. For a long time it was called the Duhem equation, but is now commonly called the Gibbs-Duhem relation.

There are important corollaries of the Gibbs-Duhem relation. If Raoult's Law holds for component 1 throughout the entire range from $x_2 = 1$ to $x_2 = 0$, then $f_1/f_1^0 = x_1$, $\ln f_1 = \ln f_1^0 + \ln x_1$, and

$\delta \ln f_1/\delta \ln x_1 = 1$; by the Gibbs-Duhem equation $\delta \ln f_2/\delta \ln x_2 = 1$. Integrating gives $\ln f_2 = \ln x_2 + \ln k_2$, where k_2 is a constant of integration. Since component 1 is assumed to conform to Raoult's Law also when $x_2 = 1$ and $f_2 = f_2^0$, therefore, $\ln k_2 = \ln f_2^0$ and $\ln f_2 = \ln x_2 + \ln f_2^0$; *i.e.*, component 2 also obeys Raoult's Law.

If, however, component 1 approximates Raoult's Law only as x_1 approaches 1 (as is usually the case) then k_2 cannot be evaluated, and $f_2 = k_2 f_2^0$ only can be written as x_2 approaches 1. This is known as Henry's Law. It is especially useful for solutions of the low-boiling gases, with solubilities at 1 atmosphere and ordinary temperature that are mole fractions of the order of 10^{-3} to 10^{-4}.

Fig. 1 and 3 show that a component approaches Raoult's Law at the end where it is the major constituent and approaches Henry's Law (*i.e.*, linearity, but of undetermined slope) at the end where it is the dilute. Van't Hoff had to assume Henry's Law for the solute to derive his equation for the osmotic pressure of the solvent; with modern thermodynamics all such equations as those used for determining molecular weights may be derived by starting with Henry's Law and the Gibbs-Duhem relation.

London Forces.—For the attractive potential ϵ_{12} between a pair of nonpolar molecules of different species London derived the equation

$$\epsilon_{12} = -\frac{3}{2}\frac{\alpha_1\alpha_2}{r^6}\cdot\frac{I_1I_2}{I_1+I_2}$$

For like molecules, this reduces to

$$\epsilon = -\frac{3}{4}\frac{\alpha^2 I}{r^6}$$

where α is the polarizability calculable from refractive index, I is ionization potential, and r is distance between molecular centres. The molecules were assumed to be spherically symmetrical and simple in electronic structure.

Most of the substances in actual solutions are much more complex than this simple model represents. They are seldom nearly spherical and their centres of attraction are in the electron clouds around peripheral atoms. They also may have very different repulsive potentials from the r^{-12} usually assumed for mathematical simplicity. In view of these departures from the simple model it is remarkable that a solubility relationship such as equation (1), based on the simple London equation, is as successful as it is for dealing with solutions of divergent components. If the components are similar in type (*e.g.*, a pair of tetrahalides or a pair of paraffins) the neglected factors have a greater tendency to cancel each other than if different types are mixed (*e.g.*, a tetrahalide with a hydrocarbon). Attempts to allow for such differences become so complicated as to be practically inapplicable.

(J. H. HD.)

SOLUTIONS OF ELECTROLYTES

Recognition of electrolytes as a distinct class of substances is best dated from 1800, when Alessandro Volta published the details of the first electric battery (*q.v.*). This new source of appreciable continuous current was shortly exploited by W. Nicholson and A. Carlisle to demonstrate the composition of water by electrolysis. Sir Humphry Davy (1807) prepared sodium and potassium (hitherto unknown elements) by electrolysis of their molten hydroxides. Michael Faraday introduced the terms still in use: electrolysis, electrolyte (for substances that are decomposed by the current), electrode, anode, and cathode. His quantitative studies led to these fundamental observations:

(a) The chemical change occurs only at the electrodes, not in the bulk of the liquid.

(b) The quantity of chemical change (*e.g.*, the mass of each new substance produced) is proportional to the product of the current intensity and the time for which it flows; *i.e.*, to the quantity of electricity passing.

(c) The mass of each new substance formed by a fixed quantity of electricity is proportional to its equivalent weight (*see* EQUIVALENT). In honour of Faraday's contribution, the quantity of electricity that liberates 1 gram equivalent of new substance at an electrode is known as a faraday (96,500 coulombs); underlying (b) and (c) above is the modern understanding that one faraday is Avogadro's number (6.02×10^{23}) of electrons.

Ions.—Faraday also introduced the term ion for the particles that he saw must exist to carry the current through the liquid; the anion being a negatively charged particle, and the cation being a positively charged particle. Ions are responsible for the special properties of electrolyte solutions. It was long believed that ions were generated by the passage of current, but the modern view that they are always present was established (against strong opposition) by S. A. Arrhenius (*c.* 1890).

This conclusion was based largely on the colligative properties of dilute electrolytes (*i.e.*, those based on the number, rather than type, of solute particles present), including freezing-point depression, osmotic pressure, boiling-point elevation, and vapour-pressure depression. While one-tenth of a gram molecule of any typical nonelectrolyte such as sucrose, glycerol, or urea dissolved in one kilogram of water lowers the freezing point by about 0.186° C, the same amount of a typical electrolyte such as sodium chloride NaCl produces nearly twice as great a drop (0.347° C).

In the limit of very low concentrations, when the ions can act nearly independently of one another, the colligative effect of NaCl (two ions: Na^+ and Cl^-) is twice that of a nonelectrolyte; that of $CaCl_2$ (three ions: Ca^{2+} and 2 Cl^-) is three times; that of $Al(NO_3)_3$ (four ions, Al^{3+} and 3 NO_3^-) is four times. With solutions of salts in water the origin of the ions is clear, for the solid salt itself has an ionic structure. The role of the solvent in this case is to surround the ions by layers of water molecules which (because of their polar nature) greatly reduce the intensity of the electric field near the ions; the normal thermal agitation of the molecules of the liquid is then sufficient to prevent the ions from coming together to reform the crystal unless there are relatively high concentrations of ions. All efficient solvents for electrolytes (*e.g.*, water, formamide $O = CH - NH_2$, hydrogen cyanide HCN) have highly polar molecules, which result in a high dielectric constant (permittivity) for the liquid. Molten salts are also substantially ionic in character; high temperature alone frees the ions from their fixed positions in the crystal lattice, producing the typical electrolytic properties of electrical conduction, with chemical change at the electrodes.

Other electrolyte solutions are formed from solutes that are not appreciably ionic in the pure state; most acids fall in this category. Hydrogen chloride gas consists of covalently bound HCl molecules; it can be liquefied by pressure or cooling, and the liquid is virtually a nonconductor. Its solution in water (hydrochloric acid) is a typical electrolyte, having an unusually high conductivity. The solution is almost entirely ionic; thus, for instance, the vapour pressure of HCl gas above a 5% solution in water is barely detectable, since the vapour is in equilibrium with only the minute concentration of molecular HCl in the solution. Acetic acid, an almost entirely nonionic liquid at ordinary temperatures, also dissolves freely in water; in this case most of the solute remains in molecular form and only a small fraction (about 0.5% in a solution containing one mole of acid per litre) exists in the form of hydrogen ions and acetate ions. This can be shown from its relatively low electrical conductivity and from the freezing-point depression of the solution. Most organic acids and bases resemble acetic acid in being so-called weak electrolytes (*i.e.*, incompletely dissociated to ions in water). In less efficient electrolytic solvents, with lower dielectric constants than water, dissociation is reduced, and even salts may exist largely as so-called ion pairs (*i.e.*, pairs [or higher aggregates] of oppositely charged ions bound together).

The laws of chemical equilibrium require that at extreme dilution any electrolyte must completely separate into its ions; in practice the dilution required for typical so-called weak electrolytes is so great that the electrolyte is undetectable. For example, acetic acid at a concentration of 3×10^{-5} moles per litre remains only about half ionized. This concentration corresponds to only two parts of the acid per million of water.

The solvent is fundamental in determining the behaviour of electrolyte solutions. Salts (typical strong electrolytes) usually have high melting points (ranging higher than 1,000° C). The

theory of ideal solutions as shown by equation (2) accordingly leads to the expectation that the solubility of salts in any solvent at room temperature should be very small. Their solubilities in nonpolar liquids such as benzene and carbon tetrachloride are indeed usually indetectable, but in water and similar highly polar solvents they are often freely soluble. Typical saturation concentrations for metallic chlorides are several moles per kilogram of water. To reconcile this observation with equation (2) consider NaCl as a typical salt. The crystal consists of unhydrated Na^+ and Cl^- ions; equation (2) accordingly gives the mole fraction of unhydrated ions in the solution at saturation. However, most of the ions in the solution have become hydrated through interaction of their strong electric fields with the electric dipoles of water molecules. Thus the mole fraction in terms of the total dissolved electrolyte is large, though the mole fraction of unhydrated ions is very low as required by theory.

The strength of ion interactions with water molecules also is shown by volume changes that occur when electrolytes are dissolved. The volume of a dilute solution of a solid electrolyte in water is nearly always appreciably less than the sum of the separate volumes of electrolyte and water. The apparent molar volume of the solute ϕ_V is computed by subtracting the volume of water from the total volume of solution and dividing the result by the number of moles of solute. This quantity is nearly always less than the volume of a mole of the solid electrolyte; in a number of cases it is actually negative. For example, when caustic soda NaOH is in a very dilute aqueous solution, it is found that $\phi_V = -5.2$ cm.^{3}/mole, compared with a molar volume of 18.8 cm.^{3}/mole for the solid caustic soda. The volume of the solution in this case is less than that of the water alone. These observations are explained in terms of electrostriction, a shrinkage in volume of the water in the vicinity of the ions. The effect increases sharply as the charge on the ions increases and the ionic radius decreases. It is shown very clearly by the volume change on neutralization of dilute solutions of strong acids and bases; *e.g.*,

$$\underset{\text{hydrochloric acid}}{H^+ + Cl^-} + \underset{\text{sodium hydroxide}}{Na^+ + OH^-} \longrightarrow \underset{\text{sodium chloride}}{Na^+ + Cl^-} + \underset{\text{water}}{H_2O}$$

There is a volume increase of 22 cm.^{3} for each mole of acid and base neutralized; this arises from the combination of the charged species H^+ and OH^- to form the neutral molecule H_2O. This reaction has an interesting consequence in the behaviour of water under high pressures. Under ordinary conditions water undergoes a very slight self-dissociation:

$$H_2O \rightleftharpoons H^+ + OH^-$$

the equilibrium concentration of ions being only about one ten-millionth of a mole per litre. The electrical conductivity of pure water is accordingly very low, though still accurately measurable. By subjecting water to the extreme pressures (several hundred thousand atmospheres) developed in explosion shock waves, S. D. Hamann found that it becomes as good a conductor as a quite concentrated salt solution. Evidently the concentration of ions is increased by a factor of about a million as the high pressure forces the volume to decrease. By the laws of chemical equilibrium, then, the self-dissociation equilibrium shifts in the direction favouring a smaller volume; *i.e.*, the extent of dissociation increases. Such effects are of great importance in geochemical studies (*see* GEOCHEMISTRY).

Activity Coefficients.—The terms activity and activity coefficient are used in representing the thermodynamic properties of electrolyte solutions (*see* THERMODYNAMICS). The mean molal activity coefficient γ of an electrolyte is so defined that it becomes unity in infinitely dilute solution, and its value at finite concentrations is related to the chemical potential μ of the electrolyte:

$$\mu = \mu^0 + RT \ln (m\gamma)^2 \quad \text{(for 1:1 electrolytes such as KCl)}$$
$$\mu = \mu^0 + RT \ln 4(m\gamma)^3 \quad \text{(for 1:2 or 2:1 electrolytes such as } Na_2SO_4 \text{ or } MgCl_2)$$

and so on, where m is the molality (gram molecules of solute per kilogram of solvent) and μ^0 is the chemical potential in a standard state. Activity coefficients may be determined by measurements of the electromotive force of suitable reversible electrical cells, or by measurements of such colligative properties as vapour-pressure lowering, freezing-point depression, or osmotic pressure. Most of the data available in the 1960s were derived from reversible cells or from isopiestic (isobaric) vapour-pressure measurements. In the latter method (R. A. Robinson and D. A. Sinclair, 1933) the solution is held in good thermal contact with a reference solution of known vapour pressure, in a vessel kept at constant temperature and evacuated of air. Distillation of water from one solution to another occurs until their vapour pressures are equal; the solutions are then weighed to determine their equilibrium concentrations. Since the solvent activity is known in the reference solution, that in the other solution is the same. From a series of such measurements at different concentrations combined with the Gibbs-Duhem equation the activity of the solute can be calculated.

The behaviour of strong electrolyte solutions attracted the attention of many of the ablest chemists of the 19th century (*e.g.*, Arrhenius, W. H. Nernst, W. Ostwald, and van't Hoff). The conductance and thermodynamic properties of weak electrolytes HA such as acetic acid (*see* HYDROGEN IONS) were satisfactorily accounted for in terms of the ionization equilibrium

$$HA \rightleftharpoons H^+ + A^-$$

with an equilibrium constant

$$K = \frac{[H^+]\,[A^-]}{[HA]}$$

Efforts were made to try to account for the behaviour of solutions of salts and strong acids and bases in the same way, but neither conductance measurements nor equilibrium studies led to satisfactorily constant values of K. This so-called anomaly of strong electrolytes resulted in the gradual development of the idea that strong electrolytes are completely ionized in solution, and that departures from the behaviour to be expected for an ideal solution of the ions arise from the relatively strong electrostatic forces between ions and from their long-range nature. The modern form of the theory of interionic forces in electrolytes was developed by P. Debye and E. Hückel (1923). In outline their approach is to fix attention on one ion in the solution, the central ion. If it is a positive ion it will tend to attract negative ions and to repel positive ions in its vicinity. Any small element of volume near the central ion therefore will have (on a time average) a slight excess of negative charge. Quantitatively, the distribution of this excess charge can be calculated from Poisson's equation of electrostatics combined with Boltzmann's equation, which states that the concentration of ions that have electrostatic energy E is proportional to $e^{-E/kT}$ (where e is the base of natural logarithms, k is Boltzmann's constant, and T is the absolute temperature). It is found that the excess negative charge density ρ around the central positive ion falls off with the distance r according to the law

$$\rho = \frac{Ae^{-\kappa r}}{r}$$

where A is a constant depending on the dielectric constant and the temperature; κ, a parameter depending on these, is proportional to the square root of the ionic strength I of the solution:

$$I = \tfrac{1}{2}c_1z_1^2 + \tfrac{1}{2}c_2z_2^2 + \cdots$$

summed over all species of ions (1, 2, 3 . . .) in the solution, where c_i is the molar concentration of ions of species i in unit volume (1 litre) and z_i is the number of unit charges carried by the ion.

From the now known charge distribution it is possible by standard electrostatic methods to calculate the extent to which the potential energy of the central ion is lowered by electrostatic interaction with its environment. This energy is a component of the chemical potential μ_i of the ion, and takes the form μ_i (electrostatic) $= -(z_i^2e^2N/2\epsilon)\,\kappa/(1 + \kappa a)$ where N is Avogadro's number, e the electronic charge, ϵ the dielectric constant of the solvent, and a is the average diameter of the ions. Assuming that the electrostatic interactions are the only basis for nonideal behaviour, μ_i (electrostatic) can be identified with $RT \ln \gamma_i$, where γ_i is the

activity coefficient of the ion. The activity coefficients of the individual ionic species cannot be measured, but the Debye-Hückel result can be converted to give the mean activity coefficient $\gamma_\pm$ of the electrolyte. The result for an electrolyte with cation 1 and anion 2 is

$$\log \gamma_\pm = -\frac{A|z_1z_2|\sqrt{I}}{1 + Ba\sqrt{I}} \tag{3}$$

Here A and B are constants involving the dielectric constant and the absolute temperature. For aqueous solutions at 25° C the value of A is 0.51 and B has the value 3.3×10^7 when I is expressed in the usual mole per litre concentration units. Since the diameter a of typical ions is a few angstrom units, the term $Ba\sqrt{I}$ in the denominator is of the order of unity at ionic strengths of about one. At very low ionic strengths, however, the term $Ba\sqrt{I}$ becomes small compared to unity, and the result takes the simpler form

$$\log \gamma_\pm = -A|z_1z_2|\sqrt{I} \tag{4}$$

A great deal of extremely precise work (mainly measurements of the electromotive forces of reversible cells and of freezing points of dilute electrolyte solutions) has established the general validity of equation (4) as a limiting law for strong electrolytes in very dilute solutions. Other thermodynamic properties such as heat of dilution and apparent molar volume for strong electrolytes have also been shown to conform at low concentrations to limiting laws demanding linearity in the square root of the ionic strength. These laws can be derived from equation (4) by partial differentiation: with respect to the temperature to give the law for the heat of dilution, and with respect to the pressure to give the limiting law for the apparent molar volume.

The actual behaviour of the activity coefficients of strong electrolytes is more complicated. Conformity with the limiting law is general at extreme dilutions. Equation (3), in which the ion-size term in the denominator is retained, is adequate up to a few hundredths of a mole per litre in many cases, provided that the ionic diameter a is treated as a parameter that can be adjusted to give the best fit between theory and experiment. Beyond this rather low limit of concentration, however, electrolytes may deviate empirically in two ways from the Debye-Hückel behaviour. What may be called a negative deviation is found with many salts in which the product of the ionic charges z_1z_2 is large, and more especially those in which the negative ion is highly charged; *e.g.*, sulfates SO_4^{2-} and ferrocyanides $Fe(CN)_6^{4-}$. These salts have lower activity coefficients than the Debye-Hückel theory predicts; likewise their equivalent conductances are lower than the values given by the Debye-Hückel-Onsager equation (*see* ELECTRICITY, CONDUCTION OF: *Conduction in Liquids*). N. Bjerrum (1926) explained this type of deviation as arising from ion-pairing, the formation of comparatively stable pairs of ions held together by electrostatic forces. The thermodynamic effects of such ion pairs are similar to those found in the presence of undissociated molecules (as in weak electrolytes) and they can be accounted for by superimposing on the Debye-Hückel model a treatment in terms of a chemical association constant. Bjerrum developed a method of calculating this association constant from electrostatic theory; the calculation usually gives at least the right order of magnitude for the association constant.

The other common type of deviation is found with many of the alkali halides, and in very marked form with the salts of divalent metals and univalent anions (such as magnesium chloride). In these cases as concentration increases, the activity coefficient rises above the Debye-Hückel value; frequently this results in a minimum in the curve and a rapid rise beyond the minimum. The activity coefficient often exceeds the value (unity) appropriate to an ideal solution; as an extreme example, in 5.5 molal uranyl perchlorate $UO_2(ClO_4)_2$ the mean activity coefficient is 1,457. Significantly in these cases the electrical conductance shows no abnormality; the usually accepted explanation of such superideal behaviour is that it is a result of ion-solvent interactions. In the Debye-Hückel treatment the solvent appears only as a dielectric medium in which the ions are immersed.

There is a great deal of evidence, however, that many ions (small and highly charged cations particularly) carry a rather firmly attached sheath of water molecules held in place by ion-dipole forces. Thus the mobilities of such ions are much lower than would be expected from their known sizes in crystals and from hydrodynamic arguments; the viscosities of the solutions are often higher than would be expected from similar hydrodynamic arguments, indicating entities larger than the bare ions. The compressibilities of the solutions indicate the presence of a region of already tightly packed water near the ions. Also, the dielectric constants of the solutions (measured by microwave techniques) suggest that a fraction of the water molecules (proportional to the ion concentration) is immobilized and fails to contribute to the dipole rotation that finds expression in the dielectric constant. Though attempts to determine the number of such bound water molecules by various methods do not agree quantitatively, there is general agreement that the order of hydration numbers of the following cations is $Li^+ > Na^+ > K^+$; and $Mg^{2+} > Ca^{2+} > Ba^{2+}$. In a series of salts such as the chlorides, this is also the order of departures from Debye-Hückel behaviour of the activity coefficients. Following early work by Bjerrum (1918), R. A. Robinson and R. H. Stokes (1949) showed that by making the proper allowance for the presence of hydrated ions, the activity coefficients could be reconciled with the Debye-Hückel equation up to concentrations of several moles per litre.

Transport Processes.—While much knowledge of electrolytes has come from the study of equilibrium properties such as activity coefficients, transport properties are of at least equal importance. Transport properties may be defined as those in which matter or energy flows irreversibly from one place to another; in accordance with the second law of thermodynamics, they are accompanied by an increase of entropy if they occur in an isolated system, or by a decrease in Gibbs free energy if they occur in a system held at constant temperature and pressure (*see* THERMOCHEMISTRY). Those of most interest in considering electrolytes are electrical conduction, isothermal diffusion, viscous flow, and thermal diffusion.

Electrolytic Conduction.—Studies of their electrical conductivity have contributed greatly to the understanding of electrolyte solutions, and also find many applications in chemical analysis. When a chemical reaction involves an increase or decrease in the number of ions in a solution or the replacement of one ion by another of different mobility, the conductivity of the solution changes in a well-understood manner. Thus in the neutralization of sodium hydroxide by hydrochloric acid, the highly mobile hydroxide ion is replaced by a chloride ion, with a decrease in conductivity. When the equivalence point is reached, the solution consists of sodium chloride only, and the conductivity reaches a minimum. As more hydrochloric acid is added, the extremely mobile hydrogen ion of the acid can no longer be neutralized, and the conductivity rises rapidly.

Isothermal Diffusion.—If two solutions of the same solute at different concentrations are placed in contact at the same temperature the solute diffuses from the higher to the lower concentration (the solvent in the opposite direction) until the solution is uniform throughout. The rate at which this isothermal diffusion occurs for a given concentration difference is measured by the diffusion coefficient D. Nernst showed that in the limit of extreme dilution the diffusion coefficient of an electrolyte was simply related to the mobilities of the ions; *e.g.*, for sodium chloride

$$D^0 = \frac{2RT}{F^2}\left(\frac{\lambda^0_{Na^+}\lambda^0_{Cl^-}}{\lambda^0_{Na^+} + \lambda^0_{Cl^-}}\right)$$

where R is the universal gas constant, F the faraday, T the absolute temperature, and λ^0 the limiting equivalent conductance. The form of this relation results from the fact that no net flow of current occurs in the process (in contrast to electrical conduction), so that the sodium and chloride ions are constrained to move in the same direction and with equal speeds. For the same reason, the process at finite concentrations is modified by the electrophoretic effect, but not by the relaxation effect (*see* ELECTROPHORESIS; RELAXATION PHENOMENA).

Viscosity.—In viscous flow each layer of the solution is in motion with respect to neighbouring layers. The external force producing this relative motion is opposed by the viscosity (*q.v.*) or internal friction of the liquid. Electrolytes added to water and other solvents produce a small increase of viscosity arising from the electrostatic forces between ions in neighbouring layers. This effect, which is proportional to the square root of the concentration, was evaluated from the Debye-Hückel-Onsager theory by H. Falkenhagen and M. Dole, and results for very dilute solutions are in accord with their predictions. At more usual concentrations this small change is masked by larger effects arising from interactions between the ions and the solvent; these are approximately proportional to the first power of the electrolyte concentration. When the ions are large compared to water molecules, they act as obstructions in the stream lines of the flow; this effect occurs also with neutral solute molecules such as sugars. A. Einstein showed that this effect should obey the equation

$$\eta_{sol}/\eta_{H_2O} = 1 + 5\phi/2$$

where ϕ is the fraction of the volume occupied by the solute particles, and η is the viscosity of the solution or the water. Smaller ions (about the size of water molecules or smaller) sometimes show a similar effect; this can be explained in terms of a permanent solvation sheath of solvent bound to the ion, increasing its effective size. A few ions (notably iodide ion) produce a marked decrease in the viscosity of water. It is believed that the electric field of such ions, though too weak to establish a permanent solvation sheath, is enough to break down the loose bonding between nearby water molecules and their neighbours, resulting in an increased fluidity.

Thermal Diffusion.—Two solutions of the same concentration but at different temperatures will diffuse when placed in contact (with suitable precautions to prevent convective flow due to density differences). After a considerable time a steady state known as a Soret equilibrium is reached in which the solute is more concentrated at either the hot end or the cold end. In electrolyte solutions the approach to this state can be followed by measurements of the conductivity of the liquid at each end, or of the potential difference between suitable electrodes in each solution. The complex theory of this process involves the simultaneous flow of heat and matter, and along with other coupled-flow processes is one of the important points of growth in the modern theory of steady-state thermodynamics.

See also references under "Solutions" in the Index.

(Ro. H. S.)

Bibliography.—D. A. MacInnes, *Principles of Electrochemistry*, 2nd ed. (1947); J. H. Hildebrand and R. L. Scott, *The Solubility of Non-electrolytes* (1950, 1964), *Regular Solutions* (1962); *Annual Review of Physical Chemistry* (1950 *et seq.*); E. A. Guggenheim, *Mixtures* (1952); I. Prigogine, *Molecular Theory of Solutions* (1957); H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions*, 3rd ed. (1958); R. A. Robinson and R. H. Stokes, *Electrolyte Solutions*, 2nd ed. (1959); G. N. Lewis and M. Randall, *Thermodynamics*, 2nd ed. rev. by K. S. Pitzer and L. Brewer (1961); H. C. Van Ness, *Classical Thermodynamics of Non-electrolyte Solutions* (1964). (J. H. Hd.; Ro. H. S.)

SOLVAY, ERNEST (1838–1922), Belgian industrial chemist and social reformer, who developed an important process for making alkali, was born at Rebecq on April 16, 1838. The son of a salt refiner, he attended school in Malonne and later worked for his father at Rebecq. Solvay read books on elementary chemistry and physics and in 1860 became an assistant to an uncle who was head of a small gasworks at Schaerbeek.

In 1863 Solvay developed his process for the production of soda ash, which is considered one of the most important 19th-century advances in industrial chemistry. Replacing the old Leblanc "black-ash" process, Solvay's process was widely utilized in many countries. (*See* also Chemical Industry and Alkali Manufacture.) Solvay became wealthy and contributed to various philanthropies and research organizations. In 1894 he established in Brussels the Institut des Sciences Sociales, which became the Institut de Sociologie in 1901. He opposed the amassing of inherited wealth and proposed a system of social accounting which would replace money. His works include *Science versus Religion* (1879) and *Questions of Social Energetics* (1910). He died at Brussels on May 26, 1922.

SOLWAY FIRTH, a shallow inlet of the Irish Sea between England and Scotland. From St. Bees Head in Cumberland to Abbey Head in Kirkcudbrightshire it is 22 mi. (35 km.) across, and from this line it stretches about 35 mi. farther northeast to the estuaries of the Rivers Esk and Eden at its head. West of its constriction between Bowness and Annan, the firth opens like a funnel, with ramified reentrants, notably at Moricambe Bay on the Cumbrian side and at the mouths of the Nith and Ure Water on the Scottish side. Wide stretches of sand are exposed at low tide, and the estuarine parts are choked by shifting sandbanks. The firth derives its name from the *sulwathe* ("muddy ford") across the Esk Estuary. Naturally the inner part or head of the firth is avoided by shipping, but small vessels from Ireland penetrate as far as Silloth, an artificial port established in 1859 to serve Carlisle. An earlier, short-lived attempt had been made to develop Port Carlisle, east of Bowness, with canal connections from Carlisle. Hadrian's Wall, crossing the isthmian portion of north England from the Tyne Estuary to the Solway Firth, ends on the south shore at Bowness by the lowest ford. The tide advances quickly up the estuary and sometimes forms a bore.

(Ar. E. S.)

SOMA, in Vedic India, the personification of the fermented juice of a plant most probably to be identified with one or more species of *Ephedra* (*q.v.*). In the Rigveda, the entire ninth book is devoted to hymns in praise of Soma, revered as a god with invigorating qualities. The juice was drunk by priests at the sacrifices and its uses go back to Iranian times; it is called *haoma* in the Avesta (*see* Zoroastrianism). Although Soma was no doubt originally equal to Indra and Agni in importance, his place as a divine link or messenger between man and heaven is gradually assumed by Agni. In the post-Vedic classical period Soma is identified with the moon, which wanes when drunk by the gods but is replenished by the sun. *See* also Vedic Religion.

(J. E. B. G.)

SOMALILAND, FRENCH (Côte française des Somalis), an overseas territory of France on the East African coast where the Red Sea meets the Gulf of Aden, is bounded by Ethiopia to the north, west, and south, and by the Somali Republic to the southeast. Its area is 8,960 sq.mi. (23,206 sq.km.) and its capital Djibouti (*q.v.*).

Physical Features.—The country consists mostly of sand and stone desert, broken in places by salt lakes and lava streams and by lofty, usually sterile mountain ranges (Musa or Moussa Ali, 6,630 ft. [2,020 m.]) which extend from a coastal plain inland to the Ethiopian highlands. The coast line is deeply indented by the Gulf of Tajura (Tadjoura). The drainage is partly eastward to the coast and partly inland to the lakes, but few streams flow above the surface except after rain. French Somaliland is notable, indeed, for its extreme aridity and torrid climate. From May to September Djibouti usually has an average maximum temperature of 40° C (104° F) and a minimum of 23° (73°); the annual rainfall is less than 5 in. (127 mm.). The cool season is from November to April. The vegetation is generally sparse, much of the country being clad with acacia scrub, and supports a similar fauna to that found in the Somali Republic.

Population.—The chief inhabitants are the Somali and Afar (or Danakil) peoples of the southeastern Cushitic or Hamitic family. They share similar physical features, although the Somali are usually taller than the Afar. The total population (1964) is 82,100, including 32,358 Somali of the Ise (Esa), Habar Awal and Gadabursi clans, 31,975 Afar, 7,271 Arabs, and 7,817 Europeans. Djibouti, the capital, had a population (1962) of 43,682, half of whom were Somali.

Apart from those in Djibouti, the Afar and Somali are essentially a pastoral, nomadic people, moving over wide areas with their flocks of sheep and goats, herds of camels, and in some areas cattle. Both peoples are hardy and extremely individualistic, being organized into kinship groups, which among the Somali are more dispersed than among the Afar. War and feud are common, particularly on account of the acute competition for access to grazing

and water; and compensation is paid between kinship groups according to a standard tariff. The two peoples differ considerably in languages and culture, but both practise polygyny within the limits set by Islamic law and, despite their traditional hostility, some intermarriage takes place between them. Both are Sunni Muslims, mainly of the Shafi'ite rite, although Hanafites are also found among the Afar. In general, Afar society is more hierarchically organized and more closely akin to that of the Arabs than is the Somali social system. Among the Afar the traditional Muslim sultanates are more developed, there is a pervasive class distinction between the Asaimara nobles and the Adoimara tributary commoners, and some age-grade organizations occur.

History.—Although France had gained Obock as early as 1862, it was not until 1888 that the Côte française des Somalis was established and its frontier with the British Somaliland Protectorate defined. In 1897, by treaty with Ethiopia, the territory was somewhat reduced in extent. Djibouti had become the official capital in 1892, and shortly afterward work began on the construction of the railway to the Ethiopian hinterland which reached Dire Dawa in 1903 and connected with Addis Ababa in 1917. Djibouti was thus assured of prosperity, and its population of 15,000 (1925) more than doubled by the 1960s. The interior was opened up between 1924 and 1934 by the construction of roads and administrative posts. After the Italian conquest of Ethiopia in 1936, the territory was heavily fortified, but in 1940 a neutral Vichy regime was established. Thereafter Djibouti was blockaded until the end of 1942 when the Côte declared for Gen. Charles de Gaulle and the Free French. After World War II, despite developments, Djibouti increasingly lost trade to the port of Assab in Eritrea, which was then federated with Ethiopia. To safeguard the economy, Ethiopia was given a larger interest in the railway by a treaty of November 1959.

French Somaliland acquired the status of overseas territory in 1946; since the *loi cadre* ("organic law") of 1956 it has enjoyed internal autonomy. In 1958 the Côte voted to become an overseas territorial member of the French Community under the Fifth Republic. Those who voted in favour were mainly Afar and Europeans; the Somali electorate as a whole either abstained or voted against the continuation of ties with France.

Administration and Social Conditions.—French Somaliland is administered by an Executive Council, presided over by the governor and elected by the Territorial Assembly. The assembly, itself elected, was set up in 1957 with 32 members. The 1958 elections were won by the party for the Défense des Intérêts Économiques et Sociaux du Territoire, and the new Executive Council consisted of four Somali, three Afar, and one European. The territory also elects a representative in both the French National Assembly and the Senate.

Political realignments in the late 1950s led to the formation of two main parties: the Union Démocratique representing the Afar electorate and the Mouvement Populaire, the Somali. In 1962 these joined forces to campaign for independence within the French Community.

Each of the administrative districts (*cercles*), Djibouti, Tadjoura, Ali Sabieh, and Dikhil, is in the charge of a French colonial officer. Government-appointed headmen (*akil*) and local authorities form auxiliary tribunals in the districts, and Islamic law affecting matters of personal status is administered by stipended cadis.

Social and economic development is financed by grants-in-aid from France which provide an annual contribution of about 239,130,000 Djibouti francs. This is used to finance road construction, port development and maintenance, health services, housing, and other urban and rural social services. The Côte has one 550-bed hospital, seven dispensaries, and five rural sanitary posts. In the mid-1960s more than 3,000 pupils attended schools including some that offered postelementary and technical education.

Economy.—Apart from the pastoral resources of the nomads (estimated at 25,000 camels, 700,000 sheep and goats, and 78,000 cattle), the territory's main economic asset is the free port of Djibouti. In the mid-1960s this was entered by more than 2,000 vessels annually and provided port revenues amounting to about 290,000,000 Djibouti fr. The port serves the Ethiopian hinterland

TRANSPORTATION ROUTES OF FRENCH SOMALILAND

to which it is connected by the Addis Ababa railway. The territory's own exports are mainly marine salt (about 6,500 metric tons annually) and hides and skins. The chief imports are cotton yarns and cotton goods, cattle, coal, and sugar. Djibouti Airport is served by international air routes from Europe, East Africa, the Malagasy Republic, the Arabian Peninsula, and the India-Pakistan Subcontinent. Radio Djibouti broadcasts daily in the French, Arabic, Somali, and Afar languages.

Unlike other French overseas territories the Côte does not belong to the franc area. From March 21, 1949, it had a freely convertible currency called the Djibouti franc (100 Djibouti francs = 0.414507 g. of fine gold). This is guaranteed by a deposit in U.S. dollars with the French American Corporation of New York ($1 = 214.39 Djibouti francs) and has a value equivalent to .23 new francs.

Bibliography.—A. de la Rüe, *La Somalie française* (1937); H. Deschamps *et al.*, *L'Union française, Côte des Somalis, Réunion, Inde* (1948); Notes et Etudes Documentaires, no. 2,774, *La Côte Française des Somalis* (1961); M. Albospeyre *et al.*, *Mer Rouge-Afrique Orientale. Cahiers de l'Afrique et de l'Asie*, V (1959); I. M. Lewis, *The Modern History of Somaliland* (1965); current history and statistics are summarized annually in *Britannica Book of the Year*. (I. M. L.)

SOMALI REPUBLIC (Somalia), a state of northeast Africa formed in 1960 by the union of the former British Somaliland Protectorate and the Italian-administered Trust Territory of Somalia. With an area of 246,154 sq.mi. (637,541 sq.km.), it covers much of the so-called Horn of Africa, extending from the border of French Somaliland along the Gulf of Aden and around the disputed eastern borders of Ethiopia, and down the coast of the Indian Ocean to the Kenya border. The capital is Mogadishu (*q.v.*).

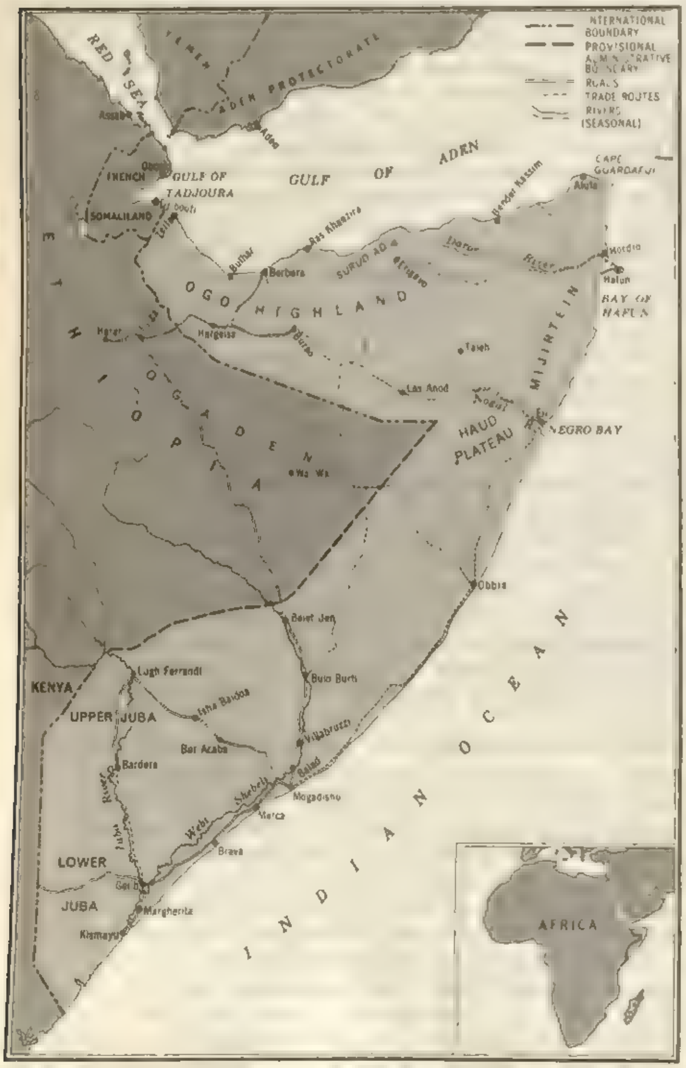
Physical Geography.—As a whole the country presents a harsh environment, part of it being semidesert. The northern coastal plains, which stretch from the Gulf of Tadjoura (Tajura) along the Gulf of Aden into Mijirtein region, are particularly arid, and from their scorched appearance in the dry seasons are known locally as *guban* ("burned"). Inland this barren coastal strip gives place to the maritime range, the Ogo Highland (Surud Ad, near Erigavo, 7,894 ft. [2,406 m.]), running from French Somaliland parallel with the coast and descending gradually southward into the vast tilting Haud Plateau. This plateau has an average height of 3,000 ft. (900 m.) in the centre but rises much higher in the Harar highlands to the west in Ethiopia. The highlands slope gradually and fairly regularly from the northwest and are intersected by

low-lying plains and valleys which are much more widely spaced in the south than in the precipitate north. The Juba River and the Webi Shebeli contain water all the year round and provide the main river system of the whole area. Both rivers descend from the Ethiopian highlands.

The Shebeli, with a total length of about 1,250 mi. (2,000 km.), runs parallel to and north of the Juba through the southern part of the Ogaden to Balad, about 20 mi. (30 km.) from the Indian Ocean, where it turns southwest and, after a further 170 mi. (270 km.), disappears in a series of marshes and sandflats a few miles north of Gelib on the Juba. With exceptionally heavy rain the Shebeli meets the Juba near Margherita and thus enters the sea; otherwise it has no outlet. In the north the largest seasonal streams are the Daror and the Nogal, which run only after heavy rain but contain permanent water in their lower reaches. The Daror enters the Indian Ocean in the Bay of Hafun, and the Nogal farther south in Negro Bay.

Geologically, the plateau consists mainly of gneisses and schists. In the north it is overlaid with red and purple unfossiliferous sandstone, capped near its edge with cherty limestone, also unfossiliferous but possibly of Lower Cretaceous Age. The plains inland from Berbera and the maritime margins between the coast and the foot of the plateau consist of Lower Oölitic limestone.

Climate.—From June to September, especially, the northern coast is unbearably hot and virtually deserted. Berbera and Bender Kassim have average maximum temperatures of 42° C (108° F) and 40° (104°), respectively, with a minimum of 21° (70°). The main wet season is in the cooler months between November and March, but the annual rainfall rarely exceeds 3 in. (77 mm.). Nevertheless, the seasonal freshets which scar the



PRINCIPAL TOWNS AND PHYSIOGRAPHIC REGIONS IN THE SOMALI REPUBLIC

maritime hills facing the Gulf of Aden make water abundant beneath the sandy topsoil. Elsewhere on the plateau water is obtained from wells sunk in dry river valleys, and on high ground it is usually abundant. Some of the inland plains, particularly the Haud, are dry and have little water except in seasonally filled basins. On high ground the rainfall reaches as much as 20 in. (500 mm.), especially in the west, and the climate is pleasant with temperatures considerably lower than along the northern coast.

Compared with the north, the south with its two permanent rivers and their tributaries is relatively well watered. The annual rainfall varies between 13 and 20 in. (330 and 500 mm.) both on the coast and inland. The mean annual temperature on the coast which is very humid in the hot seasons, is about 28° C (82° F), and only a few degrees higher inland. In both northern and southern Somaliland there are two main rainy seasons: *Gu*, falling between March and June, and *Dayr*, between September and December. In the south the hottest period is during March and April, and the coolest from July to August.

Vegetation.—Most of the country is covered by scrub bush, in places almost impenetrably dense, but intersected by wide grassy plains. Vegetation patterns correspond generally to the distribution of saline and nonsaline soils. Along the salty coast the hadun bush or seepweed (*Suaeda*) predominates; and in the lower and saltier parts of the gypseous valleys inland daran (*Limonium* species) abounds. In the more widespread nonsaline areas the most important trees are dosok, one of the box trees (*Buxus hildebrandtii*), and a wide range of acacias. Galol (*Acacia bussei*) is probably the most important tree in northern and central Somaliland. Its long roots provide the framework of the nomadic hut; mats and ropes are made from its bark; the ripe fruit is eaten; and the leaves and young branches are browsed by livestock. It is also used for firewood and charcoal burning. Above 2,000 ft. (600 m.) aloes are plentiful; and in most arid areas of the country the umbrella Euphorbia is common. The Boswellia, which produce frankincense, and other gum-bearing trees flourish in the arid regions of the northeast, particularly in the Erigavo and Mijirtein areas. Some species of figs also occur; and in a few areas above 5,000 ft. (about 1,500 m.) there are small forests of cedar of Lebanon. Along the larger permanent watercourses other trees such as the mangroves and the sycamore fig occur. The commonest grasses are daremo (*Chrysopogon aucheri*), gugangub (*Eragrostis haraensis*), dir (*Andropogon* spp.), and sadehho (*Dactyloctenium* spp.).

Animal Life.—Common among the larger mammals are leopards (widely distributed), hyenas, jackals, foxes, badgers, wild dogs, giraffes (mainly in the south), and a variety of antelopes. The latter include the beisa oryx, the greater and lesser koodoo (kudu), Swayne's hartebeest, waterbuck (mainly in the Shebeli and Juba valleys), Clarke's gazelle (dibatag), Waller's gazelle (gerenuk), Soemmerring's gazelle, the dero (*Gazella speki*), and the extremely common dik-dik or sakaro antelope. The wild ass is plentiful in the north, and the zebra occurs mainly south of the Juba. Lions are found in most parts, but the rhinoceros is rare. There are elephants and hippopotamuses along the Juba and Shebeli, and crocodiles abound in these rivers. There are wart hogs, baboons, a tree monkey, jumping shrews, two varieties of squirrels, a small hare, rock rabbits, stoats, and mongooses. Ostriches are common in open country.

Birds of prey include eagles, vultures, kites, ravens, and the carrion stork. Bustard, guinea fowl, partridge, sand grouse, and wild goose are widely distributed. There are many kinds of snakes, the more dangerous being pythons, puff adders, spitting cobras, and kraits. Scorpions are common after the rains. Mosquitoes, sand flies, houseflies, ticks, and locusts are temporary seasonal pests, usually in limited areas except for mosquitoes near permanent water. Termites are important soil makers except on gypsum.

There are six game reserves in the territory.

The People.—The Somali peoples are classed, largely on linguistic grounds, as southeastern Cushites (*see* CUSHITIC PEOPLES) or Hamites (*q.v.*). Physically they are generally tall and handsome, often quite dark in skin colour, and many have facial

characteristics that suggest an Arab or Semitic strain; this was confirmed by blood-group studies. There has long been trade between Somaliland and Arabia, and Islam is the state religion. The Somali are devout Sunni Muslims, adhering mainly to the Shafi'ite rite. Sufism, the mystical movement of Islam, is well developed among them.

In the arid northern, central, and trans-Juban areas live the Dir, Isaq (Ishaak), Hawiya, and Darod clan-families (the "Samaale," or Somali proper), who are predominantly pastoral nomads. The Digil and Rahanwein ("Sab" groups) live largely as cultivators in the more fertile country between the Shebeli and Juba rivers where they have mixed with earlier Negroid and Bantu settlers and also with Galla (*q.v.*). Some residual Bantu groups remain as cultivators in small enclaves along the rivers and in the hinterland between them. The chief communities of the Shebeli are the Shidle, Kabole, Makanne, and Shebeli; and on the Juba, the Gosha and Gobawein.

Northern Somali is the language of most of the people. Arabic, which is widely understood, is the language of religion and after Italian and English the chief written medium, although a local Somali script, called Osmania, also exists.

The widest difference in social organization occurs between the northern pastoral Samaale groups and the southern cultivating Sab. The former are warlike and frequently driven to feud and violence by competition for scarce grazing and water. A standard indemnity exists for a man's life in the form of blood money (Arabic *dia*, Somali *mag*), rated in most areas at 100 camels. There is also a tariff of compensation for lesser injuries. The basic units of pastoral society are *dia*-paying groups of related agnatic kinsmen, paying and receiving blood money in concert.

Policy decisions are made democratically by all the adult men concerned, and though some of the larger pastoral groups have clan-heads styled sultans, these have little constitutional authority. In contrast, the society of the southern cultivating Sab is more hierarchically organized; chiefs hold greater power and local territorial ties are of more importance in political unity than among the nomads.

The Somali are generally polygamous within the limits of Islamic law; and, among the pastoralists particularly, men like to marry outside their own clan to establish bonds which may be of advantage to them. Although there is no caste structure there are a number of dispersed groups of low status with whom other Somali traditionally do not marry. In the north these are known collectively as *sab*, and in the south where the cleavage is less marked as *bon*. The main groups are the Midgans (leather-workers), the Tumals (blacksmiths), and the Yibirs (soothsayers and pedlars).

History.—*Exploration.*—From their connection with the Ethiopian hinterland, their proximity to Arabia, and their export of precious gums, ostrich feathers, ghee (clarified butter), and other animal produce as well as slaves from farther inland, the northern and eastern Somali coasts have for centuries been open to the outside world. This area probably formed part of Punt, "the land of aromatics and incense," mentioned in ancient Egyptian writings. Between the 7th and 10th centuries, immigrant Muslim Arabs and Persians developed a series of trading posts along the Gulf of Aden and Indian Ocean coasts. Many of the early Arab geographers mentioned these and the sultanates that grew out of them, but they rarely described the interior of the country in detail. In fact intensive exploration really began only after the occupation of Aden by the British in 1839 and the ensuing scramble for Somali possessions by Britain, France, and Italy (see *The Imperial Partition* below). In 1854, while R. F. Burton (later Sir Richard) was exploring the country to the northwest in the course of his famous journey from Berbera to Harar, his colleague J. H. Speke was making his way along the Makhir Coast in the northeast. This region had previously been visited by Charles Guillain, captain of the brig "Ducouedid," between 1846 and 1848. Guillain also sailed down the Indian Ocean coast and went ashore at Mogadishu, Merca, and Brava, penetrating some distance inland and collecting valuable geographical and ethnographic information. In 1865 the German explorer Baron Karl Klaus von der Decken sailed up the Juba River as far as Bardera in the small steamship "Welf" which foundered in rapids above the town. Decken was killed by Somali but much valuable information collected by his expedition survived.

In 1883 a party of Englishmen (F. L. and W. D. James, G. P. V. Aylmer, and E. Lort-Phillips) penetrated from Berbera as far as the Shebeli, and between 1886 and 1892 H. G. C. and E. J. E. Swayne surveyed the country between the coast and the Shebeli and also reached farther east toward the Nogal Valley. During 1894–95 A. Donaldson-Smith explored the headwaters of the Shebeli, reached Lake Rudolf, and eventually descended the Tana River to the sea. In 1891 the Italian Luigi Robecchi-Bricchetti trekked from Mogadishu to Obbia, and then crossed the Ogaden to Berbera. About the same time further explorations were made by another Italian, Capt. Vittorio Bottego. In the 20th century several extensive surveys were made, especially in the former British Protectorate, by J. A. Hunt between 1944 and 1950, and much of the country was mapped by aerial survey.

The Coast and Hinterland Before Partition.—Until recent times the history of the Horn of Africa had been dominated by two great themes: the southward expansion of the Somali from the Gulf of Aden littoral; and the development by Arab and Persian Muslim settlers of a ring of coastal trading towns dating from at least the 10th century A.D. By this time Islam was firmly established in the northern ports of Zeila and Berbera; and at Merca, Brava, and Mogadishu on the Indian Ocean coast in the south. These centres were engaged in a lively trade, with connections as far afield as China. Initially the trend of pressure was from these coastal centres inland, especially in the north.

Probably by the 10th century the country from the Gulf of Aden coast inland was occupied first by Somali nomads and then, to their south and west, by various groups of pastoral Galla who, apparently, had expanded from their traditional homelands in southwest Ethiopia. To the south of these Hamitic Somali and Galla—the "Berberi" of classical times and of the Arab geographers—the fertile lands between the Shebeli and Juba rivers were occupied, partly at least, by sedentary Bantu tribes of the Nyika confederacy whose ancient capital was Shungwaya. Remnants of these Zanj, as they were known to the Arab geographers, still survive in this region; but their strongest contemporary representatives are found among the northeastern coastal Bantu of whom the Pokomo live along the Tana River in northern Kenya. Another smaller allied population consisted of the ancestors of the scattered bands of hunters of northern Kenya and Jubaland known as Wa-Ribi or Wa-Boni, a people whose appearance and mode of existence recall those of the Bushmen of other areas of Africa.

With this distribution of peoples in the 10th century, the stage was set for the great movements of expansion of the Somali toward the south and of the Galla to the south and west. The first known major impetus to Somali migration was the arrival (*c.* 11th century) of Sheikh Ismail Jabarti, ancestor of the Darod Somali, who came from Arabia to settle in the northeast corner of the Somali Peninsula. This was followed, perhaps two centuries later, by the settlement of Sheikh Isaq, founder of the Isaq Somali. As the Darod and Isaq clans grew in numbers and territory in the northeast, they began to vie with each other, with earlier Somali groups, and with their Galla neighbours, thus creating a general thrust toward the south and west. By the 16th century the movements which followed seem to have established much of the present distribution of Somali clans in northern Somaliland. Other Somali pressed farther south, and some, according to the Arab geographer Ibn Said, had already reached the region of Merca by as early as the 13th century.

In the meantime, farther to the west, a ring of militant Muslim sultanates had grown up around the Christian kingdom of Abyssinia and the two sides were engaged in a protracted struggle for supremacy. Somali clansmen regularly formed part of the Muslim armies: the name "Somali" first occurs in an Abyssinian song of victory early in the 15th century. In the 16th century the Muslim state of Adal, whose port was Zeila, assumed the lead in the holy wars against the Christian Amhara. The turning point in the struggle between Christians and Muslims was reached with the Abyssinian victory in 1542, with Portuguese support, over the remarkable Muslim leader Mohammed Grañ, who, with his Somali

armies, had harried Abyssinia almost to the point of collapse. This victory, which saved Abyssinia, also closed the door to Somali expansion westward and increased the pressure of the Somali and Galla thrust south. With this stimulus, the main mass of the Galla swept into Ethiopia from the south and southwest, and streamed in conquering hordes as far north as the ancient city of Harar, which was laid waste in 1567.

This massive invasion left something of a political vacuum in the south of the horn that new Somali settlers were quick to fill. By the 17th century, the influx of new migrants, competing and jostling with each other, had become considerable. The old Ajuran Somali sultanate, linked with the port of Mogadishu, was overthrown and Mogadishu itself invaded and split into two rival quarters. Some of the earlier Somali groups found refuge in northern Kenya. The continuing Somali thrust south—largely at the expense of Galla and Zanj predecessors—was ultimately only effectively halted at the Tana River by the establishment of administrative posts about 1912.

Thus by the latter part of the 19th century the coastal and hinterland traditions had merged and the centre of pressure had swung from the coast to the interior. In the north, the ancient ports of Berbera and Zeila, much reduced in prosperity and importance, were now controlled by Somali nomads and the position with the old ports of Merca, Brava, and Mogadishu was very similar. These towns had all been penetrated by various Somali clans and the dominant political influence became that exercised by the Geledi clan ruling the lower reaches of the Shebeli. Commercial and political links which provided an opening for European infiltration had, however, also been forged between these two coasts and the outside world. Part of the northern Somali coast including Zeila was then nominally under Turkish suzerainty, the Turkish claim going back to the 16th century when Turkish forces had aided Mohammed Grañ in his campaigns against Ethiopia. The southern coastal towns, on the other hand, acknowledged the overlordship of the sultan of Zanzibar, although the latter's authority was slight in comparison with that exercised locally by the Geledi Somalis.

The Imperial Partition.—About the middle of the 19th century the Somali Peninsula became a theatre of competition between Great Britain, Italy, and France. On the African continent itself Egypt was also involved, and later Abyssinia, expanding and consolidating its realm under the guiding genius of Menelik (later emperor of Ethiopia). Britain's interest in the northern Somali coast followed the establishment (1839) of the British coaling station at Aden on the short route to India. The Aden garrison relied upon the importation of meat from the adjacent Somali coast. France sought its own coaling station and obtained Obock on the Danakil coast in 1862, later thrusting eastward and developing the Somali port of Djibouti. Farther north, Italy opened a station in 1869 at Assab which with later acquisitions became the colony of Eritrea. Stimulated by these European maneuvers, Egypt revived Turkey's ancient claims to the Red Sea coast. In 1870 the Egyptian flag was raised at Bulhar and Berbera.

Britain at first protested these Egyptian moves, but by 1877 had come to regard the Egyptian occupation as a convenient bulwark against the encroachments of European rivals. With the disorganization caused by the revolt in the Sudan, however, Egypt was obliged to curtail its colonial responsibilities, and evacuated Harar and its Somali possessions in 1885. In these circumstances, the British government reluctantly decided to fill the gap left by Egypt. Between 1884 and 1886, accordingly, treaties of protection were drawn up with the main northern Somali clans guaranteeing them their "independence." Somali territory was not fully ceded to Britain, but a British protectorate was proclaimed and vice-consuls appointed to maintain order and control trade at Zeila, Berbera, and Bulhar. The interior of the country was left undisturbed, only the coast being affected.

Meanwhile, France had been assiduously extending its colony from Obock and a clash with Britain was only narrowly averted when an Anglo-French agreement on the boundaries of the two powers' Somali possessions was signed in 1888. In the same period the Italians were also actively extending their Eritrean colony and encroaching upon Ethiopian territory. Not to be outdone Menelik took the opportunity of seizing the Muslim city of Harar, left independent after the Egyptian withdrawal. In 1889 Ethiopia and Italy concluded the Treaty of Uccialli, which in the Italian view established an Italian protectorate over Ethiopia. Arms and capital were poured into the country and Menelik was able to apply these new resources to bring pressure to bear on the Somali clansmen around Harar. In 1889 Italy also acquired two protectorates in the northeast corner of Somaliland; and by the end of the year the southern part of the Somali coast leased by the British East Africa Company from the sultan of Zanzibar was sublet to an Italian company.

Italy had thus acquired a Somali colony. From 1892 the lease was held directly from Zanzibar for an annual rent of 160,000 rupees, and after the failure of two Italian companies by 1905, the Italian government assumed direct responsibility for its colony of Somalia. To the south of the Juba River the British East Africa Company held Jubaland until 1895, when this became part of Britain's East Africa Protectorate. Britain and Italy reached agreement in 1884 on the extent of their respective Somali territories, but the Battle of Aduwa (1896), at which the infiltrating Italian armies were crushingly defeated by Ethiopian forces, radically changed the position. Ethiopia, then independent of Italy, was plainly master of the hinterland, and in 1896–97 Italy, France, and Britain all signed treaties with the new emperor Menelik, curtailing their Somali possessions. Italy gave up the Somali Ogaden, and Britain excised much of the western Haud from its protectorate. Although the land and Somali clansmen (who were not consulted), so abandoned, were not recognized as belonging to Ethiopia, there was nothing then to stop their gradual acquisition by Ethiopia.

These arrangements had scarcely been completed when the British Somaliland Protectorate administration found its modest rule threatened by the religious rebellion led by Mohammed bin Abdullah Hassan. This Somali sheikh (known to the British as the "Mad Mullah") of the Ogaden clan, living with his mother's people in the east of the Protectorate, was an adherent of the Salihiya religious order whose reformist message he preached with messianic zeal. He quickly achieved wide recognition for his learning, piety, and skill as a mediator, and initially cooperated with the authorities. In 1899, however, Sheikh Mohammed came into conflict with the recently established Christian mission, and was also involved in a petty dispute with the administration. With the current European and Ethiopian encroachment, and with the example of the Sudanese mahdi, these two incidents provided the seeds which rapidly developed into a major Somali insurrection.

Sheikh Mohammed assumed the title of sayyid (not of mahdi) and his followers were known as the dervishes. He displayed great skill in employing all the traditional tactics of Somali clan politics in building up his following, strengthening these with the call to national Muslim solidarity against the infidel colonizers. Arms and ammunition, denied to Somalis in the past, became easily available through the ports of Djibouti and the northeastern coast, and the dervishes, although opposed by many Somalis who were branded as traitors to Islam, successfully weathered four major English campaigns between 1900 and 1904. The cumbersome British armies, hampered by their supply and water requirements, found the dervish guerrilla tactics hard to combat effectively, and when in 1910 the British government decided to abandon its inconclusive and extremely expensive operations and withdrew to the coast leaving chaos in the interior, Sayyid Mohammed seemed to have gained the day. A new policy was subsequently adopted, however, and with the aid of an increasingly effective camel constabulary (whose founder, Richard Corfield, was killed at the Battle of Dul Madobe in 1913), the dervishes were kept at bay until 1920, when a combined air, sea, and land operation finally routed them. The formidable dervish stronghold at Taleh was bombed but the sayyid escaped, as so often before, only to die of influenza a few months later while desperately seeking to rally his scattered followers.

After 1920 administrative control (under the colonial office since 1905) was gradually restored in the protectorate. In So-

malia, where the Italians had been gradually extending their hold on the country, the sayyid's rebellion had caused less disruption, and the appointment in 1923 of the first fascist governor marked a new positive phase in the life of the colony. Two years later, Britain ceded Jubaland with the port of Kismayu (Chisimaio); and in 1926, after a bitter military campaign, the two northern Italian protectorates were firmly incorporated. Italian settlement was encouraged, and fruit plantations were developed along the Shebeli and Juba valleys. Although agreements of 1897 and 1908 had defined the border with Ethiopia, this had not been demarcated, except for a stretch of about 18 mi. (28 km.) delimited in 1910, and remained in dispute, thus facilitating the gradual Italian infiltration into Ethiopia. In 1934 the celebrated Walwal incident (*see* ETHIOPIA: *History*) occurred in the eastern part of the Ogaden claimed by both Italy and Ethiopia. The Italian conquest of Ethiopia which followed brought the Ethiopian and Italian Somali territories together within the framework of Italy's short-lived East African empire.

Establishment of the Somali Republic.—During World War II the British Protectorate was evacuated (1940) but was recaptured with Italian Somalia in 1941 when Ethiopia was also liberated. With the exception of French Somaliland, all the Somali territories were then united under British military administration. In 1948 the Protectorate reverted to the colonial office; the Ogaden and the Haud were gradually surrendered to Ethiopia; and in 1950 the Italians returned to Somalia with ten years to prepare the country for independence under UN trusteeship.

Taking advantage of the modest progress which the British military administration had effected, social and political advancement was rapidly pursued, although economic development proved much more difficult. The effect of these changes was naturally felt also in the British Protectorate which, in the event, became independent on June 26, 1960. On July 1 Somalia followed suit and the two territories joined to form the Somali Republic.

Prior to independence the grazing movements of nomads into Ethiopian territory were recognized by the Anglo-Ethiopian Treaty of 1897 and the agreement of 1954, but the Somali government refused to take over the treaty since it did not recognize Ethiopia's rights to these grazing lands.

The dispute over this question caused relations with Ethiopia to remain strained in the early 1960s, although with the independence of Kenya, interest also focused on the future of the Somali-inhabited northeastern areas of Kenya. The Northern Frontier District Commission of October 1962 established that most of the tribesmen in these areas wished to join the Somali Republic. The failure of the 1963 Kenya constitution to meet this desire led the Somali inhabitants to boycott the Kenya elections and the Somali Republic to sever diplomatic relations with the U.K. In pan-African and international affairs the republic pursued a moderate policy seeking to maintain strict neutrality. In 1961 delegates attended the third All Africa Peoples Conference in Cairo and the Monrovia and Belgrade conferences. At the Addis Ababa Conference in February 1962 the republic indicated its desire to join an East African federation of African states.

Population.—Censuses of population have been attempted in parts of the Somali Republic but because of the difficulties and expense involved in enumerating a largely nomadic people they have not been completed. The Somali population (1963 est.) totals 2,250,000. This estimate gives an overall Somali density of about nine persons per square mile (over three per square kilometre), but in parts of the country, such as the Mijirtein region, the density falls to less than half that figure. There are also about 80,000 somalized Bantu and Negroid inhabitants, concentrated between the two main rivers, about 40,000 Arab, Persian, Indian, Pakistani, Eritrean, and Ethiopian immigrants, and about 4,500 Europeans, mainly Italian.

By clan-family the Somali population is divided approximately as follows: Darod 750,000; Hawiya 520,000; Isaq 400,000; Dir 150,000; Digil and Rahanwein 430,000. Despite a high incidence of tuberculosis, and seasonal malaria, the population was increasing in the mid-20th century. A net reproduction rate of 1.3 was calculated from data on the municipal populations of the south.

The largest towns, with 1965 estimates of population, [illegible] Mogadishu, capital of the republic (127,119); Merca [illegible] and Hargeisa [illegible] and Berbera [illegible] 219), capital and port of the Northern region, respectivel[illegible]

Administration.—The Somali Republic is a democratic unitary state. The president, the head of state, is elected for six years by the National Assembly, and he selects from its members a prime minister to form a government. The prime m[illegible] sisted by a Council of Ministers. With independence [illegible] Legislative Council of the Protectorate was joined to the [illegible]tive Assembly at Mogadishu to form a unitary National Assembly this has 123 members. Although legislative government was introduced in Somalia only in 1956 and in the protectorate in 1957, the political parties have a longer history. The party with the widest following is the Somali Youth League, formed in 1943 during the British military administration of Somalia. This party won the elections in Somalia in 1959 and, on the unification of the two territories, formed a coalition government with the Somaliland National League and the United Somali Party, the parties then in power in the protectorate. The government, which represents all the main clan and tribal groupings in the republic, is concerned to promote the full realization of the "Greater Somalia" ideal—the further union of the republic with the remaining Somali territories, French Somaliland, Ethiopian Somaliland, and the Northern Province of Kenya. Of these, Harar province of Ethiopia contains the largest number of Somali (about 500,000). Voters in a referendum held in June 1961 approved the provisional constitution (framed by the government of the former Italian Trust Territory before its union with British Somaliland) by a large majority—1,760,540 in favour and 182,911 opposed. However, voters in former British Somaliland opposed it 54,284 to 49,527. In the national elections held in March 1964 the Somali Youth League returned to power with 69 seats.

Local Government.—The republic is divided into eight regions: Western (Hargeisa), Eastern (Burao), Mijirtein (Migiurtina), Mudugh, Hiran, Benadir, Upper Juba (Alto Giuba), and Lower Juba (Bassa Giuba), each divided into administrative districts. Control is exercised from district headquarters through stipended headmen, local authorities (*akils*), or recognized chiefs (sultans). However, because of the traditional absence of strong chiefly authority, the system of administration is in practice direct rather than indirect, except in the arable areas of the south. District, municipal, and semiurban councils exist throughout the republic, but these have developed fiscal, executive, and legislative powers more effectively in the towns and certain agricultural areas than in the nomadic areas.

Living Conditions and Social Services.—In the 1960s United Nations community development plans were in progress. Health services were rapidly expanded despite a continued lack of doctors. Member states of the European Economic Community (EEC) built and equipped a modern 800-bed hospital near Mogadishu; two rural 50-bed hospitals were being built by Soviet engineers. The fight against tuberculosis, one of the most serious scourges in the country, was intensified and a new tuberculosis centre was opened with WHO aid in Mogadishu in 1961. The control of seasonal malaria was also improving through extended antimalaria services.

Justice.—Although the administration of justice is primarily the function of the judiciary, subject to the minister of justice, administrative officials in charge of districts and regions sometimes have also to act as magistrates. The operation of the official penal code is to some extent modified by the existence of the customary *dia*-paying group agreements which help to define the legal status of the individual in a dispute. A religious judiciary also exists in the form of government-appointed *qadis* administering the Shari'ah (mainly the Shafi'ite code) in matters of personal status such as marriage, divorce, inheritance (in the towns), and the assessment of physical injuries. This last function is important in a country where, particularly among the nomads, physical violence is not uncommon. Normally, however, although the *qadis* [illegible] compensation for injuries, the resulting claims for damag[illegible] dea[illegible] with by the secular courts.

A serious difficulty was the existence of two separate legal codes, that established by the British in the Northern regions and that by the Italians in the south. With UN expert advice a new legal code was drafted amalgamating and modifying the laws. An excellent police force was built up on the basis of the British-trained force of the former protectorate and that trained by the Italians in the south. This was fully integrated under a southern Somali commandant and northern Somali deputy. The U.S., Italy, and Britain assisted in the further training and equipping of the force.

Education.—There are about 400 elementary, intermediate, and assisted Koranic schools, but few secondary schools. The school of politics and administration, opened at Mogadishu in 1950, was transformed in 1957 into a technical and commercial institute. A higher institute of law, economics, and social studies was founded in 1954, providing a two-year Rome University diploma course, and in 1960 became a university institute charged with coordinating higher education throughout the country. In the 1960s this was being further developed to provide a wider range of courses. Other developments included a secondary boarding school in Mogadishu, built with Soviet aid, and a teacher training institute at Afgoi built with American aid. Egypt has also provided funds for schools and in the mid-1960s there were about 200 Egyptian teachers in the republic, the Egyptian government's main contribution to education. Several hundred Somali students attend overseas universities and technical institutes in Communist and Western countries. In the south, where Italian was formerly the most-favoured foreign language, English has been increasingly taught. It is estimated, however, that more than 90% of the population remains illiterate.

Defense.—A small but impressive national army, equipped with Egyptian, British, Russian, and Italian aid, was formed from the Somali Scouts (the military establishment of the former protectorate) and certain Italian-trained police units in Somalia. In the mid-1960s it was planned to enlarge and modernize the Army and build up the Air Force with Soviet aid.

The Economy.—This is largely a subsistence economy although livestock, hides, and skins have long been exported, and cash crops were introduced by the Italians. The Somali are chiefly a nomadic, pastoral people. More than two-thirds are pastoralists, some of whom practise subsidiary cultivation; about one-sixth are engaged in sedentary agriculture; and the remainder are occupied in commerce, industry, fishing, and the government service.

Livestock.—The pastoralists move widely in search of grazing for their flocks of sheep and goats and herds of camels, and also cattle in some cultivated areas, especially in the western part of the Northern regions and in the south between the Juba and Shebeli, except in tsetse-infected areas along the rivers. Since in the dry seasons sheep, goats, and cattle require watering every few days, they are herded and managed separately from the camels, which can go without water for many days. Small groups of families, each consisting of a man, his wife or wives, their unmarried daughters, and infant sons, move with their flocks and burden camels, setting up camp in areas of pasture within easy reach of water. The grazing camels, herded by boys from the age of about six years upward and young unmarried men, are often several hundred miles distant from the nomadic camps with their flocks, especially in the dry seasons, and the two herding groups move through the pastures to a great extent independently of each other. Prescriptive rights are generally not asserted to pasturage, and people move with their stock according to the rhythm of the seasons and the availability of grass and water and in accordance with the pattern of relations between groups. In the dry seasons people and stock are concentrated near the wells, competition for access to which is often so acute that fighting results; but in the rainy seasons more dispersed grazing patterns are assumed.

Crops.—In the western part of the Northern regions about 140,000 ac. (56,650 ha.) are tilled, the main crop being sorghum sown after the *Gu* rains of April and May and harvested about September or October. The country's most important agricultural area, however, is that between the Juba and Shebeli rivers, occupied principally by the Digil and Rahanwein tribes. About 30,000 sq.mi. (77,700 sq.km.) are estimated to be suitable for cultivation, but only about one-tenth of this area is exploited. The southern crops are cereals, durra, beans, finger millet (*Eleusine coracana*), sesame, cotton, manioc, papaya, bananas, and some other fruits. In the regions of black alluvial soil along the rivers where tsetse fly militates against mixed farming, and in the high plateau region behind the *Doi*, durra and maize (corn) are the chief crops; and in the red and white soils durra, millets, sesame, sweet potatoes, and manioc predominate. Traditionally, southern cultivation is by hoe, in contrast to the north where the cultivators use a local form of plow, and there are two durra crops, one sown in May and harvested in August and the second sown in September or October for harvesting about four months later. Yields are said to average 7–9 cwt. an acre.

The chief export crop is the banana, grown by a number of Italian and Somali concessionaires and sold mainly to Italy; in the mid-1960s the market was widening. Italian plantations and Somali farmers in association with Italian concessionaires also produce cotton for export. The production of sugar by a Somali government–Italian consortium on the Shebeli meets most of the country's needs; this has been under expansion.

Development Plans and Industries.—Agricultural and irrigation development plans were initiated after World War II, especially between the Juba and Shebeli rivers, and in the sorghum-growing areas of the Northern regions. In the first area a large-scale UN Special Fund hydrological and resources survey was begun in 1962, and in the Northern regions a mechanized cultivation project was launched in the Tug Wajale plain in 1960. Water-drilling and pastoral betterment projects and the extension of veterinary services were undertaken among the nomads who were being encouraged to move less extensively. Egypt was providing an abattoir and a weaving and spinning factory and was promoting new sugar and cement industries. The U.S.S.R. was financing the creation of a cotton ginnery, a meat- and fish-canning factory, and a dairy produce factory. The banana industry subsidized by Italy was expanded and improved. Iron ore deposits were utilized with UN assistance, and the vast gypsum deposits near Berbera, with the aid of hydroelectric power, offered considerable potential. Oil prospecting (a main field of private investment) and the search for minerals continued. Established minor local industries produced alcohol, edible oil, preserves, food-paste, canned meats and fish, mineral waters (soft drinks), shoes, furniture, chemicals, and mechanical appliances as well as electricity. In 1963 a five-year development plan was inaugurated with a budget of So.Sh.1,400,000,000.

Trade and Finance.—The main exports are bananas, livestock on the hoof, hides and skins, gums, cereals, and cotton. Since independence the livestock industry has steadily improved and banana exports have also increased. In the mid-1960s the total annual value of exports was So.Sh.180,200,000, while imports were valued at about So.Sh.270,800,000 and consisted chiefly of foodstuffs, textile, and mechanized equipment. The bulk of trade is with Italy.

Although in the mid-1960s the republic was still dependent upon Italy and Britain for budgetary aid (as well as for development aid), it was anticipated that fiscal and accounting improvements and vigorous control of public expenditure would soon make it possible to balance the national budget. Foreign aid from both the Western and Communist blocs and from neutral countries was being coordinated by a central planning commission.

The official unit of currency is the Somali shilling divided into 100 cents: 7.14 So.Sh. = U.S. $1; 20 So.Sh. = 1 pound sterling. The Credito Somalo acts as a development bank; there is also a national bank and branches of the banks of Italy, Rome, and Naples, and of National and Grindlays Bank.

Communications.—There are no railways in the republic. EEC funds were used in the early 1960s to build tarmac roads to link the ports of Merca and Kismayu with the nearby banana plantations. Work began on a $7,000,000 U.S. development project to improve Kismayu's harbour facilities, and the port of Mogadishu was also extended. In the mid-1960s international grants and credits were to provide for the development of roads. Hargeisa and Mogadishu are linked by Aden Airways, while smaller local centres are served by irregular police flights. Somalia Aviation

(state owned) operates internal passenger and cargo services. Since independence the scope of information services and broadcasting has widened. Radio Somali at Mogadishu and Hargeisa broadcasts in Somali, Arabic, Italian, and English and there are public radio sets in towns and villages.

BIBLIOGRAPHY.—*General, History, and Exploration:* R. F. Burton, *First Footsteps in East Africa* (1856); F. L. James, *The Unknown Horn of Africa* (1888); L. Robecchi-Bricchetti, *Somalia e Benadir* (1899); U. Ferrandi, *Lugh: emporio commerciale sul Giuba* (1903); H. G. C. Swayne, *Seventeen Trips Through Somaliland*, 3rd ed. (1903); D. Jardine, *The Mad Mullah of Somaliland* (1923); L. A. di Savoia-Aosta, *La esplorazione dello Uabi-Uebi Scebeli* (1932); G. Corni, *Somalia Italiana*, 2 vol. (1937); E. S. Pankhurst, *Ex-Italian Somaliland* (1951); J. A. Hunt, *A General Survey of the Somaliland Protectorate, 1944–1950* (1951); G. Natali, *Le prime elezioni politiche in Somalia* (1956); *Bibliografia Somala* (1958); A. A. Castagno, Jr., *Somalia* (1959); M. Karp, *The Economics of Trusteeship in Somalia* (1960); I. M. Lewis, *A Pastoral Democracy* (1961), *The Modern History of Somaliland* (1965).

Flora, Fauna, and Geology: R. E. Drake-Brockman, *The Mammals of Somaliland* (1910); E. Chiovenda, *Flora Somala* (1929); W. A. Macfadyen, *The Geology of British Somaliland* (1933); G. F. Archer and E. M. Godman, *The Birds of British Somaliland and the Gulf of Aden*, 2 vol. (1937); G. Dainelli, *Geologia dell'Africa orientale*, 4 vol. (1943); P. E. Glover, *A Provisional Check-List of British and Italian Somaliland Trees, Shrubs and Herbs*, vol. i (1947).

Ethnology and Archaeology: N. Puccioni, *Le popolazioni indigene della Somalia italiana* (1937); A. T. Curle, "The Ruined Towns of Somaliland," *Antiquity*, 11:315 (1937); J. D. Clark, *The Prehistoric Cultures of the Horn of Africa* (1954); I. M. Lewis, *Peoples of the Horn of Africa* (1955); E. Cerulli, *Somalia*, 2 vol. (1957, 1959); B. W. Andrzejewski and I. M. Lewis, *Somali Poetry* (1964).

Current history and statistics are summarized annually in *Britannica Book of the Year*. (I. M. L.)

SOMASCHI (ORDO CLERICORUM REGULARIUM A SOMASCHA; C.R.S.), a Roman Catholic order of clerks regular, following the rule of St. Augustine, founded in 1528 by St. Jerome Emiliani to work among the poor. It was approved in 1540. The Somaschi conduct schools and orphanages, mainly in Italy.

See also JEROME EMILIANI, SAINT; ORDERS AND CONGREGATIONS, RELIGIOUS.

SOMBOR, a town in the autonomous region of Vojvodina, in the Socialist Republic of Serbia, Yugoslavia, lies east of the Danube near the Hungarian border, 34 mi. (55 km.) SW of Subotica. Pop. (1961) 37,802. The town is a junction of seven railways, an important agricultural centre, and has minor industries. Sombor was occupied by the Turks in 1552. After a century and a half of Turkish rule, it became a fortress of the Habsburg military frontier. From 1747 it became a free town; it was severely damaged in 1848–49 during the Hungarian Revolution; since 1918 it has been part of Yugoslavia, though during World War II it was occupied by the Hungarians. (V. DE.)

SOMERS, JOHN SOMERS (SOMMERS), BARON (1651–1716), English statesman and lawyer, lord chancellor of England from 1697 to 1700, was born on March 4, 1651, near Worcester, the son of John Somers, an attorney. Educated at Trinity College, Oxford, he studied law under Sir Francis Winnington, and in the 1670s associated himself with the leaders of the Country Party, especially with the earl of Essex, William Russell, and Algernon Sidney, but he never entered into their plans so far as to commit himself beyond recall. Algernon Sidney's *Just and Modest Vindication of the Two Last Parliaments* (1681) was attributed to Somers, and he may have assisted in the final drafting. His part as junior counsel in the successful defense of the seven bishops (June 1688) made his reputation. After the Revolution in November 1688 he was elected for Worcester to the Convention Parliament in January 1689. One of the managers for the Commons in the conferences between the two Houses, he was instrumental in persuading the Lords to accept the statement that James II had "abdicated" (not "vacated") the throne. He was chairman of the committee which drew up the Bill of Rights (1689), and is usually credited with a major share in its composition.

Somers was made solicitor general in May 1689 and was knighted in October. He was chairman of the select committee on the Corporation Bill (1690) for the restoration of the borough charters forfeited to Charles II and James II. During the next three years he became closely associated with Edward Russell (afterward earl of Orford), Thomas Wharton, and Charles Montague (afterward earl of Halifax), who with him were to comprise the future Whig Junto; in May 1692 he was appointed attorney general, and in March 1693 lord keeper of the great seal. With the disgrace of the duke of Leeds in 1695 and the virtual retirement of the duke of Shrewsbury in 1696, Somers emerged as the king's chief minister, the leader of the Junto, and the spokesman of Whig opinion in general. His prestige was enhanced by his sweetness of character, combined with strength, and his reputation for disinterestedness. In 1696 he pushed through an act of association which enabled him to purge local government of the more inveterate Tory gentry, and in 1697 he was appointed lord chancellor with the title of Baron Somers of Evesham.

But with the coming of peace with France, by the Treaty of Rijswijk in 1697, the Junto ministry began to break up, and in 1698 and 1699 Somers' relations with William III became increasingly strained. Meanwhile the Commons opposition harried him unmercifully, especially for his grant of a privateering licence to the notorious Capt. William Kidd and his investment in Kidd's expedition to New England. All attempts to remove Somers by parliamentary vote failed, but on April 17, 1700, at the end of the stormiest session of the reign, he resigned at William's command. In April 1701 he was impeached by the Commons for his share in the negotiations which led to the treaty of 1698 for the partition of the Spanish Empire. He defended himself most ably before the House and the charges were dismissed by the Lords in June. Later that year William asked him to form another ministry, but these plans were dashed by the king's death in March 1702.

In the reign of Anne, Somers and the other Junto peers cooperated with the Marlborough-Godolphin government. In 1706 he was one of the managers of the negotiations which resulted in the union with Scotland (1707), and in November 1708, as part of the price for the Junto's continued support of the ministry, he was made lord president of the council. He won over Queen Anne, who had hitherto detested him, but he resigned with the other Whigs on Godolphin's dismissal in September 1710. On George I's accession (August 1714) he was at once reappointed to the Privy Council, but he was too ill to attend regularly, let alone accept the high office that would have been his for the asking. Unmarried, Somers died at his villa, Brookmans, near North Mimms, Hertfordshire, on April 26, 1716, and his title became extinct. His works, which included a few poems, several pamphlets, and an authoritative treatise on the succession to the crown, are listed in H. Walpole, *A Catalogue of the Royal and Noble Authors of England*, volume ii (1758). Most of Somers' papers were accidentally destroyed later in the 18th century.

BIBLIOGRAPHY.—J. Campbell, *The Lives of the Lord Chancellors*, vol. iv (1846); for contemporary accounts, *see* J. Addison's paper in *The Freeholder*, May 14, 1716, and the anonymous *Memoirs of the Life of John, Lord Somers* (1716). (J. P. K.)

SOMERSET, EARLS AND DUKES OF, English titles held in the later Middle Ages by members of the Beaufort family and since 1660 by the descendants of Edward Seymour, duke of Somerset, known as the Protector.

WILLIAM DE MOHUN (d. in or before 1155), lord of Dunster, was heir and possibly grandson of William de Moion, seigneur of Moion or Moyon in the Cotentin, Normandy, who went to England in or after 1066 and became sheriff of Somerset (1084 and 1086). He was created earl of Somerset by the empress Matilda in 1141, and apparently supported her throughout the civil war. Stephen, however, did not recognize William's earldom and no new charter was granted by Henry II. William was sometimes also called earl of Dorset, since the two shires were once administered by one sheriff.

The earldom of Somerset in the Beaufort family dated from February 1397, when it was granted by Richard II to JOHN BEAUFORT (*c.* 1371–1410), eldest of the three sons of John of Gaunt, duke of Lancaster, by his liaison with Catherine Swynford. He was succeeded in turn by his three sons, HENRY (1401–1418), JOHN (1404–1444), created duke of Somerset in 1443, and EDMUND (1406–1455), created duke of Somerset in 1448 (*see* SOMERSET, EDMUND BEAUFORT). The family were staunch Lancas-

trians, and Edmund's son Henry (1436–1464) was attainted (1461) under Edward IV. His brother Edmund (*c.* 1439–1471), thus "duke of Somerset" to Lancastrians only, was himself attainted (1465). On his execution (May 6, 1471) after the Battle of Tewkesbury, the male line became extinct. (For details of the careers of the Beaufort earls and dukes, *see* Beaufort.)

The title of duke of Somerset was next held by Henry Fitzroy (*c.* 1519–1536), illegitimate son of King Henry VIII, who was created earl of Nottingham, duke of Richmond, and duke of Somerset in June 1525. He died without issue.

Edward Seymour (*c.* 1500–1552), duke of Somerset, known as Protector Somerset, was the first of the line of dukes to which the present holder of the title belongs. Brother of Henry VIII's third wife, Jane, and uncle of Edward VI, he was created Viscount Beauchamp of Hache (in Somerset) in 1536; earl of Hertford in 1537; and in 1547 Baron Seymour and duke of Somerset (*see* Somerset, Edward Seymour, 1st Duke of). His honours, which were entailed on the issue of his second marriage in priority to that of his first, were forfeited by attainder in 1552. In 1613 Robert Carr (*c.* 1587–1645), favourite of James I, already (1611) Viscount Rochester, was created earl of Somerset (*see* Somerset, Robert Carr, Earl of). On his death without male issue his titles became extinct.

In September 1660 the titles of Baron Seymour and duke of Somerset were restored to William Seymour (1587–1660), great-grandson of the Protector. In 1621 he had inherited the titles of earl of Hertford and Baron Beauchamp which had been granted to his grandfather Edward Seymour in 1559 and he was himself created marquess of Hertford in 1641. In 1610 he had incurred the serious displeasure of James I by his marriage with Arabella Stuart, the king's cousin. Both were imprisoned, but William made a successful escape, remaining abroad until after Arabella's death (1615). All but John, the youngest of his five sons by his second marriage, predeceased him; he was succeeded by his grandson William (1652–1671), 3rd duke of Somerset. William died unmarried and his uncle John (*c.* 1628–1675) became 4th duke. As John died without issue, the marquessate of Hertford became extinct and his cousin Francis Seymour (1658–1678), 2nd Baron Seymour of Trowbridge, grandson of a younger brother of the 2nd duke, became 5th duke of Somerset. Francis, who died unmarried, was succeeded by his brother Charles (1662–1748), known as "the Proud Duke." His marriage (1682) with Elizabeth, sole heiress of Joceline, 11th duke of Northumberland, brought him immense estates, including Alnwick Castle in Northumberland, Petworth, Sussex, Syon House in Middlesex, and Northumberland House in London. He held household appointments under Charles II and James II, but bore arms for William of Orange (afterward William III). In the memorable crisis when Queen Anne was at the point of death, Somerset, acting with the dukes of Argyll and Shrewsbury, secured the Hanoverian succession to the crown. His son Algernon (1684–1750), 7th duke, fought (1708–11) in Flanders during the War of the Spanish Succession. He had no sons, and was created (October 1749) Baron Warkworth of Warkworth Castle, Northumberland, and earl of Northumberland with special remainder to his son-in-law, Sir Hugh Smithson, bart., and to his heirs male by Somerset's daughter Elizabeth. On the 7th duke's death the earldom of Hertford and the baronies of Beauchamp and Seymour of Trowbridge became extinct. The dukedom of Somerset and the barony of Seymour passed to the heir male, Sir Edward Seymour, 6th bart. (1695–1757), 8th duke, a descendant of the Protector by his first marriage, the issue of which had been excluded from succession to the titles and estates until after the failure of the issue by his second marriage (*see* above). The 8th duke's sons, Edward (1718–1792) and Webb (1718–1793), succeeded in turn, as 9th and 10th dukes. The 11th duke, Edward Adolphus (1775–1855), third but first surviving son of Webb, changed his name from Seymour to St. Maur. His son, another Edward Adolphus (1804–1885), 12th duke, held many minor posts in Liberal governments, and was first lord of the admiralty from 1859 to 1866. Leaving no male issue, he was succeeded in turn by his brothers Archibald Henry Algernon (1810–1891) and Algernon Percy Banks (1813–1894), 13th and 14th dukes, and then by Algernon St. Maur (1846–1923), son of the 14th duke. The 15th duke died without issue and the 16th duke was his kinsman Edward Hamilton Seymour (1860–1931). His grandson Percy Hamilton Seymour (1910–   ) is the 18th duke.

SOMERSET, EDMUND BEAUFORT, Duke of (c. 1406–1455), leader of the Lancastrians on the outbreak of the Wars of the Roses (*q.v.*), the most notable of the second generation of the Beaufort (*q.v.*) family, was the son of John, earl of Somerset. His early career was military: a prisoner in France (1421–1427), on Cardinal Beaufort's Bohemian crusade (1429), and again in the French wars in the 1430s; he was created a Knight of the Garter (1436), earl of Dorset (1442), and marquess of Dorset (1443). In 1447 he was appointed captain general in France and Guienne; in 1448 he was made duke of Somerset and in the following years he presided over the loss of English conquests in northern France.

Although the Beauforts had been barred from the succession (1407) by Henry IV, Somerset was, nevertheless, after the king the senior surviving male member of the House of Lancaster. With the death (1450) of William de la Pole, duke of Suffolk, he was the natural leader of the old court faction against the new opposition of Richard, duke of York. Through his handling of the king he did not do badly in the confused years 1450–55, during which the first sides were taken for the conflict. He survived an appeal (1450–51), was committed to the Tower (*c.* December 1453) during the king's first imbecility, and emerged early in 1455 with the king's returning sanity. York seems to have determined to redress the political balance by force of arms; and at the first Battle of St. Albans on May 22, 1455, Somerset was sought out and killed.

Bibliography.—K. B. McFarlane, "England: The Lancastrian Kings, 1399–1461," *The Cambridge Medieval History*, vol. viii (1936); C. A. J. Armstrong, "Politics and the Battle of St. Albans, 1455," *Bulletin of the Institute of Historical Research*, vol. xxxiii (1960). (P. S. Le)

SOMERSET, EDWARD SEYMOUR, 1st Duke of (c. 1500–1552), was protector of England in the reign of Edward VI. Born about 1500, he was the eldest surviving son of Sir John Seymour (d. 1536) of Wolf hall, Wiltshire. The Seymours claimed descent from a companion of William the Conqueror, who took his name from St. Maur-sur-Loire in Touraine; and the protector's mother was descended from Edward III. Edward was page of honour to Mary Tudor at her marriage with Louis XII in 1514, served in the duke of Suffolk's campaign in France in 1523 and accompanied Cardinal Wolsey on his embassy to France in 1527. Henry VIII, to whom he became (1530) squire of the body, married his sister Jane in 1536, and Edward was created Viscount Beauchamp, and, after the birth of Edward VI, earl of Hertford.

In 1541, during Henry's absence in the north, Hertford, Thomas Cranmer and Baron Audley had the chief management of affairs in London; in Sept. 1542 Hertford was appointed warden of the Scottish marches, and in December lord high admiral, a post which he almost immediately relinquished in favour of John Dudley, Viscount Lisle, later earl of Warwick and duke of Northumberland (*q.v.*). In Feb. 1544 Hertford was made lieutenant general in the north and instructed to punish the Scots for their repudiation of the treaty of marriage between Prince Edward and the infant Mary Queen of Scots. He landed at Leith in May, captured and pillaged Edinburgh, and returned a month later. He then served at Boulogne, returning to Scotland in Sept. 1545 to avenge the Scottish victory at Ancrum moor. In March 1546 he was sent to Boulogne to supersede Henry Howard, earl of Surrey (*q.v.*); and from October to the end of Henry's reign he was in attendance on the king, engaged in that unrecorded struggle for predominance which was to determine the complexion of the government during the coming minority. Personal, political and religious rivalry separated him and Lisle from the Howards, and Surrey's attempt to secure the predominance of his family led to his own execution and his father's imprisonment.

Their overthrow had barely been accomplished when Henry VIII died. He had had no statutory power to appoint a protector, but, in the council of regency which he nominated, Hertford and Lisle enjoyed a decisive preponderance; and the council deter-

mined to follow precedent and appoint a protector. They chose Hertford (now duke of Somerset), and he quickly emancipated himself from the trammels originally imposed on him; he was king in everything but name. He used his authority to divest the government of the apparatus of absolutism perfected by Thomas Cromwell. In his first parliament, which met in Nov. 1547, he procured the repeal of all the heresy laws and nearly all the treason laws passed since Edward III. In foreign affairs, also, he at first adopted a more liberal policy. He had protested against his instruction in 1544, and now dropped the claim to suzerainty over Scotland which Henry VIII had revived. He sought to win over the Scots by promises of autonomy, free trade and equal privileges with England. But Scottish sentiment, backed by Roman Catholic influence and by French intrigues, money and men, proved too strong for Somerset's amiable invitations. The Scots turned a deaf ear to his persuasions; the protector led another army into Scotland in Sept. 1547 and won the battle of Pinkie (Sept. 10).

Somerset apparently thought that the religious question could be settled by public discussion and throughout 1547 and 1548 England went as it pleased so far as church services were concerned. All sorts of experiments were tried and the country was involved in a grand theological debate, in which Protestant refugees from abroad hastened to join. The result convinced the protector that the government must prescribe one uniform order which all should be persuaded or constrained to obey; but the first Book of Common Prayer, which was imposed by the first Act of Uniformity in 1549, was a studious compromise between the new and the old learning, very different from the Protestantism of the second book imposed in 1552 after Somerset had been removed. The Catholic risings in the west in 1549 added to Somerset's difficulties, but were not the cause of his fall. The factious and treasonable conduct of his brother Thomas Seymour (*q.v.*), lord high admiral, in whose execution (March 20, 1549) he weakly acquiesced, also impaired his authority. But the main cause of his ruin was the divergence between him and the majority of the council over the questions of constitutional liberty and enclosures of the commons. His efforts to check enclosures were foiled, and the popular revolts which resulted from their failure weakened his position. He was divided in mind between his sympathy with the rebels and his duty to maintain law and order. France seized the opportunity to declare war on Aug. 8; and the outlying forts in the Boulonnais fell into their hands, while the Scots captured Haddington.

These misfortunes gave a handle to Somerset's enemies. Warwick combined on the same temporary platform Catholics who resented the Book of Common Prayer, Protestants who thought Somerset's mildness paltering with God's truth and the wealthy classes as a whole. In September he concerted measures with the former lord chancellor Wriothesley; and in October, after a vain effort to rouse the masses in his favour, Somerset was deprived of the protectorate and sent to the Tower. But the hostile coalition broke up as soon as it had to frame a constructive policy. Warwick jockeyed the Catholics out of the council and prepared to advance along Protestant lines. He could hardly combine proscription of the Catholics with that of Somerset, and the duke was released in Feb. 1550. For a time the rivals seemed to agree, and Warwick's son married Somerset's daughter. But growing discontent with Warwick made Somerset too dangerous. In Oct. 1551, after Warwick had been created duke of Northumberland, Somerset was sent to the Tower on an exaggerated charge of treason, which broke down at his trial. He was, however, as a sort of compromise, condemned on a charge of felony for having sought to effect a change of government. Few expected that the sentence would be carried out, and apparently Northumberland found it necessary to forge an instruction from Edward VI to that effect. Somerset was executed on Jan. 22, 1552.

See A. F. Pollard, *England Under Protector Somerset* (1900), also his article in the *Dictionary of National Biography* and vol. vi of *Political History of England* (1910). (A. F. Po.; R. B. Wm.)

SOMERSET, ROBERT CARR (or Ker), Earl of (*c.* 1590–1645), Scottish politician who stood high in the favour of King James I. He was the youngest son of Sir Thomas Ker of Ferniehurst by his second wife, Janet, sister of Sir Walter Scott of Buccleuch, and may have accompanied James as page to England, but his early history is obscure. In December 1607 he was knighted and made a gentleman of the bedchamber and thereafter rose rapidly in royal favour. Entirely devoid of all high intellectual qualities, Carr was endowed with good looks, excellent spirits, and considerable personal accomplishments. These advantages were sufficient for James. In 1609 the king conferred on Carr Sir Walter Ralegh's forfeited manor of Sherborne. Carr's influence was already such that in 1610–11 he persuaded the king to dissolve his first parliament, which had shown signs of attacking the Scottish favourites. On March 25, 1611, he was created Viscount Rochester, and subsequently a privy councilor (April 1612), and on Lord Salisbury's death in 1612 he began in July to act as the king's secretary. On Nov. 3, 1613, he was advanced to the earldom of Somerset, in December was appointed treasurer of Scotland, and in 1614 lord chamberlain. Somerset fell from favour in 1615, when the circumstances of the murder of Sir Thomas Overbury in 1613 were disclosed, and he and his wife, who had secured a divorce from the earl of Essex to marry him, were implicated. For this story *see* Overbury, Sir Thomas. Possibly Somerset was no more than an accessory after the event. He was pardoned in 1624, and from that time disappears from history. He died in July 1645.

See the article by S. R. Gardiner in the *Dictionary of National Biography*, with authorities there cited; G. E. Cokayne, *Complete Peerage*, xii (1953). (R. B. Wm.)

SOMERSET, a southwestern county of England, bounded north by the Bristol Channel, the Avon River, and Gloucestershire, east by Wiltshire, south and west by Dorset and Devon. The area of the geographical county is 1,614 sq.mi. (4,179 sq.km.).

Physical Geography.—The county consists of a basin surrounded on three sides by hills and limited on the fourth by the sea. The northern hills are the Mendips, composed of Carboniferous Limestone, stretching from Nunney to the sea and appearing again in the islands of Steep Holme and Flat Holme, which link the structure of Somerset with that of South Wales (*see* also Mendip Hills). The summit of the Mendips is a long tableland between 500 and 1,000 ft. (150 and 300 m.) in height, but rising in the west to just over the latter figure. To the north they die away gently, as a number of low hills, toward the Avon. There the Limestone is covered in places by Coal Measures, but most of the rocks are Triassic. Southward the Mendip Hills drop steeply in an abrupt line broken by many coombs; *e.g.*, the gorge of Cheddar. The basin to the south is composed mainly of Triassic rocks, which, near the sea and along the valleys, are covered by recent alluvium. The basin is usually lower in its western than in its eastern part, which is known generally as Sedgemoor. The large basin is subdivided into those of the Parrett and the Brue by the Polden Hills, which run parallel to the Mendips from Butleigh to Puriton. To the west of the Parrett rise the Quantock Hills, which are outliers of the Devonian moorlands of Exmoor and the Brendon Hills. These three hill groups consist almost entirely of Devonian rocks, and their highest points respectively are Will's Neck (1,261 ft.; 384 m.), Dunkery Beacon (1,706 ft.; 520 m.), and Lype Hill (1,391 ft.; 424 m.). Exmoor (*q.v.*) forest covers the western extremity of Somerset and a large tract of Devon adjoining it. From Crewkerne along the southern and eastern borders of the county as far as the Avon runs a more or less continuous line of low Jurassic hills, while around Chard in the south there is a fair extent of Cretaceous rocks.

The climate of Somerset is mild and equable except on the high ground of the western hills and Mendips. The annual rainfall averages 34 in. (864 mm.) but there is a wide variation within the county from over 72 in. (1,829 mm.) on Exmoor to 25 in. (635 mm.) in the Taunton-Langport area.

The diversity of the geological formations in Somerset gives the county an unusual variety of landscape with dramatic changes of scenery; it also gives rise to a corresponding variety of soil fertility and agricultural pattern, from the market garden areas near South Petherton, Bridgwater, and Cheddar to the rough grazings of the highlands of Exmoor and the Quantocks.

The flat alluvial levels and peat moors of central Somerset support few trees but over 1,000 ac. (405 ha.) produce withies for the manufacture of baskets. The rest of the county is well wooded and the majority of the woodlands are small and scattered and predominantly composed of the hardwoods oak, ash, and beech. The Forestry Commission owns a total of 12,738 ac. (5,155 ha.) in Somerset (1963) and approximately 44,000 ac. (18,000 ha.) are privately owned. There is an increasing interest in the county's woodland which is attracting new investment to finance the replanting of woods.

Considerable areas of Somerset are only sparsely inhabited and wildlife flourishes. Wild red deer are still found on Exmoor and the Quantocks; also a peculiar breed of ponies, hardy and small. Badgers and otters are not uncommon in many areas. Stag and fox hunting is popular. There is a limited amount of sea fishing in the Bristol Channel and salmon and trout are taken in the rivers and reservoirs. Steep Holme, an island opposite Weston-super-Mare, is a nature reserve and bird sanctuary where the wild peony is well established. Sharpham Moor and Stert Island are also nature reserves and long stretches of the coast in Bridgwater Bay are protected areas for birds. Exmoor is a national park (created 1954), and the Quantocks were designated an area of outstanding natural beauty in 1956. The National Trust owned 16,074 ac. (6,505 ha.) and protected 41 ac. (17 ha.) in 1963.

History.—In early postglacial times the lowlands were morasses and the claylands forested, and so man, when he came to the district, settled on the open heights—the less fertile old rocks of the west, the chalk of the south, the limestone of the Mendips and the oolites on the east—and the caves of Mendip have yielded valuable evidence of early prehistoric cultures. In the Mendips are the Cheddar caves and Wookey Hole. Investigations have indicated the existence of Mesolithic sites on the moors and in the western hills. Excavations of tumuli and hilltop earthworks have produced evidence of occupation during the Bronze and early Iron ages, but the most remarkable settlement of the pre-Roman period was that of the lake-village at Glastonbury (*q.v.*).

The Romans overran Somerset after the Claudian conquest of A.D. 43, and remains of the period of the Roman occupation are numerous, particularly between the Parrett and the Avon. Bath (Aquae Sulis), which was probably a settlement in earlier times, became, on account of the medicinal properties of its waters, an important Roman centre where the Fosse Way from Cirencester met another road from Silchester. The Fosse Way was continued from Bath to near Seaton, Devon, and on it was a station at Ilchester (Lindinae) where it crossed the Yeo. Remains of about 50 Roman villas or related structures were discovered in the county. Charterhouse in the Mendips was a centre of the Roman lead mining industry. In the 6th century Somerset was the debatable borderland between the Welsh and Saxons, the latter of whom pushed their way slowly westward. By 658 it had been conquered by the West Saxons as far as the Parrett, and there followed a struggle between the kingdoms of Wessex and Mercia, which led to the organization of the lands east of the Parrett as part of the kingdom of Wessex. About this time the monastery of Glastonbury was restored by Ine who completed the conquest of west Somerset. In the 7th century Somerset, as part of the kingdom of Wessex, was included in the diocese of Winchester. The new bishopric of Sherborne, founded in 705, contained the county until 909, when the see was divided into the dioceses of Salisbury, Exeter, and Wells, the latter including the whole of Somerset. The diocese was divided into three archdeaconries. Disputes later between the chapters of Bath and Wells as to the election of the bishop led to a compromise in 1245, the election being by the chapters jointly, and the see being known as the bishopric of Bath and Wells.

King Alfred's victory in 878, followed by the Peace of Wedmore, ended the incursions of the Danes for a time, but 100 years later they were again a great danger and made frequent raids on the west coast of Somerset. At the Conquest Somerset was divided into about 700 fiefs, held almost entirely by the Normans; the king's lands were of great extent and importance and the bishop of Winchester owned a vast property of which Taunton was the centre. In the 11th century or a little later many Norman castles were built, some of which have survived (*e.g.*, Dunster, Taunton).

The chief families of the county in the Middle Ages were those of De Mohun, Malet, Revel, De Courci, Montacute, Beauchamp and Beaufort, who bore the titles of earls or dukes of Somerset from 1397 to 1472. Edward Seymour was made duke of Somerset in 1547, and in 1660 the title was restored to the Seymour family by whom it was still held in modern times. The marquess of Bath is the male representative of the Thynne family, which has long been settled in the county.

Somerset was too distant and isolated to take much share in the early baronial rebellions or in the Wars of the Roses, and was really without political history until the end of the Middle Ages. In 1547 and 1549 there were rebellions against enclosures. Somerset took a considerable part in the Civil War, and with the exception of Taunton was Royalist, the strongholds being garrisoned and held for the king, but they all fell in 1645, and the county was subdued by the Parliamentary forces. It was the theatre of Monmouth's Rebellion, and he was proclaimed king at Taunton in 1685. The Battle of Sedgemoor on July 6 was followed in the autumn by the "bloody assizes" held by Judge George Jeffreys, in Taunton Castle and at Wells.

Examples of medieval ecclesiastical architecture include the Cathedral Church of Wells, Bath Abbey, and the ruins of Glastonbury Abbey, but Somerset's greatest wealth lies in its parish churches. Of these some have Norman and Early English features, but the most interesting, especially for their magnificent towers, are those in the Perpendicular style, of which the outstanding examples are at Taunton, Glastonbury, Huish Episcopi, Leigh upon Mendip, and Kingsbury. Dunster Castle, which dates from 1070, has been in the Luttrell family since 1376. Fine examples of domestic architecture of the medieval and Tudor periods are at Lytes Cary, Meare, Martock, Montacute, Barrington, and Brympton; the imposing 18th-century architecture of Bath found echoes in simpler work in other towns in the county.

The county was represented in the parliament of 1290; in 1295 it was represented by two knights, and 12 boroughs returned two burgesses each. It continued to return two members until 1832, when it was divided into Somerset East and Somerset West, each division returning two members. Two additional members were returned after 1867 for a third (the Mid-Somerset) division until by the act of 1885 the whole county was divided into seven.

Population and Administration.—The administrative area of the county is 1,603.7 sq.mi. (4,152 sq.km.). Taunton (pop. [1961] 35,192) is the county town and a municipal borough. Bath (80,901) is a city and county borough. The other municipal boroughs are Bridgwater (25,600), Chard (5,779), Wells (6,715), Weston-super-Mare (43,938), and Yeovil (24,598). There are 13 urban and 16 rural districts. The county is in the western circuit, assizes are held at Taunton and Wells, and it has one court of quarter sessions. The boroughs of Bath and Bridgwater have separate courts of quarter sessions. Somerset is in the diocese of Bath and Wells, excepting small parts in the diocese of Bristol.

Besides the county borough of Bath, there are six parliamentary divisions, each returning one member; the county includes part of the parliamentary borough of Bristol.

The Economy.—The main industry is agriculture. Other industries centre on consumer goods in food and drink, clothing, textiles, boots and shoes, paper and printing, furniture and woodworking, but there are no great concentrations of industry. The years of World War II strengthened the engineering and metal-using industries in the southeastern part of the county and new industries including the manufacture of chemicals and scientific instruments were introduced.

Somerset is the leading dairying county of England. Permanent grassland predominates, especially on the lower grounds, but there are areas where arable farming reaches a very high standard. Market gardening and fruit growing flourish in certain favourable districts. Cider making has been practised for centuries, and Cheddar has given its name to a famous cheese. Stock raising is mainly concentrated on the western hills, Exmoor, the Brendons, and Quantocks being known for their calves raised for beef. Sheep

breeding is similarly concentrated mainly in the west.

The chief mineral product is limestone from the Mendips. Peat is found in the Glastonbury area, clay for tiles and brickmaking near Bridgwater, and fuller's earth near Bath. There is coal mining in the Radstock area.

Traditional manufactures include gloves at Yeovil and Wincanton, lace at Chard, linen at Crewkerne, silk at Taunton and Shepton Mallet, and serge at Wellington. A new development in a remote rural setting was the construction (1957–65) of a nuclear power station at Hinkley Point, about 10 mi. (16 km.) NW of Bridgwater on the north coast of Somerset.

The holiday and tourist industry is economically important not only at the seaside towns such as Weston-super-Mare, Burnham-on-Sea and Minehead, but also inland at Taunton, Wells, Cheddar, on the Quantocks, and on Exmoor.

Somerset is served by the Western region of British railways. There is direct connection with London (Paddington) via Westbury and Reading, with the Midlands and the north via Bristol; and with the west. An air service between Weston-super-Mare and Cardiff supplements the summer steamship services. Harbours on the lower Parrett include those at Combwich for supplies for the nuclear power station at Hinkley Point and at Dunball for petroleum from South Wales refineries. At Watchet, coal and raw materials for the paper mills are imported.

BIBLIOGRAPHY.—J. Collinson, *The History and Antiquities of the County of Somerset*, 3 vol. (1791); *Victoria County History of Somerset*, 2 vol. (1906–11); *Proceedings of the Somersetshire Archaeological and Natural History Society* (1849–); A. K. Wickham, *Churches of Somerset* (1952); Land Utilisation Survey of Britain, *The Land of Britain*, no. 86, *Somerset*, by T. Stuart-Menteath *et al.* (1938); G. W. and J. H. Wade, *Somerset*, "The Little Guides" (1949); C. H. D. Cullingford (ed.), *British Caving* (1953); British Association for the Advancement of Science, *Bristol and Its Adjoining Counties* (1955); N. Pevsner, *South and West Somerset, North Somerset and Bristol*, "Buildings of England Series," nos. 13 and 14 (1958).

(T. J. HT.)

SOMERVILE, WILLIAM (1675–1742), English author best known for his poems on hunting, was born Sept. 2, 1675, at Colwich, Staffordshire. He was educated first at Winchester, whence he proceeded in 1694 to New College, Oxford, where he was a fellow. He was admitted at the Middle Temple, but retained his New College fellowship until 1705, when upon his father's death he resigned it and settled at Edstone, near Henley-in-Arden, Warwickshire. He remained there until his death on July 17, 1742.

Somervile lived the life of a country gentleman, serving with praise as justice, and indulging in the field sports that were to make up the subject matter of his best-known poems. His best-known work, *The Chase* (1735), traces the history of hunting up to the Norman Conquest, and instructs on kennel design, hare hunting, stag hunting, the breeding and training of puppies, dog diseases and bites, with frequent digressions including one on Oriental hunting. *The Chase* went through 11 editions in the 18th century; the 1796 edition with a critical essay by John Aikin was illustrated by Thomas Bewick. *Field Sports* (1742), on hawking, supplements *The Chase*. He also published *The Two Springs* (1725), a fable; *Occasional Poems, Translations, Fables, Tales* (1727), which include appreciations of Pope and Addison and some lines to James Thomson with whose *The Seasons* his *The Chase* has some affinities; and *Hobbinol, or the Rural Games* (1740), a burlesque of rustic May games at Evesham. Somervile enjoyed popularity with literate sportsmen into the 19th century and was included in Samuel Johnson's *Lives of the Poets* (1781).

See A. H. Higginson (ed.), *The Chase* (1929).

SOMERVILLE, a city of Middlesex County, Mass., U.S., on the Mystic River, is surrounded by Boston (Charlestown), Cambridge, Arlington, and Medford. The last of the Mystic Valley towns to separate from the early Charlestown colony, it was originally known as the Cow Commons and was entirely fenced in until 1685. In the city stands the Old Powder House, from which British Gen. Thomas Gage seized 250 half-barrels of gunpowder in 1774. In 1775 it became the magazine of the American forces besieging Boston. After the American Revolution brickmaking grew to be an important industry, and the opening of the Middlesex Canal (1803) which passed through Somerville added to its desirability as an industrial site. It was incorporated as a town in 1842 and became a city in 1871.

The greatest growth in population and industrial activity occurred after 1900, when slaughtering and packing became the principal industries. By mid-20th century Somerville had become a working and middle-class residential suburb of Boston with substantial industrial development and importance as a distribution centre. Pop. (1960) 94,697; for comparative population figures *see* table in MASSACHUSETTS: *Population*. (L. G. BA.)

SOMERVILLE, a borough of north-central New Jersey, U.S., on the Raritan River, is located 36 mi. SW of New York City; the seat of Somerset County. It is a retail-trade centre for a large area of farms, manufacturing plants and housing developments. First settled in the late 17th century by farmers, it became the county seat when the county was established in 1783. It was incorporated as a town in 1863 and as a borough in 1909. In the 1750s the Rev. John Frelinghuysen founded the first seminary of the Dutch Reformed Church in America. From these humble origins came Queens College, chartered in 1766 and now Rutgers—the State University (*see* NEW JERSEY: *Education*), and the New Brunswick Theological seminary. The Old Dutch parsonage, where Frelinghuysen lived, is open to the public, as is the Wallace House, used as headquarters by George Washington while his troops were quartered at Camp Middlebrook. The Duke estate, established in 1893 by James B. Duke, the tobacco magnate, is a research and exhibition centre for the New York Horticultural Society. For comparative population figures *see* table in NEW JERSEY: *Population*. (E. R. D.)

SOMERVILLE AND ROSS: *see* MARTIN, VIOLET FLORENCE.

SOMME, a *département* of northern France, formed in 1790 from a large part of the province of Picardy, including the districts of Ponthieu, Vimeu, Amiénois, Santerre, and Vermandois, and also from a neighbouring small portion of Artois. It is bounded north by the *département* of Pas-de-Calais, east by Aisne, south by Oise, and southwest by Seine-Maritime. To the west it fronts the English Channel for about 30 mi. (48 km.). Area, 2,384 sq.mi. (6,175 sq.km.). Pop. (1962) 488,225.

The *département* consists largely of tracts of hedgeless, arable land, typical of the rolling chalk country that occupies the northern part of the Paris Basin. The flat, alluvial-floored trench occupied by the Somme River (*q.v.*) is the largest of the valleys, aligned southeast-northwest, that groove these dry chalk plateaus. It forms the axis of the *département*, which stretches north to the Authie River and southwest to the Bresle River. The Somme, in its marshy valley, forms a historic line of defense across the approach to Paris from the Low Countries, and the *département* was the scene of long and heavy fighting during World War I. Many war cemeteries in the countryside mark its battlefields.

Much of the chalk upland is covered with limon, rendering it very fertile. Sugar beet is especially a crop of the limon soil, cultivated in association with wheat. Wheat is everywhere the pivot of the crop rotations, supported by other cereal and fodder crops. Some chicory, tobacco, and flax are grown, but the *département* lies wholly north of the limit of viticulture. In the coastal districts there is more pasture for grazing, but the arable farms everywhere carry a high density of stall-fed cattle. Fallows have been replaced by rotation crops and sheep have therefore greatly diminished in number.

The longshore drift sweeps sediment against the coast as it trends north from Le Tréport, and the Somme estuary suffers from severe silting. Between the estuaries the chalk cliffs of the former coastline are now separated from the sea by a strip of marshland behind the piled up sand dunes. This marshland, now drained by dikes and converted into market gardens and meadows, is the Bas Champs of Marquenterre. Market gardens also occupy the peaty floors of the river valleys. These are areas of small holdings, but on the chalk upland holdings of medium size prevail and are farmed by their owners. The farms, typically consisting of a substantial group of buildings surrounding an open courtyard, are clustered in villages. Rural handicrafts, once widely practised,

included locksmithery in Vimeu and hosiery in Santerre. These handicrafts have declined since the early 19th century. In the market towns textile industries have survived, especially at Amiens, which manufactures velours and heavy woolens. There are also engineering works, sugar refineries, breweries, distilleries, and flourmills.

Amiens (*q.v.*), the ancient capital of Picardy (*q.v.*), situated at the main crossing of the Somme valley, has become the chief focus of rail and road routes. It is the largest town and the *préfecture* of the *département*, which is divided into four *arrondissements*, centred upon Amiens, Abbeville, Montdidier, and Péronne. Abbeville, at the head of the Somme estuary, with the lowest bridge, is another important route centre. Poix, Roye, Ham, Albert, and Doullens are the other market towns. The *département* forms the diocese of Amiens, whose great Gothic cathedral of Notre Dame is the largest in France. Amiens also has the court of appeal, but the *département* comes under the *académie* at Lille for the administration of education. Few old buildings survived the devastation of two world wars, and the towns have been largely rebuilt. (Ar. E. S.)

SOMME RIVER, in northern France, 152 mi. (245 km.) long, flows northwestward to the English Channel across the *département* of Somme to which it gives its name, and the ancient province of Picardy. Its headstreams, including the Avre, from the rich arable country of Vermandois and Santerre, converge upon Amiens, whence the Somme follows a braided course along the flat floor of a trench cut across the chalk country. Its valley, floored with alluvium, is a marshy belt, and its line has been an important barrier in the approach from Flanders toward Paris, and as such has been of great strategic importance. The upper basin of the Somme was the scene of heavy fighting during World War I. The valley, which has been extensively reclaimed, is now occupied by meadows and market gardens (*hortillonnages*) served by a network of drainage canals; the flanking chalk uplands, with *limon* soils, are the domain of wheat and sugar beet. Below its lowest bridge, at Abbeville, the Somme enters a sand-encumbered estuary, with the little port of Saint-Valery at its mouth. The river has been canalized, and from the upper valley there are canals to the Oise (Canal de Somme) and the Escaut or Scheldt (Canal de Saint-Quentin) connecting it with the navigable waterways that link Paris and Flanders. (Ar. E. S.)

SOMNAMBULISM: *see* Hypnosis; Hysteria.

SOMNATH (Patan Somnath, Somnatha-Patan), an ancient ruined city on the southern coast of Gujarat, India, 110 mi. (177 km.) S of Rajkot. It is the site of the famous temple of Siva as Somanatha, "lord of the soma," a sacred intoxicating plant, and, by extension, "lord of the moon." The temple was sacked and its linga or phallic emblem broken into pieces by Mahmud of Ghazni in A.D. 1024–25. Reconstructed for the second time in 1169, the temple was destroyed in the final Muslim invasions of the late 13th century. It was subsequently rebuilt and destroyed on several occasions; the last reconstruction was begun in 1951. According to an ancient tradition in the Mahabharata Somnath was the scene of the internecine massacre of the Yadava clan and of the subsequent death of Krishna (*q.v.*). Recent excavations there have revealed a settlement going back to *c.* 1500 B.C. Patan (pop. [1961] 13,927), a port on the old city site, is now overshadowed by the adjacent port of Veraval (pop. [1961] 46,637).

(F. R. A.)

SON, a river of northeastern central India, 487 mi. (784 km.) long, flowing in general east-northeast. It is the chief tributary received by the Ganges from the south. Although the valley of the Son is almost a continuation of that of the Narmada (Narbada) and geologists refer to the "Narmada-Son trough," unlike the Narmada it does not serve as a main line of communication. The valley is largely forested—especially with sal (*Shorea*)—and sparsely populated, and has only occasional small patches of cultivation. To the north of the valley is the Kaimur Range; on the south the river collects the drainage of the northern parts of the Chota Nagpur Plateau. There rainfall is restricted to the rainy season (June–September); for much of the year the Son carries little water. The river cuts through the eastern end of the Kaimur Range to flow northeastward across the plain to enter the Ganges 10 mi. above Patna. An outcrop of limestone has permitted the development of a cement works with coal from Daltonganj. Dams have been constructed on some of the tributaries and where the river disgorges onto the plain is the important weir at Dehri, from which 600,000 ac. (240,000 ha.) are irrigated. (L. D. S.)

SONATA, in music (Italian, "sounded," *i.e.*, by instruments as opposed to *cantata*, "sung"), a term first used to describe a specific instrumental work by Giacomo Gorzanis, who, in his lute book of 1561, applied it to a suite of dances. At the beginning of the 17th century the term was associated with an instrumental composition, closely allied to the fantasia (*q.v.*). Loosely fugal in texture, and sometimes with brief sections in contrasting tempos or rhythms, it was an instrumental counterpart to the contemporary madrigal. An example of this form is Giovanni Gabrieli's *Canzoni et sonate . . . per sonare con ogni sorte de instrumenti* ("to be played on all kinds of instruments"), published in Venice, 1615.

Subsequent uses of the term sonata have varied so greatly that the different forms covered by the term must be dealt with separately. It has been customary, however, to confine the term to small groups of instruments playing not more than one to a part; and to regard as sonatas only those works in more than one movement—though the latter qualification does not always hold good for works of earlier periods (*e.g.*, Scarlatti's sonatas for harpsichord; *see* below, *The Transition to True Sonata Form*).

THE BAROQUE SONATA

In Italy at the beginning of the 17th century the fantasia developed a tendency to be built up in brief sections, and it was these sections which were made slightly longer and more distinct to form sonatas of several movements. Some of these sonata movements continued to be given the loosely fugal texture characteristic of the fantasia; others were in the form of dances such as the saraband or the coranto (*see* also Dance Forms in Music). Such dances were familiar in the court balls and other fashionable entertainments of the period, to which the newly developed family of violins was regarded as a particularly suitable accompaniment because of their brilliant tone and assertive articulation. The sonata, like the violin family itself, was first cultivated by the Italian musicians, and developments both of the form and of the instruments were interdependent.

The old equal-voiced polyphony of the 16th-century fantasia continued to be developed by English composers of the early 17th century, whose idiom was thoroughly instrumental, into a magnificent school of contrapuntal chamber music for viols (a family of instruments whose lighter tone and more transparent colouring enable them to perform their own elaborate counterpoint with impressive clarity). Elsewhere—and above all in Italy—the trend in music passed from equal-voiced polyphony in the direction of solo melody, to which the viols were less suited than the violins. Here is one of the distinguishing characteristics of the sonata as opposed to the fantasia. Although the texture of the Baroque sonata remained essentially contrapuntal, and although some of its movements (especially the second) tended to be fugal, there is no equal-voice polyphony in the 16th-century sense; the emphasis is on solo melody.

This solo melody is ordinarily supported by a bass line which is almost, if not quite, as melodic in quality as the main melody itself. A further development greatly favoured as the 17th century progressed consisted in writing two main melodies at about the same level of pitch; these continually cross and recross in the treble register until it is hard to say which, if either, takes primacy. It is as if the solo melody had divided into a pair of copartners still supported at some distance below by the strongly melodic bass line. The harmony in between is not filled out with independent parts as it is in the fantasia, that is to say built up out of equal-voiced polyphony. On the contrary, it is left to the continuo player to improvise upon the bass line, which, with a few figures to inform him of the harmony, is all that the composer will normally have given him to work upon (*see* Thorough Bass). There could be no clearer indication of the new inclination to ex-

ploit melody at the expense of equal-voiced polyphony.

These dual-melody sonatas became known as trio sonatas. Two treble melodies and one bass melody add up to three; and though a keyboard, lute, or other continuo accompaniment was also normally used, the work was regarded as complete without this fourth instrument. The music can be performed without continuo, though this leaves it less rich and brilliant. When a continuo instrument was used, it was possible to dispense with a melody instrument playing the bass, but this, too, was felt to be an undesirable impoverishment. Solo sonatas for violin or other instruments were also developed through the 17th century, but the trio sonata became the standard form of serious chamber music, and held a position corresponding to that of the string quartet from the time of Haydn onward. Keyboard sonatas appeared intermittently in the 17th century, notably those of Johann Kunhau (1660–1722).

The sonata as understood in the Baroque period owes most to the famous violinist-composers of the Italian school. These were successfully active very early in the 17th century, but the school reached a peak in the solo and trio sonatas of Arcangelo Corelli (1653–1713). In the same generation, Henry Purcell (1659–95), who had already composed the last great English fantasias, became acquainted with the Italian idiom, and produced the first English sonatas of enduring significance. These, for two violins and continuo and for three violins and continuo, have not the classic perfection of form of the Corelli sonatas; but the harmonic interest is greater in Purcell's music, and the feeling is warmer and more heart-searching. These features are partly due to Purcell's individual genius, and partly to the singular intensity of the English musical tradition which he inherited.

The Italian sonata as perfected by Corelli took two slightly different forms: one designed for musical performance in church; the other for secular occasions. The former variety was the *sonata da chiesa* ("church sonata"), the latter the *sonata de camera* ("chamber sonata"). Dance movements were favoured in chamber sonatas, but not in church sonatas, the style of which might be a little more solemn; in other respects there is no difference which strikes the modern ear.

A further distinction greatly stressed at the time, but which has become less noticeable to modern ears, is that between the Italian and French styles. This distinction provoked much controversy in the second half of the 17th century and most of the 18th. The Italian style brought the melody into prominence and employed rich harmonies characterized by suspensions. A flowing tempo was frequently desirable, the parts were often to be quiet-moving and the movements, normally without a descriptive title, were designed to follow each other in an established order. The French instrumental style was notable for precise, clearly marked rhythms, short phrases, and, in the slow movements, elegant ornamentation and poignant sentiment. The movements, often with literary or pictorial associations, were assembled in many different ways. The Italian style was regarded as brilliant by its admirers, but as intemperate by its detractors; the French style was regarded as tenderly expressive by its admirers, but as stiff and insipid by its detractors. The genius of François Couperin (1668–1733) rose above this embittered quarrel; he not only composed in both national styles, but combined them in a series of splendid trio sonatas which are to the French tradition what Purcell's are to the English.

The Germans were early followers of the Italians and produced solo violin sonatas and trio sonatas during the 17th century, as well as sonatas in which the viola da gamba takes a middle part. G. P. Telemann (1681–1767), G. F. Handel (1685–1759), and others composed many sonatas of value. The greatest Baroque sonatas come from J. S. Bach, particularly those for unaccompanied violin (for which there are German precedents) or violoncello, and for violin with harpsichord obbligato or viola da gamba with harpsichord obbligato; he also wrote a small number of trio sonatas. His writing for the violin owed something to Antonio Vivaldi, the greatest of all the 18th-century Italian violinist-composers, whom he greatly admired.

Though these Baroque works were called sonatas, and are still rightly known by this name, they can be more accurately described as suites rather than as sonatas in the later sense of the term. Moreover, there is no essential difference between the Baroque sonata and the Baroque suite: they are basically the same form, for which other names, such as *partita* or *ordre*, were also freely in use.

The principle is to contrast a succession of movements each of which is more or less homogeneous in itself. The contrast may be of tempo, of rhythm, of major or minor mode, or of general style; an underlying unity is provided by retaining wholly or mainly the same key, occasionally varied by its near relatives. The degree of internal contrast within the movement varies, but it was never the determining element in the construction of the sonata, as it became from the time of Haydn onward.

The Transition to True Sonata Form.—The transition from the Baroque sonata, which is in effect a suite, to the classical sonata, which is in true sonata form, occupied the middle of the 18th century. An outstanding pioneer was Domenico Scarlatti (1685–1757). His father, Alessandro, was in the front rank of Italian opera composers in his day; the son, Domenico, was in advance of his. Though born in the same year as Handel and J. S. Bach, Domenico Scarlatti did much to prepare the change of style that was to follow those two last geniuses of the Baroque period. Most of his sonatas for harpsichord are untypical in one respect: they are essentially single-movement works, two of which, however, were frequently paired in performance. Within the single movement they make a deliberate and effective use of internal contrast, and it is this feature which became decisive in sonata form as understood from Haydn onward. Domenico Scarlatti combined astonishing brilliance with a remarkable depth of poetry, and his sonatas, though best classified as transitional from the formal viewpoint, are masterpieces in their own right.

J. S. Bach's sons continued the development away from the Baroque toward the subsequent conception of the sonata. In particular C. P. E. Bach (1714–88) and J. C. Bach (1735–82) contributed to the use of contrasted keys, not only between movements, but still more significantly within the movement. It is here that the history of true sonata form begins.

TRUE SONATA FORM

The term sonata form in its stricter modern use denotes both (1) the organization of tonal material within the movement and (2) the organization of the relationship between movements.

Material Within the Movement.—The main divisions within a movement in sonata form are: (1) the exposition; (2) the development; and (3) the recapitulation.

The Exposition.—The tonal material is not single, *i.e.*, of more or less uniform character (rhythm, tonality, etc.) as in a characteristic suite movement; it is dual, *i.e.*, it includes an inherent element of contrast. Its themes, though not necessarily two in number, present a conflict of mood which tends to divide the material into two fields of force. The inherent tension between these two fields is deliberately heightened by a conflict of key: the opening material is in the tonic or "home" key; the contrasting material is in a contrasting key, which, though normally in a close relationship to the "home" key, is sufficiently distinct to mark a certain distance from the starting point. The listener feels a sense of opposites: very often the familiar human opposition of a masculine with a feminine mood, or of an energetic with a reflective temperament, or of strength with tenderness. The establishment of this opposition is known as the exposition; it may have been preceded by an introduction. The exposition is commonly marked with repeat signs, but these are optional. Some expositions seem to gain from being heard twice, others not; the modern tendency is to omit the repeat.

The Development.—In the development, any portion of the material so far heard, and in many cases new, or seemingly new, material may appear in any order and in any combination. Frequently the material is broken down into its constituent fragments, and these fragments are themselves combined and developed. Opposing moods are illustrated and seen to interact. Tonality becomes as fluctuating as the argument. The key is in constant modulation; if it settles, it is likely to be in a tonality remote from either of

those established in the exposition. Toward the end of the development section a sense of premonition is introduced which is fulfilled by a return of the "home" key, and with it the original opening material.

The Recapitulation.—What now follows is the recapitulation. Both the opposites which set up their contrasting tonalities and their contrasting fields of force in the exposition are heard again and in the same order. They may show considerable differences on this second appearance; there may be felicitous new touches showing how little a second appearance need be mechanical; but the opposites are still recognizably and often literally the same. Their effect on the listener, however, is not the same. He hears them with new ears because he has lived through the revealing experience of the development. He is prepared, in consequence, to find their initial opposition reconciled, as opposing principles and people can become reconciled by being brought into relationship. This sense of inner reconciliation is confirmed, like the initial opposition, by the tonality. Instead of appearing in a key or keys contrasted with the "home" key, the opposing material now appears in the same, or mainly in the same, key as the opening material. What began as a conflict in the exposition proceeds as dialectic in the development and ends as synthesis in the recapitulation. The tension of the opposites is not so much relaxed as brought into balance. A coda may then follow, sometimes of such elaborate proportions as almost to amount to a further development; but whether this is so or not, a sonata-form movement is complete if, and only if, it ends with reconciliation.

A form is not a mold to be filled by a composer pouring suitable material into it; a vital form is based on the principle of organic growth. This principle is illustrated in many different ways. Mozart's sonata-form movements show different inclinations from Haydn's, and Beethoven's from Mozart's. But they all show the same principle based primarily on a conflict of mood and tonality which leads through development and interplay to reconciliation. It is a drama so fundamental to human experience that it may be readily seen why the expression of it in music produced a form of such enduring conviction. This reached its perfection in the three classical masters just mentioned, but it served some of the Romantics of the later 19th century equally well.

Relationship Between Movements.—The unified contrasts in the first aspect of sonata form (the organized exposition, development, and recapitulation of tonal material within the movement) continue into its second aspect (the organized relationship between the movements). The "sonata-form movement" as described above is the first movement of the sonata at its most characteristic: so much so that it is alternatively described as "first-movement form." This, however, is not a desirable alternative, since the first movement is not invariably in sonata form; and later movements quite often are. These normally comprise a slow movement in the middle, and a quick movement at the end, with the optional insertion of a lighter movement either before or after the slow movement. In the sonata as developed by Haydn the keys in which the separate movements are composed are contrasted with the same attention to dramatic effect as the key sequences within the movement.

This was not a new principle in music; but it was pressed to new lengths. Whereas a Baroque suite (even when passing under the name of a sonata) or indeed a late Baroque opera, with its suites of arias, is a series of arresting scenes, a sonata by Beethoven or an opera by Wagner is in the form of a dramatically developed plot. The dramatic element is greater in 19th-century music than in Baroque music; and the sonata is the chief instrumental form which gave strength and coherence to this new sense of drama.

While in the classical sonata there is not usually a conscious thematic connection between the material of one movement and another, some Romantic composers from Schumann onward experimented extensively in the use of transformed material. Liszt transformed his material in such a way that he was able to maintain the original elements of his themes throughout all movements of his sonata forms, but, though his poetic intentions were undeniable, his inspiration did not always sustain his preconceived plan. Deliberate use of the same material to fill out the different movements is alien to the spirit of sonata form, though fleeting references from one movement to another can add to the sense of coherence, and are often moving.

In sonata form, unity is achieved within a movement by relating the tonalities of the contrasting sections to an underlying key. The tonalities of the different movements are also related. The slow movement and the trio are normally contrasted in key with the other movements.

The slow movement does not usually follow the plan of the first-movement sonata form, since its slow-moving tempo would not allow an expanded development without unreasonably prolonging the length. A common form for the slow movement is the "aria" or *Lied* form, ABA, often called song form. This consists of: a section complete in itself; a section contrasting both in mood and tonality; and the first section repeated, often with substantial variation. An alternative, which can be as elaborate as the composer wishes, is variation form: a melody complete in itself; variations upon it; and, sometimes, the plain melody again. Some of Beethoven's slow movements are in variation form (*e.g.*, the piano sonatas op. 109 and op. 111; the string quartets op. 127 and op. 135).

The slow movement may be preceded or followed by a movement the purpose of which is to lighten the mood and relax the tension. In the 18th century the minuet was used for this purpose. This was often followed in suites and divertimentos by a second minuet (perhaps in the minor if the first was in the major), the first minuet then returning without its repeats. As the classical sonata developed, a "trio" became usual in place of the second minuet (the name probably deriving from the convention by which four-part writing was commonly replaced by three-part writing at this point). The convention was not obligatory and later declined. Many trios have an elegiac charm, almost a naïveté, providing relief in an otherwise tense composition. Later, the scherzo (*q.v.*) supplanted the minuet; the scherzo was sometimes followed by a trio, elements of which could be repeated. In all these schemes the function of this movement remains the same: to provide either light relief, or at least relief in the sense of some decisive change of mood (some of Beethoven's scherzos are notoriously ferocious).

The function of the last movement in a sonata is variable. Some works start with a first movement so powerful that the entire sequel becomes a prolonged commentary on it without detracting from its prominence. This is more commonly found in the works of Haydn or Mozart than in those of Beethoven. Another pattern is to set the main weight in the slow movement followed by a slighter but brilliant finale. It is also possible to lead up to a massive last movement to which the others seem no more than an elaborate preparation. A quiet and contemplative ending, again, can sometimes carry profound conviction. Rondo form and variation form are common choices; they are sometimes so handled (especially rondo form) as to contain elements of sonata form, while true sonata form (perhaps somewhat modified) is a further possibility for the last movement of a sonata, particularly if the first movement has been in some other form.

Symphonies, string quartets, and to some extent concertos of the 18th and 19th centuries are commonly, in effect, sonatas. At the beginning of the 20th century the concept which reached maturity with Haydn became transfigured by Sibelius, whose symphonies, though suffused with sonata principles, often grow from small germs, instead of developing self-sufficient material of contrasting elements ultimately reconciled.

The drama of key and modulation is at the basis of sonata form. If it is found to be of less value to 20th-century composers, this is not because they would not wish music to express such fundamental ideas as conflict, growth, and reconciliation, but because, tonality having lost much of its significance. other technical means must be found to express these perennial concepts.

Bibliography.—E. Borrel, *La Sonate* (1951); M. Bukofzer, *Music in the Baroque Era* (1947); W. Mellers, "The Sonata Principle from c. 1750," in *Man and His Music* (1957); W. S. Newman, *The Sonata in the Baroque Era* (1959), *The Sonata in the Classic Era* (1963).

(R. Do)

SONDERBORG: *see* Aabenraa-Sonderborg.

SONDRIO, a town in Lombardy, Italy, and capital of Sondrio Province, lies in the Valtellina at an altitude of 1,017 ft. (310 m.) and on the Adda River, 82 mi. (132 km.) NE of Milan by rail. Pop. (1961) 19,455 (commune). Of interest are the archaeological museum and the Masegra castle (now barracks). The town, well known for its wines, is an agricultural and commercial centre, and has a cotton textile industry. It is connected by rail and road with Switzerland via Tirano and the Bernina Pass, and by road with the Stelvio Pass (9,052 ft. [2,759 m.]) and with Bolzano. Cableways link it with the winter-sports resorts of Madesimo, Bormio, Aprica, Caspoggio, Teglio, Chiesa in Val Malenco, Livigno, and Santa Caterina Valfurva.

SONDRIO PROVINCE, bounded north by Switzerland, comprises the Valtellina and the Liro and lower Mera valleys. Area 1,240 sq.mi. (3,212 sq.km.). Pop. (1961) 153,592. There are extensive vineyards and hydroelectric power is highly developed. (L. B.)

SON ET LUMIÈRE (SOUND AND LIGHT), a nighttime entertainment conceived by Paul Robert-Houdin, curator of the Château de Chambord on the Cosson River, France, where the first one was presented in 1952. Multicoloured lights of changing intensity are directed against the facade of a historic building or ruin. The changes of light are synchronized with a sound track (relayed through loudspeakers) carrying music and the dramatized story of the site. Usually no participants appear. Live effects such as smokebombs or fireworks are occasionally used for appropriate episodes.

The medium rapidly became popular in France, where by the middle 1960s about 50 annual productions took place, notably in the Loire Valley and also at Versailles and Les Invalides. Productions elsewhere on the European continent included Rome (the Forum) and Athens (the Parthenon), both in multilingual versions. The first British production was in 1957 (Greenwich Palace). Subsequent productions included Cardiff Castle, Portsmouth (H.M.S. "Victory"), Henley-on-Thames (Grey's Court), the Tower of London, Stirling Castle and the cathedrals at Gloucester and Norwich. The first presentation in the United States was given at Independence Hall, Philadelphia, Pa., in July 1962. The first African production took place at Cairo (the Giza pyramids) in April 1961; the first Asian production was at Delhi (the Red Fort) in March 1965. (A. A. BU.)

SONG, properly, a piece of music performed by a single voice, with or without instrumental accompaniment; songs for two or more voices are more accurately described as duets, trios, quartets, etc. The music of a song is normally a setting of intelligible words; examples of wordless songs (*vocalises*) occur in the works of 20th-century composers, but they are exceptional. Such a conception of song is possible only in a civilized society. To primitive man words and tune were inseparable, and this is largely true of European and American folk song. Primitive songs are also closely associated with nature and with the supernatural inherent in nature; hence they always have a definite purpose. In European and American folk songs purpose survives in the shape of lullabies, work songs, sea shanties, and the like; but a great many songs do no more than tell a story or sing of the joys and disappointments of love (*see* FOLK MUSIC).

Early European Song.—The records of European song, as distinct from folk song, are scanty before the 12th century. For recreation, monks and students sang songs in Latin, since this was the language of their education. The words of many of these have been preserved, but the tunes were rarely written down, either because they were well known or because the skill to record them was lacking, and possibly in some cases because the composers were anxious to preserve them for themselves. It is to the 12th-century troubadours (*q.v.*) in the south of France that we owe not only what appears to be a new art of lyrical poetry in the vernacular but also a large number of melodies wedded to the poems. This art was imitated by the *trouvères* (*q.v.*) in northern France, who have left a considerably larger corpus of songs with music, and also in Spain, Italy, and what is now Germany and Austria. Whether the melodies were written by the poets themselves or by musicians associated with them is a matter of dispute; but they are all highly organized, in most cases with an exact or modified repetition of the first section, and many of them reveal a subtlety of expression that has not been surpassed by any later composers. Songs of this kind could have been composed only in an aristocratic society where there was sufficient leisure for invention and where expert scribes were available to copy them into the beautiful manuscripts that have survived. No accompaniment is indicated in the manuscripts; if any was used it was presumably improvised.

Polyphony: Solo Voice and Instrumental Accompaniment.—It was inevitable that this flood of melody should have influenced the composers of polyphonic music. The 13th-century motet, even though founded on a fragment of plainsong, shows this influence very clearly; and it becomes even more marked in settings of French words intended for secular use. Furthermore, delight in the sound of the solo voice encouraged 14th-century French and Italian composers to assign all but one of the parts in a polyphonic composition to instruments. In this way a new art of solo song with accompaniment arose and continued to be practised throughout the 15th century. This art was inevitably more elaborate than the melodies of the *trouvères* and in some cases became highly mannered, creating a problem which has faced all songwriters—how to do justice to the words while at the same time allowing the music to develop to its fullest capacity.

Though there was often considerable elaboration in the instrumental parts of these compositions, on the whole they were dominated by the vocal line. But in the course of the 15th century composers of polyphonic music for voices came more and more to integrate their works by the use of imitation, and this practice spread to the composition of solo songs, with the tune sung either by an inner voice (particularly popular in Germany) or by the soprano, accompanied by instrumental parts based on fragments of the melody. At the same time, about the end of the 15th century, there was a reaction against elaborate polyphony in the shape of songs with the simplest possible accompaniment, a large number of which were composed in Italy (under the name *frottola* [*q.v.*]) and also in Spain. In songs of this kind, accompaniment by an instrumental ensemble was hardly necessary and it was often replaced by a single part for the lute. No doubt it was the popularity of these arrangements, many of which were published, that encouraged Spanish composers in the 1530s to write original songs with an accompaniment for the *vihuela* (a flat-backed instrument, tuned and played in the same way as the lute). This accompaniment could be quite simple, but the virtuosity of performers was an inducement to write varied accompaniments which offered a challenge to the accompanist without in any way interfering with the simplicity of the vocal line.

French composers of the 16th century remained faithful to the *chanson* (*q.v.*) for several voices, though many of these were also published in transcriptions for voice and lute. In England the old tradition of solo song with imitative accompaniment for an instrumental ensemble survives in William Byrd's *Psalmes, Sonets, and songs of Sadnes and pietie* (1588), though these were published with words added to the instrumental parts. Other English composers of the period, around the end of the 16th century, wrote original songs with lute accompaniment, some of which were also made available for singing by four voices.

Recitative and Aria.—By this time the Italians had developed a different manner of setting and accompanying words. This arose from the fact that composers of madrigals in Italy (*see* MADRIGAL), unlike their English imitators, had generally set distinguished poetry. The importance of the words in singing is suggested by B. Castiglione's reference in 1528 to *cantare alla viola per recitare; i.e.*, singing a solo song with instrumental accompaniment. In the course of the 16th century there was a growing feeling that the intricacy of the madrigal did less than justice to the words, though composers often avoided complication for short periods by making all the voices sing the same words at the same time. But homophonic writing, as opposed to polyphonic elaboration, did not solve the problem of allowing the words to be "recited" as if they were spoken. Hence toward the end of the 16th century a new style of word setting developed in which a single voice had complete freedom of time and expression, and the accompaniment, the simplest possible, was provided by a continuo instrument, on which

the accompanist improvised the chords indicated by a figured bass (*see* Thorough Bass). This style was known as *stile recitativo* and subsequently simply as *recitativo* or "recitative."

This manner of singing was ideal for dramatic performance and it was adopted in the first operas, at the end of the 16th century. It was not intended to be simply a translation of speech into song; the composers were too near to the madrigal, which continued to be written in the early 17th century, to ignore the possibilities of expression in vocal music. At the same time it was realized that occasional lyrical pieces were appropriate in the pastoral setting of these early operas; and when recitative was employed in the concert room it became obvious that the variety achieved by alternating recitative with song (or *aria*) was indispensable. Though recitative could be intensely moving when performed by a first-rate actor, the purely musical pleasure afforded by the aria was obviously more attractive to the average members of an audience, and it also offered more opportunities to a singer. Hence the aria came more and more to dominate opera and also oratorio; it also became very much extended by the use of instruments other than the continuo to alternate with the voice or to work in close association with it. The need for a significant musical structure on a large scale led to the *da capo* aria in three sections, the third of which was the same as the first, apart from ornamentation or elaboration of the vocal melody, which was left to the singer. The fact that this repetition of the first section sometimes made nonsense of the words does not seem to have unduly disturbed either composers or audiences, though protests against it were heard from time to time in the 18th century. The *da capo* aria, however, was not universal; there are many examples of shorter arias without this repetition.

Strophic Settings.—Outside the world of opera, the simple strophic song, with keyboard accompaniment, continued to thrive—notably in England, France, and Germany. In England it earned a new lease of life through the popularity of *The Beggar's Opera* (1728) and its successors, and in France its introduction into popular comedies in the early 18th century led to the creation of *opéra comique* (*q.v.*), which, unlike Italian *opera buffa*, remained faithful to spoken dialogue. The idea that such songs, as opposed to the operatic style of the solo cantata, could be a medium for artistic expression, was hardly realized in the 18th century. The songs of Haydn and Mozart are for the most part little more than simple ditties, though Mozart's "Das Veilchen" and "Abendempfindung an Laura" pointed to the possibilities of a deeper art of expression. It was Schubert who, beginning by imitating the narrative ballads of his immediate predecessors, showed what could be done to turn even the simplest strophic song into something more than an agreeable entertainment, and in the songs of his last year, such as "Der Doppelgänger," outstripped all his contemporaries in his power of dramatic realization. His example was not lost on the Romantic composers of the 19th century—among them Mendelssohn, Robert Schumann, Franz Liszt, and Brahms—who learned from him not only the art of varying a strophic melody but also the significance that could be given to the accompaniment. For Hugo Wolf the accompaniment often became a piece of music in its own right, allowing the voice to follow a rhythm dictated by the words and to achieve its own expressive melody without being fettered to the piano.

Problems of Songwriting.—Since that time there have been few composers of any country who have not at one time or another felt impelled to face the challenge of songwriting. The nature of the challenge is obvious enough. Inspired music, as Schubert showed, can often redeem inferior words; at the same time fine poetry may appear so self-sufficient that to add music to it is to spoil it. There is also the problem that the rhythm of spoken verse is normally more subtle than that of music, however ingeniously the composer may try to match it; and the more ingeniously he tries, the greater is the danger that music may abrogate its rights and become merely a vehicle for the poem. Italian recitative of the early 17th century solved the problem by allowing the singer complete freedom, adding only the simplest accompaniment. But however effective this may be in an opera, it risks undermining the musical structure of an independent piece.

In the long run there is no ideal solution. A continuous texture in the accompaniment may serve to unify the music, but it may also deny the singer the opportunity to make the words sound convincing as music. At first sight a strophic setting seems almost like an avoidance of the problem and may easily be regarded as much the easiest way of writing a song. In fact it is one of the hardest forms of songwriting, if it is to achieve anything like fidelity to the words. There are, however, ways of modifying a simple strophic scheme by varying the melody or the key or both; in exploring these possibilities Schubert showed incomparable skill.

The influence of language on song is not unimportant, particularly as songwriting in its turn often influences instrumental composition. The *da capo* aria grew out of the Italian language, for which it was the ideal medium. When the same style was borrowed by German composers the result was apt to be less successful. All French songwriting has a character of its own, since the language, unlike Italian, German, and English, has no regular tonic accent. This means that syllables can be set equally well on accented or unaccented beats; it also results in a peculiar fluidity which is absent from the regular accents of German song. It would seem likely that this fluidity is at least partly responsible for the shifting, kaleidoscopic harmonies which are one of the features of Gabriel Fauré's songwriting. In the same way the dramatic realism of Mussorgsky's songs is not merely an expression of his own temperament but also a consequence of the actual sounds of the Russian language. Something similar is to be found in Czech songwriting, where the lightweight syllables of the language suggest a similar treatment in the music.

Whatever the origins of song may be in remote antiquity—as a means of attracting attention, or telling a story, or providing encouragement in war or the chase—in civilized society it has developed as a natural response to the beauty of the human voice. Hence the melodic style of a song is liable to sound distorted if the capacity of the voice for phrasing and for blending syllables is ignored. Wide leaps are within the capacity of any trained singer and have often been used by composers of many different periods in the interests of dramatic intensity. But a vocal line which consists almost entirely of leaps can hardly be described as vocal and is certainly not a line. There is a tendency among mid-20th-century composers to treat the voice as if it were an instrument capable of negotiating a leap and still preserving its characteristic tone. A song of this kind, however dramatically it may express the words, is often ineffective because the singer has no opportunity to use his greatest asset—beauty of tone. Few composers have also been solo singers, but it is open to any composer to acquire an intimate acquaintance with the capacity and the sound of the human voice. The most successful modern songwriters have been those who either had such an acquaintance or were gifted with the instinct that made it unnecessary.

See also Singing and references under "Song" in the Index.

(Ja. A. W.)

SONGBIRD. Besides the application to any bird that sings, the term songbird has a restricted, precise meaning in bird classification. The true songbirds comprise the more than 50 families of the suborder Passeres or Oscines, which includes most of the great group of perching birds called the Order Passeriformes. They are well represented by many common garden birds such as sparrows, warblers, thrushes and wrens.

General Description and Natural History.—These birds vary in size from tiny kinglets and sunbirds to comparatively large crows. They are land birds that live in a wide variety of situations, from open grassland to forest. While songbirds include some of our best songsters such as thrushes, some have harsh voices like crows, and some do little or no singing. Their main food includes insects, worms, seeds and fruit picked up from the ground or from among vegetation, depending on the species. However, some like the swallows are very capable flyers and catch insects on the wing. Others, as the dippers, seek their food underwater in streams, and shrikes may catch mice and small birds.

In many species the birds pair at nesting time. The nest is

most often a cup-shaped structure. However, a hanging nest is made by some species and a tunnel excavated by others. The number of eggs may vary from one, in some tropical species, to as many as a dozen in some titmice. In colour the eggs range, depending on the species, from immaculate chalky white and light tints to deeper tones, and many are speckled and marked. The young hatch with a scant covering of down at most and in a blind, nearly helpless condition. They are cared for in the nest for a time and reach full size and are able to fly well a short time after leaving the nest. Many northern species migrate to a milder climate for the winter. The greatest number of species of songbirds is in the tropics, but they are also well represented in temperate zones. A few live in the arctic, but there are none in the antarctic.

Songbirds are distinguished from the other perching birds by certain anatomical characteristics, especially the more complicated vocal organ or syrinx (*see* below).

Bird Song.—Vocalization in birds includes a wide variety of calls besides the song proper and provides a means of social communication. Man is the only other animal that exceeds birds in the use of voice, and probably no other animals approach birds in this ability. Bird song is best considered the vocalization that is used in courtship and breeding, chiefly by the male, to advertise that he is ready to mate, to attract the female and perhaps stimulate her sexually, to keep the pair together and to inform rival males that he has established a territory from which they will be excluded. The male's calls are also part of a threat display that takes the place of actual combat in repelling intruding rivals. However, similar song is sometimes given spontaneously when there is no obvious use for it. Occasionally females sing, and especially in tropical species pairs may duet, again perhaps a method of reinforcing the bond between the pair. Often the song is delivered from a series of regularly used perches. Some species, especially those that live in grasslands, have flight songs.

In some birds other forms of sound-making have taken the place of vocalization in part at least, as in the woodpeckers which drum out a tattoo by pounding on a dead branch with their bill. Some grouse drum with their wings, others stamp their feet. The winnowing or bleating of the snipe in courtship flight is caused by air passing between the outer tail feathers; peacocks rattle their quills in courtship. (*See* COURTSHIP, ANIMAL.)

Song need not be pleasing to the human ears. The hooting of an owl; the monotonously repeated phrases of the North American whippoorwill; the crazed, repeated whistle of a Malayan cuckoo that has given it the name of the brain-fever bird; and the African tinker bird's repeated notes, which from their resemblance to hammering on metal, have given the bird its name—all must be called songs. While the best songsters are true songbirds, some birds of other groups have pleasing or musical utterances, like the quavering trill of the screech owl and the cheery whistle of the bobwhite quail.

Which birds are the best songsters is a question that is subjective in part. Although national pride and pleasant associations surely have an effect, there is general agreement on certain species. The nightingale of Europe (*Luscinia megarhynchos*), which is a small thrush, perhaps heads the list of famous songsters of literature. Its beautiful brilliant song, full of sweet trills, fluted notes and liquid phrases in surprising variety, is rich and loud. The melody is often enhanced by being delivered at night when all else is still. The European skylark of the poets (*Alaudia arvensis*) soars up, singing as it goes, until it is a mere speck in the sky alternately fluttering and sailing, pouring out its joyful and melodious song. Another outstanding songster is the European blackbird (*Turdus merula*), one of the thrushes.

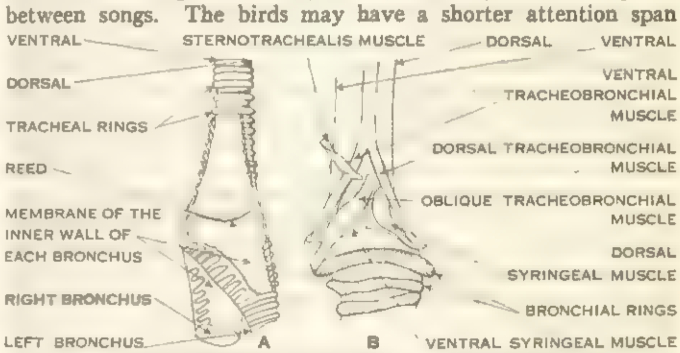
In North America the mockingbird (*Mimus polyglottos*) is a wonderful performer with a rich, melodious, long-continued song. It consists of a great many phrases with notes repeated two or three times. Imitations of other birds are often incorporated into the mockingbird's song, and on moonlit nights the bird may sing almost until dawn. Among the finest songsters are the cardinal (*Richmondena cardinalis*), with its loud clear whistles in crescendo and accelerando; the hermit thrush (*Hylocichla guttata*), with silver bell-like quality to its serene song. Townsend's solitaire (*Myadestes townsendi*) is another thrush, whose superb song consists of clear, brilliant, ringing notes that rise and fall in pitch and volume as the bird warbles and trills.

In Australia the lyrebirds, which are not true songbirds, have songs that are superlative in variety and intensity and have a dramatic quality. The pied butcherbird has a richly musical flutelike song. The brown honey eater has a song remarkably like that of the European nightingale's, and divides his song into proper stanzas.

C. Hartshorne has elaborated the concept of bird song as music. He shows there is a measure of congruity between songbirds' feeling for sound patterns and ours. Birds have in unequal degree an actual musical ability and sensibility, and their song can be considered a primitive form of music, an evolutionary anticipation of human music. Of the approximately 9,000 species of birds perhaps only a third should be considered singing species. Of these perhaps 1,200 might be said to have fairly good songs and only 225 species, scattered over the world, have really outstanding songs. Of the outstanding songsters there are 26 in Europe, 21 in North America and about 20 in Australia. It is possible that there are more outstanding bird songs in the mountains of the tropics than elsewhere.

Music consists predominantly of tones rather than noises, *i.e.*, of relatively pure sounds of a single frequency or with natural harmonies instead of miscellaneous blends of frequencies. While birds produce both noises and tones, the latter are more conspicuous in the better songs. Flutelike notes are common in bird song, though often of higher pitch than flutes. There are truly chime or bell-like, guitarlike and even organlike tones in bird song.

Bird song tends to be high pitched (this is perhaps correlated with birds' high rate of metabolism) and their hearing range somewhat surpasses ours. Notes tend to be very short and arranged in some kind of pattern. It is true that some songs are repeated over and over with slight variation, but almost always there is a pause between songs. The birds may have a shorter attention span



FROM FOWLER, "SUMMER STUDIES OF BIRDS AND BOOKS" (MACMILLAN, LTD.)

FIG. 1.—INTRICATE STRUCTURE OF A SONGBIRD'S VOCAL MECHANISM

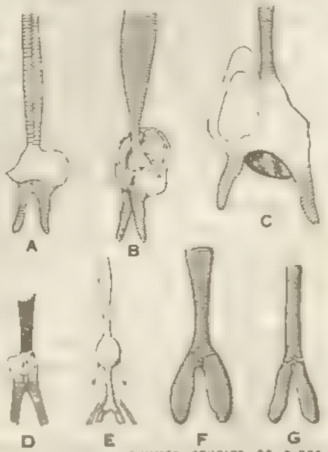
(A) Interior view of the syrinx and (B) an exterior side view of the syrinx

and weaker memory than we have, and the monotony of certain of their songs is, therefore, heard by our ears rather than the bird's. It is significant, too, that the birds' songs that rate highest with us have a certain randomness, a quality of unpredictability. The wood thrush (*Hylocichla mustelina*) has about ten songs, which can follow one another in about 90 combinations; the bird improvises scores of different contrasts in a few minutes' singing. According to Hartshorne there exists one radical inferiority in the best bird songs, when compared with human music, and that is their great simplicity in length, shown by the brief time span of the motifs or musical units. The longest seems to be about ten seconds and only a few extend to six seconds. The average is probably less than three seconds. It is true that some birds have a repertoire that extends through a minute or two of uninterrupted singing, but these repertoires consist of discrete elements each lasting a few seconds at most. Other birds have a single song lasting a minute or longer, but this, like some insect songs, is a relatively patternless prolongation of a single sound, trill or buzz.

As well as having a primary biological function in the breeding season, when bird song reaches its height, some bird vocalizing

appears to be secondarily a leisure activity or play, an effusive burst of song out of sheer joy. This is especially evident in such birds as the white-throated sparrow (*Zonotrichia albicollis*), which sings on migration; the many species that sing in the autumn; and such birds as the Carolina wren (*Thryothorus ludovicianus*), which may sing every day of the year. This off-season singing may be the species' song at its very best.

Bird song shows every elementary rhythmic effect: accelerando (field sparrow, *Spizella pusilla*), ritardando (yellow-billed cuckoo, *Coccyzus americanus*), crescendo (some thrushes), interval inversion, simple harmonic relations, retention of melody with change in key and theme with variations. Bird song, according to Hartshorne, is always intelligible as simple music. The best singers are those with more variety and complexity in the use of elementary musical devices, and they seem to sing most.



FROM FOWLER, "SUMMER STUDIES OF BIRDS AND BOOKS" (MACMILLAN, LTD.)

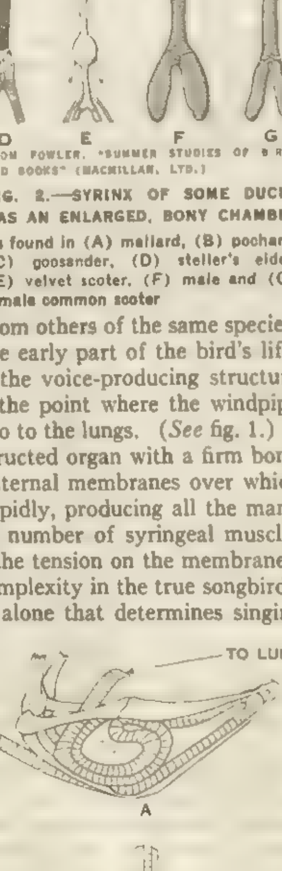
FIG. 2.—SYRINX OF SOME DUCKS HAS AN ENLARGED, BONY CHAMBER

As found in (A) mallard, (B) pochard, (C) goosander, (D) steller's eider, (E) velvet scoter, (F) male and (G) female common scoter

Bird songs have been recorded in various ways, by musical notation, by graphic notation, by fitting words to the song, by describing it in words and by capturing it on tape or phonograph disc. But a most useful way for analyzing and comparing bird song is the sound spectrograph which turns the song into a visual pattern recorded on paper. One of the important uses of this, by W. H. Thorpe of Cambridge university, was in a study demonstrating that in some birds the song is innate, while in others it is partly innate and partly learned from others of the same species. This learning takes place only in the early part of the bird's life.

Vocal Organs.—The syrinx, as the voice-producing structure or song box is called, is located at the point where the windpipe divides into two bronchial tubes to go to the lungs. (*See* fig. 1.)

The syrinx is an intricately constructed organ with a firm bony framework and filmlike vibrating internal membranes over which the air, during exhalation, passes rapidly, producing all the many utterances of the bird. A variable number of syringeal muscles and their controlling nerves adjust the tension on the membranes. The song box reaches its greatest complexity in the true songbirds. But it is not a complicated syrinx alone that determines singing ability, for some true songbirds hardly sing at all.



FROM PYCRAFT, "HISTORY OF BIRDS" (METHUEN)

FIG. 3.—ELONGATED WINDPIPE

(A) Enclosed in the sternum of a crane and (B) elaborately coiled in a manucode, one of the birds of paradise

In certain ducks, especially in the males, the syrinx has an enlarged, bony chamber (*see* fig. 2). However, none of these birds have elaborate voices, though the male's voice may be different from that of the female. Sometimes the windpipe is elongated and elaborately coiled. In some cranes and swans this elongation is enclosed within the breast bone or sternum. In certain of the birds of paradise known as manucodes the elongated windpipe is coiled on the breast between the skin and the flesh (*see* fig. 3). Presumably this lengthening of the windpipe gives resonance to the voice.

Bibliography.—J. E. Harting, *Birds of Middlesex* (1866); J. Muller, *Certain Variations in the Vocal Organs of the Passeres* (1888); C. A. Witchell, *The Evolution of Bird-Song* (1896); F. S. Mathews, *Field Book of Wild Birds and their Music* (1921); H. Oldys, "The Meaning of Bird Music," *Am. Mus. J.*, vol. xvii, pp. 123–127 (1917); H. E. Howard, *Territory in Bird Life* (1921); G. I. Hartley, *The Importance of Bird Life* (1922); E. H. Forbush, *Birds of Massachusetts and Other New England States* (1925 *et seq.*); E. M. Nicholson, *How Birds Live* (1927); A. A. Allen, *The Book of Bird Life* (1930); Cornell University, "American Bird Songs," phonograph records (1940); A. A. Saunders, *A Guide to Bird Songs* (1951), L. W. Wing, *Natural History of Birds*, ch. 17 (1956); W. H. Thorpe, *Learning and Instinct in Animals*, ch. 14 and 15 (1956); C. Hartshorne, "The Relation of Bird Song to Music," *The Ibis*, vol. 100, pp. 421–445 (1958). (A. L. Rd.)

SONGHAI (Songhoi, Songhay, Sonrhai), a tall, dolichocephalic Negroid people of probable Caucasoid admixture. Probably originating in the region of Lake Chad, they inhabit the northern area in the bend of the Niger River in the Republic of Mali, extending from Lake Débo to the mouth of the Buni N'kebbi River in Nigeria. Composed of many related groups, the most important being the Djerma (Zerma), they numbered about 800,000 in the 1960s. Berbers founded (7th century A.D.) the ruling dynasty of the Songhai; their Gao empire passed under the domination of the Mali empire (*see* Malinke) in the 14th century, and was consolidated in the 15th century by the Songhai prince of Gao, Sonni Ali, and his successors. A Moroccan invasion in 1591 began crumbling this empire, and was followed by Bambara and Fulani (*qq.v.*) invasions, and finally by French occupation of Tombouctou (1893). Songhai conversion to Islam dates from the 11th century, influencing their culture and laws, as well as religion (*see* Africa: *History: Sahara and Sudan*). Songhai language forms an independent stock.

Based on family relationships, Songhai society also has social groupings of slaves, artisans and *griots* (bards and chroniclers), freemen and headmen. Cultivation, largely based on cereals, is practised intensively only during the rainy season from June to November; cattle are reared on a small scale and fishing is the quasi monopoly of the Mandingo-speaking Bozo (Sorko). Each craft of workers in iron, wood, pottery, leather, or cotton, belongs exclusively to one particular caste. The wealth of the Songhai is a result of their privileged trading position at the crossroads of west and central Africa. In the 1960s many young Songhai were leaving home for the coast, especially Ghana. *See* also Mali, Republic of; Niger, Republic of the; Tombouctou.

See J. Rouch, *Les Songhay* (1954), *La Religion et la magie Songhay* (1960). (J.-L. Bo.)

SONGKHLA (formerly called Singora) is a port city and capital of a *changwad* (province) of the same name on the east coast of peninsular Thailand. Pop. (1960) 31,014. Situated at the outlet of Thale Sap, the largest freshwater lake in Thailand, Songkhla is one of the largest cities in Thailand. It serves as a regional centre for the coastal area fronting on the South China Sea, as a port of call with a roadstead harbour, and as a centre for both fresh- and salt-water fishing. Trucked to Ban Hat Yai (Haad Yai), fish are shipped fresh by rail 133 mi. (214 km.) to Penang, Malaya. There are several ice manufacturing plants in Songkhla.

Located on a rail spur about 20 mi. from Ban Hat Yai but about 450 mi. from Bangkok, Songkhla is oriented commercially to the ports of Penang off the west coast of Malaya and Singapore at the southern end of the Malay Peninsula. Chief exports are rubber, tin, coconuts, peanuts, and forest and fish products.

In addition to being an economic and transportation centre, Songkhla has cultural, educational, and resort facilities. Among the educational institutions are a two-year teachers' college and a four-year technical institute.

With an area of 2,576 sq.mi. (6,672 sq.km.) and a population of 500,285 in 1960, Songkhla *changwad* has a population density of about 150 persons per sq.mi. It receives about 90 in. (2,286 mm.) of rainfall annually and has no distinct dry season. Rice is grown in the eastern part, and rubber trees are concentrated in the western section on the higher and better-drained soils. Some tin is mined. Songkhla is one of the most prosperous *changwads* in Thailand. (T. F. B.)

SONG OF SONGS: *see* Solomon, Song of.

SONNEBERG, a *Kreisstadt* (county capital) of East Germany in the *Bezirk* (district) of Suhl, German Democratic Republic. Pop. (1964) 29,811. The town lies at 1,330 ft. (405 m.) above sea level on the southern slope of the Thüringer Wald

(Thuringian Forest) about 31 mi. (50 km.) SE of the district capital Suhl and 50 mi. (80 km.) S of Erfurt. It is a railway junction and a centre of the toy industry which is located in the nearby towns and villages. It has a toy museum in which toys from many countries and centuries are on display. Other light industries produce items for entertainments and festivals, porcelain, and electrotechnical goods. There are slate quarries nearby. Sonneberg received its municipal charter in 1317. The town district of Sonneberg-Neufang has the highest observatory in Germany, built in 1925 and 2,093 ft. (638 m.) above sea level.

SONNET. The sonnet is unique among the poetic forms that originated in Western Europe in that it has never since its inception gone entirely out of favour.

It appears to have taken form among 13th-century Sicilian poets experimenting with troubadour verses from Provence (*see* ITALIAN LITERATURE: *The Sicilian School*). From Sicily it spread to Tuscany and was taken over from Guittone d'Arezzo (*c.* 1235–94) by Dante for poems in the *Vita Nuova* (*c.* 1293) and by Petrarch, who, in his *Rime* for Laura, wrote probably the most famous of all sonnets. These established what came to be recognized as the "Italian" or Petrarchan form of the 14-line poem: an octave, rhyming *a b b a a b b a*, and a sestet with rhymes variously disposed but never including a final couplet. The line was the staple Italian hendecasyllabic. With few variants and aberrations, the verse or line of the sonnet has always been the staple line of the verse of its language: Alexandrine in French, iambic pentameter in English, and so on.

With other poetic forms evolved during the Italian Renaissance, the sonnet traveled through much of Europe. It was handled in Spain by the marqués de Santillana in the 15th century, and established there in the 16th by Juan Boscán and Garcilaso de la Vega. In Portugal, Luís Vaz de Camões treated it with a fine balance of fire and style. Jan Kochanowski introduced it to Poland, whence it passed, often generations later, to other Slavonic literatures. Like other Renaissance forms, the sonnet arrived comparatively late in Germany: through Georg Rudolf Weckherlin, who was much influenced by the poets of England, where he passed much of his life in the 17th century; and Martin Opitz, whose formulation (in 1624) of metrical and other poetic rules roused many other poets to imitation and emulation of sonnets among other forms admired in Italian, French, and other foreign literatures.

It was in the 16th century that the sonnet was introduced into England, first by Sir Thomas Wyat and then by Henry Howard, earl of Surrey (both of whom translated some of Petrarch's sonnets), and to France by Clément Marot and Mellin de Saint-Gelais. From such French sonneteers as Philippe Desportes and Guillaume Du Bellay, whose sonnets were translated by several English poets, including Edmund Spenser, a fresh impulse was given to the sonnet in England, and it entered on further developments. Both Wyat, whose handling of the sonnet otherwise contributed little to its characteristic history, and Surrey began to use a rhymed couplet at the end of the poem, and this practice, adopted by Spenser and others, led to the full "English" or "Shakespearean" form, consisting of three alternately rhymed quatrains and a closing couplet. Spenser, Sir Philip Sidney, Samuel Daniel, Michael Drayton, and others also developed the practice of composing sonnets in sequences, in which a general connecting notion of love for a lady is sometimes loosely organized by a hinted story of the progress of the love; occasionally, the sequential sonnets are further linked by repetitions of lines or rhyme sounds. In all the sequences, however, the finest achievement is usually in a few individually strong sonnets, such as Spenser's "Most glorious Lord of Life," which points toward the later extension of subject matter. The greatest of all sequences, Shakespeare's to the young man and the dark lady, is held together less by its supposed story, which different editors have been able to rearrange almost to taste, than by its underlying meditation on time and art, emotionally and philosophically considered; the love relationships are images, compared and contrasted with images drawn from the non-human natural world, relevant to this meditation.

By the time John Donne had written his religious sonnets (probably 1609–10), and Milton, reverting to a version of the Italian form, his sonnets on political, religious, social, and personal occasions (1642–58), the sonnet had been extended to embrace all the subjects of which it is capable—that is, nearly all the subjects of poetry. It is the astonishing genius of this short form that it can range from "light conceits of lovers" to considerations of man, death, time, and eternity without doing injustice to any of them. It is, of course, only a single aspect or experience of a theme that can be thus considered, and unity of substance, as of form, in the sonnet is one of its strengths. In English, the difficulty of rhyming and the resultant insistence on shape marked out by stressed rhyming syllables make of the form a clear, firm structure seemingly capable of holding any weight of thought or feeling. Milton's practice did much to emphasize this insistence on structure. In several sonnets, he limits the number of allowed rhyme sounds even further by using the same vowel sound in different rhymes. Thus, in "On the Late Massacre in Piedmont," the only two vowel sounds in the rhyming words throughout the sonnet are long *o* and long *a*—"bones"/"stones" and "cold"/"old" in the octave; "they"/"sway" and "sow"/"grow" in the sestet. In "Methought I saw . . ." the rhyme sounds in the octave are "(s)*ai*(nt)" and "(gr)*a*(ve)," and in the sestet "(m)*i*(nd)" and "(s)*i*(ght)"; and in "When I consider how my light is spent" they are "(sp)*e*(nt)" and "(w)*i*(de)" in the octave and "(n)*ee*(d)" and "(st)*a*(te)" in the sestet. Thus, though he often runs over and obscures the *volta* or "turn" between strict octave and sestet, Milton nevertheless hardens rather than softens the strong outline of unitary form in the Italian model.

At first sight the Italian form with its few allowed rhymes is the more difficult; its satisfying organization of octave and sestet has been compared to the rising and breaking of a wave. The greater freedom of the three quatrains in the English form, which may have as many as six rhyme sounds, is, however, outweighed by the extreme difficulty of the concluding couplet, which may be estimated from the fact that even Shakespeare often failed with it. This couplet must somehow seem to account for or correspond to the accumulated force of the preceding 12 lines and therefore needs rather the compressed power of Greek epigram than the sharp neatness of, for example, a satiric couplet. The effect when successful has been compared to the firing of an arrow after the tensing of the bow, as the mounting tension of Shakespeare's "Th'expense of spirit in a waste of shame" is startlingly released in

> All this the world well knows; yet none knows well
> To shun the heaven that leads men to this hell.

At its lightest a graceful gesture, at its most profound the sonnet can concentrate immense feeling into a shape which, making its way easily by its grace into the consciousness, there releases and expands its meaning.

The 18th century was not one of the full flowering periods of the sonnet, but William Walsh wrote some early in the century, Thomas Warton attempted a revival, William Cowper practised the form, and Thomas Gray on occasion (as on the death of Richard West) found its formal strictness a happy challenge to his temperament.

The English Romantics who had felt "the weight of too much liberty" found it

> pastime to be bound
> Within the sonnet's scanty plot of ground,

and Wordsworth's best sonnets (usually in the Italian form) are among the finest examples of the genre.

Keats's uneasiness with the "pouncing rimes" did not prevent him from writing some beautiful sonnets, but for him one of the main values of the sonnet was the influence of its organic structure in developing the great stanzas of his odes. In Shelley's "Ode to the West Wind," too, the sonnet form seems to have influenced the 14-line stanzas consisting of four sections of *terza rima* followed by a couplet using the fifth rhyme. Keats and Shelley thus drew attention to the seminal, rather than the actual, powers of the form. The strict sonnet was later beautifully handled in France by Baudelaire and J. M. de Heredia, and in England, Elizabeth Barrett Browning, in the once very popular *Sonnets from the Portuguese*, returned some way toward Elizabethan se-

quences. Experiment, however, was lively in England in the later 19th century. The 16-line poems of George Meredith's sequence *Modern Love* are developed from the sonnet, and Gerard Manley Hopkins wrote sonnets in sprung rhythm with "outrides" and extra "codas." Some of Hopkins' best poetry, however, is contained in sonnets of strict Italian form. The most distinguished 20th-century work in the kind is Rainer Maria Rilke's *Sonette an Orpheus* (1922). Some modern poets, among them E. A. Robinson, Robert Frost, and Sidney Keyes, have written sonnets, and it seems unlikely that this powerfully moving and evocative form of poetry is yet moribund.

See also articles on various writers mentioned; on ALEXANDRINE VERSE; TERZA RIMA; and on the national literatures.

BIBLIOGRAPHY.—*The Book of the Sonnet*, ed. by Leigh Hunt and S. A. Lee (1867); *Elizabethan Sonnets*, ed. by Sir Sidney Lee, 2 vol. (1904); J. W. Lever, *The Elizabethan Love Sonnet* (1956). The work of individual sonnet writers is best studied in the publications of the writers concerned and in critical studies of them. (C. M. I.)

SONNINO, (GIORGIO) SIDNEY, BARONE (1847–1922), Italian statesman and financial expert who was twice prime minister and who as foreign minister promoted his country's entry into World War I. He was born at Pisa on March 11, 1847; his mother was English. Entering the diplomatic service in 1867, he left it in 1873 to devote himself to political and social studies. On his own initiative he made exhaustive researches on the conditions of life of the Sicilian peasantry and the Tuscan métayage system. In 1877, in cooperation with Leopoldo Franchetti, he published a masterly work, *La Sicilia nel 1876*. In 1878 he founded a weekly economic review, *La Rassegna Settimanale*, which four years later he converted into a political daily newspaper, *La Rassegna*.

Elected deputy in 1880, Sonnino served for some months in 1889 as undersecretary of state for the treasury. In December 1893, at a time of severe financial crisis, he was entrusted by Francesco Crispi with the portfolio of finance, and his energetic measures averted national bankruptcy: when parliament objected, he did not hesitate to impose taxes by decree. He had to find money for the Ethiopian War and fell from office with Crispi after the disaster of Aduwa (March 1896). Assuming then the leadership of the opposition, he combated the Marchese di Rudinì's government and opposed the "democratic" finance of Luigi Pelloux's first administration. After the modification of the Pelloux cabinet (May 1899) he became leader of the ministerial majority and supported the attempt to introduce a more authoritarian form of government. On the formation of Giuseppe Zanardelli's cabinet (February 1901) he continued in opposition against the new left-wing Liberalism. He was prime minister from Feb. 8 to May 27, 1906, and again from Dec. 10, 1909, to March 29, 1910, but his temperament was too solitary and severe to win him much support in Parliament.

On Nov. 5, 1914, three months after the outbreak of World War I, Sonnino became foreign minister in Antonio Salandra's cabinet. He negotiated first with the Central Powers, hoping for territorial concessions from Austria-Hungary. When this approach failed, he turned to the Allies, though he saw that Serbia's plans for expansion interfered with Italy's. Without consulting Parliament he convinced the government that Italy should join the war on the side of Great Britain and France; and in May 1915 war was declared against Austria-Hungary. His territorial ambitions for Italy made him object to the Allies' bringing Greece into the war on their side; but he remained foreign minister both in Paolo Boselli's cabinet (under which Italy declared war on Germany in August 1916) and in V. E. Orlando's (from October 1917). He attended the Paris conference as second Italian delegate. Disappointed in his attempt to annex Yugoslav territory, he left Paris in a moment of temper. On the fall of Orlando's cabinet (June 19, 1919) he retired to private life. He died in Rome on Nov. 24, 1922. His *Discorsi per la guerra* appeared in 1922 and his *Discorsi parlamentari* in 1925.

See C. Montalcini, *Sidney Sonnino* (1926). (D. M. SM.)

SON OF MAN. In the Gospels Jesus often refers to himself as the Son of man in such a way that it becomes clear that this must have been a term that was well known and loaded with meaning before his time. There is no agreement, however, as to the origin of the idea and as to what extent these sayings of Jesus are to be considered as authentic.

In the Synoptic Gospels (Matthew, Mark, Luke) the sayings in which the term Son of man is used fall naturally into three categories:

1. Eschatological sayings in which the Son of man is depicted as the judge who is to come in glory with the clouds of heaven in order to judge the world (*e.g.*, Mark 8:38; 13:26; 14:62; Matt. 24:27, 37, 39, 44). In these sayings the Son of man is always referred to in the third person, and there is no clear indication that Jesus is identical with that eschatological figure. Only once (Luke 12:8 ff.; *cf.* Mark 8:38) is any relationship established between Jesus and the Son of man: "everyone who acknowledges me before men, the Son of man also will acknowledge before the angels of God."

2. Passages concerned with the suffering, death and resurrection of Jesus (*e.g.*, Mark 8:31; 9:31; 10:33 ff.): "the Son of man must suffer many things . . . and be killed, and after three days rise again."

3. A few sayings which refer to various aspects of the activity of Jesus: he "has authority on earth to forgive sins" (Mark 2:10), he "is lord even of the sabbath" (Mark 2:28), he is called "a friend of tax collectors and sinners" (Matt. 11:19) and he is homeless, having "nowhere to lay his head" (Matt. 8:20; Luke 9:58).

There is no relation whatsoever among these groups.

The first question that should be asked concerning these passages is that of their authenticity. In some cases it can be observed that parallel passages use a simple "I" instead of the title Son of man. Thus Matt. 10:32 uses "I" to replace the "Son of man" of Mark 8:38 and Luke 12:8 ff.; Matthew introduces "Son of man" in 16:13, a saying where Mark has "I" (8:27) But in most cases tradition is firm on the use of the term. Few scholars question the authenticity of the group 1 sayings. As for group 2, it is necessary to remember that they presuppose the events of Easter and may express the faith of the early church although afterward put into the mouth of Jesus. Other scholars admit the possibility that Jesus uttered these words also, because he must have foreseen the outcome of his conflict with the Jewish authorities. The third group also presents some difficulties, but for the most part the attempts to question the authenticity have not been successful.

As far as the first group is concerned, the implications of the term are comparatively clear on the background of Jewish apocalyptics. In Dan. 7:13 there is a reference to "one like a son of man" who was given dominion so that all nations should serve him, and mention is made of a judgment of the evil; later in the same chapter this figure is interpreted as being "the saints of the Most High" (*i.e.*, ideal Israel). It is a matter of dispute whether "son of man" here means just "a man" in general, which would be in accordance with common Aramaic and Hebrew usage, or is already an eschatological term. In later apocalyptics the latter is definitely the case.

The book of Enoch uses the title son of man to denote an eschatological figure who is to come at the end of the present age to judge the world and establish the Kingdom of God. He is referred to as existing before the creation of the world and as being kept hidden with God until his appearance at the consummation of the age. Certain motifs connected with this figure seem to be influenced by the Iranian idea of the First Man, Gayomart (*see* ZOROASTRIANISM: *The Zoroastrian Religion: Cosmology*) which is also reflected in certain (later) Gnostic speculations. It is a disturbing fact, however, that the son of man passages in the book of Enoch are all in the part of the book commonly referred to as the Similitudes, of which there remain no Greek fragments and which does not seem to be represented among the fragments from the Qumran caves. (*See* ENOCH, BOOKS OF: *I Enoch*.) This suggests to some scholars that these passages may be later, perhaps Christian, interpolations. However, II Esdras, which derives from the time immediately after the fall of Jerusalem in A.D. 70, testifies to the existence of the term in Jewish apocalyptics, and it would probably not have been introduced there after the title had been used by Jesus.

In the third group, one or two of the passages might be explained by assuming the meaning of "man" in general, or especially "mortal man," as is the case in a number of passages in Ezekiel, where the prophet is addressed as "son of man." Matt. 11:19 is associated with the humble state of Jesus and may be connected with the suffering passages.

The idea of the suffering of the Son of man, represented by the second group of sayings, presents the most difficult problem. It is a matter of dispute whether or not the book of Enoch knew the idea of a suffering Son of man. There is no unambiguous evidence, and those scholars who find this idea represented reach their conclusion only on the basis of the fact that the Son of man has several features in common with the Servant of the Lord as described in Isaiah 40–55 (Deutero-Isaiah; *see* ISAIAH, BOOK OF). But these common features are positive—his election, his function as a lawgiver and as a light to the nations and the hope of the holy ones, etc. The suffering of the Servant is not referred to in connection with the Son of man, and the contemporary Jewish method of biblical interpretation does not favour the assumption that the figure of the Servant was adopted in its totality; therefore, nothing warrants that the glory of the Son of man presupposes his suffering. The mention of "the blood of the righteous one" in Enoch 47:1–4, immediately following a chapter dealing with the glory of the son of man, is no conclusive evidence, and the text itself is not even certain.

Thus, if Jesus did really use the title "Son of man" in connection with his suffering, it would seem that it was he himself who made the combination of the Son of man of apocalyptics with the suffering Servant of Deutero-Isaiah. He may even not have known the book of Enoch but interpreted his mission in terms of Dan. 7, Isa. 53 and perhaps also the suffering righteous of Wisdom 2–5.

It has also been suggested that the term might derive from Ps. 8:4 and 80:17, where it is taken to refer to the king; since the messianic hope developed out of the idea of an ideal king, the term therefore from the beginning would have been a "messianic" title. (*See* MESSIAH.) Ps. 8:4 is quoted in the New Testament as referring to Christ, but there is no evidence that Jesus applied these psalms to himself.

As a title of Christ the term "Son of man" seems to have lost currency rather soon. By far the most instances occur in the Synoptic Gospels. There are a few instances in the Gospel of John, referring to his preexistence in heaven (3:13), his death (3:14; 8:28; 12:23, 34, etc.) and his function as a judge (5:27). Paul does not use it. It seems probable that it was introduced by Jesus himself, but that the early church found other titles more suitable to express its faith in Jesus as the saviour.

BIBLIOGRAPHY.—R. Otto, *The Kingdom of God and the Son of Man*, Eng. trans. (1943); C. H. Kraeling, *Anthropos and Son of Man* (1927); G. S. Duncan, *Jesus, Son of Man* (1949); R. Fuller, *The Mission and Achievement of Jesus* (1954); E. Sjöberg, *Der verborgene Menschensohn in den Evangelien* (1955); T. W. Manson, *The Teaching of Jesus* (1931). For the Old Testament background, *see* A. Bentzen, *King and Messiah* (1955); S. Mowinckel, *He That Cometh* (1956); E. Sjöberg, *Der Menschensohn im äthiopischen Henochbuch* (1946). (H. Ri.)

SONORA, a northwest state of Mexico. Pop. (1960) 783,378; area, 71,403 sq.mi. (184,934 sq.km.). Sonora is bounded north by the United States, east by Chihuahua, south by Sinaloa, and west by Lower California and the Gulf of Lower California. Until the 1940s Sonora was primarily a mining area, producing copper, gold, and silver. Climatic conditions are generally arid to semiarid, and irrigation is necessary to produce winter vegetables, cereals, cotton, tobacco, and maize.

The state's population is clustered around Nogales, at the border, Hermosillo the state capital, a neat and thriving air, highway, and rail centre in the midst of a flourishing cotton district, Guaymas, a coastal deep-sea fishing resort, and Ciudad Obregón, in the heart of the Yaqui Valley. The latter grew in the 1950s from a barren crossroads town to a booming city of 70,000, with modern architecture, airport, and a large number of cotton-enriched farmers. Irrigation provided by the Alvaro Obregón Dam (completed in 1952) on the Yaqui River made this expansion possible by supplying water to more than 500,000 ac. (202,000 ha.) of desert land.

The area of Sonora was explored in the 1530s, and later became a famous colonial mining district. The small town of Alamos is a quaint reminder of this colonial past. In 1830 Sonora became a Mexican state, but the unruly Yaqui tribes were not finally subdued until the 20th century. The primitive Seris still follow their old way of life on the offshore island of Tiburón. The state derives its name (sonorous) from its deposits of grayish marble, which when struck emits bell-like sounds. (J. A. Cw.)

SONSONATE, a department in western El Salvador established in 1855, bordering the Pacific Ocean. Area 459 sq.mi. (1,189 sq.km.). Pop. (1961) 166,932. The population is concentrated on southern mountain slopes, the hot coastal lowlands having few people. In the north, Izalco volcano, a black symmetrical cone without vegetation (altitude 6,183 ft. [1,885 m.]) and the most active in Central America, erupts at fairly regular intervals. The coastal lowlands produce 50% of the nation's balsam used in medicines. Sonsonate is important in the production of cheese, tobacco, coffee, sugar, pineapples, and cotton.

Sonsonate City (pop. [1961] 35,531), the departmental capital, was founded in 1524. Interesting features are El Pilar Colonial Church, an old cathedral with 17 cupolas of all sizes, and nearby the San Antonio colonial shrine visited every year by thousands of pilgrims seeking cures. Sonsonate is 41 mi. (66 km.) W of San Salvador by highway and 12 mi. by road and railway from Acajutla, an open roadstead port which handles about 20% of the country's foreign trade. (C. F. J.)

SOOCHOW (SU-CHOU, WU-HSIEN), a city in Kiangsu Province, China, situated on the Yangtze delta, 36 mi. S of the river, 55 mi. W of Shanghai and close to the eastern lake shore of T'ai Hu. It probably dates from about 484 B.C., and, like other old towns nearby, was located upon onetime hilly coastal islands, which were later joined together by Yangtze silting plus man's diking and draining for at least 3,000 years. The scenic mass of lakes and canals, including the Grand Canal, caused Europeans to name it the "Venice of China," while the Chinese esteemed its beautiful gardens and fair women. It was the capital of the Wu Kingdom in the 5th century B.C. The fertile rice and silk country created wealth which over the centuries was poured into the construction of more than 200 private gardens, many restored and maintained for public enjoyment. Water, fertile soil, a mild climate for vegetative growth, nearby eroded lake stone, and leisure for scholarly study and aesthetic development raised gardening in Soochow to one of China's great arts. It was a place of official retirement, inspiration to poets and painters and a luxury handicraft centre for jade and ivory carving, silk weaving, and embroidery. Under T'ai P'ing control (1860–63) Li Hsiu-ch'eng instituted numerous social and economic measures that, unlike T'ai P'ing programs elsewhere, won the support of the local people and foreshadowed many of the changes of the Chinese People's Republic. It was a treaty port in 1896 and was held by the Japanese (1937–45). In modern times, silk filatures, weaving mills and a sericulture institute were set up. It is the seat of South Kiangsu Technical College. The city is on the Shanghai–Nanking railway and has a spur line south to Chia-hsing, Chekiang, augmented by highways for local traffic; some of the city's streets were widened for motor traffic. Pop. (1957 est.) 633,000. (TE. H.)

SOPHIA (SOPHIA ALEKSEEVNA) (1657–1704), regent of Russia from 1682 to 1689, the first woman to rule in Moscow since the time of Ivan IV's mother, Elena Glinskaya. She was born on Sept. 27 (new style; 17, old style), 1657, the sixth child of Tsar Alexis and his first wife, Maria Ilinishna Miloslavskaya. She received what was by Russian standards a remarkable education at the hands of Simeon of Polotsk, a Belorussian monk nurtured on the Jesuit humanism of the Polish academies. On May 7 (N.S.; April 27, O.S.), 1682, at the death of her brother Tsar Fedor III, her half-brother Peter I (*q.v.*) was proclaimed tsar; but Sophia could not abide the prospect of Russia's being ruled on Peter's behalf by the family of Naryshkin (*q.v.*), to which his mother belonged. The Miloslavskis, who shared Sophia's feeling, directed the discontent

of the *streltsy* (*q.v.*) or household troops against the Naryshkins while Sophia contrived to insinuate to both the Miloslavskis and the *streltsy* that the joint rule of her younger brother Ivan V (*q.v.*) and Peter, with herself as regent, was the best solution. On June 5 (N.S.; May 26, O.S.), what remained of the *Zemski Sobor* (Assembly of the Realm) ratified this proposal. Sophia began by bringing to heel the turbulent household troops. On Sept. 27 their self-appointed commander, Ivan Andreevich Khovanski, was executed on a charge of high treason; and the next month 12 out of the 19 Moscow regiments were sent to guard the frontier.

Reserving for herself the sphere of machination and intrigue, on all major issues of home and foreign policy Sophia consulted V. V. Golitsyn (*q.v.*) who was also her lover. In his absence she bestowed her favours on the new head of the household troops, Fedor Leontievich Shaklovity, the mainstay of her rule. In most respects her reign was one of continuance and compromise, as witness the treaty of "permanent" peace and alliance with Poland concluded in Moscow on May 6 (N.S.), 1686, and the Crimean expeditions of 1687 and 1689, both necessitated by the revival of the Turkish peril. The abolition of the customs barrier between Russia and the Ukraine was a further step in the incorporation of the Ukraine; Sophia also promoted industry and encouraged foreign craftsmen. In the "academy" established in Moscow in 1687 the Greek orientation triumphed over the Latin-Polish influence of Sophia's adviser, Silvestr Medvedev. In June 1687 Sophia adopted the style of autocrat; but her authority was shaken by the failure of the Crimean expeditions, the second of which was openly criticized by Peter. Knowing of Shaklovity's readiness to wipe out the Naryshkins, she complained of their enmity before the *streltsy*, hoping to provoke a repetition of the outburst of 1682. But in August 1689 enough of the *streltsy* colonels threw in their lot with Peter to ensure the Naryshkins' victory over the Miloslavskis. Shaklovity was executed, and on Sept. 17 (N.S.), Sophia's name was removed from the tsars' joint title. A week later she was obliged to enter the Novo-Devichi convent.

An abortive revolt of the *streltsy* in 1698, during Peter's absence, though not engineered by Sophia, was intended to achieve her reinstatement. On Oct. 22 (N.S.), 1698, she was tried by a special tribunal whose verdict is unknown, and on Oct. 31 she was compelled to take the veil. She died on July 14, 1704.

See C. Bickford O'Brien, *Russia under Two Tsars, 1682–1689* (1952). (L. R. Lr.)

SOPHISTS, the name given to certain Greek lecturers, writers, and teachers in the 5th and 4th centuries B.C., most of whom traveled about the Greek-speaking world giving instruction in a wide range of subjects in return for fees.

History of the Name.—The term "sophist" (Greek *sophistes*) had earlier applications. It is sometimes said to have meant originally simply "clever" or "skilled man," but the list of those to whom Greek authors applied the term in its earlier sense makes it probable that it was rather more restricted in meaning. Seers, diviners, and poets predominate in the list, and the earliest sophists were probably the "sages" in early Greek societies. This would explain the subsequent application of the term to the Seven Wise Men (7th–6th century B.C.) and to Pre-Socratic philosophers generally. When Protagoras is made by Plato (*Protagoras*, 317 a–b) to say that, unlike others, he is willing to call himself a sophist, he is using the term in its new sense of "professional teacher," but he wishes also to claim a continuous development from earlier sages to his own function as a teacher of wisdom. Plato and Aristotle altered the meaning again when they claimed that professional teachers such as Protagoras were not seeking the truth but only victory in debate, and were prepared to use dishonest means to achieve it. This produced the sense "captious or fallacious reasoner or quibbler" which has remained dominant down to the present day. Finally, under the Roman Empire the term came to be applied to professors of rhetoric, to orators, and to prose writers generally, all of whom are sometimes regarded as constituting what is now called the Second Sophistic movement (*see* below).

The 5th-Century Sophists.—The names survive of nearly 30 sophists properly so-called, of whom the most important were Protagoras, Gorgias, Prodicus, Antiphon (*qq.v.*), and Thrasymachus. Plato protested strongly that Socrates was in no sense a sophist—he took no fees and his devotion to the truth was beyond question. But from many points of view he is rightly to be regarded as a rather special member of the movement. The actual number of sophists was clearly very much larger than 30, and for about 70 years, until *c.* 380 B.C., they were the sole source of higher education in the more advanced Greek cities. Thereafter, at least at Athens, they were largely replaced by the new philosophic schools, such as those of Plato and Isocrates. Plato's dialogue, *Protagoras*, describes something like a conference of sophists at the house of Callias in Athens just before the Peloponnesian War. Antimoerus of Mende, described as one of the most distinguished of Protagoras' pupils, is there receiving professional instruction in order himself to become a sophist (*Protagoras*, 315 a) and it is clear that this was already a normal way of entering the profession.

Most of the major sophists were not Athenians but they made Athens the main centre for their activities, although traveling continuously. The importance of Athens in the movement was due in part no doubt to the greater freedom of speech there prevailing, in part to the patronage of wealthy men like Callias, and the positive encouragement of Pericles, who was said to have held long discussions with sophists in his own house. But primarily the sophists doubtless congregated at Athens because there they found the greatest demand for what they had to offer, namely instruction to young men, and the extent of this demand was not an accident. Athens itself was a democracy; while its limits were such that Thucydides could say that in fact Athens was governed by one man, Pericles, nonetheless it gave opportunities for a successful political career to citizens of the most diverse backgrounds, provided they could impress their audiences sufficiently in the council and the assembly. After Pericles' death this avenue became the highroad to political success.

The training which the sophists gave taught men how to speak and what arguments to use in public debate. A sophistic education was increasingly sought after both by members of the oldest families and by aspiring newcomers without family backing. The changing pattern of Athenian society made merely traditional attitudes in many cases no longer adequate. The criticism of such attitudes and the attempt to replace them by rational arguments was part of the especial attraction of the sophistic movement for the young, and it explains the violent distaste which they aroused in traditionalists. Plato thought that much of the sophistic attack upon traditional values was unfair and unjustified. But he learned at least one thing from the sophists—if the older values were to be defended it must be by reasoned argument and not by appeals to tradition and unreflecting faith.

Seen from this point of view the sophistic movement should be regarded above all as a function of Athenian democracy in the 5th century B.C. It offered an education designed to facilitate and promote success in public life. All the sophists appear to have provided a training in rhetoric and the art of speaking, and the sophistic movement was responsible for large advances in rhetorical theory and contributed greatly to the development of style in oratory. Sometimes in modern times the view has been adopted that this was the sophists' only concern. But the range of topics dealt with by the major sophists makes this unlikely, and even if success in this direction was the ultimate aim of their teaching in each case, the means adopted to promote such success were surely as much indirect as direct. For the pupils of the major sophists were offered instruction not merely in the art of speaking, but in grammar, in the nature of virtue (*arete*) and the bases of morality, in the history of society and the arts, in poetry, music, and mathematics, and also in astronomy and the physical sciences. Naturally the balance and emphasis differed from sophist to sophist, and some offered wider curricula than others. But this was an individual matter, and older attempts to divide the sophistic movement into periods in which the nature of the instruction given was altered are now seen to fail for lack of evidence. The 5th-century sophists inaugurated a method of higher education which in range and method can claim to have anticipated the modern humanistic approach inaugurated or revived at the Renaissance.

The Sophists and Truth.—A question still discussed is whether the sophists as a whole had any real regard for the truth and whether they taught their pupils that truth was unimportant compared with success in argument. Plato found against them on both counts and his hostile judgment is still frequently repeated without question. The Platonic writings make frequent reference to what Plato calls "eristic" and also to what he calls "antilogic"; the two have often incorrectly been treated as identical. Eristic for Plato consists in arguments aimed at victory rather than truth. Antilogic involves the assignment to any argument of a counterargument which negates it, and the full technical sense of the term includes the contention that both argument and counterargument are equally true. Antilogic in this sense was especially associated with Protagoras, but Plato, no doubt correctly, attributes its use widely to other sophists. He regards the use of antilogic as essentially eristic, whether it be used to silence an opponent by making his position seem self-contradictory, or whether it be used mechanically to establish the negation of any proposition put forward in debate. He concludes that the widespread use of antilogic in sophistic discussions is evidence that sophists had no real regard for truth, which must itself be free from antilogic.

But Plato himself believed, for much or possibly all of his life, that the phenomenal world was essentially antilogic in character, inasmuch as no statement could be made about it which would rank any higher in truth than the negation of that statement. For example, if a man is tall in relation to one object, he will also be short, and so not-tall, in relation to another object. In so characterizing the phenomenal world Plato would certainly not wish to be called eristic—he would regard the application of antilogic to the description of the phenomenal world as an essential preliminary to the search for the truth residing in the Platonic Forms, which are themselves free from antilogic.

This affects our judgment about the sophistic use of antilogic. To the extent that it was used irresponsibly to secure success in debate it can fairly be classed as eristic and the temptation so to use it must often have arisen. But where it was invoked because it was genuinely believed that what was being described in fact involved antilogic characteristics, or where it was used as a device for analyzing a complex situation in order to reveal its complexity, then antilogic was in no way inconsistent with devotion to truth. This raises the question to what extent the sophists possessed any general view of the world and to what extent they gave expression to any genuine philosophic views, whether original or derived.

Philosophy.—Ancient writers, influenced by Plato and Aristotle, seem to have excluded the sophists, apart from Protagoras, from their schematized accounts of the succession of early Greek thinkers. Modern writers have frequently maintained that, whatever else they were, the sophists were in no sense philosophers. Even those who acknowledge the philosophical interest of certain particular doctrines attributed to individual sophists often tend to regard these as exceptions, and claim that, as the sophists were not a school but only independent teachers and writers, as a class they must be denied the title of philosophers. Two questions are involved, whether the sophists held intellectual doctrines in common and whether some or all of these could fairly be called philosophic.

Among moderns, Hegel was one of the first to reinsert the sophists into the history of Greek philosophy. He did so within the terms of his own dialectic by treating them as representing the antithesis to the thesis of the group of philosophers known collectively as the Pre-Socratics. The Pre-Socratics sought the truth about the external world with a bold enthusiasm which produced a series of explanations, each laying claim to correctness. None of these explanations of the physical world paid any attention to the observer and each was driven to reject more and more of the phenomenal world itself as unreal. With the Eleatics this culminated in the paradox that little or nothing of the phenomenal world was left as real. This in turn produced a growing distrust of the power of human beings to attain knowledge of the ultimate basis of natural phenomena. Philosophy had reached an impasse and there was a danger of complete skepticism about the possibility of knowledge. This provoked the "antithesis" of the sophistic movement, which rejected the "thesis" of the objectivists and concentrated attention upon man rather than upon nature. The sophists were subjective idealists, and so philosophy could make a step forward by turning its attention to the subjective element in knowing. Reflection upon the contrast between the thought of the sophists and their predecessors produced the "syntheses" of Plato and Aristotle.

Hegel's schematization of Greek thought was rooted in his own theory of dialectic, and the details of his dialectic made little appeal to subsequent historians of Greek philosophy. But the essentials of the scheme which Hegel propounded have endured and are embodied in the majority of full-scale histories of early Greek thought. Whether any of the sophists were subjective idealists may be doubted. It depends in part on the vexed question whether Protagoras held that phenomena had subjective existence only, or thought that all things perceived had objective existence, although their perception depends upon the nature of the percipient and their relation to him. But it is fairly clear that the sophists did concentrate attention very largely upon man and human society, upon questions of semantics and epistemology, upon the importance of the observer and the subjective element in reality and in the correct understanding of reality. To that extent Hegel's appreciation may be accepted as valid.

This goes a long way to explain the philosophic hostility of Plato and Aristotle. Above all in the eyes of Plato, anyone who looks for the truth in phenomena alone, whether he interprets it subjectively or relativistically, cannot hope to find it there, and his persistence in turning away from the right direction virtually amounts to a rejection of philosophy and the search for truth. Many a subsequent thinker for whom metaphysics is the crowning achievement of philosophy has felt with Plato that the sophists are so antimetaphysical in their approach that they have no claim to rank as philosophers. But in a period when to many who would wish to call themselves philosophers metaphysics is no longer the most important part of philosophy and to some it is not a part at all, there is a growing appreciation of the interest to philosophers of a number of problems and doctrines which recur in discussions initiated by sophists in the 5th and 4th centuries B.C.

Morality.—In the 18th and early 19th centuries the standard view of the sophists was that they were charlatans. Coupled with the attempt to discredit their intellectual honesty was the claim that their doctrines were immoral and that they were responsible for the moral corruption of the youth of Athens and even for the beginnings of the decline of Greece from the high position reached in the 5th century. The basis for the charge was two contentions, both correct: first that many of the sophists leveled damaging attacks against the traditionally accepted moral code; and secondly that they explored and even commended alternative approaches to morality which would condone or allow behaviour of a kind inadmissible under the stricter traditional code.

Much less weight has been attached to these charges since about the mid-19th century. First, many of the attacks on the traditional morality were in the name of a new morality which claimed to be of greater validity. The attacks upon particular doctrines often claimed that accepted views should be abandoned as being morally defective. Secondly, even when socially unacceptable action appeared to be commended this was frequently done in order to introduce a point of principle which must be taken into account in any satisfactory moral theory. Thus when Thrasymachus in the first book of Plato's *Republic* takes up the position that justice is unwarranted when it merely contributes to another's good and not to the good of the doer, he is making a point which Plato accepts. In fact Plato devotes the rest of the *Republic* to the attempt to show that justice is essential for the achievement of the good of the person who is required to be just. Thirdly, there is no evidence that any of the sophists were immoral in their own lives or that any of their pupils were induced to immoral actions by their teaching. Rather the serious discussion of moral problems and the theory of morality tends to improve behaviour and not to corrupt it.

Writings.—In addition to their teaching the sophists wrote many books, whose titles are known from lists preserved by

writers such as Diogenes Laërtius, which probably derive from library catalogues. It has usually been supposed that the writings themselves hardly survived beyond the period of Plato and Aristotle, but this view requires some modification in the light of papyrus finds, admittedly few, that were copied from sophistic writings in the Christian era. It has been possible to identify also certain imitations or summaries of 5th-century sophistic writers, whose names cannot be discovered, in the works of later writers, of which the most important are the discussion of law in the *Protrepticus* of Iamblichus, and the so-called *Dissoi Logoi* found in the manuscripts of Sextus Empiricus. All this evidence suggests that while most later writers took their accounts of the sophists from earlier writers, especially Plato, the original writings did in many cases survive and were at least occasionally consulted.

Particular Doctrines.—As part of his defense of the sophists against the charge of immoral teachings, the historian George Grote maintained that they had nothing in common with each other except their profession, as paid teachers, qualifying young men to think, speak, and act with credit to themselves as citizens. This denial of common doctrines to the sophists cannot be sustained—the evidence is against it. While the sophists were not a sect with a set of obligatory beliefs or doctrines, they had in common an interest in a whole series of questions to which they sought to apply solutions along certain clearly defined lines.

Before attempting to state what some of these were, however, it is important to remark that there are great difficulties in the precise statement of individual sophistic doctrines. No complete writings survive from any of the sophists to check the accounts found in Plato, and later writers were often but not always dependent upon what they found in Plato. Plato doubtless knew well the doctrines of individual sophists but he was writing for those to whom these doctrines were already well known, and he was always more interested in following the argument where it led than in providing precise statements of other people's views for the sake of posterity. Consequently almost everything that is said about particular sophistic doctrines is subject to controversy and disagreement among scholars.

Relativism and Skepticism.—These have often been regarded as a common feature of the sophistic movement as a whole. But it was early pointed out that it is only in Protagoras and Gorgias that there is any suggestion of a radical skepticism about the possibility of knowledge, and even in their case Sextus Empiricus in his discussion of skepticism is probably right when he declares that neither was really a skeptic. Protagoras does seem to have restricted knowledge to what was derived from the senses, but he did believe emphatically that whatever was perceived by the senses was certainly true. This led him to assert that the tangent does not touch the circle at a point only, but along a definite length of the circumference; clearly he is referring to our perceptions of drawn tangents and circles. Gorgias, who claimed that nothing exists, or if it does exist, it cannot be known, or if it exists and is knowable it cannot be communicated to another, has often been treated as denying both all reality and all knowledge. Yet he also seems to have appealed in his very discussion of these themes to the certainty of perceived facts about the physical world, *e.g.*, that chariots do not race across the sea. Others dismiss his whole thesis as a satire or joke against philosophers.

Probably neither view is correct; what Gorgias seems to have been attacking was not perceived reality nor our power to perceive it, but the attempt to assign existence or nonexistence with the metaphysical implications of such an operation to what we perceive around us. There is evidence that others of the sophists, *e.g.*, Hippias, were interested in questions of this kind, and the most likely conclusion is that they were all to a greater or lesser degree concerned to reject the claims of any nonsensible existence after the Eleatic fashion. The sophists in fact were attempting to explain the phenomenal world without appealing to any principles outside phenomena. They believed that this could be done by including the observer within the phenomenal world. This refusal to go beyond phenomena was in Plato's eyes the great weakness in all their thinking.

Science.—A second common generalization about the sophists has been that they represent a revolt against science and the study of the physical world. The evidence is against this, inasmuch as for Hippias, Prodicus, Gorgias, and Protagoras there are records of a definite interest in questions of this kind. The truth is rather that they were in revolt against attempts to explain the physical world by appeals to principles which could not be perceived by the senses, and instead of framing new "objective" explanations they attempted to explain where explanation was required by introducing the perceiver as one element in the perceptual situation.

Nature and Law.—One of the most famous doctrines associated with the sophistic movement was the opposition between nature and custom or convention in morals. It is probable that the antithesis did not originate in sophistic circles but was rather earlier, but it was clearly very popular and figured largely in sophistic discussions. The commoner form of the doctrine involved an appeal from conventional laws to supposedly higher laws based on nature. Sometimes these higher laws were invoked to remedy defects in actual laws and to impose more stringent obligations, but usually it was in order to free men from restrictions unjustifiably imposed by human laws that the appeal to nature was made. In its extreme form this appeal to nature involved the throwing off of all restraints upon self-interest and the desires of the individual, and it was this more than anything else which gave support to charges of immoral teaching against the sophists. On other occasions the terms of the antithesis were reversed and human laws were explicitly acclaimed as superior to the laws of nature, and as representing progress achieved by human endeavour. In all cases the laws of nature were regarded not as generalized descriptions of what actually happens in the natural world, and so not like the laws of physics to which no exceptions are possible, but rather as norms which people ought to follow although quite able to do otherwise. Thus the appeal to nature tended to mean an appeal to the nature of man treated as a source for norms of conduct.

This appeal was not to Greeks anything very novel. It represented rather a conscious probing and exploration into an area where their whole tradition of thought had long suggested the true source lay for norms of conduct. If Callicles in Plato's *Gorgias* represents a position actually held by a living sophist when he advocates free rein for the passions, then it was easy for Plato to argue in reply that the nature of man, if it is to be fulfilled, requires organization and restraint in the licence given to the desires of particular aspects of it; otherwise the interests of the whole will be frustrated. Both Plato and Aristotle, in basing so much of their ethics on the nature of man, are only following out the approach begun by the sophists.

Religion.—The sophists have sometimes been regarded as characterized by their attacks on the traditional religious beliefs of the Greeks. It is true that more than one of the sophists seem to have had to face prosecution for impiety, as did Socrates also. Protagoras wrote "concerning the gods, I cannot know either that they exist or that they do not exist nor what they are like in form," and Prodicus offered a sociological account of the development of religion. Critias went further when he supposed that the gods were deliberately invented to inspire fear in the evildoer. It is thus probably correct to say that the tendency of a good deal of sophistic thought was to reject the traditional doctrines about the gods. Indeed this follows almost inevitably if it is right to suppose that they were all attempting to explain the phenomenal world from within itself, while excluding all principles or entities not in phenomena. But in their attitude to the Olympian deities the sophists were probably at one with most of the Pre-Socratic philosophers of the 6th and 5th centuries, and also most thinking people toward the end of the 5th century. It is thus probably misleading to regard them as breaking new ground in their attitude toward religious beliefs.

Sociology.—The importance which the sophists attached to man meant that they were extremely interested in the history and organization of human societies. Here again most is known about Protagoras and there is a danger of wrongly treating his particular doctrines as typical of the sophistic movement as a whole. In the 5th century a very common attitude saw human history in

terms of a decline from an earlier golden age. Another view supposed that there were recurring cycles in human affairs according to which a progression from good to bad would give way to one from bad to good. The typical sophistic attitude to society rejected both these views in favour of one which saw human history in terms of progress from savagery to civilization. In a famous myth Protagoras explained how man achieved civilized society first with the aid of arts and crafts and then with the gaining of a sense of respect and justice in the ordering of his affairs. The general direction of the thinking of most of the sophists seems, as far as there is any evidence, to be along similar lines.

Virtue Teachable.—One of the most distinctive doctrines of all the sophists was that virtue can be taught. This sprang naturally from the sophists' professional claim to be the teachers of young men throughout the cities of Greece. But the word for virtue, *arete*, implied to Greeks both success in living and the qualities necessary for achieving such success, and the claim that *arete* could be taught by the kind of teaching which the sophists offered had far-ranging implications. It involved the rejection of the view that *arete* came only by birth, for example by being born a member of a noble family, and it involved also the rejection of the doctrine that *arete* was a matter of the chance occurring of specified qualities in particular individuals. *Arete* on this view was the result of known and controllable procedures, a contention of profound importance for the organization of society. Moreover, what can be taught has some relation to what can be known and understood. The extreme contention of Socrates that virtue *is* knowledge, whatever exactly it meant, went beyond the normal sophistic position. But the belief that teaching with a large intellectual content could produce success both for the individual and for governments has had a profound influence upon the subsequent history of education. Once again, it is through the acceptance of this doctrine by Plato and Aristotle that the sophistic position came to be part of subsequent humanist tradition.

The Second Sophistic Movement.—It is a historical accident that the name "sophist" came to be applied to the Second Sophistic movement. Greek literature underwent a period of eclipse during the 1st century B.C. and under the early Roman Empire. But Roman dominance did not prevent a growing interest in sophistic oratory in the Greek-speaking world during the 1st century A.D. This oratory aimed merely at instructing or interesting an audience, and had of necessity no political function. But it was based on elaborate rules and required a thorough knowledge of the poets and prose writers of antiquity. Training was provided by professional teachers of rhetoric who claimed the title of sophists just as the 5th-century sophists had adopted a name which had already been used by others before them.

The revival of the Greek spirit under Hadrian and other philhellenic emperors in the 2nd century A.D. found expression in a fresh flowering of Greek prose writing following the principles developed and applied by the professors of rhetoric in the 1st century A.D. Hence a group of Greek prose writers in the 2nd century A.D. came to be regarded as constituting the Second Sophistic movement. This was a backward-looking movement which took as its models Athenian writers of the 5th and 4th centuries B.C.; hence the label "Atticists" applied to some of its leading members. The limits of the movement have never been precisely defined. It is usually taken to include Polemon, Herodes Atticus, Aelius Aristides, Maximus of Tyre, and the group of Philostrati. Dio Chrysostom of Prusa is often included, although others would regard him as preparing the way for the main period. Other writers, like Lucian, Aelian, and Alciphron, were influenced by the movement even if not properly to be regarded as members of it, and the writers of prose romances, such as Longus and Heliodorus, and the historians Dio Cassius and Herodian, are also associated with the general trend. By the 3rd century A.D., however, its impulse was weakening and by the end of the century it was no longer distinguishable within the general stream of Greek literature.

See also references under "Sophists" in the Index.

BIBLIOGRAPHY.—*Fifth-Century Sophists: Testimonia and Fragments:* H. Diels and W. Kranz (eds.), *Die Fragmente der Vorsokratiker*, 6th ed., vol. ii, 252–416 (1952); K. Freeman, *Companion to the Pre-Socratic Philosophers* (1946), *Ancilla to the Pre-Socratic Philosophers* (1948). With Italian translation and commentary: M. Untersteiner, *Sofisti*, 4 vol. (1949–62). *General Discussions:* E. Zeller, *Die Philosophie der Griechen*, 6th ed., vol. i, 1278–1459 (1920); T. Gomperz, *Greek Thinkers*, Eng. trans. by L. Magnus, vol. i, book 3, ch. 5–7 (1901); M. Untersteiner, *The Sophists*, Eng. trans. by K. Freeman (1954; contains much speculative reconstruction of doctrine not universally accepted); W. Jaeger, *Paideia*, Eng. trans. by G. Highet, vol. i, book 2, ch. 3 (1939).

Second Sophistic: see GREEK LITERATURE: *Bibliography. See* also F. A. Wright, *A History of Later Greek Literature*, part ii, ch. 2–3 (1932); H. von Arnim, *Leben und Werke des Dion von Prusa* (1898).

(G. B. KD.)

SOPHOCLES (497 or 495–406 B.C.), one of the three great Greek tragic poets, was born at Colonus, a village just outside Athens. The only ancient biography is a short anonymous account of unknown date, compiled from earlier biographies, and prefixed to most of the Sophoclean manuscripts—a mixture of bald fact, hypothesis and gossip. It can be supplemented from other ancient sources, but the result is meagre. All give a picture of a dignified, gracious and well-loved citizen who moved in good society and enjoyed the pleasures of life. His father, otherwise unknown, was Sophillus, perhaps a manufacturer of armour. He was wealthy and gave his son a good education: his master in music was Lamprus, the most distinguished musician of the day, and he is said to have studied tragedy (including no doubt musical composition and choreography) under Aeschylus. He was a beautiful boy, a good dancer, lyre player and wrestler. He was selected to lead a chorus of boys which performed a paean in celebration of the victory over the Persians at Salamis (480 B.C.). It is not known when he first competed in the dramatic festival; he gained his first victory in 468, defeating Aeschylus. This began a career of unparalleled success. An inscription of the 3rd century B.C. credits him with 18 victories, the Suda lexicon (10th century A.D.) with 24, while the *Life*, citing a writer of the 2nd century B.C., says that he won 20 first prizes, many second prizes and was never third. (The lowest of these figures may not have included victories won in the Lesser Dionysia.) Aeschylus won 13—but he died at the age of 68, and Sophocles was still writing at 90; Euripides won only three. Sophocles' most famous play, *Oedipus Tyrannus* (*Oedipus Rex*), brought him only the second prize; the prizes however were awarded not for single plays but for a tetralogy comprising three tragedies and one satyr play, and it may well be that the plays presented with the *Oedipus* were not first-rate. According to the Suda lexicon he wrote 123 plays—no unlikely number: it would correspond to about 30 trilogies, in a productive career of 60 years, and in mere bulk would not much exceed Shakespeare's output. Of his minor poems nothing remains.

BY COURTESY OF THE VATICAN MUSEUM

SOPHOCLES, ROMAN COPY OF A GREEK STATUE OF ABOUT 340 B.C.

Unlike Euripides, he played a distinguished part in the public life of Athens. In 443–442 he was Hellenotamias (president of the board that collected the tribute from Athens' subject-allies in the Delian league; *q.v.*). In 440 he was elected one of the ten strategoi (military and naval commanders) in the war against the revolted Samians; his senior colleague was Pericles. The Argument to the *Antigone* attributed to the scholar Aristophanes of Byzantium says that Sophocles owed his election to the impression created by the play. The statement can hardly have been more than a surmise; nevertheless it dates the *Antigone*. A story in Plutarch's *Life of Nicias* implies that Sophocles was elected

strategos a second time. The *Life* says that he served on embassies to foreign states, which is likely enough; also that he founded the *Thiasos ton Mouson* (which might be rendered "The Royal Society for Music and Literature," except that the patrons were the muses). Outliving Euripides (who was about 12 years his junior) by a few months, he died just before the disastrous end of the Peloponnesian War. The *Life* of Euripides says: "Sophocles, hearing that Euripides was dead, brought on his chorus and actors without their crowns, and himself put on mourning; and the people wept." Euripides had died at the court of Archelaus, king of Macedon; Aeschylus also had died abroad (in 456 B.C.), in Sicily; it was gratefully remembered of Sophocles that he had remained in Athens. In 405 Aristophanes produced *The Frogs*, in which Aeschylus and Euripides contend for the throne of tragedy in Hades, with Sophocles sitting by, ready to take on Euripides if he should win (which he does not); Aeschylus himself had offered the throne to Sophocles, but he had declined it, "courteous in Hades, as he had been on earth."

WORKS

Of Sophocles' works, there remain seven complete tragedies, about 400 lines of a satyr play, the *Ichneutai* (*Ichneutae*, "The Investigators") discovered in Egypt in 1907 on a papyrus written in the second half of the 2nd century A.D. and several hundred fragments of plays now lost. The fragments vary in length from a single word, cited by a grammarian or a lexicographer, to passages of 12 or 15 verses. The seven complete plays derive from a selection made by a scholar unknown at a date unknown—perhaps in Alexandrian times, and perhaps for school use. All are works of Sophocles' maturity, though (apart from the *Antigone*, which must have been produced in 441 or 442) only one is dated by documentary evidence: the *Philoctetes*, 409 B.C. Tradition, which there seems no good reason to disturb, makes *Oedipus Coloneus* a work of Sophocles' extreme old age; it is said to have been produced by his grandson. For dating the other plays we must depend on stylistic and metrical evidence—some of which is more hazardous with Sophocles than with, for example, Euripides. Sophocles' dramatic style was much more supple, and in some respects changes from minute to minute rather than from decade to decade. Contemporary allusions are also much rarer and more problematic than in, say, Euripides or Shakespeare. However, it is clear that the *Ajax* is the earliest of the seven; 447 ± 3 is a date that most scholars would accept. The *Trachiniae* was tentatively placed by R. C. Jebb between 420 and 410; later stylistic analysis places it much closer to the *Antigone*. For the remaining plays, *Oedipus Rex* and *Electra*, dates between 430 and 415 would be reasonable; the *Electra* seems to be the later of the two. The *Ichneutae* could be an early work, but the evidence available is not worth much.

Dramatic Style.—Ancient authorities credit Sophocles with dramatic innovations, most of them of minor importance. He invented *skenographia* (certainly not "scene painting," but probably some kind of device applied to the wall of the building at the back of the *orchestra*), and *periactoi*, which may have been revolving drums giving some indication of *locale*. The one major innovation—which, however, some authorities ascribe to Aeschylus —was the introduction of a third actor. It had always been permissible for the two actors to "double," but the addition of a third enabled the dramatist to increase the number of his characters, thereby making plot more fluid and situation more complex: Sophocles' finest scenes are those that involve the three actors at once, as for instance the messenger scene in the *Electra* (*see* below). Aeschylus in his later plays used the third actor to great effect, but in quite a different way, as when Cassandra is silent, though dramatically powerful, in the scene between Agamemnon and Clytemnestra. The difference is that the typical Aeschylean plot is linear; the Sophoclean, complexive. In the *Oresteia*, for example, Agamemnon takes vengeance on Paris, Clytemnestra on Agamemnon and on Cassandra, and Aegisthus on Agamemnon; Orestes avenges his father upon Clytemnestra and Aegisthus, and then has to contend with the avenging spirits of his mother. In such a drama characters collide rather than struggle with each other; they destroy, and are destroyed, in accordance with cosmic, and evolving, conceptions of retribution. There is little personal interplay; the characters are impressive in size and power, and therefore are not drawn in detail. The typical Sophoclean tragic action is one that is seen, as it were, not through the telescope but by the natural vision, though a vision that is assisted, and controlled, by the pervasive presence of gods. The chief agent does something involving grave error; this affects others, each of whom reacts in his own way, and thereby, perhaps, causes the chief agent to take another step toward ruin—his own, and that of others as well. The characters do struggle with each other; therefore they must be drawn in more detail. Peripheral figures too become involved in the central issue; plot must be more complex —though Sophocles avoids anything like subplot or decorative additions; the economy and concentration shown by his plays are most remarkable. This more complexive tragedy demanded a third actor. Equally, those who were to suffer from the tragic error were none of them of the next generation; all were present at the time. Therefore Sophocles abandoned the spacious Aeschylean framework of the connected trilogy; all was comprised within the single play. From then on, with very rare exceptions, "trilogy" meant no more than three separate tragedies presented at the same festival.

The development of character and plot naturally increased the histrionic and decreased the lyrical element in drama. Counting by lines, slightly more than one-half of the *Agamemnon* is either sung or chanted; of the *Antigone*, less than one-third, though this play is by far the most lyrical of the seven. But the chorus is still integral. No one formula can sum up its functions. As Aristotle remarked, in Sophocles the chorus is "a fellow actor"; therefore it will often share the illusions of other actors in the piece, as it does for instance in the *Antigone* and *Ajax*, where it takes, more or less, the point of view of Creon and Ajax respectively—that is, the wrong one. Sometimes it becomes the "ideal spectator," making philosophic comment (as in the solemn third ode of *Oedipus Rex*); sometimes it is a purely lyrical instrument, giving release from the tragic tension through the beauty of its dance and music.

Some of the odes, *e.g.*, the one just mentioned, or the second, third or sixth of the *Antigone*, place Sophocles in the front rank of Greek lyric poets. Of the music that he wrote nothing is known; about the choric dances this much can be inferred from their metres, that they must have been conceived in a very plastic fashion. As for the style of the spoken verse, it is reminiscent of Shakespeare's, responding directly to the dramatic needs of the moment; it is as far as possible from being a standard, uniform style, like Euripides'. It will be weighty or swift, intense or easy-going, highly-wrought or perfectly plain. It is no accident that Aristophanes, who parodied Euripides' and Aeschylus' style, left Sophocles' alone.

His Tragic Thought.—Sophocles has been universally admired for the sympathy and vividness with which he draws characters; especially notable are his tragic women: Electra, Antigone, Deianeira, Tecmessa and others. Also, few dramatists have been able to handle situation and plot with more power and certainty; the frequent references in the *Poetics* to *Oedipus Rex* show that Aristotle regarded this play as a masterpiece of construction, and few later critics have dissented. Further, he is unsurpassed in his moments of high tragic tension, and in his revealing use of tragic irony. In other respects there has been some tendency to compare him unfavourably with Aeschylus and Euripides: Sophocles, it has been said, was a supreme artist and no more; he grappled neither with religious problems like Aeschylus nor with intellectual and social ones like Euripides; he accepted the gods of Greek religion in a spirit of unreflecting orthodoxy, and contented himself with presenting human characters and human conflicts. Further, two of his plays, *Ajax* and the *Trachiniae*, have been adversely criticized in precisely that aspect of his art for which, elsewhere, he has been so much admired, namely in their structure: for though the *Ajax* seems to reach its natural conclusion in the suicide of the hero, the play continues with several scenes that concern the burial of his body;

while in the *Trachiniae*, if Deianeira is the central character, she too kills herself two-thirds of the way through and is subsequently hardly mentioned, while if Heracles is taken to be the hero, he does not even appear until his wife is already dead, when he himself also is on the point of death. For the "lack of unity" in the *Ajax* several inorganic explanations have been offered, the most naïve of which is that Sophocles was running short of material; about the *Trachiniae* the worst that can be said is that he simply combined two personal tragedies into one play.

The source of these misconceptions is the same. Sophocles did not write tragedies of character about individuals, but tragic drama about Man and the Gods; the religious element in his plays is not orthodoxy, not something added, out of personal piety, to a humanistic drama already complete; it is his presentation of the way in which the universe works—and it is worthy of attention. This is the reason why his gods are so stern, sometimes baffling, rarely benevolent; and the reason why he does not criticize them is that there is not much point in grumbling at the universe. Further, he did not, like Aeschylus and Euripides, make statements about his religious ideas; his thought is made implicit in the structure of the plays, and if his most telling strokes are considered only as examples of theatrical virtuosity, naturally the thought disappears.

Electra.—Two scenes from the *Electra* illustrate the fusion of his thought with his art. It is easy to dismiss the messenger scene as being no more than a clever piece of stagecraft. Clytemnestra has been terrified by a dream portending that her son Orestes will return from exile to avenge his father upon her and her accomplice Aegisthus and to reverse their usurpation. She comes out from the palace; she offers prayer and sacrifice to Apollo that he will avert the omen of the dream, divert it upon her enemies (meaning Orestes and Electra her daughter), frustrate them and preserve for her what her murder and adultery have won. She has only just finished the sacred rite when there enters, unrecognized, Orestes' old servant with a story—prearranged by Orestes—that Orestes is dead: it is designed to throw Agamemnon's murderers off their guard. The story of his supposed death in a chariot race at Delphi is itself a brilliant narrative; the situation enhances it, for it is being told to the mother who finds her own deliverance through the death of her son, and to the sister who finds in it the ruin of all her life; the messenger's preoccupation is so to deceive Clytemnestra that he may be invited into the palace to make ready for Orestes' coming. But this is by no means all, for an audience that believed in Sophocles' gods—being in any case assured already, through the chorus, that Zeus will bring the avenger home to triumph, and that retribution follows crime—would see in the arrival of the messenger, directly after the blasphemous prayer, the god's devastating answer to it. So it is too with the *coup de théâtre* in which the play ends. Clytemnestra has summoned Aegisthus from the countryside to hear the glad news. He enters, arrogant, secure at last from any threat of vengeance. He orders those who have brought the body to bring it out and display it publicly. "Here lies one," he says, "struck down by the anger of the gods." He is right; he draws back the shroud, and finds himself looking at the face of his dead wife. Whereupon Orestes drives him in to die on the spot where he had murdered Agamemnon. So do the gods work; crime breeds its own recoil.

It is the interfusion of divine and human action that gives to Sophocles' character drawing its deep significance. Electra is seen, for example, passing through the whole range of human emotion—from passionate love to cruel hatred, from numb despair to wild joy; the audience sees, in some detail, how her daily life with the murderers has entered like iron into her soul; she can live for nothing but vengeance and liberty. Orestes too, living on charity in exile, needs, and receives, no divine prompting: he asks Apollo not whether to do it but how to do it; however the god, though he approves, does not help. Sophocles is showing not that the gods from afar intervene to punish crime, but that the retribution comes in the natural order of things.

Antigone and Trachiniae.—In the *Antigone* the prophet warns Creon that all the gods are angry at what he has done and that their Avengers are lying in wait for him. What the prophet threatens comes to pass, not through "divine intervention" but through the instinctive reactions against Creon of Antigone, Haemon his son and Eurydice his wife. It is not the personal conflict between Antigone and Creon, great though it is, that is the deepest conflict in the play, but the conflict between Creon and the gods. He, for what he thought good reasons, ordained that the body of Antigone's traitor brother should not receive burial, the last tribute that humanity pays to humanity, but should be eaten by animals; Antigone rising in total rebellion, obeying all her instincts of love, loyalty and humanity, is indeed obeying "the unwritten and unchanging laws of the gods"; Creon is defying them, and they crush him. Individual characters and personal conflicts are in no way diminished or obscured by being set in this universal framework; on the contrary, they preserve to the full their own particular sharpness, and receive the added significance of being made universals.

The *Trachiniae* becomes unintelligible in structure only when we reduce it (perhaps under the influence of Aristotle) to personal dimensions. Sophocles makes it quite clear what the scope of the play is. At the beginning, Heracles' patient and understanding wife is distracted with anxiety at the ominously long absence of her husband. Reassuring her, the chorus observes that in human life joy and sorrow alternate like day and night; they follow the same universal rhythm. Therefore, having known little but sorrow, she may now expect joy. Moreover, Heracles is a son of Zeus, and Zeus does not neglect his offspring. But at the end of the play, she is dead and he is dying—ravaged by the "shirt of Nessus" that she had innocently sent him, to charm back his love—and their son Hyllus cries out upon the cruelty of the gods, who have brought this to pass. But the play has made it clear why, in this case, the universal rhythm broke down: to get a young girl for his mistress, in utter disregard of his wife, Heracles has destroyed a city, killing all the men and enslaving the women. This *hubris* is followed by a second, when he sends the girl to live in Deianeira's house. Deianeira was certainly simple-minded in trusting to a supposed love charm given her by an enemy of Heracles; Sophocles' point however is that ignorance is an inescapable part of the human condition: later, he makes Hyllus curse his mother, in his ignorance, for a vindictive murderess, and then repent bitterly when he learns the truth. Life at its best is a precarious affair, but it was the reckless violence of Heracles, acting within this precarious framework, not the cruelty of the gods, that brought all down in ruin.

Ajax and Oedipus Rex.—The hero of the *Ajax* is like Shakespeare's Coriolanus: he is a magnificent fighting man, a great bulwark of the commonwealth, who yet ruins all through an insensate pride that drives him into treachery and crime, and so to suicide. If this were all, there would be no need of the final scenes; but in these, too, Sophocles draws attention to the human condition and the demands it makes. Twice already has the arrogance of Ajax affronted the gods; now, when the prize of valour has been adjudged not to himself but to his enemy Odysseus, his instant response is to murder the judges, his commanders. In this he is frustrated by Athena, who visits him with an attack of madness, so that he kills sheep instead. Athena—no gracious Madonna—points to the humiliated Ajax and invites Odysseus to exult over him: "You see how strong the gods are!" But he, unlike Menelaus later, will do no such thing: "I pity him in his ruin, for I see that Man is nothing but an empty shadow." "Therefore," says Athena, "shun pride; it is the prudent that the gods cherish." Nemesis comes to the great Ajax when he falls on his own sword: he has lacked understanding of the way in which Man must live. So also do the two kings, when they order that his body shall be flung out to the animals—Menelaus, out of mean vindictiveness; Agamemnon, to vindicate discipline. But this is no way in which man should treat man. Once more it is Odysseus who shows understanding: no man, he says to Agamemnon, is always loyal and good; we must remember benefits and forget injuries. Any man may some time stand in need of forgiveness. Death is the common lot, and it must be respected. Agamemnon yields, but only as a favour to Odysseus, and Ajax's brother will not allow

Odysseus to have any share in the funeral rite, fearing the anger of Ajax's spirit; so that Odysseus reluctantly departs, the only one in the play who understands the human condition and shows true reverence towards the gods.

In the *Oedipus Rex,* the most famous of the seven plays, the scene of discovery, between Oedipus and the two shepherds—only 86 verses long—is surely among the most tense in all dramatic literature. The play is sometimes described as a tragedy of blind, inexorable Fate before which human efforts, even human virtue, are of no avail. This interpretation, would, however, contradict the whole trend of Sophocles' work, and in particular the third ode in this play. Sophocles insists always on the necessity for reverence, purity, moderation—though, being a tragic poet, he never pretends that these virtues will ensure security. One of the many tragic moments in the play is when Jocasta, jubilant before she learns the truth, proclaims that since life is random it is best to abandon principle and live from hand to mouth. The unifying thought in the play seems to be that since not even the greatest and most intelligent of us can control things, we should avoid the self-confidence that disregards religious restraints and leads straight to disaster.

Philoctetes and Oedipus Coloneus.—The *Philoctetes* concerns Neoptolemus the honourable but inexperienced son of Achilles who, on a clever plea of harsh political necessity, reluctantly undertakes a repulsive crime, only to find that it is much more repulsive than he had foreseen, and also completely useless. The shifting relationship between him and his intended victim and Odysseus (who in this play is a plausible rogue) is presented with the utmost delicacy and vividness; also, in the background is the suggestion that the Greek commanders themselves, by their past inhumanity toward Philoctetes, have caused their present frustration, from which they are trying to escape by even worse inhumanity.

The last play, *Oedipus Coloneus,* is also the most imaginative, and is surpassed by none in the power and beauty of its poetry. Oedipus, now old, blind, an outcast, his very name a source of horror, is summoned by the god to his final resting place at Colonus, and becomes, in his death, a mysterious source of defense to the land that has given him refuge.

The Satyr Play.—The *Ichneutae* is based on two stories about the miraculous birth of the god Hermes: that the infant, growing to maturity in a few days, stole cattle from Apollo, baffling discovery by reversing the animals' hoof marks, and that he invented the lyre by fitting strings to a tortoise shell. In this incompletely recovered play the investigators are the chorus of Satyrs, who are looking for the cattle; they are amusingly dumbfounded at the sound of the new instrument. Enough of the play survives to give an impression of its style: it is a genial, uncomplicated travesty of the tragic manner; the antics of the chorus were apparently the chief source of fun. It escapes from the intensity of tragedy into a world in which myth is taken quite literally, for the sake of relaxation and amusement—a very suitable conclusion to a day in the theatre, most of which had been occupied by the strenuousness of three tragedies.

See also references under "Sophocles" in the Index.

Bibliography.—On the manuscripts the standard work is A. Turyn, *Studies in the Manuscript Tradition of the Tragedies of Sophocles* (1952). The text of the seven plays was edited by R. C. Jebb (1897) and by A. C. Pearson, Oxford classical texts (1924). Of the commentaries, the best and fullest are those of R. C. Jebb, 7 vol. (1889–1914), each giving with the text an introduction, commentary and translation, supplemented by A. C. Pearson's three similar volumes on the fragments (1917). Jebb's translations, though archaic in style, are faithful. The Budé text is edited, with introductions, by A. Dain, with French translation by P. Mazon (1950–). *Oedipus Tyrannus* has been edited by J. T. Sheppard with explanation and translation (1920). *See* also J. C. Kamerbeek, *Plays of Sophocles* (1953–1959).

For interpretative criticism *see:* the relevant chapters in M. Pohlenz, *Die griechische Tragödie,* 2nd ed. (1954); Gilbert Norwood, *Greek Tragedy,* 4th ed. (1948; reprinted with corrections, 1953); H. D. F. Kitto, *Greek Tragedy,* 2nd ed. (1954) and *Form and Meaning in Drama* (1956); K. Reinhardt, *Sophokles* (1933); T. B. L. Webster, *Introduction to Sophocles* (1936); C. M. Bowra, *Sophoclean Tragedy* (1944); F. R. Earp, *The Style of Sophocles* (1944); R. F. Goheen, *Imagery of Sophocles' Antigone* (1951); C. H. Whitman, *Sophocles: a Study of Heroic Humanism* (1951); F. J. H. Letters, *The Life and Work of Sophocles* (1953); G. M. Kirkwood, *A Study of Sophoclean Drama* (1958); H. D. F. Kitto, *Sophocles, Dramatist and Philosopher* (1958). (H. D. F. K.)

SOPHONIAS, PROPHECY OF: *see* Zephaniah, Book of.

SOPHRON of Syracuse (fl. *c.* 430 b.c.), was the author of rhythmical prose mimes in the Doric dialect. Although the mimes survive mostly in fragments of only a few words, it can be seen from their titles, *e.g., The Tunny-fisher, The Sempstress,* etc., that they depicted scenes from daily life. One longer fragment deals with a magical ceremony. Plato thought highly of Sophron, who had some influence on Theocritus and also on Herodas.

For fragments *see* G. Kaibel (ed.), *Comicorum Graecorum fragmenta* pp. 152–181 (1899); and for the longer fragment *see* D. L. Page (ed.), *Greek Literary Papyri,* "Loeb Series," 2nd ed., vol. 1, pp. 328–331 (1942). (E. A. B.)

SOPOT, a town of northern Poland in Gdansk (Danzig) *wojewodstwo* (province), and a popular seaside and health resort. Pop. (1960) 44,000. It stands on the wooded slopes of an elevation cut by ravines. The town is a well-developed cultural centre, with three theatres, and higher schools of economics and music. It has good road and rail communications with Gdynia and with Danzig and the larger towns to the south. It is also a port for coastal shipping. Sopot is mentioned in historical documents of the late 13th century, and was a favourite resort of the inhabitants of Danzig during the 17th and 18th centuries. Its modern development as a resort dates from 1823; it acquired town rights in 1902. Sopot was incorporated within the boundaries of the Free City of Danzig by the Treaty of Versailles in 1919; it passed to Poland in 1945. (K. M. Wi.)

SOPRANO, in music, the highest human voice. The term is generally applied only to the highest of the three types of voice of female singers (and of the 17th- and 18th-century male *castrati; see* Castrato; Eunuch), "treble" usually being used for a boy's voice of corresponding compass. *See* also Voice.

SOPRON (German Ödenburg), a town in Györ-Sopron *megye* (county), Hungary, lies in a salient projecting into the Burgenland (*q.v.*), Austria, 140 mi. (225 km.) W of Budapest by rail. Pop. (1960) 41,246 (mun.). Known to the Celts as Scarbantia (or Scarabantia), and then a Roman municipium, it later absorbed much German settlement. The town spread from a nucleus of medieval buildings, including notably the late medieval Benedictine church and the Church of St. Michael. Sopron and its district were especially associated in Hungarian history with the Esterházy and Széchenyi families. After World War I the Burgenland was allotted to Austria, but Sopron voted (1921) to remain in Hungary. This created economic problems, for Sopron's trade went mainly toward Vienna, and during the 19th century it had become a nodal point of rail and road traffic. The town has grown little in the 20th century but has some industries including fruit preserving, sugar refining, and the production of cotton goods. Franz Liszt was born at Raiding nearby. (H. G. S.)

SOPWITH, SIR THOMAS OCTAVE MURDOCH (1888–), British aircraft designer and yachtsman, was born in London, Jan. 18, 1888. He taught himself to fly in 1910 and in that year won the de Forest prize for the longest flight to the continent. In 1912 he founded Sopwith Aviation Co., Ltd., at Kingston-upon-Thames and, flying a Blériot monoplane, won the first aerial Derby. In 1914 he designed and built the seaplane that won the Schneider trophy. During World War I his firm produced many military aircraft, including the "Pup," "Camel," "1½-strutter," and "triplane." Sopwith was chairman of the Society of British Aircraft Constructors, 1925–27. In 1935 he became chairman of the Hawker Siddeley Group, Ltd., which during World War II built the Hurricane fighter and the Lancaster bomber. They also built the first British jet aircraft, the Gloster E 28/39, powered by the Whittle engine (*see* Whittle, Sir Frank). Between the wars Sopwith raced such yachts as "Endeavour" and "Endeavour II." (D. Cr.)

SORANUS (fl. 2nd century a.d.), Greek physician born at Ephesus, was the chief representative of the school of physicians known as "methodists." He practised in Alexandria and Rome. Two treatises by him are extant: *On Fractures* (in J. L. Ideler,

Physici et medici minores, i, 1841) and *On Midwifery and the Diseases of Women* (first printed in 1838). The latter was the source of many works on obstetrics; the book also contains a chapter on the care and feeding of infants. Of Soranus' most important work, *On Acute and Chronic Diseases*, only a few fragments in Greek remain, but there exists a complete 5th-century Latin translation by Caelius Aurelianus.

SORANUS, an underworld deity worshiped in antiquity on Mt. Soracte, in Etruria north of Rome. Traditionally the Hirpini (*q.v.*) got their name from the fact that once as they worshiped Soranus, wolves (*hirpi*) carried off the sacrificial entrails (Servius, *ad Aeneid*, xi, 785). As priests the *hirpi* (or *hirpini*) *Sorani* celebrated a rite in which they marched over burning coals barefooted (Pliny, *Natural History*, vii, 2, 19). Soranus was identified with Dis and Apollo (*q.v.*) and had a female partner, Feronia.

(R. B. Ld.)

SORBONNE: *see* Paris, University of.

SORBY, HENRY CLIFTON (1826–1908), English pioneer petrographer and microscopist, was born at Woodbourne, near Sheffield, on May 10, 1826. His first paper, on sulfur and phosphorus in agricultural crops, was published in 1847. An early paper dealing with the origin of valleys in Yorkshire was followed by others on the physical geography of geological periods, rock denudation and deposition and the formation of river terraces.

In 1849 Sorby began to prepare thin sections of rocks for study under the microscope, thus demonstrating the value of petrography, the descriptive branch of petrology (*q.v.*). In 1857 he proved that slaty cleavage was produced by lateral pressure which caused the rearrangement of particles in shaly rocks. His memoir "On the Microscopical Structure of Crystals," published in 1858 in the *Quarterly Journal of the Geological Society of London*, made a strong plea for the adoption of microscopic methods in geology and illustrated their value.

In 1865 Sorby announced a new type of spectrum microscope to be used in study of organic pigments, and especially in determination of minute bloodstains. He studied the metallography of iron and steel, concluding that the latter was a crystallized igneous rock, and did related work of industrial value. His later geologic studies dealt with, among other subjects, the origin of limestones and other stratified rocks, and weathering. His publications in these various fields number 240. He died in Sheffield on March 9, 1908.

(C. L. Fe.; M. A. F.)

SORDELLO (*c.* 1200–before 1269), the most renowned of Italian troubadours, was born, *c.* 1200, at Goito, near Mantua, and became famous in 1224, when at the court of Richard of Bonifazio, at Verona, he abducted his master's wife, Cunizza, at the instigation of her brother. After this act (which was primarily political in purpose, whether or not it was rewarded with Cunizza's love), he went to Treviso, married Otta di Strasso and, pursued by the hatred of several families, crossed the Alps. He traveled, in the way of a troubadour, through Spain and southern France, sang the praises of many ladies, and, *c.* 1237, settled at the court of Raymond Berengar IV of Provence. After Raymond Berengar's death in 1245 he remained in Provence, and later became a companion of Charles of Anjou, with whom he returned to his native country in 1265 when the latter became Charles I of Naples and Sicily. In 1266 he was a prisoner in Naples. He died sometime before 1269, probably in the Angevin Kingdom of Naples.

Sordello, who is made the type of patriotic pride in Dante's *Purgatorio*, is also the subject of a poem by Robert Browning. He left 1,325 lines of a didactic poem *L'Ensenhamen d'onor* and 42 lyrical pieces, including 12 *canzos*, 6 *tensons* and *partimens*, and 7 *sirventès*, mostly love songs and satirical pieces. (*See* Troubadours: *Literary Forms*.) His *planh*, or lament, on the death of his patron Blacatz (Blacas; 1237), in which he invites all Christian princes to feed on the heart of the hero, is one of the masterpieces of Provençal poetry.

Bibliography.—C. de Lollis, *Vita e poesie di Sordello di Goito* (1896); E. Benson, *Sordello and Cunizza* (1903); *Sordello: le poesie*, new critical edition by M. Boni (1954).

(A. A. F. B.)

SOREL, ALBERT (1842–1906), French historian, noted for his study of diplomatic history, whose works greatly influenced historical research on the Revolution of 1789. Born at Honfleur, Normandy, on Aug. 13, 1842, he entered the Ministry of Foreign Affairs in 1866 and was attached to the diplomatic delegation of the government at Tours during the Franco-German War (1870–71). In 1872 he helped to found the École Libre des Sciences Politiques, where he taught until his death. Nominated general secretary of the Sénat in 1875, he devoted most of his time to the study of history. In 1889 he was elected to the Académie des Sciences Morales et Politiques and, in 1894, to the Académie Française. He died in Paris on June 29, 1906.

Sorel was the author of many books, notably *L'Histoire diplomatique de la guerre franco-allemande* (1875), *La Question d'Orient au XVIII^e siècle* (1878), *Montesquieu* (1887), *Madame de Staël* (1891), *Bonaparte et Hoche* (1896), and especially *L'Europe et la Révolution française*, eight volumes (1885–1904), in which he demonstrates that the Revolution continued the traditional policy of establishing and defending "natural frontiers," especially those on the left bank of the Rhine. France, in his view, encountered an equally traditional opposition from its neighbours, particularly from England, and was finally beaten by those to whom it claimed to have brought liberty and independence.

See G. Picot, *Albert Sorel* (1907); F. Hess, *Albert Sorel als Historiker* (1932).

(J. A.-S.)

SOREL, CHARLES, Sieur de Souvigny (1602–1674), French writer, author of *Histoire comique de Francion* (1623), a picaresque novel of the road in Burgundy and Paris, remarkable chiefly for its sense of life and vigour of language; it went through 30 editions before the end of the century. He also wrote *Le Berger extravagant* (1627), *Polyandre* (1648) and many shorter satires and scholarly and religious works, including one of the first histories of French literature, *La Bibliothèque Françoise* (1664).

Sorel was born in Paris. Guy Patin in his letters described him as a fat little man with a sharp nose, leading a quiet bachelor's life with his sister, planning more works than his health would allow him to write. His chief glory is to have inspired more than one scene of Molière. He died on March 7, 1674. (W. G. Me.)

SOREL, GEORGES (1847–1922), French social philosopher, an unorthodox commentator on Marxism, and the exponent of a doctrine of violence that has been misrepresented as one of the foundations of totalitarianism. He was born at Cherbourg on Nov. 2, 1847. He received scientific training, served as a government civil engineer until 1892, and always retained a lively interest in the philosophies of science and technology, exemplified even in such works as *L'Ancienne et la Nouvelle Métaphysique* (1894), *Les Préoccupations métaphysiques des physiciens modernes* (1905), and *De l'utilité du pragmatisme* (1917). In these he expounded a relativist and pluralist conception of knowledge, but his thought was never irrationalist.

Sorel in retirement devoted himself to writing articles for various French and Italian periodicals, but his writings were unsystematic and of uneven quality. Originally he had been a liberal conservative, as is evident from his *Procès de Socrate* (1889); but in 1893 he discovered Marxism and began writing the analytical critiques that constitute his most original and valuable achievement. He strongly opposed Karl Kautsky's determinist simplifications and not only showed himself sympathetic to Eduard Bernstein's revisionism but also interpreted Marxism as a philosophy of freedom and action.

After being passionately Dreyfusard in 1897 (*see* Dreyfus, Alfred), Sorel became disgusted with certain aspects of the political exploitation of "the Affair." By 1902 his thought had grown most extreme, and he enthusiastically supported revolutionary syndicalism, a movement with anarchist leanings stressing the spontaneity of the class struggle. His most famous work, *Réflexions sur la violence* (1908; Eng. trans. 1914 and 1950 [*Reflexions on Violence*]), was a philosophic commentary partly inspired by Henri Bergson. In it he developed his notions of "myth" (modeled on the syndicalist vision of a general strike) and of "violence." Violence for Sorel was the revolutionary denial of the existing order; but in describing its creative historical role he opposed it to "force," that is, to the state's power of coercion, the abuse of which he consistently denounced.

After 1909 Sorel's thought took a less certain direction. Feeling obliged to revise his revolutionary syndicalism, he adhered, not without embarrassment and hesitation, to the monarchist movement Action Française. Next, regarding World War I as a betrayal, he declared himself for Bolshevism in his *Plaidoyer pour Lénine* (1919)—with the more enthusiasm because he thought it a lost cause. He died at Boulogne-sur-Seine on Aug. 28, 1922.

Sorel is sometimes held to have inspired Communist and Fascist dictatorship. Yet he had no influence at all on Lenin, who mentions him only once, with contempt, for Sorel denounced everything that might subject socialist action to control by party leaders. Benito Mussolini frequently professed himself to be a disciple of Sorel, whose theories of "myth" and of "violence" he took to glorify the blind motivation of the mob and to justify mere physical brutality—distortions which Sorel explicitly condemned

See R. Humphrey, *Georges Sorel, Prophet Without Honor* (1951). (G. Gy.)

SORGHUM, a cereal, forage, and sirup crop plant grown in many countries and known botanically as *Sorghum vulgare*. It probably originated in Africa. The types grown for grain (grain sorghum) are called by various names, including durra, milo, shallu, kafir corn, Egyptian corn, great millet or Indian millet. On the subcontinent of India it is known as jowar, cholum, or jonna; in the West Indies as petit mil or Guinea corn; and in China and Manchuria as kaoliang. Sorghum is a strong grass, growing to a height of from 2 to 8 ft. or even 16 ft. The stalks and leaves are coated with a white waxy bloom. The pith in the stalks of certain varieties is juicy and sweet. The leaves are sheathing, solitary, and about 2 in. broad and 2½ ft. in length; the panicles or flower clusters are loose, contracted, or dense. Self-pollination of the flowers is common but considerable cross-pollination occurs. The grains may be either free or retained in the hulls after threshing. Many varieties are awned. The seeds are ellipsoid, rounded, or flattened and of varied sizes somewhat smaller than wheat grains. The seeds may be white, yellow, red, or brown. The hulls are mostly straw-coloured, red, brown, or black.


W. H. HODGE

SORGHUM (S. VULGARE)

Sorghum is the leading cereal grain in Africa and is important also in the United States, India, Pakistan, north China, and Manchuria. It is grown to some extent in the U.S.S.R., Iran, Arabia, Argentina, Australia, and southern Europe, as well as in other regions. It is best adapted to warm conditions and is very resistant to drought and heat. Hundreds of varieties are grown. The grain is similar in composition to that of corn (maize) except in being higher in protein and lower in fat. It replaces corn as a feed grain in hot, dry regions. For food it usually is ground into a meal and made into porridge, bread, or cakes. Natives of southern Africa refer to the product as "mealies." Whole grains sometimes are popped or puffed. The grain also is used in making an edible oil, starch, dextrose, paste, and alcoholic beverages. The stalks provide fodder and building materials. The sweet sorghums (sorgos) are grown chiefly in the U.S. and southern Africa for forage or for sirup manufacture. The sweet stalks are chewed by peoples of various countries. The broomcorn plant, belonging also to the species *S. vulgare,* is similar to other sorghums in adaptation and many characteristics; it is used in making brooms. (J. H. Mn.)

SORIA, a city and province of north-central Spain. The city, the provincial capital, lies on the Duero (Douro) River, 115 mi. (185 km.) NE of Madrid. Pop. (1960) 19,301 (mun.). Restored after the Moorish invasion, medieval Soria was ruled by the unique Doce Linajes (12 noble families who governed in turn). The churches of Santo Domingo and San Nicolás, the collegiate church of San Pedro, the cloister of the convent of San Juan de Duero, and other ecclesiastical buildings are fine specimens of Romanesque work. Near the Duero are the ruined citadel and 13th-century walls. The Numantine Museum houses important relics excavated from nearby Numantia (*q.v.*). Commercial activities include food processing (flour, dairy products, meat packing, wine), sawmilling, and the manufacture of cement, tiles, and soap. There are road and rail communications with Madrid, Saragossa, Burgos, and Pamplona.

Soria Province was formed in 1833 out of Old Castile. Area 3,972 sq mi. (10,287 sq km.); pop. (1960) 147,052. It is a bleak, austere region, bounded on three sides by mountains, and has a cold, dry climate. Most of the province is watered by the Duero and its affluents and crops are grown in irrigated areas. There are forests of pine, oak, and beech, and large tracts of grazing land (mostly for sheep). (J. A. P.-R. y G.-S.)

SORIN, EDWARD FREDERICK (1814–1893), French-U.S. Roman Catholic priest and educator, founder and first president of the University of Notre Dame, was born at Ahuillé, near Laval, France, on Feb. 6, 1814. Ordained priest in 1838, he joined in 1840 the Congregation of Holy Cross, a group of priests and brothers organized at Le Mans by Abbé Basil Antoine Moreau.

At the invitation of Bishop Celestine Hailandière of Vincennes, Ind., Sorin and six brothers went to Vincennes in 1841, at first settling at St. Peter's in Daviess County. In 1842 Hailandière offered Sorin land near South Bend, in St. Joseph County, where Father Stephen Theodore Badin had formerly conducted the mission Sainte Marie des Lacs. Sorin arrived there on Nov. 26, and in 1844 he obtained from the general assembly of Indiana a charter for the University of Notre Dame. He was president of the university until 1865, was provincial superior of his community in the United States until 1868, and was from then until his death, on Oct. 31, 1893, the superior general.

In 1843 Sorin established at Bertrand, Mich., near Notre Dame, a community of French Sisters of Holy Cross, and in 1854 secured for them the site adjacent to Notre Dame on which St. Mary's College was founded. He was instrumental in bringing to the Sisters of Holy Cross Mother Angela (Eliza Maria Gillespie), who guided the community for nearly 30 years. In 1865 Sorin began publishing *Ave Maria* magazine. (T. T. McA.)

SORITES. In traditional logic, a chain of successive categorical syllogisms in the first figure (*i.e.*, where the middle term is subject of the major premiss and predicate of the minor premiss; *see* Logic, History of: *Ancient Logic: Aristotle*) may be so related that either the conclusion of each except the last is the minor premiss of the next, or the conclusion of each except the last is the major premiss of the next. If then the intermediate conclusions are suppressed (*i.e.*, the conclusions of all the successive syllogisms except the last), and only the remaining premisses and the final conclusion are stated, the resulting argument is a valid inference from the stated premisses, which may be considered independently of its analysis into syllogisms, and which is called a sorites. In the case of the first alternative we have the so-called Aristotelian sorites, and in that of the second the Goclenian sorites (enunciated by the German logician Rudolph Goclenius, 1598). Considered as independent inferences, the Aristotelian and Goclenian are identical, except for the matter of the order in which the premisses are stated. But the Aristotelian and Goclenian analyses of the sorites into syllogisms are different.

An example, taken from Lewis Carroll, is the following. The premisses are: (1) No one takes in [subscribes to] the *Times*, unless he is well-educated; (2) No hedgehogs can read; (3) Those who cannot read are not well-educated. The expected conclusion is: (5) No hedgehog takes in the *Times*. To analyze this into syllogisms, we restate the premisses, with obversion of (2) and (3), as follows: (1a) None who are not well-educated take in the *Times*; (2a) All hedgehogs are unable to read; (3a) All who are unable to read are not well-educated. In the Goclenian analysis, the first syllogism has (1a) as major premiss, and (3a) as minor premiss, and hence the intermediate conclusion, (4) None who are unable to read take in the *Times*. Then the second syllogism has

(4) as major premiss and (2a) as minor premiss, hence (5) as final conclusion.

In a more complicated example there might be more than three premisses. In general, there may be $n + 1$ premisses, and analysis then yields a chain of n successive syllogisms.

See Lewis Carroll (Charles Lutwidge Dodgson), *Symbolic Logic*, 4th ed. (1897, reprinted 1958); J. N. Keynes, *Studies and Exercises in Formal Logic* (1884 and later editions to 1906). (Ao. C.)

SORØ, an *amt* (county) in southwest Sjælland (Zealand), Denmark. Area 571 sq.mi. (1,479 sq.km.). Pop. (1960) 129,580. This prosperous farming region with its fertile clay loams is drained by the Tude River in the west and by part of the Sus River system in the east. The principal towns are Slagelse (pop. [1960] 20,562), Ringsted (9,694), and Sorø (5,494), all inland, the important ferry port of Korsør (14,276) and the minor port of Skælskør (2,889). The administrative centre is Sorø. (Ha. T.)

SOROLLA Y BASTIDA, JOAQUIN (1863–1923), Spanish painter, whose style was a conservative variant of Impressionism, was born at Valencia on Feb. 27, 1863, and studied at the academy there, in Italy, and in Paris, where he especially interested himself in the works of J. Bastien-Lepage and A. von Menzel; he was also influenced by the north European realists, particularly A. Zorn. His early paintings were of history and social realism (one of the latter kind, "Otra Margarita," 1892, being his earliest success), but he later became well known for brightly lit scenes with Valencian peasants and fisherfolk and children playing in the surf, his style after about 1903–04 becoming impressionistic and summary, with heavily impasted pigments. Between 1910 and 1920 he painted portraits of Spanish writers and a "Panorama of the Forty-nine Provinces of Spain" for the Hispanic Society of America. Sorolla died in Madrid on Aug. 10, 1923.

SORORATE, a term introduced by Sir James George Frazer to designate all marriages with a wife's sister, whether in the lifetime of the first wife or after her death. In his view, generally accepted in the 1960s, it is complementary to the custom of the levirate (*q.v.*). The concept appears in the work of Sir Edward Burnett Tylor where levirate and sororate figure as correlates of the postulate that matrimony is a bond between families rather than individuals. A. R. Radcliffe-Brown subsumed both institutions under the principle of the social equivalence of brothers and sisters, respectively. These relationship terms are to be understood in a classificatory sense; *i.e.*, a more remote relative of the same sex may serve as secondary mate instead of a blood sibling. Though related in principle, the levirate and sororate are not invariably associated, but they usually are, and appear to be the commonest of preferential secondary marriages. Either may be permissive rather than obligatory. The Maricopa (Arizona) traditionally insisted on a widow's marrying a husband's relative, whereas the replacement of a deceased wife by a kinswoman of hers was customary, but not compulsory.

Though successive and simultaneous marriage of two or more sisters falls under the same principle, some peoples (*e.g.*, the Kazakh) favour one, while tabooing the other practice. Hence it has become necessary to distinguish between sororal polygyny and sororate. The typical rule for the former is that the husband of the eldest girl in a family marries her juniors as they come of age; L. H. Morgan found this usage in at least 40 North American tribes during the 19th century; more recently Navaho men often were simultaneously married to two sisters, occasionally to three. Australian aborigines recognized the same preemptive claim, but in many tribes the husband contented himself with the two oldest girls, conveying his claims on their younger sisters to his junior brother. With remarkable unanimity aborigines explain sororal polygyny on the ground that sisters are unlikely to quarrel as co-wives.

The effect of both the sororate and sororal polygyny is to have children extend the term mother to the maternal aunt, but this terminological trait is more probably directly correlated with unilinear descent.

See also Marriage, Primitive.

Bibliography.—A. R. Radcliffe-Brown, *The Social Organisation of Australian Tribes* (1931); R. H. Lowie, *Social Organization* (1948); George P. Murdock, *Social Structure* (1949); H. G. Nutini, "Polygyny in a Tlaxcalan Community," *Ethnology*, vol. iv (April 1965).

(R. H. Lo.)

SORORITY, a social, professional or honorary organization of women, usually a secret society whose name consists of a series of Greek letters and which is connected with a college or university. *See* Fraternity and Sorority.

SORREL, any of several herbs, some weeds, of the genus *Rumex* of the buckwheat family (Polygonaceae), especially *R. acetosa*, the garden sorrel; *R. acetosella*, sheep sorrel; and *R. hastatulus*, heartwing sorrel. The leaves of garden sorrel are used as greens in soups, salads, and sauces. French sorrel, *R. scutatus*, is a hardy perennial, distributed throughout Europe.

Species of *Oxalis* (*q.v.*) are commonly called wood sorrel or lady's sorrel. The sourwood (*Oxydendrum arboreum*) is known as the sorrel tree, an attractive deciduous tree of the heath family, native to the southeastern United States; its leaves turn a brilliant scarlet in the autumn.

SORRENTO, a city of Campania and in Napoli Province, Italy, stands on a peninsula separating the Bay of Naples, which it faces, from the Gulf of Salerno, 17 mi. (27 km.) SSE of Naples. Pop. (1961) 12,531 (commune). The backbone of the peninsula is formed by the Monti Lattari, the highest of which, Monte Sant'Angelo, rises to 4,735 ft. (1,443 m.). The cathedral has a marble side portal of 1479, and the church of St. Francis of Assisi has a picturesque 14th-century cloister. The Museo Correale di Terranova contains an important collection of Campanian decorative art, medieval sculpture, paintings, and classical remains. Sorrento is the seat of an archbishop.

Sorrento is linked by a branch line to the main railway from Naples to Salerno, and by rail and road with Pompeii and Mt. Vesuvius; steamers ply to Naples and the islands of Ischia and Capri. The district produces oranges, lemons, and nuts.

The ancient Surrentum, Sorrento has been a summer and winter resort since Roman times because of its agreeable climate and situation. Many writers, artists, and musicians have stayed and worked there; the poet Torquato Tasso was born there in 1544.

SORSKI, NIL (1433–1508), Russian saint and mystic, the first Russian mystic to write about the contemplative life and to leave a guide to the spiritual path of self-perfection, and the first Russian to advocate methodically in his writings asceticism and renunciation of worldly possessions. He was born in Moscow in 1433, but little else is known of his early life. He was tonsured in the monastery of St. Cyril at Beloozero and was present at the first council against the Judaizers (1490; *see* Joseph, Saint, of Volokolamsk). He made a voyage to Constantinople and Mt. Athos, where he studied patristic writings and acquainted himself with Hesychasm (*q.v.*) and the external forms of Athonite monasticism. After his return he founded his own monastery (*skit*) on the Sora River (hence "Sorski") near Beloozero, a small community bound together more by common worship than by strict discipline. In 1503, at a council in Moscow, he allegedly spoke against monastic landownership and is consequently known by historians as the leader of the so-called nonpossessor party within the Russian church (*see* Monasticism). He died on May 7 (his feast day), 1508.

Writing of the monastic life, Nil recommended moderation in fasting, in discipline, and in church decoration, stressing always the need for mental rather than physical asceticism. His writings, most of which are derivative, include a statement on the external precepts for the running of his monastery; a monastic rule (*Ustav*) in which instruction is given on how to analyze and overcome the passions on the way of self-perfection; and several epistles, usually containing words of comfort for his spiritual children.

See F. von Lilienfeld, *Nil Sorskij und seine Schriften* (1963), which contains German trans. of Nil's works. (J. L. I. F.)

SORSOGON, the name of a municipality, provincial capital, and the southernmost province of Luzon, Philippines. Area of province, 793 sq.mi.; pop. (1960) 347,771. The province is bordered on the south by San Bernardino Strait and comprises largely volcanic cones with broad rich level areas between; the highest peak is Mt. Bulusan (5,118 ft.), a recently active cone. Rainfall

varies from 100 in. to more than 160 in., with no dry season and maximum fall in the winter. Rice, sweet potatoes, and cassava are the main food crops produced, but the province is best known as an abaca (Manila hemp) centre. Before World War II slightly over 40% of the cultivated area was devoted to abaca, but by the 1960s this area was halved. Coconuts continue to occupy a significant proportion of the farmed area. Sorsogon (pop. [1960] 35,542) is a trading town, processing station for abaca and copra, and port on the northeastern shore of Sorsogon Bay. (R. E. HE.)

SØR-TRØNDELAG, a *fylke* (county) of central Norway. Area 7,227 sq.mi. (18,718 sq.km.). Pop. (1960) 206,802. It contains valleys and uplands south of Trondheimsfjorden, extending eastward to the Swedish border, and coastal districts west of the fjord, with the large islands of Hitra and Frøya. In the eastern fjord areas agriculture and forestry are favoured by fertile soil, but in the valleys agriculture is limited by climate and relief. Pyrite occurs in the districts of Røros and Orkdal. Røros (pop. [1960] 2,904), 2,065 ft. (629 m.) above sea level, is an interesting old mining village founded in 1644. In the coastal districts soil is more scarce and the hills more barren. Fishing is an important subsidiary occupation to agriculture. There are manufacturing industries in and around Trondheim (*q.v.*), which, being the site of the greatest medieval cathedral in Scandinavia, is by tradition a regional centre for a much wider area than the county. (L. H. HG.)

SOSNOWIEC, a town of southern Poland in Katowice *wojewodztwo* (province), and part of the conurbation of the Upper Silesian industrial region, is situated on the Czarna Przemsza River, a tributary of the Vistula, 6 mi. ENE of Katowice. Pop. (1960) 131,654. The town is a railway junction. Most of the inhabitants are employed in industry; and there are three coal mines, a number of heavy industrial plants (iron foundries, smelting works, machinery), and glassware and textile factories. The town has a mining museum, the first of its kind in Poland. Sosnowiec developed from a village into a factory centre of the Polish kingdom *c.* 1877. Its development is associated with the construction of the Warsaw–Vienna railway line in 1858. It obtained town rights in 1902. (K. M. WI.)

SOTER, SAINT, pope from about 166 to about 175. He wrote to the church of Corinth and sent it aid. His letter is mentioned in the reply given by Dionysius, bishop of Corinth, and A. Harnack thought that it could be identified with the so-called Second Epistle of Clement (but *see* CLEMENTINE LITERATURE). St. Soter's feast day is April 22.

SOTHERN, EDWARD HUGH (1859–1933), U.S. actor, widely popular for his roles in romantic comedy and noted also as a Shakespearean actor, was born at New Orleans, La., on Dec. 6, 1859, the son of the English comedian Edward Askew Sothern. His first stage appearance was in a small part with his father's company at the Park Theatre in New York City in 1879. He toured England in 1882–83, became leading comedian in John McCullough's company in 1883, and under Daniel Frohman was leading man at the Lyceum Theatre in New York. He married Virginia Harned in 1896 and in 1899 formed his own company with her as his leading lady.

In 1900 Sothern appeared in the title role of *Hamlet*, in 1901 in that of *Richard Lovelace*, and in 1902–03 as Villon in *If I Were King*, three of his greatest roles. In 1904 he played opposite Julia Marlowe (*q.v.*) for the first time in *Romeo and Juliet* at Chicago, Ill. Thereafter, except for two years, 1907–09, they appeared together on the stage almost continuously until their retirement. They were married in 1911. Besides *Romeo and Juliet* they played together in *Much Ado About Nothing*, *Taming of the Shrew*, *Merchant of Venice*, *Twelfth Night*, and *Macbeth* and in *Jeanne d'Arc*, *John the Baptist*, *When Knighthood Was in Flower*, and *The Sunken Bell*. Sothern had a repertory of over 125 diverse parts, but his audiences liked him best as the dashing, romantic hero in such plays as *The Prisoner of Zenda*. He died on Oct. 28, 1933, in New York.

See his autobiography, *The Melancholy Tale of Me* (1916).

SOTHO, one of the main cultural and linguistic groupings of the high grasslands of southern Africa. These Bantu-speaking Negro people were estimated to number more than 2,979,000 in the 1960s. They are classified as the Transvaal or Northern Sotho (Pedi, Lovedu, and others); Western Sotho (Tswana [*q.v.*] tribes); and Southern Sotho of Lesotho, the Basuto or Suto (*q.v.*).

See also AFRICA: *Ethnography (Anthropology)*: *Southern Africa;* LESOTHO: *The People;* BOTSWANA: *The People;* SOUTH AFRICA, REPUBLIC OF: *The People.*

BIBLIOGRAPHY.—In addition to the bibliographies of articles cited, *see* I. Schapera, *The Bantu-Speaking Tribes of South Africa* (1937), *Government and Politics in Tribal Societies* (1956); P. Duncan, *Sotho Laws and Customs* (1960). (J. F. M. M.)

SOUL. Man's experience of his own "inwardness," including his thoughts, desires, and feelings, is a characteristic and determining feature of the human form of existence. Whatever the forms of inwardness and consciousness that may be ascribed to other biological organisms, the structure of human experience is characterized by a reflexive inwardness; its determining mark is self-consciousness rather than mere consciousness. A corollary of this self-awareness is the capacity to objectify and to experience oneself as a subject *vis-à-vis* objects. However, it is not only the outside world or his physical body and its spatially localized and publicly observable activities that are objects for man's subjectivity. Man's subjectivity itself—*i.e.*, his own inwardness—can be objectified, both imaginatively and conceptually, and corresponding nonphysical objective entities (soul, mind, psyche, spirit) posited. But the further we push the objectifying analysis of man's faculties, drives, desires, feelings, and all his other "inward parts" and processes, the more imperative becomes the question as to the nature of that ultimate subject, or essence of personality, which does all the objectifying.

The quest for this elusive subject inevitably leads to an infinite regress. To escape from this regress some have denied the notion of a discrete soul substance or transcendental personality and have maintained that the concepts of ego, soul, or self merely reify the awareness of certain processes. The task of a critical philosophy would therefore consist in dereifying and dematerializing these images and concepts.

On a less critical and less sophisticated level, however, man has always been impressed by his dimension of subjectivity and by the manifold variety of the manifestations of his inwardness. Hence it is not surprising that man should frequently have taken a multiple rather than a unified, or holistic, view of his nature. Body and mind dualism is perhaps the best-known expression, in Western philosophy and religion, of this notion of discrete entities, but it certainly is not the only or the most obvious system of division. There are many different types and kinds of inwardness or of nonbodily selfhood, and a variety of psychological functions and vital phenomena (dreams, sleep, death, etc.) may give rise to notions of multiple souls. Without committing ourselves to E. B. Tylor's views regarding the extension of the notion of soul, as derived from immediate human experience, to disincarnate beings (spirits, demons, gods) and the attribution of souls to inanimate things (animism; *q.v.*), we may, nevertheless, agree that his description covers at least part of the facts:

> It seems as though thinking men, as yet at a low level of culture, were deeply impressed by two groups of biological problems. In the first place, what is it that makes the difference between a living body and a dead one; what causes waking, sleep, trance, disease and death? In the second place, what are those human shapes which appear in dreams and visions? Looking at these two groups of phenomena, the ancient savage philosophers probably made . . . the obvious inference that every man has two things belonging to him, namely a life and a phantom (*Primitive Culture*, 5th ed., vol. i, p. 428; London, John Murray, 1913).

The connection of the soul with the principle of life is also brought out by the terminology associating soul with "breath"; the Gr. *psyche*, Lat. *anima*, Heb. *nefesh*, *neshama*, and probably Sansk. *ātman* are all so associated. Sleep is often considered as a kind of minor death. Thus according to the Koran (6:60–61) God takes the souls of men back at death as well as every night during sleep; in the latter case, however, the souls are returned again in the morning.

Tylor collected a large amount of evidence for the belief in

multiple souls. The Fiji islanders know of a dark spirit, which ultimately goes to the underworld; and a light spirit, which can be seen as a reflection in water. Belief in duality of souls is widespread among North American Indians; among the Algonkians one soul goes out in dreams while the other stays behind; at death one remains behind while the other proceeds to the land of the dead. Tylor also quotes examples of three- and fourfold divisions. C. Nimuendaju has described the psychological system of the South American Apapocuva, who ascribe the different human characteristics to the several plant and animal souls inhabiting man. The idea of a multiplicity or even hierarchy of life- and soul-principles is found in higher civilizations too; it may be seen in, for example, the *ba* and *ka* of the ancient Egyptians, the Gnostic *psyche* and *pneuma*, or the fivefold division of the Jewish Cabalists (*nefesh, ruah, neshamah, ḥayyah, yeḥidah*). The Tikopia of the Solomons, on the other hand, know of one soul only, the *mauri* or *ora*, which after death becomes an *atua* soul.

Traditional definitions of the soul have usually emphasized man's consciousness of his psychological and mental processes, and the term was generally taken to denote the principle of thought and action in man, or as that which thinks, wills, and feels, conceived as a perdurable entity and a subject of conscious spiritual experience. Modern psychology has added a further dimension by drawing attention also to unconscious faculties, mechanisms, and experiences. Great variety in terminology and in conceptual traditions makes it difficult to generalize, though it is possible, perhaps, to distinguish between the notion of the soul as (1) the whole man minus the body, and (2) a special substance or collection of several substances. In the former case it is the whole, though disincarnate, man that survives and, *e.g.*, goes to the land of the dead; in the latter case it is a specific soul substance that persists and returns to its heavenly home or, alternatively, haunts the living or is reincarnated.

Science and Psychology.—As a scientific concept, soul did not require specific justification as long as its legitimacy was taken as axiomatic and accepted on general philosophical grounds. The division of man into *soma* and *psyche*, or even of the whole of reality into *res extensa* and *res cogitans*, as by Descartes, presupposed the existence of mind as an independent principle. The dichotomy gave rise to serious philosophical and logical problems, however (*see* BODY AND MIND), which came to a head in the 17th century. One line of thought may be traced from Plato to Leibniz and to the subsequent idealistic and vitalistic philosophies, but it also includes Aristotle's concept of the soul as the form, or entelechy, of the body. Aristotle's teaching, however, also has affinities with the more positivistic and naturalistic views, since it holds the soul to be inseparable from the body and consequently permits of a behaviourist (functionalist or even epiphenomenalist) theory of mind.

This second line of thought, which leads to naturalism and empiricism, has its philosophical antecedents—as far as modern thought is concerned—in Locke, though more recent research would hardly subscribe to the view that the mind is merely a passive and purely receptive organ for ideas which enter via the sensory apparatus. Whatever the correct view about ideas, it seems certain that forms of experience and behaviour are to some extent "given" in the cortex and in the central nervous system. Neurophysiology, electroencephalography, brain surgery, psychiatry, and research into mental states under the influence of LSD and other psychopharmaca have shown the fruitfulness of the naturalistic approach while not necessarily excluding some notion of mind or psychic activity. In fact, radical scientific naturalism, or criticism of the Lockean type, simply evades certain issues (*e.g.*, human inwardness, or the experience of freedom) by declaring these "introspective" areas out of bounds for the scientist or scientific philosopher and turning them over to the "nonscientific" philosophers, who apply to them their methods of phenomenological and existential analysis.

Up to a point the quarrel is either a semantic or a methodological one, the latter bearing on the delimitation of the proper areas of meaningful scientific research. In point of fact, however, these methodological or semantic considerations spill over into philosophy and ideology. The question is not merely one of determining the relation of psychology to philosophy (viz., philosophical anthropology) or theology, but of defining the scope of psychology. Is it the study of the manifestations of the soul, or the measurement of behaviour, or something in between these extreme positions? Or are the terms soul and mind merely convenient shorthand, standing for the fact that we are aware of mental activities and processes? A conditioned reflex psychology of the Pavlovian type (*see* CONDITIONING), or a strict behaviourism (*q.v.*) as conceived by J. B. Watson can discard mind as an unnecessary hypothesis. More recently cybernetics (*q.v.*), with its self-directing and self-controlling machines, has been claimed to provide possible models for an understanding of man.

The naturalistic, reductionist, and, as it were, soulless psychology (psychology without *psyche*, as some have said) has provoked other psychologists to emphatic assertions of the "reality of the soul" (C. G. Jung). These assertions, though impressive as testimonies to a profound awareness of man's psychological dimensions, have so far exhibited little or no capacity to deal with the logical and methodological challenge of naturalism and have certainly failed to distinguish sufficiently between psychological and spiritual dimensions of existence. Philosophical anthropology, with its phenomenological and existential analyses of the human mode of being (M. Plessner, M. Scheler, M. Heidegger, A. Gehlen), tries to bring out more clearly the specific spiritual character of human existence without, however, necessarily reifying it into spirit or mind, let alone into soul or immortal soul substances.

Psychical Research and Parapsychology.—Curiously, the allegedly most scientific attempt to establish the reality of souls (or of spirit or mind) is, at the same time, the most unspiritual and crudely material. Spiritualism, psychical research, and to a considerable extent, also parapsychology or metapsychology claim to produce evidence of the manifestations of spirits (souls) in mediumistic trances, materialization phenomena, tablelifting or tapping, ouija-board writing, reincarnations. (*See* PARAPSYCHOLOGY; PSYCHICAL RESEARCH.) The nature of the findings, as well as their interpretation, is still doubtful; it seems clear from a logical point of view, however, that while certain parapsychological phenomena (*e.g.*, telepathy) could, if properly established, prove the operation of laws other than those traditionally recognized by the physical sciences, yet they would still be far from proving the existence of spirits or individual souls. Altogether the ideological bias of psychical research is evident from the openly admitted belief that its findings should save modern civilization from its materialism and reawaken it to an awareness of the reality of mind. Thus J. B. Rhine, whose parapsychology laboratory at Duke University was well known, has declared that if his studies established the existence of "discarnate spirit personalities" (man's "living spirit") then there would be "dramatic proof of spiritual realities in everyday life that Science had failed so far to find. The machine theory would be wrecked once and for all." In fact, "proof of survival would squelch forever the dreadful error of the materialistic view of man on which Communism and other gross misconceptions about humanity rest. On the other hand, certainty about it could revitalize religion [*sic*] and give to all human life a new dimension" (Rhine, in *The American Weekly*, Dec. 7, 1957, pp. 7 and 32).

Western Philosophy and Religion.—Reference has been made in the preceding paragraphs to some of the philosophical presuppositions and implications of the concept of soul. In the history of Western philosophy, ideas regarding the nature and essence of the soul derive partly from Plato and Aristotle and partly from religious traditions (themselves indebted to philosophy) and preoccupations. While Plato held the soul to be eternal and preexistent, hence also capable of reincarnation, Aristotle's more biologically oriented definition implies that the soul is inseparable from the body. In fact, if soul is the technical term for the specific "form" of an organism, then human organisms are not the only ones to be endowed with souls. The Aristotelian tradition thus distinguished between vegetative, animal, and rational souls. Whereas the animal soul (also in a human individual) is born and perishes with the body, mind, or intellect (Gr. *nous*), which

is unoriginated and imperishable, enters the soul-germ from outside and remains unaffected by the death of the body. The combination of *nous* with the animal soul constitutes the human personality. Aristotle distinguished a passive and an active soul. Only the latter is eternal, whereas in the former Aristotle evidently hoped to find the link connecting *nous* and the animal soul.

It is not quite clear how Aristotle conceived the personal survival of a disembodied *nous*. Later representatives of the Peripatetic School, such as Alexander (*q.v.*) of Aphrodisias (*c.* A.D. 200), and some medieval Arab and Jewish Aristotelians denied the immortality of the soul even more uncompromisingly and explained the active *nous* (Latin *intellectus agens*) as the divine spirit working upon the soul (*see* ARABIC PHILOSOPHY: *Philosophers;* JEWISH PHILOSOPHY: *Aristotelianism in Medieval Jewish Philosophy*). Man's original endowment therefore consists of the disposition to thought ("potential" intellect) only. Immortality has to be acquired; it depends on the degree of development of the potential intellect; *i.e.*, on the measure of "acquired intellect" realized under the influence of the divine mind.

Neoplatonism had absorbed Aristotelian elements, and it is not surprising therefore that for Plotinus (*q.v.*) *nous*, the first product of the highest Being (the One), should be above the soul. In fact, *nous* creates the world soul, which in its turn creates and comprehends a plurality of particular souls. The Neoplatonic notion of an impersonal *anima mundi* has obvious pantheistic affinities (*see* PANTHEISM), while the idea of a return of man's particular soul to its original abode in the higher All Soul has definite religious qualities. To the extent that the return of the soul to its celestial or supramundane abode is conceived as a union (or reunion) with the godhead, Neoplatonism also has affinities with mysticism. Indeed, most systems of mysticism presuppose, by implication or by explicit teaching, specific psychologico-philosophical doctrines regarding the nature of the soul.

The Stoics did not teach immortality; for them the soul consisted of matter, though matter of the purest and noblest kind, part of the divine fire substance which descended from the ether into the bodies of men.

Western Christianity was influenced by the Neoplatonic tradition in its Latin form (St. Augustine) rather than in its pagan (Plotinus) or Greek Christian (pseudo-Dionysius) forms. When Aristotelianism penetrated the Schools, it cost St. Thomas Aquinas and others considerable effort to combine this teaching with the Christian belief in a created soul and in individual immortality. A radical distinction was also made between the "essence of the soul," in which the workings of divine grace take place, and the "faculties of the soul," with which psychology deals. The substance or essence of the soul (also *fundus animae*) is hidden, and only its actions through its faculties are manifest and phenomenally accessible.

Eastern Philosophy and Religion.—Just as in Western philosophy there is no agreement about the exact connotations of such terms as soul, spirit, or mind, Indian psychology too has a variety of terms which, though not synonymous, refer to the immaterial part of the individual: *jīva, citta, mano, vijñāna, ātman.* Generally speaking the *ātman* is considered, at least from the Upanishads on, as the essence or "self" of the individual and as that which underlies his consciousness, survives his death, and constitutes the continuity of the self in the cycle of transmigrations. According to the dominant schools of thought, *ātman* as the inner essence of the human personality is identical with the *brahman*, the "world soul"—the Absolute and the divine ground of all being. While this necessarily brief and superficial summary neglects the great variety of doctrines held by the divergent schools, it may nevertheless suffice to indicate the main trend of Indian thought on the subject and to give the background for the general belief in reincarnation (*see* below) which is assumed by all the religious systems of India. (*See* further INDIAN PHILOSOPHY.)

The Buddhist scriptures deny the existence of an *ātman* (Pali *attā*), and the doctrine of nonself (Sansk. *an-ātman*, Pali *an-attā*) is generally admitted to be one of the chief tenets of Buddhism. Disregarding both the problem of the origin of this doctrine and the question whether the Buddha himself engaged in explicit polemics against the traditional Upanishadic notion of *ātman*, it may suffice here to say that Buddhism regards the human individual as an agglomeration of five *skandhas* ("heaps" of elements). Insight achieved through meditation reveals the illusory character of the self. To assume the existence of a self is merely to reify as a substance the fact that there are mental processes. In this respect Buddhist psychology merely says, in its own fashion, what Locke or Hume or Gilbert Ryle have said in the terminology of European critical philosophy. Where Buddhism goes further is in preaching the necessity of this meditative insight as an indispensable condition for salvation. However, since Buddhism, while denying an individual self or *ātman* yet shares the general Indian belief in transmigration, the problem inevitably arises as to what exactly it is that survives to be reincarnated, that suffers transmigration, and that seeks release in nirvana (*q.v.*) from the wheel of *samsara.* It may be mentioned in passing that some Hindu apologists (*e.g.*, A. K. Coomaraswamy, *Buddha and the Gospel of Buddhism*, 1916) claim that the Buddha missed the basic Hindu teaching, according to which it is precisely not the *ātman* but a kind of "soul or subtle body" that transmigrates; hence the Buddhist doctrine, far from being revolutionary, actually repeats the *true* Hindu teaching.

Transmigration and Reincarnation.—Belief in transmigration, which is taken for granted in the major Asian religions and philosophies, was never dominant in the West. The doctrine was held in several Gnostic sects but subsequently disappeared or went underground. The Platonic and so-called Pythagorean doctrines, though cultivated in the Neoplatonic schools (Plotinus), exerted little influence on a Europe that was being conquered by Christianity. *Metensomatosis* (better known under the less felicitous term metempsychosis [*q.v.*]; other common terms are Gr. *metangismos* = Syr. *taspikha*, and Lat. *revolutio*) was a major Manichaean doctrine (*see* MANICHAEISM) derived, according to the 10th–11th century Arab scholar al-Biruni, from India. Its reappearance in the south of France in the 12th century (*see* CATHARI) is generally held to be due to influences of a Manichaean type. At about the same time and in the same area the doctrine also appeared in Jewish mystical circles and henceforth became a cardinal tenet of the Cabala (*q.v.*). It was revived in Christian Europe together with other Neoplatonic ideas during the Renaissance (*e.g.*, Giordano Bruno) and was subsequently taken up by Christian Cabalists and theosophists. In fact, most theosophical systems, whether ancient or modern, evince what appears to be a natural proclivity for linking their various spiritual metaphysics and cosmologies with the doctrine of reincarnation (*see* THEOSOPHY). Modern spiritualism, though it started in America, in its subsequent theoretical elaboration and theosophical development deliberately drew on Indian philosophy and on what was held to be the occult wisdom of the East. Contrary to Indian, Manichaean, and Cabalistic theories, however, most modern protagonists of the doctrine believe in human reincarnation only and reject the idea of transmigration into animal bodies. The doctrine was rejected as heretical by the Roman Catholic Church, which also condemned modern spiritualism (reply of the Holy Office of April 24, 1917). For a rejection of spiritualism from a fundamentalist Protestant point of view *see* L. E. Froom, *The Conditionalist Faith of Our Fathers* (1965).

In primitive religions a belief in what may be termed reincarnation appears in two main forms. One assumes that souls continue the round of existences in various natural species—often in insects, because these are found in or near the grave or the wooden soul-effigies. (This idea should not be confused with the notion that the soul itself appears as a small animal, particularly as a bird.) The other is connected with ideas about the relation of the living to the dead of the clan or tribe. Then it may be believed that pregnancy is due to an ancestral soul's entering the mother's body; in order to have a child, a woman would therefore pass, or worship at, a place inhabited by or associated with the souls of ancestors.

Immortality, the Afterlife, and Resurrection of the Dead. —*Dead and Living.*—It should be clear from the above that practically all religions entertain some notion of a soul. These notions

are based partly on prephilosophical, primitive ideas and partly on elaborations of philosophical concepts and their application to the requirements and experiences of religion. Among primitives some kind of survival is usually taken for granted, and the well-being of departed souls and their orderly progress to the land of the dead, as well as the prevention of their return as harmful spirits, are assured by funerary and seasonal rites. The dead visit the living, watch over their well-being, and, conversely, can avenge themselves if not paid sufficient attention (*see* ANCESTOR WORSHIP). They can also manifest themselves in various ways and even communicate with the living, *e.g.*, by means of spirit possession and in the institutionalized forms of shamanism (*q.v.*). Mediumistic contact with the dead as practised by spiritualism is in fact a modern version of shamanism but lacking its cultural sanction and institutionalized social function. The practice of communicating with the souls of the departed has given rise to a variety of new cults both in the West (Spiritualist churches) and in the East; the Cao Dai movement in Indochina is an interesting case of an Eastern cult heavily influenced by European spiritualism. Instances are also known of marriages of a living individual to a ghost partner.

In shamanistic and mediumistic contact with the spirit world, the distinction between departed souls and spirits of a different kind (gods, demons) is not always clearly made. The souls of departed saints, good spirits, and angels may become identified, and conversely the distinction among the souls of wicked men, evil spirits, and demons may become blurred. Spirit possession may be due to either, and rites of exorcism may be held to expel a demon as well as a former human soul (*e.g.*, a *dybbuk*). For the sake of terminological clarity, the following distinctions, drawn by R. Firth (*The Fate of the Soul*), may nevertheless be found helpful: whereas "soul" refers to the more or less immaterial survival-personality of a human individual, the term "spirit"—while including the soul—also denotes other categories of immaterial entities. The term "ghost" is generally used of the human survival-personality in its apparitional form after death.

Character of Life After Death.—Not all ancient religions took the permanent personal continuity of the individual soul for granted. In many primitive religions individuality is thought to persist for a few generations only; thereafter the departed soul joins the anonymous host of ancestors. But whereas modern man may have difficulty in entertaining the notion of an immortal soul, many primitive and archaic cultures had difficulty in conceiving the possibility of sudden and complete cessation of personal existence. Even when death was regarded as the final and inexorable end of human life, this end was not thought of as complete cessation but rather as a kind of existence in a bleak world of shadows —the Hades of the Greeks or the She'ol of the Hebrews (*see* HADES; HELL). Thus the psalmist admits that, as the shadowy doubles of their former full-blooded selves, "The dead do not praise the Lord, nor do any that go down into silence" (Ps. 115: 17), and when Saul wished to consult the dead Samuel the medium at Endor "brought up" the prophet's apparition from the underworld (I Sam. 28:8–14); she did not call down his soul from the upper spheres. Likewise did the weeping apparition of Patroclus come up to Achilles from the underworld (*Iliad* xxiii).

While the folklore and the ritual practices of the ancient world testify to a belief in some kind of afterlife, they hardly permit the generalization that an eternally blissful state was the normal expectation. Some civilizations seem to have held that only the virtuous (particularly warriors fallen in battle) would go to a kind of paradise, while others would suffer in a kind of hell. The view that the moral, spiritual, or ritual achievements of man in this life are insufficient to assure salvation and a blissful state for his soul is found almost everywhere. Hence almost all religions possess a variety of funerary rites and mortuary cults; *e.g.*, Roman Catholic practices intended to obtain remission of temporal punishment for poor souls in purgatory, Hindu *antyeshti* and *shrāddha* ceremonies, and the Parsee *baj* and *afringan*. The soul of the departed can benefit not only from rituals and prayers directly concerned with it but also from other works of merit (*e.g.*, acts of worship or charity, chanting of Holy Scriptures). The idea underlying these vicarious practices is the accumulation of disposable merit that can be transferred to the soul of the deceased. Indeed, many Mahayana Buddhist rituals are simply masses for the dead.

Immortality and Resurrection.—Significantly, when the ancient Jews began to feel that their historical, collective eschatology and their belief in a national restoration under a messiah were not enough—since they left the question of the destiny of the individual unanswered and failed to give a satisfactory account of the ways of divine justice in meting out reward and punishment—they evolved, or rather borrowed from Persia, the notion of a resurrection of the dead. Clearly the doctrine of the resurrection of the body implies a very different type of anthropology from that which produced the notion of an immortal spiritual soul substance.

The typical form of the doctrine of immortality as found in, for example, the Orphic and Platonic Greek tradition and in India is opposed to the idea of resurrection in two respects. In the first place, it tends to regard the independent spiritual substance which is the soul as the essential part of man. The body is merely its material clothing, and a vile and inferior one at that. The body is a prison (*ta sōma sēma*, as the Orphic proverb has it), and a considerable amount of philosophical and mythopoeic ingenuity had to be spent on accounting for the "descent" or "fall" of the pure, spiritual, and heavenly soul into this miserable body. Where, as in Christian teaching, the soul is not thought of as preexistent, the problem is, of course, different (*see* CREATIONISM AND TRADUCIANISM). It is obvious that on these premises the best hope that can be held out to the individual is the final separation, which is tantamount to liberation, of the soul from the body and its ascent to its celestial home. In fact, many systems of mystical contemplation and Yoga-like disciplines are designed to anticipate death and to realize the liberation from the body in this life. On this view it would be a most unfortunate fate for a soul to have to rejoin a body at the end of time. A similar evaluation of bodily life also underlies most theories concerning the transmigration of souls.

The doctrine of resurrection, on the other hand, in spite of its materialism and irrationality which have scandalized so many moderns, implicitly safeguards two fundamental biblical principles. The one is the integral or "holistic" conception of the human personality as an animated body: flesh vivified by the spirit. Resurrection of the body, if it means anything at all, can mean only that there is no way of talking of an eternal individual destiny unless one is prepared to define "individual" as this particular person with this particular body. The other point about resurrection is that it adheres to the pattern of an eschatological consummation of the process of history (*see* ESCHATOLOGY). The resurrection of Christ, in traditional Christian belief, is such an eschatological event inaugurating the new *aion*. In fact, some modern Christian theologians (*e.g.*, Karl Barth, Reinhold Niebuhr) insist that the symbol of the resurrection expresses the nature of the biblical message more adequately than does the "Greek" hope of immortality. It should be added that early Christian anthropology (St. Paul) was not based on the dualism body-soul but on that of flesh-spirit. Yet in due course the latter dualism became identified with the former.

The tension between the two conflicting anthropologies is well illustrated by the compromise solution according to which resurrection means the return of the soul to a spiritual ("glorified") body. Of course these contradictions appear as such only in a rational analysis bent on distinguishing "ideal types." In actual practice and in the eyes of the ordinary believer the divergencies are as good as nonexistent: individual souls are provisionally judged after death and go to heaven or, alternatively, to hell; from this discarnate state they will be called to the final judgment and resurrection. This view is characteristic not only of traditional Judaism and Christianity (Jesus adopted the Pharisaic dogma of the resurrection, which was denied, as unscriptural, by the Sadducees) but also of Zoroastrianism. According to the latter, man's soul (*fravashi*) is judged provisionally after death, pending a final purgation and resurrection (*see* ZOROASTRIANISM). The Jewish Cabalists even combined belief in a hereafter, the

transmigration of souls, and the resurrection of the body.

Release From, and Acquiring, Immortality.—According to the more fully developed Indian systems of thought, immortality was, as it were, natural (*samsara*), but as long as it meant being bound to the wheel of *karma* (*q.v.*), it was negative. The aim was not immortality but liberation. For the ancient Egyptians, on the other hand, immortality had to be laboriously acquired, by means of a ritual assimilation of the departed to the resurrected Osiris (*q.v.*). This rebirth to immortality was at first the privilege of the king (himself divine) but subsequently became accessible to nobles and even to commoners. In fact, "immortal" was synonymous with "divine," since, in Greek as well as Egyptian religion, it was precisely their mortality that distinguished men from the gods, who otherwise could be very similar to human beings and even of common origin with them. Hence the gift of immortality, as bestowed, for example, by initiation into the mystery cults (*see* MYSTERY), was tantamount to deification. Immortality was the boon of salvation held out by the mystery religions, and some of the related notions and practices of antiquity may have affinities with the Christian formulation of salvation as rebirth to immortality by sharing in the resurrection-life of Christ.

Perfect and Imperfect.—Whatever the philosophical difficulties of justifying the notion of an immortal but not eternal (since created) soul, once the idea is accepted it almost automatically becomes the focal point of man's speculations about his spiritual, moral, and intellectual nature and destiny. Since the body, matter, and this world in general are but the temporary housing of the spiritual soul, and moreover a source of temptation and of entanglement with lower things, this preoccupation was often accompanied by contempt of, or indifference to, things bodily, material, and temporal. Religion being generally considered to be concerned with things eternal and nontemporal, its main business was the cure and the salvation of souls. This cure of souls could assume philosophical, ethical, or ritualistic (*e.g.*, sacramental) emphasis. Thus the Aristotelian tradition, which associated the immortal part of the soul with the intellect, clearly conceived of moral perfection *sub specie intellectus* (Averroës, Maimonides). For many Muslim philosophers and mystics *nafs* (soul, *psyche*) denoted man as a living organism, the bearer of appetites and drives directed toward the body, whereas *'aql* (intellect, *nous*) was the truly spiritual principle, alien to this world and striving to return to its heavenly home. Where, on the other hand, man's spiritual calling was defined in terms of piety, love of God, faith, obedience, or good works, it could happen that intellect and reason were decried as of the devil. Man's moral conscience too—God's voice in the heart of man or, in the words of St. Bonaventura, *praeco Dei et nuntius*—could, accordingly, be regarded as a function of reason (possibly of a "practical reason," as in Kant) or, on the contrary, of a special moral faculty of the soul.

In many religions the division between the pure and perfect and the impure and imperfect is not identical with that between soul (or intellect) and body. There are evil ("fallen") souls; and bad *karma*, due to earlier evil deeds, may result in a degeneration of the soul. Adam's fall, according to Christian teaching, constitutes the Original Sin as a result of which human nature, and the soul in particular, is corrupted; though Roman Catholic and Protestant dogmatics differ as to the degree of this corruption.

Summary.—The idea of a soul can be viewed from different angles. As a scientific concept it seems useless (except, perhaps, as a hypothetical construct for theorizing about the facts alleged by psychical research). In primitive and some higher religions, souls, like ghosts, are simply and naïvely taken for granted, though reasons of a philosophical and psychological kind may be adduced subsequently. From the point of view of the theologian and the moralist (*e.g.*, Kant, Butler, Royce) the assumption that souls exist and that they are immortal renders a moral view of God and of human life and suffering possible. The belief owes its strength to the fact that the notion of a soul functions as one of the symbols—concretizing and reifying, no doubt, like all symbols—of man's self-consciousness and of his awareness of the mystery of his individuality and of the spiritual dimensions (including inwardness, mentation, morality) of his existence.

Ideas about the nature or origin of the soul serve to determine man's image of himself, as an individual being or as a member of a clan, tribe, religious community, etc., and to explain his character or fate. Relations with the souls of ancestors, or the departed in general, confirm social solidarity and group consciousness and may impart a sense of continuity to society. Doctrines regarding the future of the soul function as moral sanctions not only in an individual but also in a social sense: they serve to separate the sheep from the goats in some more ultimate realm, whereas in our present world the social lines of division do not coincide with those envisaged by the moral system. Doctrines regarding the hereafter may therefore be said to assert as absolute and definite those values which society affirms although individual and historical experience of actual life may have failed to validate them.

"Now that Sartre has replaced Dante as our eschatological authority" (R. Firth, *The Fate of The Soul*, p. 14; Cambridge University Press, 1955), statements about the soul and its fate in the hereafter tend to be interpreted as statements about the human condition and as expressions of attitudes about behaviour in this world. Contemporary opinions regarding the soul range from outright denial on logical-analytical or naturalistic-positivistic grounds, through the modern, existentially demythologized interpretation of the ancient images and concepts, to acceptance of the traditional doctrines of the great religions.

See also BODY AND MIND; METEMPSYCHOSIS; IMMORTALITY; and references under "Soul" in the Index.

BIBLIOGRAPHY.—E. B. Tylor, *Primitive Culture*, 2 vol., 2nd ed. (1958); J. G. Frazer, *The Belief in Immortality and the Worship of the Dead* (1913 *et seq.*); Åke Hultkrantz, *Conceptions of the Soul Among North American Indians* (1953); R. Firth, *The Fate of the Soul* (1955); E. Rohde, *Psyche* (1953); W. F. Otto, *Die Manen*, 2nd ed. (1958); F. Cumont, *After Life in Roman Paganism* (1922); E. Abegg, *Indische Psychologie* (1945); E. J. Thomas, *The History of Buddhist Thought*, 2nd ed., pp. 92–106 (1951); H. Günther, *Das Seelenproblem im älteren Buddhismus* (1949); C. J. Bleeker (ed.), *Anthropologie religieuse* (1955); W. Wundt, *Völkerpsychologie*, vol. iv, 3rd ed. (1920); C. J. Ducasse, *A Critical Examination of the Belief in a Life after Death* (1961). (R. J. Z. W.)

SOULOUQUE, FAUSTIN ÉLIE (1789?–1867), Negro emperor of Haiti, was born a slave about 1789 while Haiti was under French rule. He participated in the successful Haitian revolt against the French in 1803 and thereafter continued as an officer in the Haitian army. He was made president of Haiti in 1847 because the mulatto leaders who had dominated the government under several figurehead Negro presidents thought that his illiteracy and ignorance would make him easy to control. He soon turned against his would-be advisers and ruled as a cruel and corrupt despot, proclaiming himself Emperor Faustin I in 1849 and creating a numerous nobility. Many of the mulatto leaders were killed or exiled. Soulouque made several costly and unsuccessful attempts to conquer the Dominican Republic, until the U.S. and France and Great Britain in 1851 demanded that he desist. He later renewed his attacks and in 1855 was defeated by the Dominican army. In 1859 he was ousted after the chief of his general staff, realizing that the emperor suspected his loyalty, led a revolt. Soulouque escaped and went into exile. He died in 1867.

See also HAITI: *History*. (D. G. Mo.)

SOULT, NICOLAS JEAN DE DIEU, DUC DE DALMATIE (1769–1851), French soldier, one of Napoleon's group of able commanders, was born on March 29, 1769, at St. Amans-la-Bastide (now St. Amans-Soult, Tarn). On the death of his father, a notary, he enlisted (April 1785) in the French infantry, and on the outbreak of the Revolution he was a sergeant at Strasbourg. He served under Adam Philippe de Custine (1792) and under Lazare Hoche (1793) on the Moselle, and from 1794 under François Joseph Lefebvre, being made a general for his conduct at the Battle of Fleurus (June 1794). In March 1799 he replaced the wounded Lefebvre in a division at the Battle of Stockach. The army regarded him so highly that André Masséna, who did not know him, called him to Switzerland as his chief lieutenant. He distinguished himself in the Battle of Zürich (September 1799), and with Masséna in the hills above Genoa in 1800 he increased his reputation for vigour, boldness, and method. He was put in charge of the southern part of the Kingdom of Naples (1800–02)

and in 1804 was made a marshal of France. He commanded a corps at Ulm (October 1805), and after his decisive attack at Austerlitz (December) he was considered "the best tactician of Europe." He was equally successful at Jena (October 1806), but in Poland his attacks at Eylau and Heilsberg (February and June 1807) were costly and indecisive. Created duc de Dalmatie after the Treaties of Tilsit, he went to Spain late in 1808 and pursued Sir John Moore to Corunna (January 1809).

For the next four years Soult remained in Spain, and his military history is that of the Peninsular War (*q.v.*). Surprised (May 1809) and driven from Oporto by Sir Arthur Wellesley (later duke of Wellington), he retreated into Spain, where his arrival in the Tagus Valley caused Wellesley to withdraw from Talavera. Napoleon ignored the Oporto failure and sent him to Madrid to direct, under Joseph Bonaparte, all the French armies in Spain. The results were brilliant; his concentration crushed the Spanish counteroffensive at Ocaña (November 1809). As governor of Andalusia he supported Masséna's attack on the English in Portugal, and took Badajoz in March 1811. He was, however, defeated in May by William Carr (later Viscount) Beresford at Albuera, southeast of Badajoz. The defeat of Auguste Marmont at Salamanca (July 1812) forced him to evacuate Andalusia, and although he subsequently recovered Madrid, he failed to give battle to Wellington at Salamanca in November. Recalled to France (early 1813), he then served in Germany, but on the news of Joseph Bonaparte's defeat at Vitoria (June 1813), Napoleon sent him back to Spain with full powers. His immediate offensive across the Pyrenees was too bold for his wearied and outnumbered troops; he subsequently held and then was forced from each river line in southwestern France, sustaining his last defeat at Toulouse (April 1814), four days after Napoleon had abdicated.

Marshal Soult's political career was less creditable, and it has been said of him that he showed character only in the face of the enemy. At the First Restoration he declared himself a royalist and was appointed minister of war, but during the Hundred Days he again supported Napoleon, acting as his chief of staff at Waterloo. Exiled at the Second Restoration, he was recalled in 1819. Under Louis Philippe he presided over three ministries (October 1832–July 1834, May 1839–March 1840, October 1840–September 1847), usually holding, as well as the office of president of the council, that of minister of war, which he had previously held under Jacques Laffitte and Casimir Périer (November 1830–May 1832). He was responsible for the conquest of Algeria carried out by Thomas Robert Bugeaud de la Piconnerie. In 1847 he was made marshal general of France, a dignity previously held only by Turenne, Villars, and Maurice de Saxe. In 1848, when Louis Philippe was overthrown, he declared himself a republican. He died at St. Amans-Soult on Nov. 26, 1851. His *Memoires* (three volumes) appeared in 1854. (I. D. E.; X.)

SOUND. This article is divided into the following sections and subsections:

I. WHAT IS SOUND?

When a person speaks he is said to emit sound stimuli. Anyone in the vicinity with normal hearing (or in default thereof a suitable hearing aid) is said to experience sound. This common experience reflects a phenomenon of the greatest significance to human beings, who may be considered to be immersed in a world of sound stimuli that have notable influence on most everyday activities. The importance of this experience called sound is reflected in the large number of words descriptive of the different kinds of sound that fill human awareness. Inanimate nature produces the thunder of the storm, the roar and pounding of the surf, the whistling of the wind, the whispering of the trees, the patter of rain, the rippling and gurgling of running water, the humming of wires, the creaking of snow. Even more rich is the vocabulary descriptive of the sounds of living things: the barking and snarling of dogs, mewing of cats, crowing of fowl, roaring of lions, hissing of snakes, the lowing and bellowing of cattle, blatting of goats, chirping of birds and insects, screaming of gulls, crying of infants, and so on.

Not content with all the racket around him over which he has little or no control, man has contrived to produce sounds of almost infinite variety giving both pleasure and pain to countless millions. The boom of cannon, the crack of the pistol, the rattle of musketry and machine-gun fire, the whine of the shell, and the blast of its explosion are unpleasantly familiar. The whir of machinery and the ticking of clocks have more agreeable connotations though they might be distracting to some people. But the melody and harmony of music are generally admitted to contribute aesthetic enjoyment.

What are these stimuli that produce so large a part of man's waking experience? This article gives the answers of physics to that question; it shows that there is far more to the experience called sound than "meets the ear" and that stimuli beyond the range of hearing are in some respects the most important for modern physics and its technical applications. Sound stimuli are physical and the branch of physics that describes them is called acoustics (from the Greek word *akoustikos*, relating to hearing). Physics as an abstract science employs mathematical analysis freely. Because general readers understandably prefer everyday language, this article introduces the subject with a general nonmathematical survey; for readers who seek a deeper understanding, a more analytical account follows.

1. Motion and Sound.—It is obvious to even the most casual observer that a thorough analysis of all that is involved in the emission of any one of the stimuli mentioned above must be complex. To make clear in detail what goes on, for example, in the head and larynx when one speaks would require explanations drawn from physiology, psychology, and physics, with, of course, the aid of mathematics. Nevertheless, it is generally agreed that the r sult is some motion of the air in front of the mouth. It is generally accepted that when a person hears (except in drea

hallucinations) there is some motion of the air at the entrance to his ear.

A careful examination of sound-producing phenomena consistently shows evidence of motion in some medium. This is satisfying to physicists, who have long sought to account for most phenomena in terms of motion, which they believe they understand. To those whose ideas of motion are confined to such solid things as automobiles and airplanes, the association of motion with sound may provide some difficulty. Realizing that in open, still air it is possible to hear a cricket chirp at a distance of half a mile, the skeptic asks how such a small insect can move the mass of air (more than 1,000,000 tons) in a hemisphere with a radius of half a mile so that the air near his ear may move sufficiently to lead to hearing. The question rests on the assumption that the sound producer must move all the surrounding air all at once to produce the experience of sound. However, air is a compressible elastic fluid (*i.e.*, it can be squeezed) so that motion can occur in one part without appearing simultaneously everywhere else. One of the interesting properties of such a medium is that if one small region is squeezed, it does not stay that way. Once squeezed it tends to expand again and compresses (*i.e.*, moves) an adjoining portion of the medium, which in turn repeats the process. The result is that portions of the medium far from the source get squeezed at a time later than the original disturbance. This kind of motion, communicated in time from one part of a medium to another, is called wave motion. Thus the sound stimulus travels in air as a compressional (or squeeze) wave.

2. Velocity and Intensity.—The standard visual illustration of wave motion is provided by the ripples produced on the surface of water by a stone. Strikingly, the water itself does not move outward from the centre of the disturbance, only the distortion of the surface (the ripple) does. The distance traveled by the ripple per unit time is called the wave velocity. All waves have a finite velocity that varies greatly from type to type. Thus ripples on the surface of water may move only a metre or so per second. Light in free space, on the other hand, travels at the enormous rate of 3×10^8 m. per sec. The velocity of sound waves in air increases with temperature, but at room temperature it is about 344 m. (1,125 ft.) per sec. (roughly, 767 miles per hour). Since its value is relatively small compared with that of light, the sound of a distant cannon is heard some time after the puff of smoke is seen.

In fresh water sound waves have a velocity of about 1,500 m. per second at room temperature.

The everyday distinction between loud and soft sounds rests on differences in the magnitude of motions involved in wave propagation. In normal conversation the pressure of the air in front of the mouth of the speaker is changed at most by only about one-millionth of the standard atmospheric pressure. At the same time the accompanying motion that produces the compression leads to an air-flow velocity with a maximum value of only two one-hundredths of a centimetre per second. That such small changes in the air lead to hearing should strengthen respect for the sensitivity of the ear structures. Since motion implies energy, sound waves represent the transmission of mechanical energy. An average rate of transfer of only 10^{-16} (1 divided by 10 raised to the 16th power) watt per square centimetre of acoustic power is enough to produce hearing in a normal young person. This means that mechanical energy of only about 10^{-10} erg (or 3×10^{-24} kw-hr.) can produce an identifiable signal in the normal ear; in this respect the ear is as sensitive as the eye. The average power transmission in a sound wave per unit area is called its intensity. The modern unit for this quantity is the decibel (abbreviated db.), which indicates intensity relative to a standard that usually corresponds to minimum audibility (average excess pressure 2×10^{-4} dynes/cm^2). On this basis conversational speech has an intensity of about 60 db. at a distance of a few feet, traffic at a busy intersection 75 db.; a boiler factory can reach a level of 130 db.

As a source of sound in the open moves farther off (for example, an airplane) the less distinctly it is heard. The intensity of sound in an unenclosed space thus decreases with distance from the source. This is basically a matter of geometry. The airplane may be thought of as emitting the same amount of energy every second in every direction, but as the distance increases this energy spreads over a larger and larger area, and hence the average flow per unit area is diminished. Thus, if the sound disturbance travels with the same velocity in every direction, it will at any given instant reach the surface of a sphere with the source at the centre and with radius equal to the product of the velocity and the elapsed time. Hence the intensity should be expected to decrease in the same ratio as the surface of the sphere increases with its radius; *i.e.*, with the square of the distance from the source. However, it is observed that the drop is greater than predicted by this inverse-square relationship. Apparently some of the energy represented by the sound disturbance is absorbed and changed into heat. This absorption is particularly noticeable in a viscous liquid like glycerol. In air it is accentuated by reflection, refraction, and scattering. The situation is unfortunately more complicated for a source of sound on the ground or under water (near the surface), and even more involved when the source is enclosed in a room.

3. Reflection and Refraction.—Sound waves manifest other properties besides velocity and intensity. Like light waves, they are reflected from surfaces separating media of different properties, and refracted (*i.e.*, bent from the original direction of propagation) in crossing such surfaces obliquely. A familiar example of reflection is the echo produced near a high wall or cliff. This is most readily interpreted as the reflection of the original sound wave from the solid surface, verified by the observation that the time elapsing is that required for the wave to reach the cliff and for the reflected disturbance to return.

Important also is the concept of acoustic image. Like its optical counterpart this is an imaginary source of sound which, if it were located exactly as far behind the reflecting surface as the initial source is in front of it and if the surface were then removed, would give the observed echo at the proper time and with the proper intensity.

It is much easier to speak audibly in a closed room than in the open air. This is another illustration of reflection; the sound waves from the speaker reach the listener not merely in the direct line between the two but also by many paths involving reflection by the walls, floor, and ceiling. Consequently much energy that would be dissipated in all directions out of doors is, so to speak, trapped in the room and helps build up the sound intensity at every point. In a not-too-large room the echoes from the reflecting surfaces merge to produce a continuous sound or reverberation that can impair distinctness of hearing if it lasts too long. Such a room needs acoustic treatment to cut down reflection; *e.g.*, by the introduction of suitable absorbing materials.

Refraction of sound waves is less obvious than in the case of light but produces well-recognized effects. Sound waves travel faster in warm air than cold and since the atmosphere almost always shows vertical temperature differences the waves rarely pass through it in straight lines from the source. Normally there is a negative temperature gradient (*i.e.*, the air becomes increasingly cooler with altitude) and the sound rays (see *Wave Fronts and Rays* below) are bent upward; hence the range of hearing in the horizontal direction is materially reduced. When the temperature gradient is positive (so-called temperature inversion since the air warms with altitude; infrequent in most localities, but seen occasionally in early daylight hours after a clear, wind-free night) upward-moving rays tend to be bent downward, giving a longer range of transmission. Refraction of light through air of varying temperature can produce a mirage (*q.v.*); acoustic mirages also occur under similar conditions, and sound may appear to come from a different direction than it would through a homogeneous medium. These effects are complicated by refraction arising from the motion of the medium (*e.g.*, winds). Also, sound waves travel better with the wind than against it since propagation velocity is changed by addition with wind velocity.

Refraction of sound waves plays a significant and embarrassing role in gun ranging, an acoustical method of locating guns by the sounds of their firing. The reflection and refraction of sound waves in water have practical importance in detecting underwater objects (*e.g.*, a submarine) acoustically. Detection of reflected

sound waves leads to an estimate of range and bearing of the object being sought; but refraction or bending produced by temperature gradients in the water introduces complications (*see* ECHO SOUNDER).

4. Diffraction and Scattering.—One of the most striking properties of sound waves is their tendency to bend around obstacles, permitting a source to be heard around a corner. This is an illustration of the general wave characteristic called diffraction. Light waves do not diffract to such a degree and a light source cannot be seen around a corner through this mechanism.

Because of diffraction objects of ordinary size cannot form sharp acoustic shadows (*see* SHADOW).

An interesting example of diffraction is provided by the human head. Since the ear structures are imbedded in a roughly spherical, more or less solid sphere, the ticking of a watch sounds different in front of the head than it does at the same distance directly behind. Theory and experiment indicate that because of diffraction a source of sound produces a greater intensity at the head if it lies on an extension of the line joining the ears than if it is at the same distance directly in front of the head. This effect has significant bearing on the testing of microphones.

When the diffracting obstacle is relatively small (*e.g.*, a fog droplet), the sound wave is said to be scattered, since the bending makes it turn in all directions from the original.

5. Pitch, Frequency, and Wavelength.—An ordinary air siren produces sound waves by interrupting jets of air at different rates by means of a rotating disk with orifices in its periphery. The more rapidly the interruption takes place (*i.e.*, the greater the number of separate puffs of air per unit time) the higher the perceived pitch. A source of sound in which the disturbance repeats regularly like the puffs of air in the siren is said to produce a harmonic sound wave characterized by a definite frequency. This is defined as the number of times per second the disturbance in the wave at any point is repeated.

The relation between pitch and frequency in a sound wave like that associated with the human voice or ordinary musical instruments is more complex than for a pure harmonic wave as emitted by a carefully made tuning fork. The fork is said to produce a pure tone of definite single frequency. In a harmonic sound wave in air the compression varies periodically at every point. The time for one complete cycle of pressure change at any point is called the period of the wave. The period in seconds is equal to the reciprocal of the frequency (represented by the number of complete cycles per second).

Sounds of all frequencies are not heard equally well. If a hand is waved back and forth periodically with a frequency less than 15 cycles per second, no sound is heard. An object vibrating faster than this will tend to be audible if the intensity is sufficiently great. When frequency is increased beyond 20,000 cycles per second, audibility ceases for most people, and with advancing age this frequency threshold diminishes decidedly. Such waves of very high frequency are inaudible no matter how great their intensity. These are called ultrasonic waves and play an important role in modern acoustics. Ultrasonic waves tend to travel in beams like light, while low-frequency sound waves tend to spread in every direction from the source. Obviously such high-frequency waves need special devices for their production and detection. Much of the progress of acoustics after about 1925 came from the development of such equipment, which produced frequencies up to 100,000,000 cycles per second.

Harmonic waves are also characterized by wavelength: the distance between successive points in a spreading wave at which the disturbance is exactly the same and doing the same thing (*i.e.*, getting larger or smaller). At two such points the phase of the wave is said to be effectively the same. Since the disturbance in a wave travels a distance of one wavelength in one period, the product of wavelength and frequency equals the wave velocity. This important relation holds for all periodic waves. It means that high-frequency acoustic waves in a given medium have smaller wavelength than do low-frequency sounds. Thus in water a wave of frequency 1,000 cycles (1 kilocycle: kc.) has a wavelength of approximately 1.5 m., whereas for a frequency of 1,000,000 cycles (1 megacycle: Mc.) the corresponding wavelength is only 1.5 mm.

Many of the properties of sound waves depend on frequency. Very few sounds in daily experience are pure tones characterized by a single frequency; most can be considered as arising from more or less complicated combinations of harmonic waves, each of definite frequency. One or more frequencies will often predominate and help give the sound its observed quality. In general a high-frequency wave is bent by diffraction less than is a low-frequency wave, and permits an obstacle to cast a sharper acoustic shadow. (A corresponding phenomenon occurs with visible light, where the frequency is very much higher than that of any sound wave.) When sound waves are scattered by objects whose dimensions are small compared with the wavelength, the high-frequency components are scattered much more effectively than those of low frequency. This is analogous to the scattering of sunlight by dust and molecular particles of the atmosphere; the short wavelength or high-frequency light (*i.e.*, the blue) is scattered much more than the long wavelength (red) to give the sky its characteristic colours.

6. Wave Fronts and Rays.—Recall that an audible wave tends to spread out in all directions from the source. However, a beam of sound can be produced if the frequency is high enough. It is helpful in describing this situation to use the concept of wave front: a surface of such kind that at any instant the disturbance characterizing the sound wave is the same at every point.

An example is an airplane propeller radiating a sound wave into the surrounding air. At any instant the state of compression of the air in this wave will be the same on the surface of a sphere with centre at the propeller. This sphere may be called a wave front with a radius that increases from the propeller at a rate equal to the velocity of the sound wave. The expanding circular ripples on water produced by a dropped stone represent a clear analogy. A spherical front will not develop if a sound wave is confined to a tube of constant cross section; *e.g.*, a speaking tube. Though the propagation of the disturbance is complicated by reflections from the walls of the tube, it is approximately accurate to assume that a plane (flat) wave front perpendicular to the direction of propagation will traverse the tube.

Plane and spherical fronts are the most important types encountered in sound-wave propagation. As distance from the source increases the effective part of a spherical wave becomes increasingly flat and plane-wavelike.

Lines perpendicular to a wave front are known as rays. These indicate at any instant the direction in which the wave front is moving at any place. For many purposes it is convenient to describe sound-wave transmission in terms of rays, as in the analogous case of light. However, the wavelength is long compared with that of visible light; consequently ray acoustics is in general by no means as accurate a representation as ray optics. It is more difficult to produce a beam of sound waves than a beam of light, but it can be done if the wavelength is sufficiently small compared with the size of the sound source. The mechanism by which this takes place is called interference and is explained below. Such sound beams are very useful in underwater detection.

II. ANALYSIS OF ACOUSTIC PROPAGATION

1. Wave Functions.—The mathematical treatment of sound waves in general is very complicated. It is simplest to begin with the kind of wave that can be produced in a long, flexible string by flicking one end. Since the kink travels along the string (fig. 1) at right angles to the direction of propagation it is called a transverse wave. This differs from a sound wave in which the compression moves in the same direction as the wave; thus sound waves are called longitudinal. Each particle of the fluid in this case is displaced from its equilibrium position in the direction of the wave propagation. However, the mathematical analysis in its more fundamental aspects is the same for both kinds of wave.

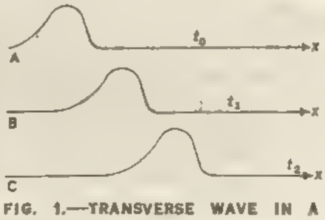


FIG. 1.—TRANSVERSE WAVE IN A STRING

If a snapshot were made of the kinked string at a given instant t_0, it might look like A in fig. 1. Here the disturbance, measured by the distance each part of the string has been displaced from its undisturbed position, can be represented by a mathematical function $f(x)$, where x denotes distance along the string. At a later time t_1 the disturbance has moved a distance $V(t_1 - t_0)$, where V is the velocity of the wave, so the same function $f(x)$ reappears, but displaced (B in fig. 1); likewise for t_2 as indicated in C. The mathematical function of x that describes motion along the positive x axis with velocity V is

$$\xi = f(x - Vt) \qquad (1)$$

The Greek letter ξ (*i.e.*, xi) denotes the displacement (*i.e.*, the measure) of the disturbance. If the wave travels in the negative x direction the corresponding wave function is

$$\xi = g(x + Vt) \qquad (2)$$

Note that there is nothing periodic about equations (1) and (2). However, most sound waves of interest have frequency and wavelength; *i.e.*, they are periodic. The simplest of all periodic waves is the harmonic, in which equation (1) takes the form

$$\xi = A \sin(\omega t - kx) = -A \sin k(x - \omega/k \cdot t) \qquad (3)$$

This is evidently in the proper form for a wave function if ω/k is interpreted as equal to the velocity V. But what are ω and k? The displacement of a particular point x_0 of the string traversed by such a wave varies with time; it is sinusoidal with period $P = 2\pi/\omega$ or frequency $\nu = \omega/2\pi$, which fixes the frequency of the wave. A picture of the string at the instant t_0 will show that the displacement varies sinusoidally with x and the distance between successive points of maximum displacement in the same direction is $2\pi/k$. This has been defined as the wavelength λ of the wave. Hence

$$\nu = \omega/2\pi, \lambda = 2\pi/k \qquad (4)$$

Since $\omega/k = V$, this leads to the important relation

$$\nu\lambda = V \qquad (5)$$

This relation holds for any harmonic wave. It is worth noting that if the cosine function were used in place of the sine function in equation (3) the physical meaning would not be altered. The quantity ω is often referred to as the angular frequency of the wave.

2. Equation of Wave Motion.—If ξ in equation (3) is differentiated twice with respect to x (keeping t constant), and then twice with respect to t (keeping x constant), it can be shown, keeping in mind equation (5), that

$$\frac{\partial^2\xi}{\partial t^2} = V^2\frac{\partial^2\xi}{\partial x^2} \qquad (6)$$

The existence of this equation also follows from the general wave functions: equations (1) and (2). It is called the general differential equation for wave motion in the x direction. Its general solution is the sum of equations (1) and (2) representing general waves progressing in both positive and negative x directions. The physical advantage of a partial differential statement like equation (6) is that it contains the description of so much in so compact a form. The task of theoretical physics in almost every branch may be said to be the development of such general equations. The whole theory of the propagation of normal-intensity sound waves is implied in equation (6) and its generalization to three dimensions (in which any point in space has the coordinates x, y, z):

$$\frac{\partial^2\xi}{\partial t^2} = V^2\left(\frac{\partial^2\xi}{\partial x^2} + \frac{\partial^2\xi}{\partial y^2} + \frac{\partial^2\xi}{\partial z^2}\right) \qquad (7)$$

Physicists feel that they understand something about the propagation of sound waves since equations (6) and (7) are a logical consequence of the behaviour of a fluid when it is compressed.

The behaviour of an ideal fluid when disturbed from its state of rest is described (*see* MECHANICS, FLUID) by three fundamental equations. The equation of continuity indicates that no matter how the fluid moves, as long as it hangs together, there must be conservation of mass. The equation of motion states that the total time rate of change of velocity of a particle of fluid must equal the force per unit mass acting on the particle. For purely compressional disturbances the rate is proportional to the negative of the rate of change of the excess pressure with distance. Finally the equation of state connects the change in pressure with the change in density. This shows that the excess pressure p_e produced by any disturbance in the fluid (*i.e.*, the difference between the actual pressure and the equilibrium pressure prevailing before the disturbance) is directly proportional to the excess density ρ_e (*i.e.*, the difference between the actual density in the disturbed fluid and the equilibrium density). The coefficient of proportionality is usually written as c^2. This idealized static equation of state takes no account of the time required for the imposition of excess pressure to produce the excess density.

Mathematical combination of the three equations leads to a statement of the same form as equation (6). When a fluid is disturbed mechanically the pressure is bound to change; and this leads to motion of the fluid as well as compression or rarefaction (change in density). This must take place in such a way that no mass is lost (continuity), and if the fluid is elastic the compression will move through the fluid as a wave. The result of the combination can be expressed conveniently in terms of excess pressure p_e as

$$\partial^2 p_e/\partial t^2 = c^2\partial^2 p_e/\partial x^2 \qquad (8)$$

having the same mathematical form as equation (6), but with p_e in place of ξ and c in place of V. Physically it means that the acoustical disturbance represented by the excess pressure is propagated as a wave along the x axis with velocity c. It can be shown that the displacement ξ and the excess density ρ_e are propagated in the same way.

In the mathematical deduction leading to equation (8) it has been assumed that the disturbance is not too extreme. Specifically this means that the ratio of the excess density to the equilibrium density (usually called the condensation) is very small compared with unity.

This also means that the gradient of the particle displacement in the fluid (*i.e.*, $\partial\xi/\partial x$) is much smaller than unity. Moreover the flow or particle velocity associated with the disturbance must be very small compared with c (the wave velocity). These conditions prevail for all normal acoustic phenomena associated with speech, sound reproduction, and even for most acoustic transmission, as in underwater sound signaling. However, they fail to hold for sounds produced by explosions, by the rush of gas in a jet airplane, or for sound waves that can produce cavitation (*i.e.*, the appearance of bubbles) in liquids. High-intensity sound (macrosonics) is describable not by equations (6) or (8), but by

$$\partial^2\xi/\partial t^2 = \frac{V^2\partial^2\xi/\partial x^2}{(1 + \partial\xi/\partial x)^{\gamma+1}} \qquad (9)$$

in which the quantity γ in the exponent, if the medium is a gas, is the ratio C_p/C_v, where C_v is the specific heat of the gas at constant volume and C_p is the specific heat of the gas at constant pressure (*see* CALORIMETRY). The equation can also hold for solid and liquid media but γ will then have a different meaning; in the simplest cases it is unity, but it can have higher values. The solution of equation (9) is not so simple as that of equation (6) (to which it reduces for $\partial\xi/\partial x << 1$). For $\gamma = 1$, for example, the solution can be put in the form

$$\partial\xi/\partial t = f[x + \xi - (V + \partial\xi/\partial t)\, t] \qquad (10)$$

where f is an arbitrary function. The physical meaning of this complicated expression is that the various parts of the wave profile (*see* fig. 1) do not move forward with the same velocity. Rather, the top of the hump moves faster than the bottom and hence a disturbance that starts out with a symmetrical profile as in fig. 2A will become asymmetric as in fig. 2B in which the front of the profile will be steeper than the rear. Eventually the front will become vertical and then lean over so that


A B
FIG. 2

its top is farther to the right than its bottom until the wave breaks, as in waves at the seashore. The only way a macrosonic wave can maintain its profile or wave form is through some kind of damping mechanism (*e.g.*, viscosity).

3. Velocity in Fluids.—Recall that the velocity of sound waves in an ideal fluid is given by c in the expression

$$c = \sqrt{p_e/\rho_e} \tag{11}$$

The evaluation of the radical for a liquid follows from the definition of the bulk modulus

$$B = \frac{p_e}{\rho_e/\rho} \tag{12}$$

whence

$$c = \sqrt{B/\rho} = \sqrt{1/\rho K} \tag{13}$$

where K is the compressibility. Since liquids have a relatively small compressibility that changes but little with the density, it is in general adequate to insert ρ_0 (the mean density) for ρ in equation (13).

For gases the situation is different; thus for an ideal gas the equation of state takes the form

$$p/\rho = RT \tag{14}$$

where R is the so-called gas constant per unit mass and T is the absolute temperature. This gives

$$p_e/\rho_e = RT = p/\rho \tag{15}$$

if the temperature remains constant. Hence if the compressional disturbance in the gas takes place isothermally, the velocity has the form

$$c = \sqrt{p/\rho} \tag{16}$$

This formula was first derived by Sir Isaac Newton. It gives, however, a value definitely smaller than that measured for the nearly ideal gases like oxygen and nitrogen. On the other hand, if it is assumed that the acoustic disturbance takes place adiabatically, equation (14) must be supplemented by

$$p/\rho^\gamma = p_0/\rho_0{}^\gamma \tag{17}$$

in which $\gamma = C_p/C_v$ is the ratio of the specific heat of the gas at constant pressure to that at constant volume. It follows then that

$$c = \sqrt{\gamma p/\rho} \tag{18}$$

This result agrees so well with experiment that the measurement of sound-wave velocity became a standard way of evaluating γ, and it is generally assumed that acoustic disturbances in an ideal fluid take place adiabatically. It must not be assumed offhand that the same holds for a viscous fluid, for example, or one that conducts heat. For liquids (equation [13]) the adiabatic compressibility must be used in computing c, but the difference between adiabatic and isothermal compressibilities here is generally small (*see* THERMODYNAMICS).

For an ideal gas it follows from equations (14) and (17) that the velocity of sound waves is independent of the pressure. On the other hand c does depend on the temperature:

$$c = \sqrt{\gamma RT} \tag{19}$$

If the velocity at 0° C is c_0, that at t° C is clearly

$$c_t = c_0\sqrt{1 + t/273} \tag{20}$$

For dry air under standard conditions $c_0 = 331.3$ m. per sec. This agrees well with measurements made in the open air, considering that they are extremely difficult to perform precisely; atmospheric conditions are rarely uniform, and sound-wave transmission in the open air is complicated by nonhomogeneity and wind.

For empirical work with gases equation (19) must be modified, since the equation of state is no longer as given in equation (14). It is now more convenient to revert to equation (13). Since the adiabatic bulk modulus must be used for B, it is seen that $B_{ad} = \gamma B_{iso}$ and by definition

$$B_{iso} = \rho(\partial p/\partial \rho)_T \tag{21}$$

The expression for the velocity c then becomes

$$c = \sqrt{\gamma(\partial p/\partial \rho)_T} \tag{22}$$

For the ideal gas as described in equation (14) this formula yields equation (18). One useful form of the equation of state of an empirical gas is

$$p/\rho = RT(1 + \beta\rho + C\rho^2 + \ldots) \tag{23}$$

The quantities β and C are known as the second and third virial coefficients per gram of gas respectively and are functions of the temperature. Substituting in equation (22) yields

$$c = \sqrt{\gamma RT(1 + 2\beta\rho + 3C\rho^2 + \ldots)} \tag{24}$$

In this form it is clear that the velocity of sound waves depends on the pressure as well as the temperature. The effect in air is small except at very low temperatures. Thus for air at 90.1° K the velocity at 1 atm. pressure is about 188 m. per sec. (A. van Itterbeek and W. van Doninck), whereas at 0.2 atm. it is 191 m. per sec. (Note that at this temperature β is negative.)

The experimental determination of the velocity of sound waves can be used as a method of measuring the virial coefficient β as well as γ. It is also possible to calculate theoretically the velocity of sound waves in a mixture of two gases and then use the formula to determine the relative concentrations of the components by a velocity measurement (H. B. Dixon and G. Greenwood).

As might be expected, the velocity is most sharply dependent on pressure in the neighbourhood of the critical point. Thus in carbon dioxide at 31° C, for example, the value of c drops from about 260 m. per sec. at 20 atm. pressure to a minimum of 150 m. per sec. at 71 atm. and rises steeply thereafter to 330 m. per sec. at 90 atm.

The equations thus far presented involve no dependence of velocity on the frequency of sound waves. This corresponds to experience as far as the so-called permanent gases are concerned at ordinary pressure and low frequencies; they show no dispersion. As the frequency-pressure ratio increases, however, all gases will show dispersion; indeed, the velocity is found to increase with increasing ν/p. Thus for hydrogen, c increases about 8% from its low-frequency standard-pressure value of 1,284 m. per sec. as ν/p is raised from 1 Mc. per atm. to 30 Mc. per atm.

The velocity of sound waves also depends on the purity of the gas. Small traces of foreign gases can produce significant changes in c. Thus for pure carbon dioxide (with only a trace of neon and argon) the velocity for $v/p = 3.2 \times 10^4$ cycles per atmosphere corresponds to a value of γ in equation (18) of approximately 1.34. With the addition of 2.8% water vapour, however, this same value of γ is associated with $v/p = 1.6 \times 10^6$ cycles per atm. The whole dispersion curve is thus displaced. Alternatively put, the sound-wave velocity at a given v/p value is much decreased by the addition of the water vapour. The velocity is affected, of course, by the motion of the medium. Thus in the open air with the wind blowing the velocity of the wave with respect to the ground will be greater with the wind than against it.

Equation (13) can still indicate the velocity of sound waves in liquids but necessarily uses the adiabatic bulk modulus or compressibility. The change in compressibility of a liquid with temperature is much more complicated than that of a gas; this effectively bars the development of helpful general formulas. Water has been carefully studied and the following equation (G. W. Willard) represents rather accurately the dependence of velocity in metres per second on temperature in degrees centigrade

$$c_t = 1{,}557 - 0.0245(74 - t)^2 \tag{25}$$

This indicates that c_t increases with temperature up to a maximum value at $t = 74°$ C, and thereafter decreases. For practically all other liquids that have been studied c decreases with temperature over the whole range in which the material can stay liquid. Salt solutions obey a rule like equation (25) with different values of temperature for maximum c_t. The case of sea water is complicated by changes in salinity, temperature, and pressure (with depth in the ocean) that are effective in changing velocity. The following empirical formula (L. E. Kinsler and A. R. Frey) holds fairly

well over the range of variables usually encountered

$$c_t = 1.41 \times 10^5 + 4.21 \times 10^2 t - 3.7t^2 + 1.1 \times 10^2 S + 1.8 \times 10^{-2} d \quad (26)$$

where t is in degrees centigrade, S is the salinity in parts per 1,000, and d is the depth in centimetres. The velocity is given in centimetres per second.

As is suggested by equation (26), sound-wave velocity in solutions is markedly dependent on the concentration of the solute. For NaCl, for example, the velocity in a 10% solution at room temperature is 1,600 m. per sec. while for a 20% solution it rises to 1,720 m. per sec. Not all solutions yield higher velocity than does the pure solvent, however.

Equation (26) also indicates that increase in pressure on a liquid raises the velocity more or less linearly. Thus for benzene P. Biquard found that the velocity is increased by about 17% in going from atmospheric pressure to 500 atm.

Unlike the situation in gases, no definite indication of sound-wave dispersion in liquids has been found experimentally, though theoretical considerations suggest it for sufficiently high frequencies, probably of the order of several hundred megacycles.

4. Wave Fronts.—The concept of wave front was introduced earlier to describe the propagation of acoustic waves. Equation (3) dealt with harmonic wave functions corresponding to propagation along the positive x axis only. But since it is the general tendency for sound waves in free space to spread in all directions the analysis must be generalized. To simplify wave propagation in three-dimensional space, the excess pressure is used as the measure of the acoustical disturbance. Consider the function

$$p_e = A(x, y, z) \sin \omega[t - \psi(x, y, z)/c_0] \quad (27)$$

in which A is the amplitude of the pressure wave and may be a function of position in space. The function $\psi(x, y, z)$ is assumed to have the dimensions of distance only, and c_0 is a constant having the dimensions of velocity.

Equation (27) is a genuine harmonic wave function (*i.e.*, a solution of the general three-dimensional equation [7] under certain conditions). The phase of the wave $\omega[t - \psi(x, y, z)/c_0]$ has the constant value ωt_0 at the time t over the surface with the equation

$$\psi(x, y, z) = c_0(t - t_0) \quad (28)$$

Hence this is the general equation for the wave front at time t. In the passage of time the wave-front equation will alter just as if the front itself were moving through space, carrying with it the same value of the disturbance. If

$$\psi(x, y, z) = \alpha x + \beta y + \gamma z \quad (29)$$

equation (28) represents a plane wave front with the direction cosines of the normal to the front equal to α, β, γ. In this case $A(x, y, z)$ reduces to a constant. If

$$\psi(x, y, z) = \sqrt{x^2 + y^2 + z^2} \quad (30)$$

equation (28) represents a spherical wave front. Equation (27) is then a harmonic pressure wave spreading out of the origin if $A(x, y, z) = A_0/r$, where

$$r = \sqrt{x^2 + y^2 + z^2} \quad (31)$$

and A_0 is constant.

Such fronts are somewhat idealized; a spherical wave front demands an obviously nonexistent precise point source of sound. A plane wave front demands an infinitely extended plane source, likewise not achievable. Nevertheless empirical wave fronts often approximate these ideal types sufficiently well to justify their use. In general the smaller the dimensions of a source compared with the wavelength the more nearly the front approximates spherical or hemispherical shape. The larger the dimensions of a plane source compared with the wavelength, the more nearly is a plane wave front realized.

5. Sound Rays; Refraction.—For many purposes the propagation of light can be effectively studied in terms of rays—curves that are always perpendicular to the family of wave fronts associated with a particular propagation. The ray concept is also useful in acoustics but only if the wavelength is sufficiently small; more accurately, if the change in wavelength (due to change in velocity from one point to another in the medium) over a distance of the order of one wavelength is very small compared with the wavelength itself (*see* OPTICS).

Moreover, the ray must not bend too rapidly as the sound wave progresses, and the change in amplitude over a like distance must be small compared with the initial amplitude. If these conditions are satisfied, it is meaningful to write the differential equations of the ray paths in any medium in which the wave velocity c varies from point to point. By analogy with optics it is customary to define the ratio c_0/c as the index of refraction n of the medium. The physical meaning of c_0 is the nonvarying velocity in some homogeneous medium bounding the given nonhomogeneous medium. Let the direction cosines of the ray path be α, β, γ. These are in general functions of position; *i.e.*, of the coordinates x, y, z. If ds is the element of distance along the ray, the differential equations of the ray paths take the form

$$d(\alpha n)/ds = \partial n/\partial x,\ d(\beta n)/ds = \partial n/\partial y,\ d(\gamma n)/ds = \partial n/\partial z \quad (32)$$

As soon as n is known as a function of x, y, z, equations (32) can be integrated to give the rays. As an illustration, suppose n is a constant, so that there is no variation of acoustic velocity in the medium. Equations (32) then lead to

$$\alpha = K_1, \beta = K_2, \gamma = K_3 \quad (33)$$

where K_1, K_2, K_3 are constants, giving then the direction cosines of a straight line. The rays are straight lines, irrespective of the character of the wave fronts (the latter, for example, may have the form of plane or spherical wave fronts).

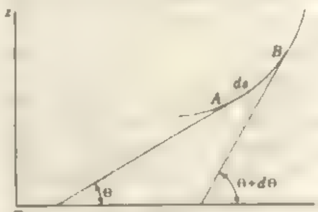


FIG. 3

A typical case of ray propagation occurs when the medium is stratified; *i.e.*, the change in velocity takes place in one direction only, for example, the z axis. It will be sufficient to consider the ray path in the xz plane. In fig. 3 let AB be the element ds of the ray and let the tangents at A and B, respectively, make the angles Θ and $(\Theta + d\Theta)$ with the x axis. Thus

$$\partial n/\partial x = \partial n/\partial y = 0$$
$$\alpha = \cos \Theta, \beta = 0, \gamma = \sin \Theta$$

Equations (32) now yield

$$d(n \cos \Theta)/ds = 0,\ d(n \sin \Theta)/ds = dn/dz \quad (34)$$

The first equation shows that along the ray

$$n \cos \Theta = \text{constant} = n_1 \cos \Theta_1 \quad (35)$$

where Θ_1 is the direction of the ray at the place where the index is n_1. If $\Phi = \pi/2 - \Theta$, where Φ is the angle the ray makes with z, equation (35) becomes

$$\sin \Phi/\sin \Phi_1 = n_1/n = c/c_1 \quad (36)$$

This has the familiar form of Snell's law for the refraction of light rays (*see* REFRACTOMETER). Here, however, it is of much more general form since the variation of velocity is continuous and does not take place abruptly at a boundary plane as in the usual elementary optical case. The abrupt change involves a discontinuity in n and therefore demands special treatment, discussed in section II.7. The second of equations (34) makes possible the evaluation of the curvature $(d\Theta/ds)$ of the ray at any point and hence the equation of the ray path. By straightforward analysis

$$d\Theta/ds = -dc/dz \cdot \cos \Theta_1/c_1 \quad (37)$$

As an illustration, suppose

$$c = c_1 + az \quad (38)$$

i.e., the velocity is a linear function of the depth z (a is a constant), with c_1 = velocity at $z = 0$; this might be the surface of the sea. Then equation (37) shows that the curvature is constant

This means that the path must be a circle with radius $|c_1/a \cos\Theta_1|$. If a is positive, the velocity increases with depth, the curvature is negative, and the path bends upward toward the surface $z = 0$. If a is negative the velocity decreases with depth (a more common situation in the ocean) and the circular path bends downward. These results can readily be followed graphically.

6. Refraction in a Moving Medium.—The propagation of sound waves through a moving medium (*e.g.*, the atmosphere with the wind blowing) is extremely complicated and its analysis will not be given here. However, the law of refraction for a stratified medium turns out to be a rather simple modification of the Snell's law equation (36). If the velocity of the medium is confined to the x direction only and is equal to u, which is a function of z (*e.g.*, height above the ground, taken as the xy plane, in the case of transmission through the atmosphere), the law of refraction becomes

$$c/\cos\Theta - c_1/\cos\Theta_1 = u_1 - u \qquad (39)$$

Here u_1 is the medium velocity at the point where the sound velocity is c_1 and the ray direction is given by Θ_1. Equation (39) reduces to equation (36) for $u = u_1 = 0$. From equation (39) and also from relatively simple graphical constructions the effect of the motion of the medium is seen to bend the wave front and ray in the direction in which the motion takes place. Thus if wind velocity increases upward from the ground, sound wave fronts are lifted to windward and depressed to leeward. This tends to decrease the range to windward and increase it to leeward. These considerations were once of considerable importance in the acoustical detection of aircraft, but were superseded by radar. They remained important in sonic gun ranging.

7. Huygens' Principle; Reflection and Refraction at an Interface.—The propagation of an acoustical wave front through a medium can also be studied by means of a principle developed by Christiaan Huygens (1629–1695). According to this principle, in its elementary form, each point on a wave front (like F_1 in fig. 4, for example) may be assumed to be the source of a hemispherical wavelet which moves outward from F_1 in the direction of propagation; the new wave front after a short time is the mathematical envelope of all these wavelets; *e.g.*, F_2 in the figure. This principle can be demonstrated by considering the solution of the general wave equation. A simple illustration is the establishment of the laws of reflection and refraction at a plane interface separating fluid media with different properties; *e.g.*, different mean densities and sound velocities.

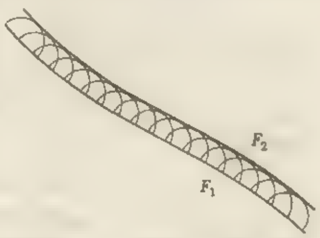


FIG. 4

Let AB in fig. 5 represent the trace in the plane of the paper of the plane interface between a fluid medium I in which the velocity of sound is c_1 and the mean density is ρ_1, and a medium II in which the corresponding quantities are c_2 and ρ_2, respectively. Let CD denote the trace of a plane wave front (perpendicular to the plane of the paper) incident on AB at angle Θ_i. This means that the incident ray travels in the direction HC perpendicular to CD. The reflected and refracted wave fronts can be constructed by means of Huygens' principle. At the instant the disturbance at one end of the wave front has reached C on the interface, the disturbance at the other end is still at D in medium I. In the time $t_1 = DE/c_1$ that it takes the disturbance at D to reach E on the interface, the disturbance will travel out in all directions from C. In particular in I it will traverse a distance equal to DE.

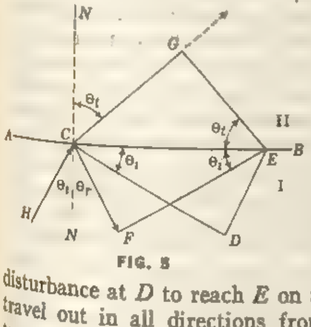


FIG. 5

If (with C as centre) a semicircle is drawn in medium I with radius DE, and a straight line from E is constructed tangent to this, it produces FE which makes the angle $\Theta_r = \Theta_i$ with AB. That this is the reflected wave front can be confirmed by making a Huygens construction for other points on the incident wave front. It is seen that the reflected wave front makes the same angle with the interface as the incident wave front. This, combined with the fact that both wave fronts are perpendicular to the same plane (the plane of the paper), constitutes the law of reflection. Equally well expressed in terms of the incident and reflected rays, it is the same law that holds for light-wave fronts and rays.

The same kind of analysis shows that after transmission across the interface the incident wave front CD is changed to the refracted wave front GE making the angle Θ_t with AB, such that

$$\sin\Theta_t/\sin\Theta_i = c_2/c_1 = n_{21} \qquad (40)$$

This is Snell's law again, with n_{21} the index of refraction of medium II with respect to medium I. If the media are in motion with respect to each other, this result must be modified (*cf.* equation [39]). Equation (40) accounts well for the passage of sound waves from air to water or vice versa.

8. Energy Propagation; Intensity.—Sound-wave propagation is basically a form of transmission of energy through a medium. When a fluid, for example, is compressed, work is done, representing the expenditure of energy. The reappearance of the disturbance at a distant point by wave propagation corresponds to the transfer of the original energy. The greater the amount of energy transported per unit area of wave front per unit of time, the greater will be the intensity of the sound wave.

To be more precise, the rate of flow of energy in the sound wave per unit time per unit area of wave front is simply calculated as the average of the product of the excess pressure p_e and the flow velocity $\partial\xi/\partial t$ in the medium. Analysis shows that the intensity I of a plane wave computed in this way can be expressed as

$$I = p_0^2/2\rho_0 c \qquad (41)$$

This important formula (G. W. Stewart and R. B. Lindsay) shows that the intensity of a plane wave is equal to the square of the excess pressure amplitude p_0 divided by twice the product of mean density and sound-wave velocity. Significantly it is independent of frequency.

The standard absolute unit of intensity is the erg/sec-cm^2. More practical is the watt/cm^2 or, in the metre-kilogram-second systems, the watt/metre^2. However, the decibel (commonly used as the standard unit of sound intensity) is a relative measure. If two sound waves have absolute intensities I_1 and I_2 respectively they are said to differ in intensity level by D decibels (or db.), where

$$D = 10 \log_{10} I_2/I_1 \qquad (42)$$

Hence if $I_2 = 2I_1$, $D = 3.01$ db., or doubling the intensity means an increase of intensity of a little more than 3 db. To use this method of measuring intensity requires the choice of a standard level. This is often taken as that corresponding to an average excess pressure of 1 dyne/cm^2 in air (close to the normal level for conversational speech). The minimum audible stimulus lies about 70 db. below this level, while the threshold at which the stimulus becomes painful lies about 70 db. above this level. These figures are approximate and depend on frequency. In modern acoustical practice the reference level is almost always taken at the average human threshold for hearing, corresponding to root-mean-square acoustic excess pressure 2×10^{-4} dyne/cm^2 or intensity 10^{-16} watt/cm^2.

From equation (41) it is clear that intensity at given pressure depends on the medium and in particular on the product $\rho_0 c$. This quantity is known as the specific acoustic resistance of the medium (for a plane wave). For water, for example, it is about 3,800 times the value for air at standard pressure. Consequently the same excess pressure in water produces a much lower intensity than in air, at first glance suggesting that it is harder to produce an intense sound wave in water than in air. However, at given frequency and pressure a solid radiating source is empirically more efficient in water (and in liquids generally) than in air.

9. Intensity of Spherical Waves.—Equation (41) was cited as applying to a plane wave. A similar formula applies to a spherical wave in the corresponding medium, but the pressure and particle velocity expressions are different. Whereas for a plane

wave the excess pressure and flow velocity (sometimes called particle velocity) amplitudes are constant in a nonabsorbing medium, for a spherical wave they both fall off with the distance from the source of the radiation. Multiplication and averaging, as in section II.8 above, yields for the intensity in the spherical wave at distance r from the source

$$I = p_0^2/2\rho_0 c r^2 \tag{43}$$

Equations (41) and (43) can be put in the same form by writing

$$I = \overline{p_e^2}/\rho_0 c \tag{44}$$

where $\overline{p_e^2}$ is the average of the square of the excess pressure at the point where the intensity is computed.

10. Flow of Energy Across a Boundary.—Recall that when a sound wave strikes the interface separating two media with different physical properties (*e.g.*, going from air to water or vice versa) some of the wave is reflected while the rest is transmitted (and in general refracted). This means that the transmission of acoustic energy is affected by the presence of the boundary.

In fig. 6, OA denotes the plane boundary separating media I and II, the x axis being perpendicular to the boundary. In medium I the mean equilibrium density is ρ_1 and the wave velocity is c_1. The corresponding quantities in medium II are ρ_2 and c_2. The arrow i denotes a plane harmonic sound wave traveling in I from left to right in the x direction and striking the boundary at $x = 0$. Similarly r denotes the wave in I reflected from the boundary and t denotes the wave transmitted into medium II across the boundary. For simplicity assume perpendicular incidence. Let the amplitude of the excess pressure in the incident wave be p_i, that in the reflected wave p_r, and that in the transmitted wave p_t. By setting up two algebraic relations connecting these three quantities, the reflected and transmitted pressure amplitudes can be expressed in terms of the incident; by doing the same for the particle velocity amplitudes, the relative amount of transmitted and reflected energy can be estimated.

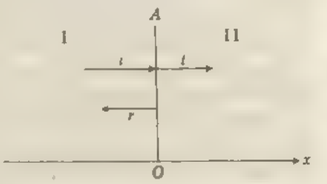


FIG. 6

These two relations are called the boundary conditions at the interface. They specify that there must be no discontinuity in the excess pressure and particle velocity at the boundary. Otherwise put, the sum of the pressure amplitudes of the incident and reflected waves at the boundary must equal the transmitted pressure amplitude, and similarly for the particle-velocity amplitudes. The result of applying these two conditions is

$$\frac{p_r}{p_i} = \frac{1 - \rho_2 c_2/\rho_1 c_1}{1 + \rho_2 c_2/\rho_1 c_1} \tag{45}$$

and

$$\frac{p_t}{p_i} = \frac{2}{1 + \rho_1 c_1/\rho_2 c_2} \tag{46}$$

Use of equation (41) yields for the power transmission ratio P_r of the intensity in the transmitted wave to that in the incident wave

$$P_r = I_t/I_i = \frac{4\rho_2 c_2/\rho_1 c_1}{(1 + \rho_2 c_2/\rho_1 c_1)^2} \tag{47}$$

It will be noted that P_r depends wholly on $\rho_2 c_2/\rho_1 c_1$, or the ratio of the specific acoustic resistances of the two media. If the resistances are equal in spite of the boundary, then $P_r = 1$; for any other ratio $P_r < 1$. The specific acoustic resistance ρc therefore controls the transmission of energy at normal incidence from one medium to another. If the boundary is between a gas and a solid, the large difference between the ρc values for the two will give a small P_r, suggesting that in this case the reflection coefficient (ratio of reflected to incident intensity) will be high. The reflection coefficient R is given by

$$R = 1 - P_r = \left(\frac{1 - \rho_2 c_2/\rho_1 c_1}{1 + \rho_2 c_2/\rho_1 c_1}\right)^2 \tag{48}$$

Equations (47) and (48) account adequately for the observation that transmission from air to water or from air to a solid wall is relatively poor, with reflection correspondingly high. For air-to-water transmission the value of P_r is only 0.12%.

In considering the more general case of oblique incidence, recall that Snell's law describes the geometrical refraction associated with oblique transmission. The power transmission ratio now depends on the angles of incidence and refraction Θ_i and Θ_t respectively. Thus

$$P_r = \frac{4\rho_2/\rho_1 \cdot \cot\Theta_t/\cot\Theta_i}{(\rho_2/\rho_1 + \cot\Theta_t/\cot\Theta_i)^2} \tag{49}$$

It is to be noted that P_r here is the actual ratio of transmitted power (*e.g.*, watts) to incident power. It is not the ratio of transmitted intensity (*e.g.*, watts/cm²) to incident intensity. To obtain the latter, P_r is divided by the ratio S_t/S_i, where S_i and S_t are respectively the areas of the incident and transmitted wave fronts. From fig. 5

$$S_t/S_i = \cos\Theta_t/\cos\Theta_i \tag{50}$$

In practice P_r is in general the more important quantity. The reflection coefficient R in this case (the wave fronts are here equal in area) can be calculated as in equation (48) and becomes

$$R = 1 - P_r = \left(\frac{\rho_2/\rho_1 - \cot\Theta_t/\cot\Theta_i}{\rho_2/\rho_1 + \cot\Theta_t/\cot\Theta_i}\right)^2 \tag{51}$$

When $\Theta_i = 0$ (normal incidence), equation (51) reduces to equation (48). When $\Theta_i = 90°$ (grazing incidence) then $P_r = 0$ and there is complete reflection. If $c_2 > c_1$ total reflection ensues for angles of incidence greater than the critical angle Θ_c, where

$$\Theta_c = \text{arc sin } c_1/c_2 \tag{52}$$

in complete analogy with geometrical optics. On the other hand, there is complete transmission and no reflection if

$$\cot\Theta_t/\cot\Theta_i = \rho_2/\rho_1 \tag{53}$$

This takes place for angle of incidence Θ_i such that

$$\cot^2\Theta_i = \frac{c_1^2/c_2^2 - 1}{\rho_2^2/\rho_1^2 - c_1^2/c_2^2} \tag{54}$$

demanding

$$\rho_2/\rho_1 > c_1/c_2 > 1 \text{ or } \rho_2/\rho_1 < c_1/c_2 < 1 \tag{55}$$

There are important applications of these formulas to transmission in the atmosphere between layers of air at different temperature and humidity, and to transmission in sea water with temperature gradients.

11. Impedance.—The importance of the product ρc (usually written $\rho_0 c$) in transmission of a plane wave between media is emphasized by its name (specific acoustic resistance).

Inevitably this suggests an electrical analogy. It is readily verified that the ratio of the excess pressure to the flow or particle velocity in a plane progressive acoustic wave in a fluid is $\rho_0 c$. In the case of direct currents in electricity, the ratio of potential difference to current flow is called the resistance of a conductor. Consider acoustic excess pressure as analogous to potential difference and acoustic particle velocity as analogous to electric current: the significance of acoustic resistance then becomes clear. Since acoustical quantities vary with time and space (sinusoidally in the most commonly used cases) it appears more relevant to make the analogy with alternating electric current. In the latter the ratio of alternating potential difference (or electromotive force, emf) to alternating current is called the electrical impedance. Moreover $\frac{\partial \xi}{\partial t}$ appears in the guise of an acoustic current density; to generalize

$$p_e/\frac{\partial X}{\partial t} = Z = \text{acoustic impedance} \tag{56}$$

where $\frac{\partial X}{\partial t} = S\frac{\partial \xi}{\partial t}$, the so-called acoustic volume current, S being the area of the acoustic wave front.

In alternating current theory the impedance Z is in general made

up of a resistive component and a reactive component; the resistive component yields the part of the current that is in phase with the alternating emf and the reactive component the part of the current that is out of phase with the emf by 90°. For an alternating current circuit containing resistance R in series with inductance L and capacitance C, the ratio of the amplitude of the applied electromotive force to the amplitude of the resulting current is

$$Z = \sqrt{R^2 + (\omega L - 1/\omega C)^2} = \sqrt{R^2 + \chi^2} \tag{57}$$

where ω ($= 2\pi\nu$) is the angular frequency of the current and χ is equivalent reactance. The current lags behind the emf by the phase angle

$$\alpha = \arctan\left(\frac{\omega L - 1/\omega c}{R}\right) = \arctan\frac{\chi}{R} \tag{58}$$

This electrical theory applies to oscillations, while the acoustical analogy concerns waves. There is no phase lag between excess pressure and volume current in a plane acoustic wave. Hence the corresponding phase angle in this case is zero. This means that the equivalent acoustic reactance vanishes and the acoustic impedance is a pure resistance; hence the name acoustic resistance for $\rho_0 c/S$. Calling $\rho_0 c$ itself the specific acoustic resistance (denoted by Z_s) also follows electrical analogy, since the specific electrical resistance of a conductor is the resistance per unit length and for unit area of cross section.

Though the specific acoustic impedance of a plane wave is a pure specific resistance, this does not hold for a spherical wave, which has both reactive and resistive components in its impedance. The physical meaning of this is connected with the spreading of the wave in all directions. The reactance of a diverging spherical wave is positive and behaves like an electric inductance, whereas the reactance of a spherical wave converging to a point is negative and thus behaves like an electric capacitance.

The impedance notation permits acoustic transmission calculations to be replaced to advantage by analogous electrical computations. Many acoustical effects can be readily estimated in terms of known values of acoustic impedance at strategic points. For example, the effectiveness of a horn (*i.e.*, a tube of variable cross section) both as a receiver and a transmitter of sound waves is measured (for given volume current) by the value of the acoustic resistance at the throat. The effective acoustic transmission through a tube in which a branch line is inserted at some point can be shown to depend in a relatively simple manner on the acoustic impedance at the branch point. Problems in architectural acoustics involving the reflection of sound from the walls of a room can often be simplified by the use of the impedance notation.

12. Selective Transmission; Filtration.—The transmission of acoustic energy in a harmonic wave across the interface separating two media is independent of frequency as equations (47) and (49) indicate. Here both media have been considered semi-infinite (medium I infinite to the left of the interface and medium II infinite to the right). A different situation arises if one of the media is finite in the direction of wave travel.

In fig. 7 let medium II, of length l, be inserted between media I and III, which are semi-infinite. The mean densities in the three media are ρ_1, ρ_2, ρ_3 respectively and the corresponding sound velocities are c_1, c_2, c_3.

FIG. 7

The analysis of section II.10 above may then be applied, remembering that because of the reflections at interfaces 1 and 2 there will be waves traveling in both positive and negative directions in medium II, even though the radiation comes from the left in I and goes to the right in III. Use of the appropriate boundary conditions as before gives for the power transmission ratio P_r (ratio of transmitted power in III to incident power in I) for normal incidence

$$P_r = \frac{4R_3/R_1}{(R_3/R_1+1)^2} \cdot \frac{1}{1 - \dfrac{(R_3^2/R_2^2-1)(R_2^2/R_1^2-1)}{(R_3/R_1+1)^2} \cdot \sin^2 k_2 l} \tag{59}$$

In this expression R_1, R_2, R_3 are the specific acoustic resistances of media I, II, III respectively, and $k_2 = \omega/c_2$. The dependence of P_r on the frequency arises through the term $\sin^2 k_2 l$. The effect can be seen most simply for the special case in which $R_1 = R_3$; *i.e.*, the third medium is the same as the first. Then (now denoting the power transmission ratio by P_r')

$$P_r' = \frac{1}{1 - \dfrac{(2 - R_2^2/R_1^2 - R_1^2/R_2^2)}{4}\sin^2 k_2 l} \tag{60}$$

If $\sin^2 k_2 l = 0$, $P_r' = 1$. If $\sin^2 k_2 l = 1$, P_r' passes through the minimum value $4/(2 + R_2^2/R_1^2 + R_1^2/R_2^2)$, which is the smaller, the greater the difference between R_1 and R_2. The plot of P_r' as a function of frequency then appears as in fig. 8.

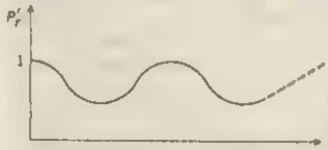


FIG. 8

The transmission is said to be selective with respect to frequency, with the frequencies $\nu_n = nc/2l$, where n is integral, yielding 100% transmission, and the frequencies $\nu_n = (2n+1)c/2l$ yielding minimum transmission. In no case does the value of P_r' become precisely zero. The situation may be described by saying that the medium II acts as if it preferred to have harmonic waves of frequencies $\nu_n = nc/2l$ traveling back and forth in it and would not tolerate others so readily.

This suggests the possibility of arranging a collection of media that will act as an acoustic filter; *i.e.*, transmit waves of certain frequencies well but fail entirely to pass waves of other frequencies. If this could be realized the curve of P_r against frequency ν would appear, for example, as in fig. 9. The transmission here has unit value between *0* and *A*, *B* and *C*, *D* and *E*, and so on. These frequency intervals should be called pass bands, whereas since in *AB*, *CD*, and so on, P_r drops to zero, they correspond to attenuation bands.

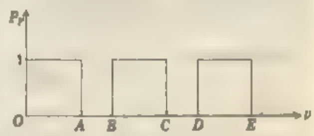


FIG. 9

Theoretical study indicates that a structure with these properties can be approximated by an iteration of two media (fig. 10). The larger the number of pairs in the iteration (or sections of the filter) the more closely will the transmission curve correspond to the sharp alternation of pass and attenuation bands as in fig. 9. In any case the structure is called a low-pass filter, since low frequencies from zero to *A* are passed.

FIG. 10

The type of acoustic filter just discussed depends on the alternation of properties of two different media. This is not necessary to achieve the same general result.

Fig. 11 represents a succession of cylindrical tubes of cross-sectional areas S_1, S_2 and of lengths $AB = CD = EF$, and so on, for the wide tubes and $BC = DE$, and so on, for the narrow tubes. Such a structure acts as a low-pass filter for harmonic sound radiation—the transmission pattern of pass and attenuation bands is very similar to that in fig. 9.

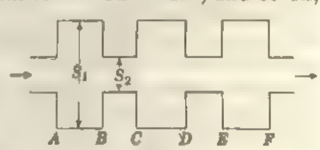


FIG. 11

Here the ratio S_2/S_1 of the cross-sectional areas of the constricted and expanded tubes serves the same role as the ratio of the specific acoustic resistances in the iterated-media filter. The lengths of the sections have a decided bearing on the frequency range of the transmission and attenuation bands. Many other types of acoustic filters have been constructed to suppress certain frequency bands in blowers, mufflers, and other equipment producing noise by the flow of air and other gases. The same type of theory applies to the filtration of elastic waves in solids; *e.g.*, rods. Many musical instruments incorporate acoustical filters.

Selective transmission as discussed in this section does not arise from a genuine frequency-dependent absorption with transforma-

tion of acoustic energy into heat. The filtration results when waves of certain frequency ranges are effectively barred; *i.e.*, not accepted from the source.

13. Absorption and Dispersion in Fluids.—The decrease in intensity from a localized source (*i.e.*, approximately at a point) because of spherical spreading fails to account for the total loss in intensity. The additional loss is manifested even when waves pass through a more or less homogeneous medium in which there is no selective transmission or filtering. Such loss is ascribed to absorption of the sound energy and its transformation into heat. An obvious source of such absorption is the internal friction of the medium through which the radiation passes. This friction is exemplified by the so-called shear viscosity of a fluid, which is a measure of the tendency of relative motion of adjacent layers of the fluid to be decreased (*see* VISCOSITY). For the simplest type of viscous fluid (*e.g.*, air and water) the shearing stress or drag per unit area of parallel layers of fluid on each other is directly proportional to the rate of change of the flow velocity (assumed parallel to the layers) with respect to distance normal to the flowing layers; the coefficient of proportionality is called the shear viscosity η.

Analysis shows that for a plane harmonic acoustic wave of angular frequency ω progressing in the x direction in a viscous medium the displacement ξ can be written

$$\xi = \xi_0 e^{-\alpha x} \cos (\omega t - kx) \tag{61}$$

in which α is known as the amplitude absorption coefficient. It represents an exponential decrease in the amplitude of the wave with distance traveled, corresponding to the absorption of the acoustic energy by the medium. Expressed in terms of the viscosity of the medium and the frequency, the absorption coefficient can be written

$$\alpha = \omega^2 \tau V/2c^2 (1 + \omega^2\tau^2) \tag{62}$$

where

$$V = \sqrt{2}c/\omega\tau \cdot [(1 + \omega^2\tau^2) (\sqrt{1 + \omega^2\tau^2} - 1)]^{\frac{1}{2}} \tag{63}$$

and

$$\tau = 4\eta/3\rho_0 c^2 \tag{64}$$

Here $V = \omega/k$ is the phase velocity of the wave. The fact that V in equation (63) depends on the frequency indicates dispersion of the wave. This recalls that the velocity of light in matter normally decreases as the frequency increases. From equation (63) it is seen that at zero frequency $V = c = \sqrt{B/\rho_0}$, where B is the adiabatic bulk modulus of the fluid, and as ω increases, V increases. For real fluids, even with considerable viscosity (*e.g.*, glycerol with $\eta = 5$ poises at room temperature), the value of τ (often referred to as a relaxation time) is extremely small and hence $\omega\tau$ is negligibly small compared with unity except for extremely high frequencies. Equations (62) and (63) for α and V do not in themselves account for the observed acoustic absorption and dispersion in fluids even at low frequencies. Other mechanisms enter into the attenuating process.

Another obvious basis for acoustic absorption in fluids is the conduction and radiation of heat. In the passage of a compressional wave the temperature is increased at points of maximum excess density (condensation). In the simple theory of wave-energy propagation set forth earlier, the heat energy corresponding to this temperature (strictly extra kinetic energy of the molecules of the fluid) is passed on to the adjacent portion of the fluid to produce a corresponding compression there. This neglects heat conduction and radiation to the surroundings. Some of the energy that is necessary to propagate the disturbance gets lost in transit, effectively corresponding to absorption of the sound wave. The complicated analysis shows that the absorption coefficient due to the combined effects of viscosity and heat conduction at not-too-high values of the ratio of frequency to pressure (*i.e.*, less than 1,000 Mc. per atm.) becomes

$$\alpha = 2\pi^2\nu^2/\gamma p_0 c \cdot [4\eta/3 + (\gamma - 1)\kappa/c_p] \tag{65}$$

where ν is the actual frequency, p_0 the mean equilibrium pressure, c_p the specific heat at constant pressure, κ the thermal conductivity, γ the ratio of the specific heat at constant pressure to that at constant volume, and the other symbols have their previously assigned meanings. The two terms in the brackets in equation (65) correspond to the contributions of viscosity and heat conduction respectively. Analysis shows that for monatomic gases the contribution of heat conduction to α is about 70% that due to viscosity. The ratio of the heat conduction to the viscosity contribution is approximately $\gamma - 1$, so that for polyatomic gases the heat conduction contributes relatively much less.

For liquids other than mercury and liquefied gases the contribution of heat conduction is such a small fraction of that due to viscosity that it may be neglected. The part played by radiation in gases and liquids is negligible except possibly at very low pressures or very high temperatures.

Equation (65), the so-called classical absorption formula, agrees very well with experimental values for monatomic gases (*e.g.*, helium, argon). The more precise formula of which this is an approximation has been found to agree in the case of helium down to pressures of the order of 1 mm. of mercury and up to frequencies of 1 Mc. However, for diatomic and other polyatomic gases and liquids there are some frequency regions in which the absorption is much greater than that predicted. For the vast majority of liquids the absorption over wide frequency ranges is found experimentally to vary more or less as the square of the frequency as equation (65) indicates. However, the experimental values are almost uniformly in excess of those predicted. Thus α/ν^2 for water at room temperature over the range from 1 to 50 Mc. is about 25×10^{-17} sec^2 cm^{-1}, whereas the formula yields only 8.5×10^{-17} sec^2 cm^{-1}. The discrepancy is even greater (approaching a factor of 100 or more) for many organic liquids like benzene.

Such deviations have led to alternative theoretical postulations for acoustic absorption. One of these is based on the idea that when a fluid is compressed in the passage of a wave the work done does not merely increase the average translational velocity of the molecules; some of it, through molecular collisions, enters into other states of energy; *e.g.*, rotational and vibrational in the case of diatomic and other polyatomic molecules. Calculation shows that there is a lag (relaxation time) in this process of energy transfer; *i.e.*, a given average time is required for the energy to come back to translational form after it has become rotational or vibrational. Because of this lag the changes in translational energy of the molecules fail to keep in step with the propagation of the wave; thus some of the original energy of the disturbance is made unavailable during each cycle for transmission. It is therefore lost in dissipation and the result is effectively absorption. Shear viscosity can be considered a kind of relaxation process and the corresponding relaxation time is the τ given by equation (64). It is evidently not an adequate mechanism. In the case of polyatomic gases vibrational (and to some extent rotational) relaxation has proved useful in accounting for the observed excess absorption. This type of relaxation is usually termed thermal, since it corresponds effectively to an apparent increase in specific heat (*see* RELAXATION PHENOMENA).

If a thermal relaxation process is at work in a gas, when the product of absorption coefficient α and wavelength λ is plotted against the frequency, the result is a curve that rises to a maximum and then falls off again. The frequency at which the maximum occurs is called the relaxation frequency and is equal to $1/2\pi\tau$, where τ is the relaxation time. Acoustic absorption measurements in gases thus serve to indicate the presence of a relaxation effect and to evaluate τ. In many gases the relaxation frequency was found in the ultrasonic range. For example, for hydrogen it is 10 Mc. at 1 atm. For lower pressures it is proportionally reduced. Measurements of relaxation frequency have shed light on the internal energy states of gas molecules, especially on the probability of transitions between these states, a quantity proportional to the reciprocal of the relaxation time.

These acoustic-absorption measurements can also be of value in engineering. In the high-speed flow of gases and vapours on obstacles (*e.g.*, steam against turbine blades), sudden compressions and rarefactions are accompanied by temperature changes. The time during which such changes take place is controlled by

the dimensions of the obstacles and the flow velocity. If the time interval is of the same order as the relaxation time for transitions to and from internal energy states of the gas molecules, translational energy will be lost (similar to acoustic absorption in the gas) with a measurable decrease of impact pressure on the obstacle. Acoustical measurement of the relaxation time in the relevant pressure and temperature range is the best way to estimate the likelihood of this process.

Absorption with regard to liquids is not so satisfactory theoretically as for gases. For many organic liquids (*e.g.*, benzene and carbon tetrachloride) the excess absorption over that provided by classical viscosity and heat conduction can probably be accounted for by thermal relaxation. This is, however, not the case for water. A theory due to L. H. Hall assumes that molecules of water can take on two kinds of arrangement. In one they are arranged more or less tetragonally as in ice; in the other they are packed more closely together (cubic close-packed structure). These arrangements are analogous to states of translational and internal energy of gas molecules and a relaxation theory can be worked out for them. Termed structural relaxation, it implies a kind of relaxational compressibility: the difference between static compressibility as ordinarily measured and so-called instantaneous compressibility corresponding to very high-frequency sound waves. This leads to a compressibility that varies with frequency and hence yields dispersion as well as absorption. By reasonable choice of parameters the theory accounts for the observed excess of absorption over that classically computed and also gives the proper variation of the absorption coefficient with temperature from 0° to 100° C. The predicted dispersion is extremely small and had not yet been measured in the 1960s. Hall's theory was criticized and modified by C. Davis and T. A. Litovitz; however, the fundamental idea of structural relaxation has been confirmed.

Whether the theory of structural relaxation can be applied to water solutions like electrolytes is uncertain. Relaxation caused by chemical dissociation probably plays a role also. As in gases, it is likely that information about the liquid state (the most difficult state to understand) could be gained from study of absorption and dispersion mainly in the ultrasonic range. Efforts to use relaxation mechanisms in accounting for acoustic absorption in solutions of electrolytes increased in the 1960s (led by E. Yeager in the U.S.).

The relaxation-time theory of acoustic absorption in liquids has suggested the concept of bulk or volume viscosity. This is associated with a resistance related to the rate of volume compression as the more common shear viscosity is related to the rate of shearing strain. According to Hall's theory this bulk viscosity for water is several times as great as the shear viscosity. It decreases as the temperature increases. A bulk viscosity of this magnitude corresponds to a very high relaxation frequency; in water, for example, of the order of 10^{12} cycles/sec and hence outside the range of direct experiment.

Wave dispersion in fluids is basically a high-frequency (ultrasonic) phenomenon. More strictly, in gases it appears at large ratios of frequency to pressure; *i.e.*, it has been observed at relatively low frequencies (though not in the audible range) if the pressure is low enough. However, dispersion is a second-order phenomenon compared with absorption; that is, wave velocity differs from its value $c = \sqrt{\gamma p/\rho}$ in the simple classical theory only by terms of the order $c\omega^2\tau^2$, whereas the absorption coefficient depends on the first power of τ. This is clear from equations (62) and (63) for the case of shear viscosity, and the result is not materially different in the case of relaxation absorption and dispersion. Hence dispersion due to relaxation may be more difficult to detect than the associated absorption.

Acoustic dispersion in liquids has been particularly difficult to detect experimentally, since the velocity is very sensitive to temperature, and it is hard to keep temperature constant in acoustic transmission. Up to the 1960s no unambiguous case of dispersion in a pure liquid had been exhibited, though dispersion had been measured in liquid suspensions.

Elementary treatments of wave propagation assume that liquids (like gases) have no shear elasticity and hence can transmit compressional waves only. This has been the basis for the analysis presented here. However, it is not empirically verified in the case of viscous liquids like polybutylene, gels, and rubber solutions; called viscoelastic substances, they have shear elasticity and can propagate shear waves. Since their effective rigidity modulus increases with the frequency, the velocity of shear waves in this case is also frequency dependent. In polybutylene solutions the rigidity may rise from 3×10^7 dynes/cm^2 at 300 kc. per sec. to 5×10^9 dynes/cm^2 at about a megacycle. The viscosity of such liquids can be determined by acoustical measurements. The relaxation-time concept was profitably applied in the 1960s to account for acoustic absorption in viscoelastic substances, largely through the work of W. P. Mason in the U.S.

Of all liquids, probably liquid helium (*q.v.*) exhibits the most interesting and anomalous properties. Thus below about 2.2° K (its so-called lambda point) it loses practically all shear viscosity and gains high thermal conductivity. This modification is known as helium II and is called a superfluid. Its acoustical characteristics also show anomalies. Above the lambda point acoustic velocity increases as the temperature is lowered, reaching a maximum of 220 m/sec at 2.5° K. It then falls sharply to about 216 m/sec at the lambda point and thereafter rises again as the temperature is lowered, appearing to level off to about 239 m/sec as absolute zero is approached. Acoustic attenuation in liquid helium above the lambda point is explainable in terms of viscosity and heat conduction; no thermal relaxation appears to be involved. However, the attenuation reaches a maximum (for given frequency) at the lambda point and then declines very rapidly to a minimum at 2° K. It then rises again as the temperature is lowered to absolute zero. This peculiar behaviour is evidently due to some kind of unusual relaxation mechanism.

Liquid helium II can transmit a peculiar kind of thermal wave; *i.e.*, pressure and density remain constant throughout the liquids, but a temperature fluctuation is propagated. The velocity of this wave is zero at the lambda point but increases as the temperature falls, reaching a maximum of 20 m/sec at about 1.75° K, and then decreasing again as the temperature is lowered further. This wave phenomenon, first reported by L. D. Landau (*q.v.*), is called second sound, though this is questionable terminology. The source of the thermal radiation is usually a thermophone: a platinum ribbon to which an alternating current is applied, producing a harmonically varying source of heat. The waves are detected by a resistance thermometer. For more information on the properties of helium at low temperatures, *see* LOW-TEMPERATURE PHYSICS.

Liquid helium is a valuable adjunct in low-temperature acoustics, being used as a coolant in transmission studies at very low temperatures, especially in solids.

14. Diffraction and Scattering.—It has been mentioned that when a sound wave strikes an obstacle of finite size some of it is reflected, some is transmitted, and in general some gets around the obstacle (or is diffracted). Because of the obstacle the original sound beam is deflected and the pattern of intensity distribution is altered; it is customary to say that the beam is scattered. The resultant intensity pattern is a superposition of the original beam and the scattered energy due to reflection, transmission, and diffraction.

In theoretically estimating harmonic-wave scattering it is assumed that from the completely rigid obstacle there spreads out in every direction a wave of arbitrary amplitude and phase. When the particle displacement due to this scattered wave is added to that corresponding to the primary undisturbed wave the result must be zero everywhere on the surface of the obstacle. (This is idealized since there apparently is no such thing as a completely rigid obstacle that sound energy cannot penetrate. Nevertheless this approximation is useful for such objects as the human head.) The application of this boundary condition suffices to evaluate the unknown amplitude of the scattered wave and hence to give the scattering pattern. This general method can be applied readily only to obstacles of simple geometrical shape, such as cylinders and spheres. But many obstacles found in practice can be approximated reasonably well by such surfaces.

The scattering pattern depends on the relation between the wave-

length of the primary wave and the dimensions of the obstacle. Consider a sphere of radius a as the obstacle. If the wavelength λ is very small compared with $2\pi a$, the scattered wave on the side of the sphere opposite to the direction of the incident primary wave combines with the latter to produce zero intensity (destructive interference). Hence the obstacle forms a shadow that is sharper the smaller the ratio $\lambda/2\pi a$. Since this is analogous to optical shadows of objects with dimensions much larger than the wavelength of light (the case of geometrical optics), it is part of geometrical acoustics. For audible waves (wavelength in air greater than about 1.7 cm.) this type of scattering is important only for relatively large obstacles. However, for ultrasonic (high-frequency) waves it can be quite significant. Geometrical acoustics in the audible region is not as useful as geometrical optics for visible light, which has a much smaller wavelength.

For the opposite extreme, if the wavelength is very large compared with $2\pi a$, the primary wave is scattered in all directions and there is no distinct shadow. The intensity of the scattered wave is directly proportional to the square of the volume of the scattering particle (in this case a sphere) and inversely proportional to the fourth power of the wavelength. This is the Rayleigh scattering law in acoustics; its optical analogue helps account for the blue colour of the sky (*q.v.*). In the scattered radiation the higher frequencies will produce the high intensities. Sound waves scattered by a grove of trees appear to be raised in pitch. A similar phenomenon results when underwater sound waves are scattered by air bubbles.

The middle range in which the wavelength and the dimensions of the scattering obstacle are comparable is much more difficult to handle computationally. In general the intensity of the scattered radiation at any particular distance from the obstacle goes through a series of maxima and minima as the azimuth is varied. Relatively near the obstacle the diffraction pattern tends to be very complex, resembling the well-known Fresnel pattern in optics. Farther away it reduces to the smoother Fraunhofer diffraction pattern (*see* LIGHT: *Diffraction*).

The special case of diffraction by the human head can become evident to any careful observer. For given audible wavelength the acoustic intensity from a speaker's mouth is greatest directly in front. It falls off at right angles to this but rises again back of the head. This secondary maximum is only comparable with that directly in front at distances large compared with the wavelength. Very close to the head, however, the variation in intensity with azimuth is marked.

The head's effect on sound waves received at the ears is also a well-marked phenomenon. The resultant apparent intensity at the ears is a maximum when the line connecting the ears lies in the direction of the sound and falls in general to a minimum when the direction of the sound lies at right angles to this line.

This result is obtained theoretically by the use of the so-called reciprocal theorem of H. L. F. von Helmholtz (Lord Rayleigh, *Theory of Sound*, vol. ii, para. 294 [1945]): "If in a space filled with air which is partly bounded by finitely extended fixed bodies and is partly unbounded, sound waves be excited at any point A, the resulting velocity potential ϕ at a second point B is the same both in magnitude and phase as it would have been at A had B been the source of sound." Though stated for the velocity potential (equal in acoustic waves to the negative of the time integral of the excess pressure divided by the mean density), the theorem also applies to acoustic intensity and plays an important role in calibrating acoustical instruments for measuring radiation.

15. Waves in Finite Spaces; Standing Waves in a Tube.—Spreading sound waves or waves progressing in one direction only require an infinitely extended medium. If the medium is surrounded by a barrier to produce a closed finite space, reflection tends to occur, producing waves in several directions.

Consider a cylindrical tube, of length l and cross-sectional area S, closed rigidly at both ends. A disturbance anywhere in the air in the tube will be reflected from both ends and produce in general a series of waves traveling in both directions along the tube. From the geometry of the situation and the finite constant value of acoustic velocity, these must be periodic waves with frequencies fixed by the boundary conditions at the ends of the tube; *i.e.*, zero resultant value. The allowed frequencies of the waves in the tube satisfy

$$\sin kl = 0 \tag{66}$$

i.e., the allowed frequencies are

$$\nu_n = nc/2l \tag{67}$$

where n is any integer and c is the acoustic velocity in the tube. These are the frequencies of harmonic waves that can exist in the tube and still satisfy the boundary conditions at the ends. They are called the characteristic frequencies or normal modes of vibration of the air column. The fundamental frequency is

$$\nu_1 = \nu_f = c/2l \tag{68}$$

The higher frequencies, called harmonics or overtones, are multiples of the fundamental. It is customary to refer to the fundamental as the first harmonic; $n = 2$ gives the second harmonic or first overtone, and so on. Approximately the same set of characteristic frequencies hold for a cylindrical tube open at both ends, though the boundary conditions are different. Ideally at the end of such a tube it may be assumed that the acoustic excess pressure remains zero since it is open to the surrounding air. This is by no means obvious since there can readily be nonvanishing excess pressure in a wave progressing through the open air. It could be reasoned that at such an open end there is nothing to check the air displacement along the length of the tube and hence the particle displacement or velocity will be a maximum there. This means $\partial\xi/\partial x = 0$, which is equivalent to the vanishing of the excess pressure. The use of a probe microphone confirms this approximately by direct experimentation. It yields, as stated, the normal modes given by equation (67) and these are in approximate agreement with experiment.

Exact agreement is not obtained since the air at the end of the tube has inertia and must be disturbed by the wave traveling along the tube. Moreover, the disturbance at the end of the tube acts effectively as an acoustic source radiating into the open air, leading to loss of energy. Empirically, this latter phenomenon has only a small effect on the characteristic frequencies of the tube, but the former can be quite noticeable. Consider a tube closed at one end and open at the other. Application of the ideal boundary conditions just mentioned then leads to the set of characteristic frequencies

$$\nu_n = (n + \tfrac{1}{2})\, c/2l \tag{69}$$

but because of the open-end effect these are modified to

$$\nu_n' = (n + \tfrac{1}{2})\, c/2(l + S/C) \tag{70}$$

where S is the area of cross section of the tube and C (with dimensions of length) is known as the acoustic conductivity of the opening. It is difficult to compute C theoretically but for a circular opening it is of the order of magnitude of the diameter. Hence the term S/C in this case is approximately $\pi a/2$, where a is the radius of the circular tube. The characteristic frequencies are thus those of a tube of equivalent length $l + \pi a/2$, which for a short wide tube can be rather different from l. In general the quantity S/C is known as the end correction. It can be evaluated by direct measurement of the normal modes and is a useful method of obtaining C.

In the actual wave disturbance corresponding to a normal mode in a tube open at both ends, there are positions in the tube, specifically for

$$\cos n\pi x/l = 0 \tag{71}$$

at which the displacement is zero at all times. This cannot take place in a progressive wave; thus the wave disturbance corresponding to a normal mode is known as a standing wave. The positions of continuous null displacement are known as nodes, while the positions for which $\cos n\pi x/l = 1$ are called antinodes or loops, corresponding to maximum displacement. The distance between successive nodes is equal to a half wavelength of the particular mode. From the relation between displacement and excess pressure a displacement node is an excess pressure antinode and

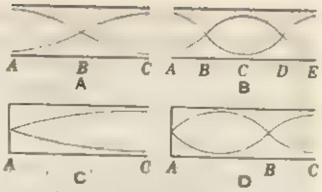


FIG. 12.—STANDING WAVES

(A) Fundamental or first harmonic and (B) second harmonic in tube open at both ends; (C) first harmonic and (D) second harmonic in tube open only at one end

vice versa. The tube is shown open at both ends in fig. 12 (A) and (B). Here in (A) A and C are antinodes and B (at centre) is a node. In (B) A, C, and E are antinodes and B and D ($\frac{1}{4}$ and $\frac{3}{4}$ of the distance along the tube) are nodes. The tube open at one end and closed at the other is exemplified in fig. 12 (C) and (D). In (C) A is a node and C an antinode. (All this refers to the phenomena of displacement.)

The resultant particle displacement in the standing wave produced by arbitrary initial conditions in a tube of length l open at both ends is taken to be the sum over all the normal modes; *i.e.*,

$$\xi = \sum_{n=0}^{\infty}(A_n \cos n\pi c/l \cdot t + B_n \sin n\pi c/l \cdot t) \cos n\pi x/l \qquad (72)$$

The assumption is known as the principle of superposition and was enunciated by D. Bernoulli in 1755. It is the basis for the Fourier analysis of an arbitrary disturbance in a finite medium into harmonic modes (*see* FOURIER SERIES).

From equation (72) the initial displacement at any point x in the tube is

$$\xi_0(x) = \sum A_n \cos n\pi x/l \qquad (73)$$

while the initial velocity is

$$\dot{\xi}_0(x) = \sum n\pi c/l \cdot B_n \cos n\pi x/l \qquad (74)$$

The dot in $\dot{\xi}_0(x)$ denotes partial differentiation with respect to time. Note that $\xi_0(x)$ and $\dot{\xi}_0(x)$, though presumably arbitrary, are here represented as sums of harmonic components in x. The evaluation of the coefficients A_n and B_n proceeds according to the Fourier rule and it is found that

$$A_n = 2/l \cdot \int_0^l \xi_0(x) \cos n\pi x/l \cdot dx \qquad (75)$$

$$B_n = 2/n\pi c \cdot \int_0^l \dot{\xi}_0(x) \cos n\pi x/l \cdot dx \qquad (76)$$

This assumes the mathematical validity of the expansion. Physically possible forms for $\xi_0(x)$ and $\dot{\xi}_0(x)$ in general satisfy the mathematical conditions.

In addition to sound disturbances propagated in a cylindrical tube, the same type of analysis can be applied to any confined space; *e.g.*, a room with or without open windows. The reflection of the disturbance from the boundaries of the room produces a three-dimensional standing wave pattern that can be analyzed into a set of normal modes or characteristic frequencies depending on the shape and dimensions of the room. Some of these modes may be much more prominent than others and have a good deal to do with the acoustics in the room; absorption at the walls and in the air itself plays an important role.

16. Wave Pulses; Group Velocity.—Thus far attention has been restricted to the propagation of continuous progressive waves like

$$\xi = A \cos(\omega t - kx) \qquad (77)$$

in which for given x there are nonvanishing displacements no matter what t is, and vice versa, or standing waves like

$$\xi = A \sin \omega t \cdot \sin kx \qquad (78)$$

where again the displacement occurs for all x and all t. These are ideal situations. Instead of finding the wave of equation (77), in practice, waves are encountered in which the disturbance at any instant is localized in a bounded region of space as in fig. 13, which pictures the disturbance at the instant t as a function of x.

FIG. 13

From infinity on the left up to A, the value of ξ is zero; from A to B, the value of ξ is given by equation (77), and from B to infinity on the right ξ is again zero. This is known as a finite wave train or wave pulse. As time passes, the train moves from left to right carrying the disturbance with it. In empirically producible wave disturbances this is what is found.

A wave pulse may be represented by a superposition of continuous and infinite harmonic wave trains, like equation (77), if the frequencies do not form a discrete set like the normal modes in a tube; the method requires infinitely many continuously distributed frequencies. The amplitude of each harmonic component depends on its frequency or wavelength and may become negligibly small for sufficiently great and small frequencies; thus the band of frequencies necessary for the harmonic decomposition of the pulse effectively will be finite if the pulse is finite in length. For example, consider the pulse

$$\begin{aligned} \xi(x - ct) &= \cos k_0(x - ct), & |x - ct| &< L/2 \\ \xi(x - ct) &= 0, & |x - ct| &> L/2 \end{aligned} \qquad (79)$$

which is the situation pictured in fig. 13. The amplitude $A(k)$ of the harmonic component corresponding to wave parameter k then is

$$A(k) \doteq \frac{\sin(k_0 - k)L/2}{\pi(k_0 - k)} \qquad (80)$$

By plotting $A(k)$ as a function of $(k_0 - k)$ it is found that the range of wave parameter k necessary to represent the traveling pulse adequately is inversely proportional to L, the length of the pulse. A very long wave train therefore needs only a very small range of parameter to represent it and in the limit of $L \rightarrow \infty$, the train becomes $\cos k_0(x - ct)$ precisely. If the pulse is very short $k_0 - k$ becomes correspondingly large.

These considerations apply only if the wave velocity c is independent of the frequency or wave parameter. If there is dispersion the situation is more complicated. On the assumption that only a finite range Δk of wave parameter about a mean parameter k_0 is adequate for the representation of the pulse, the wave train can be expressed approximately as a single sinusoidal wave of angular frequency ω_0 with an amplitude that is a function of time and space through the quantity

$$x - (d\omega/dk)_0 t \qquad (81)$$

The meaning of $(d\omega/dk)_0$ is that the dispersion of the medium is expressed by writing ω as a function of k:

$$\omega = \omega(k) \qquad (82)$$

Then differentiate ω with respect to k and take the value of the derivative corresponding to $k = k_0$. A point proceeding in the x direction with velocity

$$U = (d\omega/dk)_0 \qquad (83)$$

will always be at same amplitude. If it travels at this velocity with the maximum amplitude it will be a point where the individual harmonic components in the original wave train heap up, so to speak, to form what may be called a wave group. The velocity of progress of the group is called the group velocity and is given by equation (83).

The group velocity also can be conveniently expressed as

$$U = \frac{c_0}{1 - (\nu/c_0)(dc/d\nu)_0} \qquad (84)$$

where the phase velocity c is expressed as a function of the frequency ν, and c_0 is the mean phase velocity in the group. If $c = c_0$, a constant $U = c_0$ independent of frequency and of the group velocity coincides with the constant phase velocity. Since acoustic dispersion in fluids is generally very small over obtainable frequency ranges (save for hydrogen), the group velocity will not in general differ much from the phase velocity. At any rate in the use of short-pulse techniques for measuring acoustic velocity in fluids it is the group velocity that is measured. In dispersion associated with fluid viscosity, heat conduction, and relaxational effects, the velocity in general increases with the frequency. Hence $dc/d\nu > 0$ and $U > c_0$: the group velocity is greater than the

average phase velocity; this is opposite to the case of normal dispersion of light.

The advantage of short-wave pulses in measuring acoustic properties is that undesirable reflections can be separated from the main beam. The echo method of acoustic depth recording as well as gun ranging and the detection of submerged objects in water (sonar) depend on the use of pulse techniques.

From its definition, it may be assumed that the group velocity is that with which the energy in the wave train is transmitted. This holds approximately, but requires more consideration than will be given here.

It is often desirable to amplify an acoustic pulse by transforming it into an electrical pulse in an oscillating circuit and then feeding it into an amplifier. If the original shape of the pulse is to survive amplification undistorted the uniform frequency response of the amplifier must cover a sufficiently large band of frequencies. Thus, for the pulse considered in the example above, the frequency band must correspond to a wave-parameter band satisfying the approximate condition

$$(k_0 - k)\ L/2 \doteq \pi \tag{85}$$

17. Radiation Pressure.—Thus far when the pressure associated with an acoustic wave has been mentioned, it referred to the excess pressure produced by the disturbance and propagated through the medium. For a plane harmonic progressive wave the excess pressure is a sinusoidally varying function of time and space and hence vanishes if averaged over the time at any particular place. It does of course have a nonvanishing mean square value. However, associated with a progressive acoustic wave in any physically realizable medium there is also a pressure that does not vanish on the average, thus making a net addition to the static pressure prevailing in the medium. This is known as the radiation pressure. It is a so-called second-order quantity because it is much smaller than the root-mean-square (square root of the average of the square) acoustic excess pressure. In the ordinary acoustic equations as developed here it is therefore usually neglected. However, it is always present because of the essential nonlinearity of the acoustical equations if higher-order terms are included. This means that if the particle displacement in an acoustic wave has the form of equation (3), the corresponding acoustic excess pressure will be made up not merely of a corresponding linear term proportional to sin $(\omega t - kx)$ but also a term proportional to $\sin^2 (\omega t - kx)$, with a much smaller coefficient of proportionality. This latter term, when averaged over time, gives rise to the steady radiation pressure.

The simplest expression for the radiation pressure is

$$p_r = \frac{p_e \partial\xi/\partial t}{c} \tag{86}$$

where p_e is the root-mean-square excess acoustic pressure and $\partial\xi/\partial t$ is the root-mean-square particle velocity. Since the latter is always smaller than the wave velocity c, it follows that $p_r << p_e$. In spite of its small value relative to p_e, the radiation pressure p_r produces some important consequences. Its existence implies that any solid object in the path of an acoustic beam will undergo a steady pressure.

This may be used to estimate the acoustic intensity, since p_r is directly proportional to the intensity (*see* section V.1 below). Moreover, under some conditions p_r is associated with the production of steady streaming of the fluid medium through which the sound wave is passing. This streaming may produce interesting physical effects (section V.6 below).

18. Shock Waves.—In a macrosonic or high-intensity sound wave, a disturbance that starts out with a symmetrical profile (fig. 2) will very soon develop a steepening of the wave front. This process is opposed in fluids like air or water by the attenuation mechanisms mentioned in section II.13.

The result can be a saw-toothed curve (fig. 14) in which at points *A, B, C,* and so on, there are considerable changes in pressure over a very small space. Ideally there is a genuine discontinuity in pressure, density, and flow velocity at these points. Empirically there is no discontinuity, but the drop in pressure or density occurs over an interval order of 10^{-4} cm. in a gas under standard conditions; *i.e.*, about 10 mean free paths of the fluid. This transition state moves through the medium faster than the normal acoustic velocity. Called a shock wave it can result from an explosion or from a solid object (*e.g.*, a missile or airplane) moving through a fluid faster than acoustic velocity. Shock waves can be observed in a tube through the sudden rupture of a membrane separating a region of high pressure from one of low pressure. The thickness of a shock front, as it is called, can be estimated from its optical reflectivity (associated with the change in the index of refraction) or by the change in temperature through it. Shock fronts can be studied with an optical interferometer; the rapid change in density across the front produces a shift in the interference fringes. The strength of a shock is defined as the ratio of the pressure just ahead of the front to that just behind it.

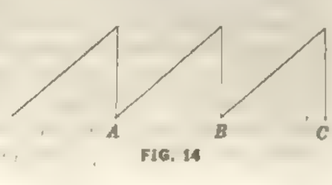


FIG. 14

A shock wave is reflected by an obstacle, but the law of reflection for ordinary low-amplitude waves is followed only for angles less than about 40° for strong shocks. For larger angles the reflection is complicated. Similar considerations apply to refraction and diffraction. By the 1960s there was an extensive literature on shock waves and their significance in physics and chemistry (*e.g.*, *see* AERODYNAMICS: *Supersonic Aerodynamics;* AERONAUTICS. *Airfoils: Shock-Wave Phenomena;* EXPLOSIVES: *Deflagration and Detonation*).

III. ACOUSTIC SOURCES

1. Introduction.—Sound-wave disturbances must be produced before they can be propagated. An appropriate change in the density at a point in an elastic medium may be said to be a source of sound. This accounts for the great variety of sound sources, since density change may be produced in many ways, including mechanical, thermal, electrical, magnetic, and chemical actions. The most common sound waves are periodic or considered as compounded of periodic (harmonic) components. The simplest and most important sources are vibrating portions of solids, liquids, and gases.

2. Damped and Forced Vibrations.—In a mechanical vibrator any portion of the system that is displaced from equilibrium tends to return to its undisplaced condition. Such a vibrator is characterized by a mass or inertia factor m; a stiffness factor f representing the elastic tendency to return to the equilibrium state, and a damping factor R, representing the tendency for the motion resulting from the original disturbance to decay instead of continuing forever. Consider a weight suspended from an ideally weightless spring fastened to a rigid ceiling. When the weight is displaced from rest position and let go, it tends to move to the original position. The inertia factor here is the mass of the weight, the stiffness factor is the applied force per unit stretch of the spring, and the damping factor is provided by the internal friction of the spring, by the viscosity of the surrounding medium (*i.e.*, the air), and by the loss of energy to the medium as sound radiation. This is not generally an efficient sound source. A membrane that presents a larger vibrating and radiating area to the medium is more practical, although assignment of the three vibrational factors is more difficult. For example, though the equivalent mass of the vibrator is proportional to the mass of the membrane it is not equal to it (*see* section III.5 below). However, there is great analytical advantage in treating any vibrator in terms of these factors.

The equation of motion of a vibrator with a displacement from equilibrium denoted ξ is

$$m\ddot{\xi} + R\dot{\xi} + f\xi = 0 \tag{87}$$

where again the dot denotes time differentiation; *i.e.*, $\ddot{\xi} = \dfrac{d^2\xi}{dt^2}$. The form of the solution depends on the interrelation of m, R, and F. If

$$f > R^2/4m \tag{88}$$

vibrations result and the solution is

$$\xi = Ae^{-Rt/2m}\cdot\cos(\omega t+\alpha) \tag{89}$$

where e is the base for natural logarithms,

$$\omega = \sqrt{f/m - R^2/4m^2} = 2\pi\nu \tag{90}$$

and ν is the frequency of the vibrations. A is called the initial amplitude and α is the initial phase angle or epoch of the motion. Because of the term $e^{-Rt/2m}$ the amplitude of the oscillations steadily diminishes with time and ξ plotted as a function of t appears as in fig. 15. This is called a damped oscillation; equation (89) describes its so-called free vibration; *i.e.,* uninfluenced by any outside force. The extent of the damping is described by $2m/R$, called the decay modulus. Alternatively, the logarithmic decrement δ is used; it is the natural logarithm of the ratio of two successive maxima of the ξ, t curve, and is given by

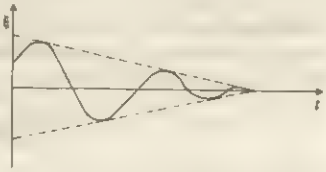


FIG. 15

$$\delta = RP/2m \tag{91}$$

where P $(= 1/\nu)$ is the period of the vibrations (*i.e.,* the time for one complete to-and-fro motion of the vibrating mass, or the time from one maximum of the curve in fig. 15 to the next. The larger R is for given P, the larger is δ and the more rapid is the decay.

For acoustic applications in which electric circuits are the source of mechanical oscillations, the so-called quality factor Q is used to describe the damping. In terms of the logarithmic decrement this takes the form

$$Q = \pi/\delta \tag{92}$$

Thus a highly damped oscillator is said to have a low Q and vice versa.

The quantities A and α depend on the initial conditions of the motion. If the value of ξ at $t = 0$ is ξ_0 and that of the displacement velocity $\dot{\xi}$ is $\dot{\xi}_0$ then

$$A = \sqrt{\xi_0^2 + (\dot{\xi}_0 + R\xi_0/2m)^2/\omega^2} \tag{93}$$

$$\alpha = \arctan\left(-\frac{\dot{\xi}_0 + R\xi_0/2m}{\omega\xi_0}\right) \tag{94}$$

If $f < R^2/4m$ (*i.e.,* if there is relatively large damping) there are no vibrations and ξ makes a single swing and decays to zero. Such overdamped systems are not considered in steady state acoustics, though they can serve as sources of sound pulses.

The average rate of dissipation of energy of the vibrating system due to the damping force is

$$\overline{\dot{E}} = -\overline{R\dot{\xi}^2} \tag{95}$$

where the bar indicates the average taken over an arbitrary time interval. If $\overline{\xi^2{}_{\max}}$ denotes the mean-square amplitude over the time interval

$$\overline{\dot{E}} \doteq -R\omega^2\overline{\xi^2_{\max}}/2 \tag{96}$$

If the damping arose entirely from radiation of acoustical energy into the surrounding medium, equation (96) would give the acoustically radiated power output. This is never the case, however; in general the power output is some proper fraction of $\dot{\mathrm{E}}$, measuring thereby the acoustical efficiency of the source.

Damped vibrations provide only a transient sound source. To maintain steady oscillations an external force must be applied. If the force varies harmonically with angular frequency ω_0 and amplitude F_0 the system will reach a steady state of vibration with the same angular frequency, which in general will differ from the free vibration frequency of equation (90). The steady state displacement has the form

$$\xi = \frac{F_0\cos(\omega_0 t - \beta)}{\sqrt{\omega_0^2R^2 + (f - m\omega_0^2)^2}} \tag{97}$$

where

$$\beta = \arctan(\omega_0R/[f - m\omega_0^2]) \tag{98}$$

is the phase angle between the force and the displacement. Other things being equal, the smaller the damping factor R, the more nearly ξ is in phase with the force or 180° out of phase with it.

In some ways the velocity is more significant. Here

$$\dot{\xi} = \frac{F_0\cos(\omega_0 t - \gamma)}{\sqrt{R^2 + (m\omega_0 - f/\omega_0)^2}} \tag{99}$$

where

$$\gamma = \arctan\left(\frac{m\omega_0 - f/\omega_0}{R}\right) \tag{100}$$

is the phase angle between force and velocity; clearly $\beta = \pi/2 - \gamma$. Both ξ and $\dot{\xi}$ are harmonic in time with frequency $\nu_0 = \omega_0/2\pi$. The amplitude of $\dot{\xi}$ is a maximum for

$$\omega_0 = \sqrt{f/m} \tag{101}$$

and this is called the resonance (angular) frequency ω_{res} of the system. In words, when the frequency of the force is adjusted to $1/2\pi\cdot\sqrt{f/m}$, the system is said to resonate with the force and maximum steady state velocity oscillations ensue. The displacement does not attain maximum steady state amplitude at resonance but at the frequency

$$\nu_1 = \tfrac{1}{2\pi}\cdot\sqrt{f/m - R^2/2m^2} \tag{102}$$

If $R^2/2m^2 << f/m$, this is close to the resonance frequency and to the free oscillation frequency of equation (90).

The analogy with electrical oscillations of an L, R, C circuit is striking. The denominator in the right-hand side of equation (99) is called the mechanical impedance of the vibrator. Thus

$$Z = \sqrt{R^2 + (m\omega_0 - f/\omega_0)^2} \tag{103}$$

is a minimum for resonance and equal to the damping factor R.

Since the harmonic force maintains the system in a steady state of vibration, the average rate of energy flow from the source of the force to the system just equals the average rate of dissipation of energy through the damping as indicated in equation (95). This average flow of energy is a maximum at resonance.

If the amplitude of $\dot{\xi}$ is plotted as a function of $\omega_0 - \omega_{\text{res}}$, a so-called resonance curve results. The maximum $\dot{\xi}$ occurs at $\omega_0 - \omega_{\text{res}} = 0$ and decreases on either side of the origin. The nature of the resonance peak is controlled by the damping factor R. If R is small the peak is high and sharp; if R is large the peak is low and broad. In the first case the response of the system to the force is more selective with respect to frequency than in the second.

3. Coupled Systems.—As has been emphasized, the vibrations of a single mass particle are only an idealized representation of sound-source motions. Actual sound sources are extended bodies or collections of particles, like strings, membranes, rods, and plates. Such sources may be thought of as collections of vibrating particles coupled together. A simple illustration is provided by two mass particles attached to a weightless, stretched, flexible, horizontal string of finite length. If the particles move in a vertical plane only, the motion of each is affected by that of the other through the tension of the string joining them. The vibrations of such a system correspond to the superposition of component vibrations of two different frequencies, even if the masses are the same and the attachment to the string is perfectly symmetrical. The generalization of this situation is that of n particles attached at regular intervals to the stretched string (the so-called loaded string). There are now n characteristic frequencies and the motion of each particle is a superposition of n harmonic components with these n different frequencies. Finally, as n increases indefinitely the system should act more and more like a continuous string. Mathematically the string is treated not as a collection of discrete particles but as a continuous medium (albeit linear) somewhat like a one-dimensional continuous elastic fluid that exhibits wave motion when disturbed from equilibrium.

4. Strings; the Helmholtz Resonator.—If an indefinitely long horizontal flexible string of mass ρ_l per unit length is stretched

with tension T along the x axis any disturbance represented by a transverse displacement ξ from the equilibrium (rest) position is propagated in accordance with the wave equation (R. B. Lindsay)

$$\frac{\partial^2 \xi}{\partial t^2} = T/\rho_l \cdot \frac{\partial^2 \xi}{\partial x^2} \quad (104)$$

the solution of which corresponds to transverse waves propagated along the string with velocity

$$c = \sqrt{T/\rho_l} \quad (105)$$

If the string is of finite length l, and fastened at both ends, the solution of equation (104) that satisfies the boundary conditions $\xi = 0$ at $x = 0$ and $x = l$ is of the same form as that discussed in section II.15 above. Thus the motion of the string is compounded of a series of harmonic standing waves or normal modes of vibration with frequencies given by

$$\nu_n = nc/2l \quad (106)$$

where n is integral. The relative importance of the various modes in any given motion of the string depends on the initial conditions; *i.e.*, the methods used to excite the string into vibration. Commonly these are plucking (harp, mandolin), striking (piano), and bowing (violin). Each of these methods in turn permits considerable variation in the nature of the acoustic output depending on where the initial disturbance is produced. When a string is plucked at the centre, for example, all the even harmonics are absent, seriously affecting the musical quality of the emitted waves. Plucking near one end restores all the harmonics and improves the quality. The nature of the plucking instrument also is of importance. The struck string involves more elaborate considerations, since the impact of the hammer is a complicated function of the time. All harmonics having a node at the point where the string is struck will have zero amplitude. Bowing is the most complicated way of exciting a string, since the communication of energy to the string is more or less continuous. The oscillation is a special kind of forced vibration that may be properly called self-maintained.

The transmission of energy to a string by bowing is not precisely like the forced vibrations considered earlier since the force in this case is not harmonic. The string receives energy from the bowing, but the frequency of vibration is conditioned by the string itself, its mounting, and the position of the bow. The resulting oscillations are often termed auto-oscillations or self-maintained oscillations. Their mathematical study can be facilitated by treating the force as a kind of negative damping. Another acoustic illustration is provided by the blown organ pipe (jet-edge tones). The electrical oscillations produced by an ordinary simple vacuum tube circuit are also of this character.

The output of sound energy from a vibrating string depends on its surroundings; *i.e.*, the presence of other resonating systems such as air cavities and solid baffles (*see* STRINGED INSTRUMENTS).

The Helmholtz resonator is a good example of a resonating air cavity. It is basically a hollow sphere with a circular opening (with a diameter of the order of one-tenth that of the sphere) and usually equipped with a very short neck. An acoustic source brought near the resonator can set the air in the opening in vibration. The action then is effectively that of a simple vibrating system of mass equal to that of the moving plug of air in the orifice, of stiffness provided by the elasticity of the air cushion inside the sphere, and of resistance largely due to the radiation of sound energy from the opening. Analysis indicates that the appropriate elements are respectively

$$\begin{aligned} m &= \rho_0 S^2/C \\ f &= \rho_0 c^2 S^2/V \\ R &= \rho_0 \omega^2 S^2/2\pi c \end{aligned} \quad (107)$$

where ρ_0 is the density of the air, V the volume of the resonator sphere, S the cross-sectional area of the opening, C the acoustic conductivity of the opening, and c the acoustic velocity in air. The resonance angular frequency of the resonator is approximately

$$\omega_{res} = c\sqrt{C/V} \quad (108)$$

The amplification A of the resonator is defined to be the ratio of the square of the excess pressure amplitude in the opening to that present at the same point in the absence of the resonator. At resonance this has the form

$$A = 4\pi^2 V/C^3 \quad (109)$$

Resonators serve as useful acoustic amplifiers (for single frequencies) for numerous types of sound sources and receivers.

5. Membranes.—A vibrating circular membrane is a more practical source of sound of high intensity than a vibrating string since it is two-dimensional. The equation of motion of a perfectly flexible membrane stretched with surface tension S (force per unit length across the membrane) and surface density ρ_s (mass per unit area) is of the form

$$\frac{S}{\rho_s}\left(\frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2}\right) = \ddot{\xi} \quad (110)$$

if the equilibrium position of the membrane is in the xy plane and ξ denotes a displacement perpendicular to this plane. If the membrane is circular it is simpler to replace x and y by the polar coordinates r and θ and write equation (110) in the form

$$\frac{\partial^2 \xi}{\partial r^2} + \frac{\partial \xi}{r \partial r} + \frac{\partial^2 \xi}{r^2 \partial \theta^2} = \frac{1}{c^2}\ddot{\xi} \quad (111)$$

with the velocity

$$c = \sqrt{S/\rho_s} \quad (112)$$

The solutions of equation (111) are complicated and will not be given here. It is of interest, however, to note the normal modes of oscillation of a membrane of radius a rigidly clamped around its periphery, since many membrane sound sources are used in this way. The three lowest harmonics have frequencies

$$\nu_1 = \frac{0.77c}{2a},\ \nu_2 = \frac{1.22c}{2a},\ \nu_3 = \frac{1.76c}{2a} \quad (113)$$

where the fundamental is ν_1. The higher modes do not form a harmonic series in the sense of the characteristic frequencies of the vibrating string or tube. The standing waves that correspond to the normal modes of the clamped vibrating membrane have nodal lines similar to the nodal points in the string or tube. These are both circles and radii that can be experimentally demonstrated by covering the vibrating membrane with lycopodium powder which tends to collect along the nodal lines.

For practical purposes it is often sufficiently accurate to consider a circular membrane clamped at the periphery and vibrating in its fundamental mode as equivalent to a vibrating piston in which all points have the same displacement and velocity at the same instant. This piston may be thought of as having effective mass m and effective stiffness f, so that its resonance frequency is given by equation (101). Using the approximation that in the fundamental mode the shape of the displaced membrane is approximately paraboloidal, the effective mass is one-third of the actual mass and the effective stiffness is 2π times the surface tension S. The fundamental frequency given by equation (101) in this case is only about 1% higher than 0.77 $c/2a$ as shown in equations (113).

6. Longitudinal Vibrations of Rods.—For a straight rod of uniform cross section, the length of which is very large compared with its diameter, the differential equation of motion for a small extensional disturbance ξ (*i.e.*, directed along the rod) is

$$\ddot{\xi} = Y/\rho_0 \cdot \partial^2\xi/\partial x^2 \quad (114)$$

in which Y is Young's modulus and ρ_0 is the actual volume density. The disturbance thus travels along the rod as a longitudinal wave with velocity

$$c = \sqrt{Y/\rho_0} \quad (115)$$

For most hard solids this velocity is of the order of 3,000 m. per sec., though it ranges from about 1,000 to 6,000 m. per sec., depending on density and elasticity.

The specific acoustic resistance for a longitudinal wave in a solid rod is $\rho_0 c$, as in the case of compressional waves in a fluid.

A finite vibrating rod serving as a source of sound must satisfy certain boundary conditions. These will lead to the production of

standing waves as in the case of air vibrations in a tube. Thus if the rod vibrates so that both ends are rigidly clamped, it requires that $\xi = 0$ at $x = 0, l$ (if l is the length of the rod). This leads to the normal modes or characteristic frequencies

$$\nu_n = nc/2l \tag{116}$$

the fundamental corresponding to $n = 1$.

On the other hand if one end of the rod only is clamped and the other end is left free, the corresponding frequencies are

$$\nu_n' = (n + \tfrac{1}{2})\, c/2l \tag{117}$$

A common case is that in which the rod is clamped at the middle. The fundamental frequency is still given by $c/2l$, since this provides for a displacement node at the centre. However, in this case the even harmonics are absent since they do not have nodes at the centre, and rigid clamping at the centre must always correspond to a node there.

The modes of oscillation of a longitudinally vibrating rod are usually studied by sprinkling the rod with lycopodium powder which gathers at the nodes, thus permitting the measurement of wavelength and (through knowledge of the driving frequency) the longitudinal wave velocity. If the transverse dimensions of the rod are not negligible compared with the length, wave transmission along the rod is much more complicated and equation (115) no longer holds. When an elastic rod is stretched there also must be a change in cross-sectional area. This leads to radial vibrations and the coupling between these and the longitudinal variety may be expected to lead to complicating effects, especially when the wavelength of the longitudinal wave is of the order of magnitude of the transverse diameter. For very low frequencies with corresponding long wavelengths the coupling between the two types of mode is weak and the velocity of equation (115) is observed. But as the frequency increases, say to 200 kc. for a rod in which the diameter is 1 cm., the effective longitudinal velocity drops noticeably from the value given by equation (115). This is a type of dispersion; *i.e.*, change of velocity with frequency. It may be called configurational, since it is associated with the finite transverse dimensions of the rod rather than with internal molecular effects. As the frequency is further increased, say to 500 kc. in the case of the rod just mentioned, the velocity falls asymptotically to the value for shear or torsional waves in the rod: $c = \sqrt{\mu/\rho_0}$, where μ is the rigidity or shear modulus. Further increase of frequency produces little change. A complete explanation of these effects demands a complicated analysis of elastic wave propagation under the appropriate boundary conditions.

Just as in elastic wave propagation in a fluid, so also in a solid there are dissipative effects that lead to absorption of the wave energy with a loss of intensity. In analogy with the discussion for equation (61), the displacement from equilibrium in a harmonic longitudinal disturbance propagated along the rod may be written

$$\xi = \xi_0 e^{-\alpha x} \cos(\omega t - kx) \tag{118}$$

where ω ($= 2\pi\nu$) is the angular frequency of the wave and α is the absorption coefficient. For most solid rods at frequencies above a few cycles per second, α is not proportional to the square of the frequency as in the case of liquids and gases but more nearly approximates a linear dependence. For polycrystalline media (with metallic grains) this is corrected by an additional term depending on the fourth power of the frequency. Thus for polycrystalline aluminum, Mason and H. J. McSkimin found

$$\alpha = a_1\nu + a_2\nu^4 \tag{119}$$

with $a_1 = 0.845 \times 10^{-9}$ sec/cm, $a_2 = 3.74 \times 10^{-30}$ sec^4/cm. The linear dependence on the frequency is thought to have its origin in internal friction or solid viscosity (which evidently operates in a different way from the viscosity of fluids), whereas the fourth-power dependence is associated with a scattering of the elastic wave by the metallic grains of which the polycrystalline material is composed. This scattering is analogous to that which in the case of light is considered to account for the blue colour of the sky. Equation (119) was found to work well in the frequency range from 2 to 15 Mc. per sec.

Equation (118) expresses absorption of a longitudinal harmonic wave in a solid in terms of the amplitude decay with distance of travel. The absorption may also be expressed in terms of decay in time. This is a more appropriate way to express it in a finite rod, set freely vibrating by some energy source and then set free from the energy source and allowed to die down. In this case the expression for the displacement at any place x at time t in a progressive harmonic wave in the positive x direction is

$$\xi = \xi_0 e^{-\beta t} \cos(\omega t - kx) \tag{120}$$

in which β is a kind of temporal absorption coefficient. At any particular point on the rod (definite value of x) the displacement behaves like a damped harmonic oscillator. The logarithmic decrement is

$$\delta = \beta P = \pi/Q \tag{121}$$

where P ($= 1/\nu$) is the period of the wave; the quality factor Q was introduced in equation (92). Still another way of referring to the attenuation is to use the average fractional loss of energy W in the radiation per unit volume for one cycle of the harmonic disturbance. Since the energy per unit volume is proportional to the square of the amplitude,

$$W = 2\beta P = 2\delta \tag{122}$$

It can be shown that independent of the precise attenuating mechanism the fractional energy loss can be expressed in the form

$$W = 2\alpha\lambda \tag{123}$$

where λ is the wavelength. This makes it possible to express the decrement and Q value in terms of the absorption coefficient α.

Attenuation also arises from the friction drag of the surrounding fluid medium on the vibrating surface of the rod. This is generally smaller than that from internal friction. The rod also radiates sound energy to the surrounding medium; this accounts for loss of energy in the vibrations. This effect is also very small except at very high frequency. In other words, at low frequencies a vibrating rod is not a very efficient acoustic source unless it is coupled to a membrane or diaphragm with a larger radiating surface.

Transmission in solid rods, mainly at ultrasonic frequencies, has been studied to learn more about the physical properties of the solid material; *e.g.*, its elastic constants. Thus equation (115) permits the evaluation of Young's modulus Y from a knowledge of the density ρ and the velocity c of longitudinal waves in the rod, provided the frequency used is such that the wavelength is considerably larger than the width or diameter of the rod (if circular). High-frequency waves are particularly suitable for this purpose, since the wavelength can be long enough to satisfy the requirement just stated and still short enough to permit its measurement by a powder pattern on the rod produced by standing waves.

For optically transparent solids the determination of elastic constants is easiest by an optical technique based on the Debye-Sears effect. If a transparent solid cube is filled with an ultrasonic standing wave pattern produced by a piezoelectric quartz oscillator (*see* section III.11) cemented to one of its faces, and a beam of light is then passed through it, an optical diffraction pattern is formed. If the solid is isotropic the pattern consists of rings that elasticity theory readily associates with the bulk and shear elastic moduli as well as Poisson's ratio σ (*see* section III. 8). For transparent crystalline solids the same method is applicable, though the patterns are more complicated.

For the study of transmission through solid rods and other extended solids, pulse techniques are often more practical than standing wave patterns, especially if the solids are optically opaque. With appropriate electronic equipment it is possible to produce ultrasonic pulses of only a few milliseconds duration and measure with great accuracy their transit time through a solid specimen; in this way velocity can be readily determined. Recall that it is the group velocity that is found in this way. The method is very simple in principle but demands many precautions against errors if high precision is sought.

Acoustic attenuation in a solid can also be measured by the

pulse method by noting the relative amplitudes of the initial pulse and that returning to the source after a double transit of the specimen under study. When rubberlike materials are investigated by this method, considerable velocity dispersion is found. Thus for natural rubber at 0° C the velocity ranges from about 1,800 m. per sec. at 0.1 Mc. to about 2,700 m. per sec. at 10 Mc. The dispersion decreases as the temperature is raised. Attenuation measurements in such materials indicate relaxation mechanisms similar to those for liquids and gases.

Solids at very low temperatures (*i.e.*, below 20° K) became available through the use of liquid helium. Particular attention has been paid to those solids that become superconducting at low temperatures, such as lead, tin, and indium. These metals appear to lose all electrical resistance at temperatures near 4° K. Thus tin becomes superconducting below 3.73° K, its so-called transition temperature. If ultrasonic radiation (*e.g.*, at 10 Mc.) is passed through a rod of such a superconducting metal, attenuation above the transition temperature increases as temperature decreases until the transition temperature is reached, when a sharp discontinuity in slope appears, and as the temperature is lowered attenuation decreases rapidly. On the other hand, if a magnetic field is imposed on the specimen, superconductivity is destroyed and attenuation continues to increase as the temperature falls below the normal transition point. It thus appears that the acoustic attenuation arises from a mechanism closely related to that responsible for the superconductivity. It should be remembered that a metal crystal consists of a more or less stable lattice of positive ions and a collection of readily mobile, more or less free conduction electrons that are responsible for electrical conductivity. The lattice ions are never wholly motionless, but vibrate about their mean positions; above the transition temperature they can exchange energy with the conduction electrons. When a high-frequency sound wave passes through the metal energy exchange is increased, and this leads to attenuation. Evidently this cannot happen in the superconducting state. It was believed in the 1960s that in this state it is more difficult to transfer energy from lattice to electrons and the attenuation mechanism ceases to work.

At temperatures well above the superconducting transition and for nonsuperconductors other relaxation mechanisms lead to acoustic attenuation in solids. These are related to imperfections in the lattice of which dislocations (*i.e.*, displacements of one layer of metal ions relative to the adjoining one) are examples (*see* CRYSTALS, DISLOCATION OF).

7. Flexural and Torsional Vibrations of Rods.—Of possibly more practical importance than longitudinal vibrations are the bending variety in which the propagated disturbance is a flexure at right angles to the length of the rod.

The disturbance in this case is no longer propagated in accordance with the simple wave equation (6). Nevertheless, it can be shown that the flexure can move along the rod as a simple harmonic wave provided the velocity is a function of frequency; *i.e.*, the bending rod acts as a dispersive medium. The phase velocity is found to have the form (H. Lamb)

$$c = \frac{kK}{\sqrt{1+k^2K^2}}\cdot\sqrt{\frac{Y}{\rho_0}} \tag{124}$$

where Y is Young's (or the stretch) modulus of the rod and K is the radius of gyration of the cross-sectional area about an axis through its centre of symmetry normal to the plane of flexure. As usual k is the wave parameter. Wave energy is transmitted by flexure along the rod with the group velocity

$$U = 2c - \frac{k^3K^3\sqrt{Y/\rho_0}}{(1+k^2K^2)^{3/2}} \tag{125}$$

For very low frequencies for which $k^2K^2 < < 1$ the wave velocity c is a small fraction of $\sqrt{Y/\rho_0}$ the velocity of longitudinal stress waves in the rod; the group velocity U is very nearly twice c. For very high frequencies for which $k^2K^2 > > 1$ wave velocity approaches $\sqrt{Y/\rho_0}$ and U approximates the same value. This neglects radial effects; *i.e.*, the change of cross section accompanying bending as well as extension. When these are taken into account, it is found that as the frequency is increased c rises steadily from zero to an asymptotic value equal again to the velocity of torsional waves in the rod $(\sqrt{\mu/\rho_0})$. This was experimentally verified by S. K. Shear and A. B. Focke (among others).

The greatest interest in the flexural vibrations of rods rests on their use as vibrators for sound sources; *i.e.*, as a finite rod clamped in some specified way. Application of the appropriate boundary conditions yields the fundamental frequencies and overtones. A few samples are indicated here (P. M. Morse):

1. Rod clamped at one end and free at the other
   Fundamental frequency $\nu_f = 0.5596\ K/l^2\cdot\sqrt{Y/\rho_0}$
   First overtone frequency $= 6.267\ \nu_f$
   Second overtone frequency $= 17.55\ \nu_f$
2. Rod clamped at both ends
   Fundamental frequency $\nu_f = 3.56\ K/l^2\cdot\sqrt{Y/\rho_0}$
   First overtone frequency $= 2.76\ \nu_f$
   Second overtone frequency $= 5.40\ \nu_f$

It is clear that the overtones are not harmonics.

A torsional wave results when a twist in a solid circular rod or tube at right angles to the length is propagated lengthwise. If the twist at distance x from a chosen origin is denoted θ (*i.e.*, the angle through which a circular section is turned from its equilibrium position), the wave equation for the torsional wave becomes

$$\ddot{\theta} = \mu/\rho_0\cdot\partial^2\theta/\partial x^2 \tag{126}$$

where μ is again the shear modulus. The velocity of torsional waves is therefore

$$c = \sqrt{\mu/\rho_0} \tag{127}$$

which is less than that for longitudinal waves, though greater in general than for flexural waves. The damping effect of the friction drag of the surrounding medium on flexural and torsional vibrations is much greater than on longitudinal vibrations. As a matter of fact the decay in torsional oscillations of a cylinder in a gas is commonly used to measure the viscosity of the gas.

Wholly aside from the dissipative attenuation produced in elastic waves in solid rods, it is possible to produce wave transmission that is selective with respect to frequency (*i.e.*, filtration) by placing appropriate equally spaced and iterated loads along the rod. Such structures are usually band-pass solid acoustic filters with the lowest frequency band a pass band.

8. Plates and Diaphragms.—A membrane is assumed to have perfect flexibility, but a solid sheet or plate with its greater elastic constants as a rule provides much greater stiffness. Consequently, the velocity of propagation of elastic disturbances in a plate is usually greater than in a membrane. The theory of the vibrations of a plate is very difficult. However, the fundamental normal mode frequency of a circular plate of radius a clamped around the periphery is

$$\nu_f = 1.62\ c^3/a^3 \tag{128}$$

where

$$c = [Yl^2/12\rho_0(1-\sigma^2)]^{1/4} \tag{129}$$

and l is the thickness of the plate, ρ_0 the density of the material of which it is made, and σ is Poisson's ratio (the ratio of the lateral contraction to lengthwise elongation of a bar of the material composing the plate, with a value ranging from 0.2 to 0.4 for most solids). For example, for an iron plate for which $a = 2.2$ cm. and $l = 0.02$ cm., $\nu_f \approx 10^3$ cycles per second. The higher characteristic frequencies are not harmonic and the first few are given (P. M. Morse) as

$$\nu_{11} = 2.09\ \nu_f,\ \nu_{21} = 3.43\ \nu_f,\ \nu_{02} = 3.91\ \nu_f,\ \nu_{31} = 5.98\ \nu_f$$

Comparison with the corresponding case of the flexible membrane shows that the overtones for the plate are more widely spread out than for the membrane.

The vibration theory of plates is so difficult that it is customary to replace the circular diaphragm by an equivalent piston vibrator, all parts of which move in phase. This technique (mentioned in connection with membranes) is strictly justified only when the diaphragm vibrates below its fundamental frequency, so there are no nodal lines or circles.

Interest in the vibrations of solid structures of more or less complicated shape has been stimulated largely through developments in aeronautics and in space vehicles. When exposed to intense aerodynamic acoustic fields, solid surfaces can vibrate with enough intensity to produce serious structural fatigue (*see* FATIGUE OF METALS). The need for reducing vibration has led to the development of ingenious vibration dampers (*e.g.*, layered plates with inserted damping material as in a sandwich).

9. Radiation Efficiency.—If a vibrating system is to serve as an acoustic source it must radiate sound energy. The radiation efficiency of such a source will depend on the properties of the medium as well as those of the source itself. Consider the ideal case of a spherical shell vibrating radially. If the outside radius of the shell is a (with surface area $S = 4\pi a^2$) and it vibrates harmonically with angular frequency ω, the radiation exerts a reaction force that tends to increase the effective mass of the sphere and to increase its effective damping coefficient. As an oscillator in a vacuum the sphere will be characterized by a mechanical mass m, a damping factor R, and a stiffness f. Since it is radiating into a medium of density ρ_0, its mass is effectively increased by

$$m_r = Z_2 S/\omega \qquad (130)$$

and its damping coefficient is effectively increased by

$$R_r = Z_1 S \qquad (131)$$

In these expressions

$$Z_1 = \frac{\omega\rho_0 k a^2}{1 + k^2a^2} \qquad (132)$$

and is called the specific radiation resistance at the surface of the sphere. Also

$$Z_2 = \frac{\omega\rho_0 a}{1 + k^2a^2} \qquad (133)$$

and is called the specific radiation reactance at the surface of the sphere. They are the components of the specific acoustic radiation impedance. The quantity m_r in equation (130) is usually referred to as the radiation inertia, while R_r is called the radiation resistance.

For low frequencies or long wavelengths ($ka << 1$) the radiation inertia becomes the mass of a volume of the medium three times the volume of the vibrating sphere and can be appreciable (affecting, for example, the resonance frequency of the vibration), while the radiation resistance is relatively small. For high frequencies or short wavelengths ($ka >> 1$) the radiation inertia becomes negligibly small and the radiation resistance approaches $4\pi a^2\rho_0 c$. The specific radiation resistance becomes $\rho_0 c$ or the value for a plane wave. The result is that for given velocity amplitude the pulsating sphere is a more powerful acoustic radiator at high frequencies than at low, for dissipation of energy in the form of radiation increases with frequency, as the behaviour of the radiation resistance shows. This is one reason for the interest in ultrasonics.

The pulsating sphere usually is not a practical sound source. However, similar considerations apply to the membrane or diaphragm source in loud-speakers. The result is that for sound sources of all kinds radiation efficiency tends to be greater at high frequencies than at low. At all frequencies the average rate of radiation depends directly on the density of the surrounding medium. This makes it increasingly difficult to get radiation from any source into a medium of much smaller density. This accounts for the famous bell-in-evacuated-jar experiment, commonly explained by saying that sound is not transmitted through a "vacuum"; waves will still travel without abnormal dissipation in gases at low pressure. The principal difficulty is to get the wave into the gas in the first place.

10. Beams.—A relatively sharp beam can be produced from a light source with mirrors and lenses. Common experience indicates that most ordinary audible sources radiate in all directions. This is particularly noticeable in the human voice. However, ultrasonic (inaudible) sources can produce beam radiation. Qualitatively, consider a plane sound wave traveling in the x direction and striking the infinite rigid baffle W placed at right angles to the direction of propagation (fig. 16). Let a hole be cut in the baffle with centre at 0 on the x axis, so that some of the wave can go through. By Huygens' principle the sound field to the right of the barrier will depend on the wave front at the hole. In other words, every point in the hole will act effectively as an individual source of sound, sending out waves whose superposition at any point (*e.g.*, P) will determine the acoustic intensity there. Calculations of this superposition must consider that the effect of each part of the sound disturbance in the hole at the point P will depend on the distance of this part from P. This is not simply because intensity falls off with distance, but more particularly that wave disturbances reaching P from different parts of the hole can arrive out of phase with each other. Thus if the distances to P from two points of the hole differ by a half wavelength or any odd multiple thereof, two disturbances starting together from these points will be effectively a half wavelength apart on arrival and hence will cancel (interfere destructively with each other). If the two distances differ by a whole number of wavelengths, the disturbances will arrive in phase to give constructive interference; *i.e.*, there will be a reinforcement of the resulting disturbance at P. If the wavelength is small compared with the diameter of the hole, the magnitude of the radiated disturbance will tend to be larger in the right circular cylinder whose axis is the x axis and whose diameter is equal to that of the hole. At points like P' well outside this cylinder the intensity will be decidedly less and grow more so as P' recedes from the axis. This will not follow if the wavelength is large compared with the size of the hole. If it is large enough radiation will tend to spread more or less uniformly in all directions to the right of the baffle; *i.e.*, act effectively as a hemispherical wave. Thus with short wavelength or high-frequency radiation acoustic beams something like beams of light can be formed. Since ultrasonic waves have frequencies much smaller than those of visible light, their beams are by no means so sharp as light beams.

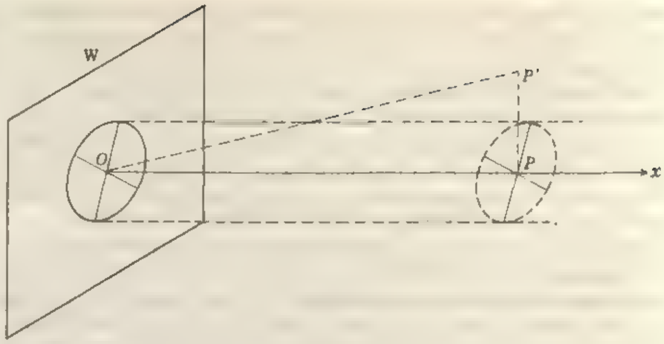


FIG. 16

Such considerations also hold if the hole is filled with a vibrating piston source. If the amplitude of the velocity at each point of the piston has the constant value ξ_0, and the wavelength λ and radius a of the piston are connected by the relation $a = m\lambda$ (where m is any number), the intensity I at any point P on the x axis distant x from the piston is given by

$$I = 2\rho_0 c\xi_0^2 \sin^2(\pi m^2\lambda/2x) \qquad (134)$$

Thus the intensity on the axis runs through a succession of maxima and minima over the region for which $m^2\lambda > x > m\lambda$. However, when $x > m^2\lambda$, I effectively falls off inversely as the square of the distance. This shows that no matter what the wavelength is, cylindrical beam radiation will not persist indefinitely far from the source; instead, there will be conical divergence after $x = m^2\lambda$. The order of magnitude of the conical or solid angle in this case is π/m^2. More exactly, for x twice the effective diameter of the piston the circle at the periphery of which the intensity falls to 1/10 that at the axis in the same plane will subtend at the centre of the source a solid angle corresponding to a plane angular spread of 2 arc tan (.45λ/a). Thus for a piston source with $a = 10$ cm. emitting radiation of frequency 50 kc. the plane angular spread of the beam is about 15°.

An interesting special case is that of a concave piston with

radial motion of its parts. Here there exists the possibility of getting focused radiation with consequent increased intensity at some points.

11. Practical Sources.—*Voice and Speech.*—The human voice operates by forcing air from the trachea to vibrate the vocal cords. This in turn sets into vibration the air in the cavities of the throat and mouth and the resulting disturbance emerges from the lips. The part the vocal cords play in singing and speaking has been established by motion pictures of their vibration (*see* VOICE).

Work on the acoustical power output of conversational speech (for example, that by H. Fletcher) showed that in the case of an average speaker the overall level directly in front of the mouth can vary from 81 db. at 15 cm. from the mouth to 65 db. at 100 cm. These decibel levels are relative to a standard reference level of 10^{-16} w/cm^2. The mouth is not precisely a point source giving rise to spherical waves, for at 100 cm. directly back of the head the intensity in the case mentioned was 62 db. This effect is due to diffraction by the head and is present to some extent in all sound sources that have appreciable size.

The power in speech sound waves varies, being much larger for vowels than for consonants. Ordinary speech involves the superposition of many waves of different frequencies. Moreover, the recognition of speech as mediated by the ear is not at all as simple as the detection of ordinary sound sources by suitable receivers. In the 1960s acoustical studies of speech were directed toward improved methods of analysis (especially of vowels and consonants). Machine synthesis and recognition of speech (*e.g.*, a typewriter that would take spoken dictation) were under development for practical application.

Sirens.—One of the simplest sources of intense sound is the siren, which in principle consists of a disk containing a number of small holes placed at regular intervals around the same periphery. If this is rotated past an orifice lined up with the holes and compressed air is driven through the orifice, the result will be a succession of puffs of air. If the number of puffs per second (governed by the speed of rotation of the disk) is sufficiently large the result will be a periodic sound wave whose frequency is equal to the number of puffs per second. By varying the rotational speed of the disk, variations in frequency can be produced. Sirens have been built to produce both sinusoidal and nonsinusoidal output. In one type of siren acoustic power output as high as 176 w. with an air pressure of 0.2 atm. has been achieved in the frequency range from 3 to 19 kc. (about 150 db. relative to the minimum audible level at 3 kc. at a distance from the siren of 25 cm. on its axis). With air pressure at 2 atm., a power output of 2,000 w. is feasible with an overall efficiency of 20%.

Jet-Edge Tones and Organ Pipes.—Another standard sound source operated by blown air is the jet-edge system coupled to an air column as in the organ pipe. In the familiar type of pipe a narrow jet of air impinges on a narrow edge of the bottom of the pipe. The resulting vibrations are coupled to those of the air in the pipe. The frequency of the wave emitted depends on the velocity of efflux of the air. The natural frequency of the edge tone uncoupled to the pipe increases about linearly with the velocity, but the coupling forces the fundamental tone of the pipe on the system for small velocities. When flow velocity reaches the value for an isolated edge tone of frequency equal to the second harmonic of the pipe, the emitted tone rises to this harmonic and the pipe is said to be overblown. The theory of edge tones was uncertain in the 1960s, though their production is clearly associated with turbulence (vortex motion) in the air flow in the vicinity of the edge.

Closely allied to the jet edge is the Hartmann oscillator in which a high-velocity jet is directed at the orifice of a small bottle-shaped resonator. Usually the output is mainly ultrasonic radiation of high intensity.

Electroacoustic Sources.—The most important sound sources in practical use employ electrical energy. These may be divided into two main classes: irreversible and reversible. In irreversible sources the energy for sound radiation is supplied by a flow of fluid (air or water) under pressure, and the electrical energy is used for control. For example, a variable magnetic field can produce vibrations in a reed or diaphragm that will control the flow of compressed air through variable orifices. Such a device can produce very intense sound waves with high efficiency (*i.e.*, high acoustic energy output relative to the electrical and mechanical input). In this respect it resembles the siren, though it is not a single-frequency source.

A reversible electroacoustic source transforms electrical energy directly into mechanical energy and thence into acoustical radiation. As its name implies, it can function in reverse to transform acoustical into electrical energy via some mechanical structure. Such sources can therefore be used equally well as receivers. What is called the reciprocal or reciprocity theorem (*see* section II.14 above) states that under specified conditions a good reversible source of sound will be a good receiver. It was shown by W. Schottky that the ratio of the reception efficiency to the emission efficiency decreases with the square of the frequency. This means that low frequencies are relatively better received than emitted.

Any system that transforms electrical to acoustical energy is called an electroacoustic transducer, whether used for emission or reception. The important characteristics of a transducer used as a source are its acoustical output, its efficiency, and its departure from linearity. This departure is the extent to which the mechanical force producing the vibrations is not strictly proportional to the electrical field or voltage applied in the input. Such nonlinearity (lack of proportionality) leads to distortion in the radiated output. Another characteristic that is often significant is the directivity; that is, the intensity of the radiation in different directions from the source.

Electroacoustic transducers develop an oscillating force operating on a mechanical vibrator; *i.e.*, a rod or diaphragm. In the electromagnetic transducer, an alternating current of the frequency desired for the acoustic output is passed through a coil that actuates an electromagnet to produce an alternating force on a diaphragm of iron or some magnetic alloy. This is the type employed in the usual telephone. In the more common electrodynamic transducer, the oscillating current passes through a coil suspended in a permanent magnetic field and undergoes the usual ponderomotive force. The coil (called the voice coil) is fastened mechanically to the radiating surface. In the well-known loud-speaker (*q.v.*) the latter is usually a cone attached to the voice coil at or near its vertex. The cone is the acoustical element, radiating directly to the surrounding medium. In earlier types the conical diaphragm was coupled to the air by means of a flaring horn, a method that continues to be used in large speakers for auditorium use. Mounting the cone on a baffle or in a cabinet can also improve its radiating efficiency. The cone is a resonant system; this can lead to distorted output at certain frequencies unless mechanical or electrical filters are employed. Compromises can be made in the size of the conical diaphragm. If it is small, the number of resonance frequencies in the audible range is small; but it takes more power to drive it for a desired output because of the small radiating surface. A larger cone is a more efficient radiator, although it has more resonance frequencies. If the latter are numerous enough they may be distributed over the usable frequency range to give good reproduction. All the general properties of acoustic radiators discussed in section III.9 above apply here. The efficiency of a dynamic loud-speaker tends to vary directly as its linear dimensions. For speakers in common use (as in radio and television sets) the efficiency runs from 2 to 5%. By the use of horns this can be increased to as much as 50%; these figures are dependent on the frequency range. In a typical small cone speaker 10 cm. in radius the efficiency is 2% only over the frequency range from 75 to 1,000 cycles per second. At 10,000 cycles per second its efficiency has dropped practically to zero.

The total acoustical power output of a dynamic loud-speaker can have maximum values ranging from 0.5 to 150 w., depending on size and the associated electrical circuit. Driving the speaker at high power risks nonlinear behaviour; *i.e.*, large amplitude diaphragm motion which no longer is proportional to the electrical voltage.

Another kind of reversible electroacoustic transducer is the electrostatic or condenser type. In this an alternating electrostatic

field of high intensity (*i.e.*, many volts per centimetre) is set up between a fixed metal plate and a thin metal diaphragm separated by a very small distance. The diaphragm is set into vibration by alternate attraction to and repulsion from the fixed plate. By proper choice of mounting and dimensions such transducers can give rather uniform output over a wide frequency range. However, their radiation efficiency is low except at very high (ultrasonic) frequencies, and they are more commonly used as receivers (microphones).

For high-frequency radiation the most important electroacoustic radiators employ piezoelectric (sometimes called electrostrictive) and magnetostrictive effects. Some asymmetric crystals like quartz and Rochelle salt generate electrical charges of differing sign on opposite faces when subjected to mechanical stress. This is the piezoelectric effect (*see* further PIEZOELECTRICITY). If such a crystal is made the dielectric in a condenser subjected to an alternating electrical field the crystal will vibrate with the frequency of the field. When a plate of quartz is cut with its faces parallel to the optic axis of the crystal (so-called X cut) the highest fundamental frequency will correspond to vibrations in the direction of its thickness. This frequency is given approximately by the empirical formula

$$\nu = \frac{286}{d} \times 10^3 \qquad (135)$$

in cycles per second, where d is the thickness in centimetres. By making the quartz oscillator thin enough, ultrasonic frequencies may be generated. Moreover, harmonics of the fundamental can provide even higher frequencies, and are often used, since there is a practical lower limit to d. The resonance in a quartz crystal is very sharp.

For ultrasonic radiation usually a number of X-cut crystals are arranged in the form of a mosaic to provide a larger radiating surface. The piezoelectric effect can also be used in a loud-speaker in the audio range. In this case it is customary to cement two or more crystal plates together and arrange the electrodes so that the whole structure will bend or twist when electrically excited. Large amplitudes can be obtained in this way, although the system is not so practicable as the electrodynamic speaker. In the early 1960s the practical upper frequency limit of ultrasonic radiation obtainable from piezoelectric generators was 10^{10} cycles per second (10,000 Mc.). By 1964 this had been raised by applying laser techniques to 6×10^{10} cycles per second.

The magnetostrictive effect is a change in length of a rod of magnetic material (iron, nickel, or ferromagnetic alloy) when the magnetic field to which it is exposed varies in magnitude (*see* MAGNETISM). If an alternating current passes through a coil surrounding such a rod (not yet magnetized), the rod will vibrate lengthwise with a frequency double that of the current; the frequency will equal that of the current if the rod is already magnetized. If the frequency of the current coincides with one of the resonant modes of the rod, particularly large vibrations result and sound waves of this frequency are radiated. As an example, a nickel rod 10 cm. long emits a fundamental frequency of 24.3 kc. Magnetostrictive sources are less efficient than piezoelectric devices largely because of the loss of energy through hysteresis in the magnetized rod. This can be reduced by using as a sender laminated nickel sheets fastened together as in the core of a transformer. Such transducers are used in producing underwater acoustic beams, usually beyond the audible range.

12. Doppler Effect.—In the discussion of sound sources it has been tacitly assumed that the source is at rest relative to the medium. However, moving sources are important in application; *e.g.*, airplanes, submarines, and projectiles. A significant effect of such motion on acoustic reception is the perceived change in pitch, in accordance with the Doppler principle. If a harmonic source of frequency ν moves with linear velocity v_s in a stationary medium, and a receiver is moving with velocity v_r along the same line, the receiver will estimate the frequency as ν' where

$$\nu' = \nu(c - v_r)/(c - v_s) \qquad (136)$$

where c is the acoustic velocity. Here v_r and v_s are reckoned as positive in the direction from the source to the receiver and negative otherwise. For the special case of a receiver at rest and the source in motion, the apparent frequency at the receiver is lower than the frequency if the source is receding, and is higher than the source frequency if the source is approaching. Knowledge of ν and measurement of ν' thus permit the determination of the velocity of the source, a matter of significance in military applications.

IV. ACOUSTIC RECEPTION

1. Introduction.—To be studied and applied, acoustic energy must be detected; *i.e.*, transformed to make it perceptible to people or sensible to cybernetic devices. This is the function of a sound receiver; any scheme for transforming the energy of elastic deformation of a medium can serve the purpose. Every reversible sound source can in principle be used as a sound receiver.

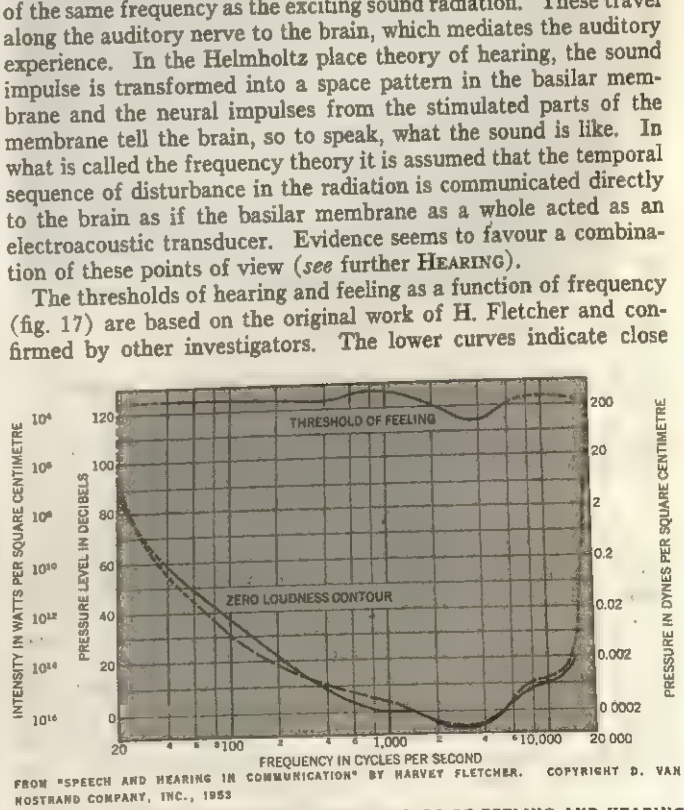
2. The Ear and Hearing.—Normal people can detect acoustic intensity as low as 10^{-16} w/cm², and can stand intensities up to about 10^{-4} w/cm² before pain ensues. The empirical values depend on the frequency, and those cited represent minima and maxima over the range of audibility: approximately 16 to 20,000 cycles per second. The presence of two ears leads to the binaural effect that serves to detect the direction of a distant acoustic source.

The detailed account in EAR, ANATOMY OF, describes the external ear (auricle). The effect of the auricle in gathering sound waves is negligible for most audible frequencies and becomes noticeable only for frequencies higher than 10 kc.

Sound waves set the eardrum (tympanic membrane) into vibration, and this motion is communicated via the bony ossicles (a kind of solid acoustic filter) to the oval window of the cochlea, a spiral cavity. The flexible basilar membrane in the cochlea can vibrate under the impact of motions of the cochlear fluid. Fine hairs in the adjacent organ of Corti in the cochlea communicate these vibrations to terminals of the auditory nerve. The system functions as a transducer, converting mechanical energy to neural energy.

The theory of hearing offered by Helmholtz was that the fibres of the basilar membrane act like resonators tuned to various frequencies over the audible range. Experimental stimulation of the nerve endings in the basilar membrane produces electrical impulses of the same frequency as the exciting sound radiation. These travel along the auditory nerve to the brain, which mediates the auditory experience. In the Helmholtz place theory of hearing, the sound impulse is transformed into a space pattern in the basilar membrane and the neural impulses from the stimulated parts of the membrane tell the brain, so to speak, what the sound is like. In what is called the frequency theory it is assumed that the temporal sequence of disturbance in the radiation is communicated directly to the brain as if the basilar membrane as a whole acted as an electroacoustic transducer. Evidence seems to favour a combination of these points of view (*see* further HEARING).

The thresholds of hearing and feeling as a function of frequency (fig. 17) are based on the original work of H. Fletcher and confirmed by other investigators. The lower curves indicate close



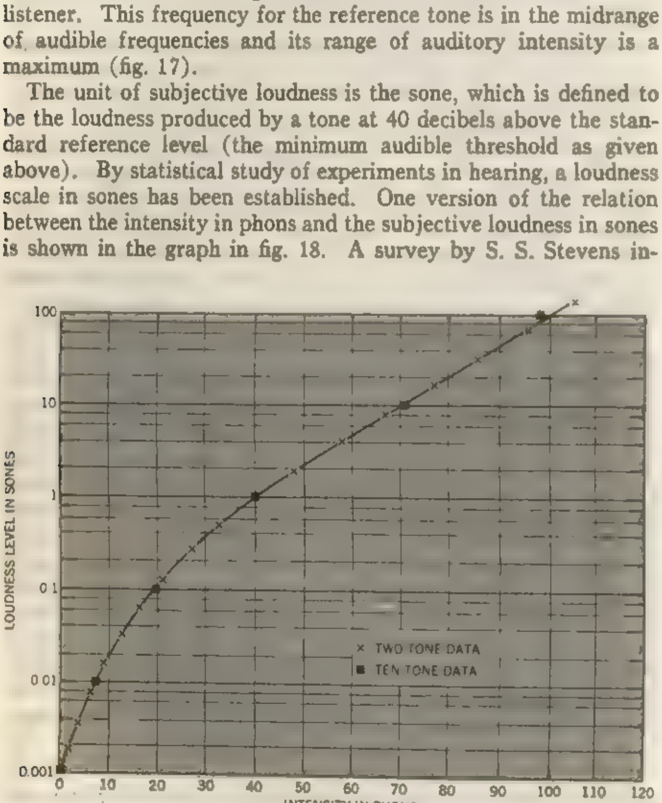
FROM "SPEECH AND HEARING IN COMMUNICATION" BY HARVEY FLETCHER. COPYRIGHT D. VAN NOSTRAND COMPANY, INC., 1953

FIG. 17.—AUDITORY AREA BETWEEN THRESHOLDS OF FEELING AND HEARING

agreement between different investigators in relating frequency to minimum audible intensity. This rises at the ends of the audible range and has a minimum at about 3,000 cycles per second. The upper curve is the corresponding plot of the threshold of feeling; intensities greater than those along this curve produce pain. The total audible range is therefore that included between the curves for loudness and feeling. People show considerable variation in thresholds, and the curves represent averages. Advance in age results in general in hearing loss, which would appear in fig. 17 as a rise in the lower curve, though its shape remains roughly the same. The tendency is for the upper frequency limit to drop.

Loudness is an experience that is related to the intensity of the sound wave, but not directly proportional to it. While intensity is an objective physical quantity, loudness is subjective and is defined in terms of the so-called average listener. In one common method of measuring loudness a reference tone of variable intensity is adjusted until it appears to the listener to be as loud as the sound stimulus under test. The number of decibels by which the reference tone has been raised above the minimum audible threshold is said to be the intensity in phons. This applies, strictly speaking, only if the reference tone has the frequency 1,000 cycles/sec and is at a distance greater than 1 m. from the head of the listener. This frequency for the reference tone is in the midrange of audible frequencies and its range of auditory intensity is a maximum (fig. 17).

The unit of subjective loudness is the sone, which is defined to be the loudness produced by a tone at 40 decibels above the standard reference level (the minimum audible threshold as given above). By statistical study of experiments in hearing, a loudness scale in sones has been established. One version of the relation between the intensity in phons and the subjective loudness in sones is shown in the graph in fig. 18. A survey by S. S. Stevens in-



FROM "SPEECH AND HEARING IN COMMUNICATION" BY HARVEY FLETCHER. COPYRIGHT D. VAN NOSTRAND COMPANY, INC., 1953

FIG. 18.—RELATION BETWEEN LOUDNESS LEVEL (INTENSITY) IN PHONS AND SUBJECTIVE LOUDNESS LEVEL IN SONES

dicated the following relation between the loudness L (in sones), and intensity P (in phons)

$$\log_{10} L = 0.03P - k \tag{137}$$

where k is a constant depending on the frequency. For a reference tone of 1,000 cycles/sec, $k = 1.2$. This means that L is proportional to the intensity raised to the power 0.3.

Hearing with two ears leads to the ability to detect the direction of sound waves. Thus a source of sound on the right of the plane bisecting the line joining the ears and perpendicular to it will be recognized as on the right. It was once believed that this arose from the difference in intensity at the two ears, but this theory proved inadequate. Rather, this binaural effect arises from a difference in phase at the two ears. This effect has proved useful in acoustic detectors for gun ranging and underwater ranging.

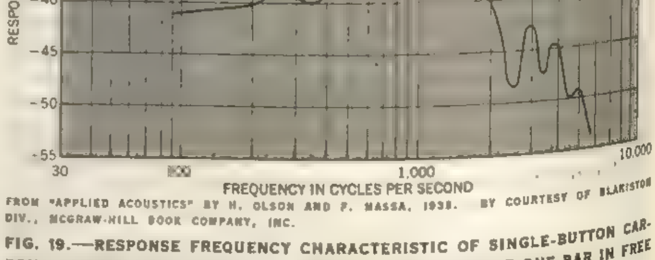
Since the eardrum is an asymmetrical vibrator (*i.e.*, its displacement for given applied pressure is not the same in the positive as in the negative direction), people can hear summation and difference tones. If two tones of frequencies ν_1 and ν_2 are presented to the ear and if $\nu_1 - \nu_2$ is small compared with either frequency, the usual phenomenon of beats ensues; *i.e.*, the intensity of the perceived sound fluctuates $\nu_1 - \nu_2$ times a second. On the other hand, if ν_1 and ν_2 are widely different a new tone of frequency $\nu_1 - \nu_2$ will be heard. So will a tone of frequency $\nu_1 + \nu_2$. The difference tone is sometimes called a beat-frequency tone. And indeed, tones of frequency $j\nu_1 \pm k\nu_2$ where j and k are arbitrary small integers, have been detected. These effects are all predicted theoretically.

3. Microphones.—Any physical system that can transform a pressure change at a point in a medium into some other form of disturbance (*e.g.*, electrical or optical) can be used as an acoustic receiver. For example, a sensitive flame, formed by burning gas issuing at high speed from a very small orifice and keeping it just below the point of flaring, was an early form of sound receiver. The flame jumps in response to acoustic waves of rather high frequency. In the Phonodeik of D. C. Miller the incident wave vibrates a fine membrane that communicates its motion to a movable mirror from which a spot of light is reflected (*see* HARMONIC ANALYSIS).

Practically all modern sound receivers are electroacoustic transducers, which when used for reception are known as microphones (sometimes termed hydrophones when used in water). Recall that reversible transducers may be used interchangeably as sources and receivers. When used as a microphone such a transducer transforms pressure changes into alternating electrical currents that can be amplified and visualized on an oscillograph (*q.v.*).

One of the important properties of any microphone is its sensitivity or response—defined as the ratio of the electrical output (usually measured in terms of open-circuit voltage) to the excess acoustic pressure on the microphone. The free-field response is that which refers to the pressure at a definite point in the acoustic field before the microphone is inserted. The term pressure response is sometimes used to refer to the case in which the pressure is that acting directly on the microphone when it is in place. The two responses differ because of acoustic diffraction by the microphone, which serves to distort the acoustic field prevailing before the insertion of the microphone. The response is commonly expressed in decibels relative to some arbitrary reference level, which is often chosen to be 1 v. per dyne/cm^2 excess pressure.

Since the microphone diaphragm has mechanical elements its vibration depends on the frequency of the incident radiation; the same holds for the response. A typical response *v.* frequency curve for the simple carbon microphone used widely in telephony is shown in fig. 19. In this figure zero decibel corresponds to 1 v. per dyne/cm^2. Hence -40 db. means 40 db. in intensity level below 1 v. per dyne/cm^2. Since power is proportional to the square of voltage, the number of decibels corresponding to the voltage ratio e_1/e_2 is $20 \log_{10} e_1/e_2$. Hence -40 db. relative to 1 v. per dyne/cm^2 is 10^{-2} v. per dyne/cm^2. From the figure the response stays



FROM "APPLIED ACOUSTICS" BY H. OLSON AND F. MASSA, 1939. BY COURTESY OF BLAKISTON DIV., MCGRAW-HILL BOOK COMPANY, INC.

FIG. 19.—RESPONSE FREQUENCY CHARACTERISTIC OF SINGLE-BUTTON CARBON MICROPHONE FOR CONSTANT ACOUSTIC PRESSURE OF ONE BAR IN FREE SPACE. RESISTANCE OF BUTTON = 100 OHMS

near this value from 100 to 1,000 cycles per second. Between 1,000 and 2,000 cycles per second it rises to between −30 and −35 db. and thereafter falls off sharply. The uneven response is due to the various modes of vibration of the microphone diaphragm.

The response discussed here has reference to the normal incidence of a plane acoustic wave. Another important characteristic of a microphone is its directivity pattern. This is the plot of the free-field response at a given frequency as a function of the angle of incidence of the radiation. If a microphone is completely nondirectional, the directivity pattern for every plane will be a circle. On the other hand the ideal pattern for the so-called ribbon microphone has the appearance of fig. 20. In the x or $-x$ direction the response is a maximum, falling to zero or near zero in the y or perpendicular direction.

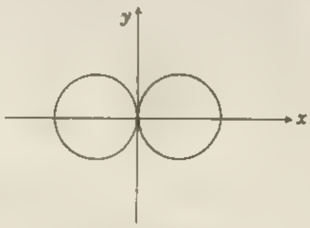


FIG. 20

Electroacoustic microphones may be classed either as pressure or pressure-gradient (velocity) devices. The carbon microphone referred to above is of the pressure type. Another is the condenser microphone, mentioned earlier as the electrostatic or condenser transducer. By appropriate mounting of the condenser diaphragm a response characteristic that is very uniform with frequency can be obtained, though the level is considerably below that of the carbon microphone; for practical use the instrument needs an associated amplifier to increase output voltage. Other examples of the pressure type are crystal microphones and the dynamic or moving-conductor microphone, analogous to the electrodynamic loud-speaker. In the pressure-gradient type of microphone, the electrical voltage produced corresponds to the difference in pressure between two points in the instrument. When this distance is a small fraction of the acoustical wavelength, the pressure gradient corresponds to the particle velocity in the medium. Hence such a device, in which a loosely stretched metallic ribbon is suspended in the air gap between two magnetic pole pieces, may be considered a velocity microphone.

The ribbon vibrates under the influence of the pressure difference produced by the sound wave on its two sides. Its vibration in the magnetic field induces an electromotive force between its two ends. Its response is best at low frequency. Its directivity pattern has been presented in fig. 20, where the x axis is normal to the plane of the ribbon.

Another interesting example of a velocity microphone is the hot-wire variety in which a fine wire is heated electrically. When a sound wave impinges on it the cooling effect alters the electrical resistance in two ways: as a steady change and as an alternating change of twice the frequency of the sound radiation. The sensitivity is materially increased by mounting the wire in the opening of a Helmholtz resonator.

Great increase in the directionality of a microphone may be obtained by mounting it at the focus of a suitably large parabolic reflector.

Waves received by a microphone can be recorded in a variety of ways. The voltage produced can be fed directly into a suitable voltmeter via an amplifier and measured directly in volts or in decibels. Special continuously recording meters have been devised for this purpose. Often it is desirable to reproduce the signal from the microphone in direct visual form and in this case a cathode-ray oscilloscope may be used. This is particularly helpful when the radiation being received is in the form of a succession of pulses. Considerable electronic circuitry is usually employed in modern acoustical reception techniques. By the 1960s the application of semiconductors had led to the construction of transducers (*e.g.*, the transistor microphone) for the detection and measurement of small stresses at acoustic frequencies. *See* also MICROPHONE; TRANSISTOR.

V. APPLICATIONS

1. Measurements.—Acoustical phenomena can be studied in terms of a number of fundamental quantities. These are velocity, frequency, sound pressure (intensity), and impedance. Common methods of measuring these quantities are discussed below.

Velocity.—Velocity in air was the earliest acoustical quantity to be measured. Marin Mersenne is usually credited with making the first estimation (1640) with a pendulum that measured the time between the uttering of a sound and the return of the echo from a distant reflector. He found the value 316 m. per sec.; there is no record of the mean temperature. This is an example of the direct method of velocity measurement, based on the time it takes a sound wave to travel a known distance. Another illustration is the cannon-flash-and-sound method popular in the 17th and 18th centuries. By this method G. A. Borelli and V. Viviani in Florence, Italy (1660), obtained the value of 361 m. per sec.; however, its accuracy is questionable because of individual differences in reaction time among observers. Moreover, velocity in air varies with temperature, chemical composition, wind, and humidity. Automatic time recording was introduced by H. V. Regnault and F. P. Leroux in the 1860s in the course of measuring the velocity in air confined in pipes with rigid walls. This velocity is always somewhat lower than that in the free atmosphere.

The direct method has also been used in liquids and solids. In the 20th century the accepted method employs pulses produced by magnetostriction or by piezoelectric oscillators. For example, a radio-frequency pulse of, say, 6 Mc. per sec. and about 10 microseconds duration excites a quartz crystal. The crystal radiates a corresponding acoustic pulse that strikes a reflector at a distance from the crystal. In the meantime another pulse is emitted by the crystal. If the distance is adjusted so that the first echo coincides with this pulse, the time for the first pulse to travel to the reflector and back equals the time between successively emitted pulses. Since the latter time (the pulse-repetition rate) can be accurately measured the velocity is readily determined as the distance traveled by the pulse divided by the time.

In a more elaborate variant of this method the emitted pulse is received at a second crystal placed at a fixed distance from the emitting crystal, and is there reconverted into an electric signal and sent back through an electronic circuit to the oscillator driving the emitting crystal. There it triggers another pulse which repeats the process, driving a succession of pulses around the loop consisting of an electronic path and an acoustic path. Each triggering pulse is counted and the total time for a large number of round trips is measured by an electronic clock. The total time for the given number of pulses to traverse the acoustic path yields the velocity. The delay in traversing the electronic path must be considered, and can be eliminated by having the same train of pulses pass through two acoustic paths of different lengths in the same medium. With a method of this kind velocity measurements in liquids and solids can be carried out to an accuracy of 1 part in 10^5; it is the group velocity that is measured by this method.

Indirect methods of measuring acoustic velocity may use standing waves. If a standing wave vibration mode is set up in a column of gas, liquid, or solid of fixed length the distance between two successive nodes is a half wavelength. Measurement of this quantity and the frequency of the vibration leads at once to the velocity through the product of the wavelength and frequency. The tube developed by A. A. E. E. Kundt (*q.v.*) applies this method to gases. Here the nodes are indicated by piles of fine powder that become arranged in layers when the tube is excited to a particular resonance frequency. In solids the nodes are usually also indicated by patterns made with lycopodium powder.

Another example of the indirect method is provided by the acoustic interferometer. Waves produced usually by a piezoelectric source are radiated continuously into a fluid medium and strike a movable reflector placed accurately parallel to the source. They are then reflected back to the source. As the distance between source and reflector is varied the current in the plate circuit of the oscillator driving the source varies periodically, the period corresponding to changes in the source-reflector distance equal to one-half wavelength. From a knowledge of the oscillator frequency the velocity is then immediately determined. This method is usually limited to liquids and gases. It has been brought to a high degree of accuracy, although probably it is not susceptible of

quite the same precision as the pulse technique. (*See* also INTERFEROMETER.)

An interesting optical technique for measuring acoustic velocity was devised by P. Debye and F. W. Sears, who discovered that a train of plane sound waves progressing through a liquid can act as a diffraction grating for a beam of monochromatic light traveling through the liquid at right angles to the direction of propagation. The distance l between the lines of the acoustic grating is the acoustic wavelength and can be determined by the usual grating formula for the diffraction angle θ of the light:

$$\sin \theta = \pm \lambda_0/2l \tag{138}$$

where λ_0 is the wavelength of the light. The method has been applied to transparent liquids and solids. It can even be applied indirectly to optically opaque solids.

Frequency.—Since harmonic waves are heavily used in practice, the measurement of frequency is of major importance. To do so requires a frequency standard with respect to which any periodic sound wave can be compared.

A common frequency standard is a tuning fork driven by a vacuum-tube circuit. The fork can be calibrated by having it drive a synchronous motor. The latter in turn rotates a stroboscopic disk illuminated by a flashing lamp controlled by a modulated radio signal, as from the station maintained by the National Bureau of Standards in Washington, D.C. The accuracy of the radio signals is estimated as better than 1 part in 10^7. They are modulated to produce acoustic frequencies of 1,440 and 4,000 cycles per second.

In practical cases it is sufficient to compare the sound wave of unknown frequency (assumed to be a pure tone; *i.e.*, of single frequency only) with a calibrated oscillator that can produce any frequency in a given range. The oscillator is set for the approximate frequency of the unknown. From the beats associated with the two waves the unknown can be accurately determined.

Still more precise is the transformation of an unknown acoustic frequency into an equivalent electrical oscillation which is then placed across one pair of plates of a cathode-ray oscilloscope; across the other pair is placed an oscillation of continuously variable frequency. The resulting Lissajous pattern (*see* CURVES, SPECIAL) on the oscilloscope can be adjusted by varying the frequency of the known electrical oscillation. When it becomes an ellipse the two frequencies are equal. The carrier frequency of the usual radio station ordinarily is not so precisely controlled that it can safely be employed to monitor the frequency of electrical oscillators used to drive electroacoustic transducers.

Intensity.—Recall that the intensity of acoustic radiation may be given either in terms of the acoustic excess pressure p_e or the particle velocity $\dot{\xi}$ in the medium. Hence the measurement of either of these quantities suffices to yield the intensity. Properly calibrated pressure or velocity microphones can therefore serve as intensity meters. However, they must be checked against absolute standards. An absolute standard for the determination of particle velocity is the Rayleigh disk, a light circular plate suspended by a fine fibre in the path of the sound radiation, usually in a tube. If the equilibrium position of the disk in the absence of sound waves is such that the normal to its plane makes an angle θ with the axis of the tube, the torque on the disk in the presence of sound waves for which the mean square particle velocity is $\overline{\dot{\xi}^2}$ becomes

$$L = \frac{4}{3}\rho_0 a^3 \overline{\dot{\xi}^2} \sin 2\theta \tag{139}$$

where a is the radius of the disk and ρ_0 is the mean density of the medium. Usually θ is chosen as 45° and the torque, measured by a torsion head on the suspension, suffices to give $\overline{\dot{\xi}^2}$ and hence the intensity in absolute measure.

Among numerous direct attempts to measure acoustic excess pressure may be mentioned the pinhole probe of C. Barus. He found that if a pinhole orifice in the shape of a cone is inserted in an air column at a pressure loop and the other end is connected to one arm of a sensitive manometer, the excess pressure can be measured directly with an optical interferometer.

E. Gerlach devised a sensitive electrical method for measuring acoustic excess pressure. He compensated the pressure on a diaphragm due to an incident sound wave by measurable electrodynamic forces so that the diaphragm remains motionless under the joint action. Since the required electrodynamic effect can be measured very accurately this provides a precise measure of sound pressure.

Sound waves also produce a radiation pressure on any object on which they impinge. This is a steady pressure like that produced by light and other electromagnetic radiation, and has a magnitude directly proportional to the intensity of the radiation. In practice the so-called Rayleigh radiation pressure is distinguished from the Langevin radiation pressure. Suppose the mean equilibrium hydrostatic pressure in a fluid at rest is p_0. If $\overline{p}$ is the time average pressure in this fluid due to the passage of a sound wave, then $\overline{p} - p_0$ is the Rayleigh radiation pressure. For an elastic liquid it is zero, whereas for an ideal gas it becomes $(\gamma + 1)/2 \cdot I/c$, where I is the intensity of the sound, c its velocity, and γ the ratio of the specific heats of the gas. By the Langevin radiation pressure is meant the difference between the average acoustic pressure at a wall or obstacle (whether absorbing or totally reflecting) and that in the medium at rest behind the obstacle. Calculation shows this to be equal to I/c in every case. This is the radiation pressure usually measured experimentally. For example in the balance method of measuring acoustic absorption in liquids, a vertical sound beam impinges on an obstacle attached to one arm of a balance and can be thereby directly "weighed." Differences in radiation pressure at different points in a sound beam yield corresponding differences in acoustic intensity, hence to the decay of intensity with distance, and ultimately to the absorption coefficient.

The absolute measurement of the amplitude of particle displacement under the action of a sound wave would also provide an accurate measure of intensity. E. N. da C. Andrade attempted to do this by observing the motion of smoke particles in an acoustic field. He predicted theoretically the relation between the acoustic motion of the fluid and the corresponding motion of the smoke particle. Acoustic amplitudes can be measured thereby to an accuracy of about 2% up to a frequency of 5 kc.

The reciprocity theorem mentioned earlier can be used in the calibration of a microphone used to measure acoustic pressure. Let A and B be any two points in a homogeneous isotropic medium of very large extent (a so-called free field) and let a transducer T be located at A while a receiving microphone R is placed at B (fig. 21). If an emf E applied to T produces an acoustic excess pressure p_e at R, and if at the same time an acoustic source of strength Q placed at B produces a current I in the transducer at A, the following equation holds (having due regard for consistent units)

FIG. 21

$$E/p_e = Q/I \tag{140}$$

Assuming that the transducer and microphone are identical instruments, the open-circuit sensitivity of this device may be found through equation (140) in current and voltage. This constitutes the calibration of the instrument as an intensity-measuring device. Since it is rare to have identical transducers, three transducers A, B, C are employed in practical calibration. A is used only as a transmitter, B only as a receiver, and C is used once as a transmitter and once as a receiver. Three measurements are performed using the pairs AB, AC, CB (the first named is always the transmitter). By successive application of the reciprocity theorem the sensitivity of each transducer can be determined in terms of electrical quantities.

Impedance.—Usually denoted Z, impedance was defined earlier as the ratio (generally complex) of the acoustic pressure to the acoustic volume current. Its analogy with electrical impedance makes it a useful quantity in acoustical calculations. A knowledge of Z in every part of an acoustical system permits that system to be replaced with an equivalent electrical circuit or transmission line. This is common practice in acoustical engineering. The impedance of any element in an acoustical circuit (*e.g.*, that of a side

tube or orifice in a horn or filter or that of the termination of a tube by acoustical absorbing material) reveals what is practically important about the acoustical behaviour of the element. The electroacoustic analogy commonly is used to relate acoustic excess pressure to emf and acoustic volume current to electric current. It is sometimes advantageous to use the inverse analogy in which excess pressure is taken as electric current and volume current as emf. Introduced by F. A. Firestone the approach has been applied with particular success to magnetostrictive elements.

Strictly speaking, the direct way of measuring Z would be to find the pressure and volume current independently at the same place. This is rarely practical, nor is it necessary. In most methods theoretical expressions are used that give the impedance at a point in the medium (that is, at some acoustical absorbing material) in terms of the pressure in phase and magnitude at other points. This is standard practice in the so-called transmission-line method to measure the impedance of material placed at the end of a long tube or that of a branch line inserted at a point in the side of the tube. Often the pressure measurements are those of maximum and minimum values in the standing wave setup in the impedance tube.

2. Recording and Reproduction.—Acoustic recording and reproduction became a major industry dating from Thomas A. Edison's invention of the cylindrical phonograph record (1877). Also, a visible record may be made with an oscilloscope and preserved photographically. The analysis of complex sound waves has been facilitated with such visible forms.

For audible reproduction, three types of recording systems are in common use: mechanical, as in the disk phonograph; optical, as in the sound film; and magnetic, as in the tape or wire recorder. In the mechanical system the record is cut into a rotating disk of wax or foil with a cutting stylus. The stylus is usually driven by electroacoustic coupling with the source being recorded. The record then consists of a spiral groove extending from the circumference of the disk to the centre. In most cases the stylus vibrates transversely to the groove, although in the 1960s some interest continued in the vertically vibrating cutter originally employed in Edison's phonograph.

Standard 78 rpm records have about four grooves per millimetre and the maximum excursion from the average line is about 0.06 mm. The long-playing $33\frac{1}{3}$ rpm microgroove record has about eight grooves to the millimetre with a maximum excursion of 0.035 mm. From a metal reproduction of the original cutting, copies can be made in large numbers on shellac or plastic disks.

For proper audible reproduction, the disk must be rotated at the same speed at which the recording was made and a needle must follow the groove. This needle is part of a pickup that usually transforms mechanical vibrations into electrical oscillations and then back again into mechanical vibrations in a loud-speaker. An important requirement in cutting the original groove and in reproducing from the disk is uniformity of frequency response to avoid distortion. The modern high-fidelity phonographic recording system provides fairly uniform frequency response over the range from 25 to 10,000 cycles per second (*see* PHONOGRAPH).

Optical systems of recording modulate a beam of light by means of the sound waves. A photographic record of the light then constitutes the sound track—the basic plan used in sound motion pictures. The sound waves may modulate the brightness of the light source (rarely used) or control the beam electromechanically by a chopper. The sound track may be either of variable density or variable area. In sound movies it is usually placed directly on the film. The sound track passes a beam of light that next falls on a photocell connected through an amplifier to a loud-speaker (*see* MOTION PICTURES).

In magnetic recording a steel wire or film strip coated with iron oxide is moved past an electromagnet (the so-called sound head) through which microphone output is passed. The current through the head magnetizes the wire or tape and provides a more or less permanent record. For reproduction the tape is passed by another sound head in which it induces a variable current that is amplified and fed into a loud-speaker. The record on the wire or tape can be readily "erased" by using a larger demagnetizing field. Much used in commercial broadcasting and in many industrial operations (*e.g.*, to record data from high-speed computers), the magnetic tape system was beginning to compete with the ordinary phonograph in the 1960s for home reproduction (*see* further TAPE RECORDING, MAGNETIC).

Consideration of direction is important in achieving realistic reproduction from a spatially distributed source (*e.g.*, a large orchestra). This is the aim of stereophonic reproduction; for example, microphones spaced at suitable intervals each activate a separate loudspeaker correspondingly spaced in the listening room. In reproduction the sound track from each microphone is played back through a separate loudspeaker.

3. Architectural.—Architectural acoustics is concerned with closed spaces such as lecture halls, theatres, church auditoriums, classrooms, and homes. About 1895 W. C. Sabine studied reverberation time (that is, the time a sound impulse in a room takes to be reduced to one-millionth of its initial intensity). He discovered the relation of this quantity to the volume of the room and the absorbing power of its surfaces; modern architectural acoustics dates from this work.

Acoustic Requirements.—A primary requirement in a living room is that it be free of noise that masks conversation. In a lecture hall speech should be heard clearly in every seat. In a concert hall devoted to instrumental music the listener wants to hear the contribution of the individual instruments. There should be no marked inequalities in reception from one seat to another. Each room, depending on its acoustic use, has an optimum reverberation time; each must be treated to control the normal modes of sound vibration associated with its geometrical shape. Except for some rooms of unusual and complicated shape, the acoustical design of most buildings can be reliably carried out before construction. Concert halls tend to give more difficulty than do rooms used for speakers because of the subtleties involved in aesthetic appreciation (*see* ACOUSTICS OF BUILDINGS).

Acoustic Distribution.—Mathematically, the distribution of sound waves in a closed space involves the solution of equation (6) subject to boundary conditions. This is unfeasible in closed form for any but very simple geometrical shapes. However, the sound source will excite the normal modes of the geometrical space in question in exactly the same way that blowing excites them in an organ pipe.

A room (with three dimensions) has more modes than a linear system like a pipe or finite string, or a two-dimensional vibrator like a membrane. The simplest case is a rectangular enclosure with sides l_1, l_2, l_3. The characteristic frequencies or normal modes are then given by

$$\nu_n = \frac{c}{2}\sqrt{(n_1/l_1)^2 + (n_2/l_2)^2 + (n_3/l_3)^2} \tag{141}$$

where c is the acoustic velocity as usual and n_1, n_2, n_3 are any integers. The modes of lowest frequency correspond to small values of n_1, n_2, n_3. There are no harmonics in the sense of the vibrating string or organ pipe. Between any two frequencies the number of modes increases with average frequency; as a square function if the frequency is high enough.

After the sound source in the room is turned on, all the normal modes will be excited, but eventually only those with frequencies equal to those of the source will persist. The process is complicated by interference between the waves reflected from the room surfaces (producing the normal modes) and the waves emitted from the source. Absorption at the room surfaces limits the buildup of acoustic energy. If reflection from the surfaces were complete, the energy in the resonant modes would increase indefinitely. To make the problem simpler geometrical acoustics is usually employed, and it is assumed that all ray directions are equally probable, so that acoustic diffusion in the room is perfect. Then it can be shown that the root-mean-square excess acoustic pressure p_{ave} in the room in dynes/cm^2 as a function of the time t from the turning on of the source is

$$p_{ave} = 1{,}300\,(W/a)^{1/2}(1 - e^{-cat/4V})^{1/2} \tag{142}$$

where W is the acoustical power output of the source in watts, V is the volume of the room in cubic feet, c is velocity in feet per

second, and a is total surface absorption in square-foot units. If the fractional amount α is absorbed per unit of time, then α is called the absorption coefficient of the surface. Thus if 30% of the incident sound energy is reflected back into the room, the absorption coefficient of the surface is 0.7. Each square foot thus acts as the equivalent of 0.7 sq.ft. of a perfect absorber. It is customary to use the sabin (after W. C. Sabine) as the unit of acoustic absorption at a surface. One sabin is the equivalent of one square foot of perfectly absorbing material. When α is known for each element dS of surface of the room, the total absorption is given by

$$a = \int \alpha \, dS \tag{143}$$

in sabins or square-foot units, where the integral is taken over all the surfaces of the room. It is customary in practical calculations to replace the integral by the sum of products

$$a = \sum \alpha_i S_i \tag{144}$$

where α_i is the average absorption coefficient over the surface S_i, and the sum is taken over all surfaces of the room.

According to equation (142) the average sound pressure in the room continues to grow for an infinite time, approaching the steady-state value

$$p = 1{,}300\,(W/a)^{1/2} \tag{145}$$

For most rooms encountered in practice $ca/4V$ is considerably greater than one sec.$^{-1}$ (*e.g.*, for a cubical room of side l, $ca/4V = 3\bar{\alpha}c/l$, where $\bar{\alpha}$ is the average absorption coefficient) and hence the steady state is effectively attained in a few seconds (unless the room is very large and absorption is very low, as in a cathedral). The pressure in equation (145) is normally expressed in terms of the number of decibels for the equivalent intensity above that corresponding to the standard (average minimum audible) 2×10^{-4} dynes/cm.^{2} This level in decibels above minimum audible then becomes

$$L = 10 \log_{10} (W/a) + 136 \tag{146}$$

The same kind of analysis that leads to equation (142) indicates that if the sound source is stopped when the effective steady state has been reached, the average pressure will decay in accordance with

$$p = 1{,}300\,(W/a)^{1/2} e^{-cat/8V} \tag{147}$$

The time for the pressure to drop to $\frac{1}{1{,}000}$ of its initial value (or 60 db. equivalent drop in intensity) is defined to be the reverberation time of the room. From equation (147) this becomes

$$T = 0.161\, V/a \tag{148}$$

if V is in cubic metres and a is in equivalent square metres of absorption. If V is in cubic feet and a in sabins the corresponding result is

$$T = 0.049\, V/a \tag{149}$$

The two preceding formulas contain Sabine's law of reverberation time, first established on experimental grounds. Actually this law is not completely accurate, particularly for rooms with abnormally high absorption (what are called dead rooms). By considering multiple reflections at the walls more carefully, C. F. Eyring showed that a more accurate expression for the reverberation time is (using feet for dimensions)

$$T = \frac{0.05\, V}{-2.30\, S \log_{10}(1 - \bar{\alpha})} \tag{150}$$

where S is the total surface area of the room and $\bar{\alpha}$ is the average absorption coefficient. If $\bar{\alpha} < < 1$, the expansion of the logarithm gives equation (149) as the first approximation. However, if $\bar{\alpha} > 0.5$, as in a dead room (*e.g.*, a broadcasting studio), equation (150) is a better representation of the facts. It has a definite economic value since it requires much less absorbing material to make a room of specified reverberation time than would be calculated by Sabine's formula.

Both reverberation formulas depend on the assumption of uniform diffusion in the room. However, if complete diffusion were realized there would be no average fluctuation of intensity with frequency at any point, and a measure of diffusion was sought in the quantity

$$\psi = \frac{\Sigma p_{\max} - \Sigma p_{\min}}{n \Delta\nu} \tag{151}$$

where $\Sigma p_{\max}$ is the sum of the pressure maxima in the frequency interval $\Delta\nu$, presumably covering the frequency range of greatest interest in the room in question; $\Sigma p_{\min}$ denotes the sum of the corresponding pressure minima. The n is the number of pressure maxima in the frequency interval.

When diffusing surfaces in the form of wall irregularities (like cylindrical bosses or straight wall sections meeting at obtuse angles close to 180°) are introduced into the room (leading to a subjective impression of better diffusion) the quantity ψ becomes smaller. It therefore seems likely that ψ may be used as a criterion of diffusion; it has been called the diffusity of the room.

Optimum Conditions.—The optimum acoustical conditions in a room depend considerably on its use. If a room is used mainly for speaking a relatively short reverberation time tends to ensure good articulation. With long reverberation successive utterances overlap unless the speech is abnormally slow.

In a standard articulation test the percentage of meaningless monosyllables produced by a speaker that are written correctly (suitably averaged over the room and preferably by several observers) is called the percentage articulation. This simple measure has proved very useful.

Empirically, if the articulation is 85%, an average of about 97% of normal speech will be understood. If the articulation falls to 75%, about 6% of ordinary speech will be misunderstood; these are relatively good conditions, though attentive listening is required. If the articulation drops to 65%, hearing conditions are bad, with nearly 10% of all speech being missed.

Experiments in auditoriums that have a volume of the order of 2×10^5 to 5×10^5 cu.ft. indicate a maximum reverberation time of about 2.5 sec. for satisfactory articulation; best listening conditions would be at about 1 sec. These figures are somewhat idealized since they assume a low general noise level (*i.e.*, about 30 db. above minimum audible) and good sound diffusion. In very large rooms (10^6 cu.ft. and more) no speaker can be heard well without a public-address system no matter how low the reverberation time. A room used primarily for music requires a longer reverberation time than does one used for speech. For an auditorium of volume between 2×10^5 and 5×10^5 cu.ft., music usually requires an optimum reverberation time of about 1.7 sec.; for organ music it is closer to 2 sec. These figures are for a frequency of about 500 cycles per second. The Sabine law indicates that, since absorption a tends to increase with frequency up to about 1,000 cycles per second, the reverberation time varies inversely. Most reverberation times are given for 500 cycles per second; for lower frequencies they will tend to be higher and for higher frequencies slightly lower (though often not appreciably so).

Experiments indicate that surface reflections that reach an auditor within the first 0.05 sec. after the wave coming directly from the source keep the sound from appearing too directional and not sufficiently mixed. They also give the impression of greater loudness. Places in a room where such reflections reach an auditor seem to be preferred; the reason for this was not fully understood in the 1960s. Of course reflected waves arriving too long after the direct stimulus produce unpleasant echoes and should be avoided.

Absorbing Materials.—Since acoustic absorption is the primary factor controlling reverberation time, numerous porous materials are in use. These include cork products, textiles, and materials manufactured with perforations or fissures. Theoretical studies show that a hard smooth surface is a poor absorber compared with one that has many indentations and air pores.

Amplification Systems.—Even a room designed carefully for good acoustics may need artificial amplification if it is very large (or if articulation must be uniformly good, as in a theatre). Amplification systems consist of one or more microphones, appropriate electronic amplifiers, and loudspeakers. In a single-channel

system all microphones, though spread over the source area (*e.g.*, the stage of a theatre), feed the same amplifier for all loudspeakers (placed so that their output seems to originate at the source). For rooms with a volume V of 5×10^4 to 5×10^5 cu.ft. the total acoustical power output in watts for satisfactory reproduction of speech is about $V \times 10^{-7}$.

In a multichannel or stereophonic system each microphone feeds its own loudspeaker. When properly installed the system can, for example, make stimuli from the left side of a stage appear to be coming from that direction.

4. Musical.—In the past, the sounding pitch associated with the various note names has varied greatly; but until relatively recent times acoustical theory as developed by physicists had relatively little influence in such matters. For example, a conference of physicists in Stuttgart, Ger., in 1834 proposed that the musical note A be standardized at 440 cycles per second, yet great variety in practice continued; not until 1938 was this standard adopted internationally. Undoubtedly the advent of large-scale radio broadcasting hastened this move.

Musical instruments are complicated examples of coupled vibrating systems (*i.e.*, systems of many degrees of freedom). They function as acoustical filters, emitting only certain frequencies or combinations thereof (*see* MUSICAL INSTRUMENTS).

Organ pipes and horns are special cases of vibrating air columns. The behaviour of a horn is largely a function of its resonance frequencies; the flare (or bell, as in a trumpet) improves the acoustic impedance match with the surrounding air to yield more radiation than does an equally long tube of uniform cross section (*see* WIND INSTRUMENTS).

The musical possibilities of electrophonic instruments are illustrated by the Theremin, essentially a beat-frequency oscillator connected to a loudspeaker through an amplifier. The capacity controlling the frequency is alterable by the motion of the hand (or a baton) relative to a metal loop. A continuous range of single frequencies is thus available. More elaborate is the Novachord or Hammond organ, a keyboard instrument in which each key controls a rotating wheel that electromagnetically induces a current pulse in a coil. The pulses are mixed, amplified, and fed to a loudspeaker. A similar type of instrument employs variation in electrostatic capacity produced by a series of rotating disks tied in with the grid circuits of a set of vacuum tubes. Another example is the photoelectric organ, in which light from a source of controllable intensity falls on the edge of a rotating disk bearing a sinusoidal pattern in black paint. After being chopped by the disk the light falls on a photocell and the resulting current after amplification is fed to a loudspeaker (*see* ELECTRONIC MUSIC).

5. Military.—Many military acoustical applications have been subject to security classification, but the basic ideas have been published. Technical problems arising in these applications stimulated fundamental research; *e.g.*, inquiries into the nature of abnormal acoustic absorption in liquids.

The principal military use of acoustics has been in the detection of objects at a distance in the atmosphere and under water.

Ranging in Air.—If an impulsive sound wave from a source at O (fig. 22) is received at microphones placed at A and B, knowledge of the acoustic velocity and the differences in time of arrival at the two microphones is sufficient to establish the direction of the source. For if C is midway between A and B on the base line of the microphones

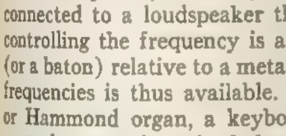


FIG. 22

$$\cos \theta = c(t_A - t_B)/l \qquad (152)$$

where l is the distance between A and B, t_A and t_B are the arrival times, and c is the acoustic velocity. It is assumed that l is very short compared with either ct_A or ct_B. Another microphone along the same base line provides a second bearing on the source. The intersection of the two bearings yields the distance of the source from the microphones; this provides in principle a method of artillery ranging. However, the projection of a shell from a gun involves the explosive discharge of the gun; the bow wave from the projectile which comes as a sharp crack; and the sound wave from the bursting of the shell. Successful ranging of the gun requires that the microphone distinguish these three; hot-wire and condenser microphones with resonance in the region of low frequencies are employed for the purpose. Since acoustic velocity is dependent on wind, temperature, and humidity concomitant meteorological measurements are required. Errors in location are reduced if more than three microphones are employed on a longer base line.

Detection of airplanes with large horns using the binaural effect was practised with some success in World War I, but abandoned in favour of electromagnetic wave methods in World War II (*see* RADAR).

Underwater Detection.—Acoustics has been extensively applied to the detection of underwater objects like submarines. Electromagnetic radiation is too rapidly damped (sea water being a good conductor of electricity) to be of much use. However, the sea is a good transmitter of sound radiation; since its characteristic impedance is much closer to that of solid sound sources than that of air, it is easier to radiate sound waves into water than into air.

One general method of underwater detection, like sound ranging in air, is based on passive reception of sound waves from an object. The waves from a ship or a submarine come mainly from the cavitation produced in the water by the propellers. Some engine vibration will be transmitted into the water and the hull will itself vibrate as a diaphragm or solid shell. The hull wave is mainly of low frequency, while the cavitation wave will have sonic and ultrasonic components. For localization in passive listening the binaural effect is employed, usually with an array of underwater microphones termed hydrophones. The latter are usually magnetostrictive or piezoelectric. They are connected to an elaborate electronic circuit that permits phase differences of the sound waves at the various hydrophones to be compensated electrically so that the whole array can be "steered" in the direction of the sound source without actually rotating.

To locate an underwater object that may be producing no sound waves, an echo method was devised similar to echo-sounding in which an ultrasonic pulse is emitted straight down from near the surface. Some of the pulse is reflected from the bottom and returns to a receiver at the surface that may be the same transducer as the sender. From the time difference between the direct and reflected pulse and knowledge of acoustic velocity in sea water, the depth may be calculated or read from a specially designed recorder. When used to detect distant objects this method is known as sonar (or asdic in Great Britain). If the emitted pulse is of high frequency, it travels in a more or less narrow beam. By rotation of the transducer this beam sweeps the surrounding medium and a reflected pulse returns from any obstacle to the sender, who can locate the object from bearing and distance. This method is the acoustic analogue of radar. Since the acoustic absorption coefficient in water increases roughly with the square of the frequency, the method fails in long-range detection at high frequency; at very low frequency the beamlike quality is lost unless the source is large. Hence a compromise must be worked out. The pulse is subject to disturbing effects if the water is nonhomogeneous because of variations in temperature, currents, changes in salinity, and so on. The sound beam then suffers refraction, diffraction, and scattering as well as absorption. Corrections must be made if the echo method is to give accurate results. In the 1960s sonar was becoming increasingly important in navigation and undersea exploration (*see* OCEAN AND OCEANOGRAPHY: *Sound in the Sea*).

6. Ultrasonics.—The velocity, attenuation, and scattering of sound waves depend on the physical properties of the medium. Recall the dependence of velocity on elastic moduli and density. Attenuation or absorption in a fluid can depend on viscosity, thermal conductivity, and on internal energy changes in the molecules of the fluid (thermal relaxation). Scattering of sound waves by a suspension depends on the size and distribution of the suspended particles. Thus acoustical measurements can shed light on the properties of the medium. Means are available for producing and detecting high-frequency waves extending far beyond the upper

limit of audibility. This branch of acoustics is known as ultrasonics and has been a principal tool in the study of matter. Just as X rays (ultrahigh-frequency electromagnetic radiation) can tell more about the structure of crystals than can visible light (of lower frequency), so ultrasonic radiation of frequency up to 10,000 Mc. can provide many more details of structure than can audible radiation. Ultrasonic radiation also has valuable applications in technology, chemistry, and medicine (*see* ULTRASONICS).

7. Noise Control.—Noise or unwanted auditory experience was a vital aspect of practical acoustics in the 1960s. Noise control had developed into a major branch of acoustical engineering.

Every advance in the transformation of heat energy into mechanical energy has involved a noise problem, and in general it increases with power production. Large-scale turbulence of exhaust gases from a jet airplane is an unusually intense source of sound. Control of the noise is more difficult than that from an automobile engine in which a relatively simple acoustic filter or muffler is acceptable.

Noise control problems involve the standard acoustical elements: source, transmission path, and receiver. Common noise sources include machinery, some so-called musical instruments, jets, and living things, including people. Transmission may be through the open air, as street noises or airplane racket; through the sea, as acoustic waves from fish that interfere with underwater signaling; through the walls of buildings, and so on. The human ear is a most vulnerable receiver; however, noisy vibration may be enough to damage heavy machinery (*see* VIBRATION CONTROL).

As an example of problems in noise control, consider factory noise, which, if produced by intense and sustained sources, can permanently damage the hearing of a worker. Various methods of control are used (as shown by the work of L. L. Beranek); in one, a partial enclosure may be built around a noisy machine and lined with acoustic absorbing material. This can provide an attenuation of from 5 to 10 db. for a worker on the outside. Full enclosure of the machine can provide perhaps 40 or 50 db. attenuation. But if the worker must function right at the machine, ear defenders (small acoustic filters inserted in the ear canal) may be required (*see* NOISE AND ITS CONTROL).

BIBLIOGRAPHY.—J. Tyndall, *On Sound* (1867); W. H. Bragg, *The World of Sound* (1920); H. Lamb, *Dynamical Theory of Sound*, 2nd ed. (1925); Lord Rayleigh, *Theory of Sound*, 2 vol. rev. (1926), 1 vol. ed. (1945); G. W. Stewart and R. B. Lindsay, *Acoustics* (1930); W. P. Mason, *Electromechanical Transducers and Wave Filters* (1942), (ed.) *Physical Acoustics: Principles and Methods*, 2 vol. (1963–65); L. L. Beranek, *Acoustic Measurements* (1949), *Acoustics* (1954), (ed.) *Noise Reduction* (1960), *Music, Acoustics and Architecture* (1962); P. M. Morse, *Vibration and Sound*, 2nd ed. (1948); V. O. Knudson and C. M. Harris, *Acoustical Designing in Architecture* (1950); E. G. Richardson, *Ultrasonic Physics* (1952), *Sound*, 5th ed. (1953), (ed.) *Technical Aspects of Sound*, 2 vol. (1953–56); H. Fletcher, *Speech and Hearing in Communication*, 2nd ed. (1953); F. V. Hunt, *Electroacoustics* (1954); L. Bergmann, *Der Ultraschall*, 6th ed. (1954); T. F. Hueter and R. H. Bolt, *Sonics* (1955); A. B. Wood, *Textbook of Sound*, 3rd rev. ed. (1955); C. A. Culver, *Musical Acoustics*, 4th ed. (1956); H. F. Olson, *Acoustical Engineering*, 3rd ed. (1957); R. B. Lindsay, *Mechanical Radiation* (1960); W. A. van Bergeijk, J. R. Pierce, and E. A. David, *Waves and the Ear* (1960); L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics*, 2nd ed. (1962); E. G. Richardson and E. Meyer (eds.), *Technical Aspects of Sound*, vol. 3 (1962); J. N. Bradley, *Shock Waves in Chemistry and Physics* (1962); V. M. Albers, *Underwater Acoustics*, 2nd ed. (1965). Important journals include the *Journal of the Acoustical Society of America*, *Revue d'Acoustique*, and *Acustica* (*International Journal of Acoustics*). *Soviet Physics: Acoustics* in English is available through the American Institute of Physics, and useful reviews of acoustics are found in *Reports on Progress in Physics*, published annually by the Physical Society of London. *See* also the *Proceedings* of congresses sponsored by the International Commission on Acoustics. (R. B. L.)

SOUND, THE (ØRESUND). One of the busiest sea-lanes in the world, the Sound separates Sjælland (Zealand), the largest island of Denmark, from Skåne, the southern tip of Sweden. Its extreme length, between Kullen and Falsterbo on the eastern shore, is 70 mi. (113 km.). The most landlocked portion between Helsingør-Hälsingborg in the north (width 3 mi. [5 km.]) and Copenhagen-Malmö in the south (width 9 mi. [14.5 km.]) is 33 mi. (53 km.) long. In the past, political control of this seaway, the shortest route from the Kattegat to the Baltic, conferred great commercial benefits. Before 1658 Denmark controlled both shores and exacted tolls from all shipping passing through. Even after Skåne was annexed by Sweden in 1658 the great Danish coastal fortress of Kronborg at Helsingør (Elsinore) still levied the Sound toll until 1857. Three large islands lie in the Sound: Amager (partly embraced by Copenhagen), Saltholm, and Ven. Between the two former lies the channel of Drogden, while the broader channel of Flinterenden separates Saltholm from Skåne. The shipping channel has a minimum depth of 23 ft. (7 m.). Ice in the almost tideless strait may impede navigation in severe winters. Copenhagen and Helsingør are the principal ports on the Danish side, Malmö and Hälsingborg on the Swedish side. (HA. T.)

SOUND BARRIER: *see* AERODYNAMICS: *Supersonic Aerodynamics*.

SOUNDING, the act of determining the depth of the ocean or any body of water. The number representing the depth of water at any position is "a sounding"; and a ship is said to be "on soundings" when she is in a portion of the ocean, as over the continental shelves, where bottom is easily reached, and "off soundings" when in deeper water.

Sounding by Pole.—The first Stone Age man who shoved off his crude dugout into a muddy estuary no doubt quickly discovered that he needed a way to determine the depth of water, and he doubtless soon made use of a marked pole which he poked vertically toward the bottom. A sounding pole is shown in an Egyptian carving of 1500 B.C. Although they are still employed in some parts of the world, long poles are dangerous and tiring devices to use from a boat in any but calm waters, and the development of the lead line probably came quite early in the history of navigation.

Sounding by Line.—The lead line essentially is a length of fishing line equipped with a sinker (the sounding lead) and marked at appropriate intervals with knots, coloured tags, or similar indicators. Its use is mentioned by Herodotus (430 B.C.) and in Acts xxvii, 28 ("So they sounded and found twenty fathoms . . ."), and it is still employed under suitable conditions. Formerly the sounding lead was provided with a hollow at the bottom, into which tallow was placed. A lead so fitted was described as "armed," and with it a sample of the bottom could be obtained at the same time as the depth was determined. Since in many areas there is more variation in the type and colour of bottom than there is in depth, the navigator had an additional bit of information to help in ascertaining his position.

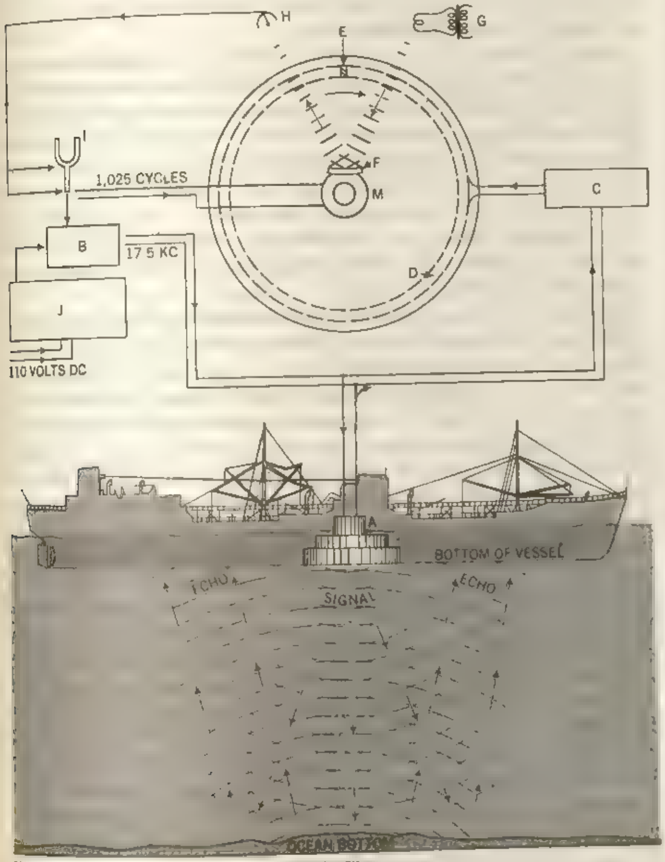
To obtain an accurate sounding the lead line must be vertical; hence the ship must be nearly motionless or else in very shallow water. In order to permit a ship to obtain a sounding while underway, various expedients were tried that measured the depth of water independently of the length of line paid out. One device was a vertically oriented propeller, which rotated as the lead sank to the bottom and thereby actuated a counter. It was difficult to haul this device back in at high speeds, and in unskilled hands it could give erroneous results; nevertheless it was fairly widely used in the 19th century. As steamers came into wide use and the speed and size of vessels increased, investigation of causes of shipwrecks showed more and more cases where soundings had been neglected, the captains preferring to take chances rather than risk delays through slowing or stopping to obtain soundings.

The problem was presented to Sir William Thomson (later Lord Kelvin; *q.v.*), professor of natural philosophy at Glasgow University, in 1878, and his solution received wide use as the Kelvin-White sounding machine. This device consisted of a small hand winch holding a reel of piano wire, which was as strong as the hemp lines previously used and much easier to retrieve. To the sounding lead was attached a glass tube, coated on the inside with silver chromate and sealed at the upper end but open at the lower. As the tube descended in the water, hydrostatic pressure compressed the air and permitted the water to rise. A chemical reaction with the seawater changed red silver chromate to white silver chloride, so that after the tube was hauled back aboard the depth could readily be determined. Thus, if the seawater had come halfway up the tube, the hydrostatic pressure had been two atm., corresponding to a depth of 33 ft.; two-thirds of the way up was three atm. or 66 ft.; and so on. As long as the lead reached

the bottom, a good record of the depth was obtained regardless of the amount of wire that had been paid out.

The Kelvin-White system, with improvements such as the use of power on the winch and stranded wires, was standard on most vessels until after 1930. The echo sounder, which measures the time required for sound waves to return to the ship after being reflected from the bottom, gradually replaced it.

Echo Sounders.—A typical echo sounder, the Dorsey Fathometer, is shown schematically in the figure.



BY COURTESY OF U.S. COAST AND GEODETIC SURVEY

SCHEMATIC DIAGRAM OF DORSEY FATHOMETER

The transceiver A is a combined transmitting and receiving unit, mounted in a hole cut in the bottom of the vessel. It receives a powerful pulse of electrical energy from a 17.5-kc. oscillator and amplifier B and then converts this pulse to an acoustic pressure wave directed downward into the water. The received echo signals in turn are converted from acoustic to electrical energy and then fed to the echo amplifier C, which amplifies the weak voltage and flashes the neon tube D. An aluminum disk rotated at 1,230 r.p.m. by a synchronous motor M hides the neon tube except for a narrow slot E. The motor also rotates a mirror F. Once each revolution the mirror reflects the beam from a light G onto a photocell H, which keys the oscillator B into vibrating for 0.7 milliseconds. Thus, each time the index slot E is vertical, a signal is generated in the oscillator-amplifier B and transmitted into the water by the transceiver A. Assuming that the water is 120 ft. deep, the signal, traveling at 4,800 ft. a second, reaches the bottom in $\frac{1}{40}$ sec., and the echo returns in another $\frac{1}{40}$ sec., or a total of $\frac{1}{20}$ sec. Since the disk rotates at 1,230 r.p.m. or 20.5 r.p.s., the index slot will have completed almost one complete revolution, and the flash of the neon tube will be seen just to the left of the zero mark. In this configuration, the echo sounder has a scale range of 120 ft. (20 fathoms), but deeper ranges can be provided by driving index disks through suitable reduction gears. Since the accuracy of the depth determination depends entirely on accuracy of timing the revolutions, a tuning fork I is provided to regulate the frequency of the power supply. A converter J produces alternating current from the ship's direct current supply.

Numerous modifications of the echo sounder have been made. The Fathometer in the figure uses ultrasonic frequencies; others use sound in the audible range, which is capable of being transmitted for greater distances in seawater. This Fathometer has a magnetostrictive transmitter, but piezoelectric, electromagnetic or percussion sound sources can be used. For survey purposes it is desirable to have a permanent record of the soundings, and echo sounders are used that draw a continuous profile of the bottom on a roll of paper.

The echo sounder is an aspect of the general underwater navigation system called sonar (*see* NAVIGATION: *Electronic Aids for Marine Use*).

Deep-Sea Sounding.—The sounding methods described above are adequate for the ordinary purposes of navigation "on soundings." Commencing about 1850, with the promotion of submarine cables, there arose an interest in the exact determination of deep oceanic depths, which could be met only by the use of much more elaborate equipment. Devices were developed that released the lead on contact with the bottom, so that the line could be hauled in more readily, while accurate meters measured the amount of wire paid out. Even so, a sounding might require several hours to complete, and although the ship was maneuvered to permit the line to descend vertically as it entered the water, there was the possibility that different currents at depth might cause the line to depart from the vertical. Echo sounding has proved as feasible in deep water as in shoal, when signals of suitable power are sent out. (J. L.)

SOURIS, an important tributary river flowing into the right bank of the Assiniboine River, about 20 mi. SE of Brandon, Man., Can. The river is 600 mi. in length and drains an area of about 22,000 sq.mi. in the provinces of Manitoba and Saskatchewan and the state of North Dakota. Rising in the southern part of the Regina plains near Weyburn, Sask., it receives drainage from Moose Mountain to the east, and the Coteau du Missouri on the west. Flowing southeasterly, it crosses the international boundary into North Dakota, making a long loop before recrossing into Manitoba and turning northeasterly to join the Assiniboine River. A considerable part of the drainage basin is floored by silt and clay deposited on the bottom of an ancient glacial lake known to geologists as Lake Souris. The river follows the valley cut by the outfall from the old lake for some distance near Souris, Man., but suddenly turns northeastward, cutting a deep, narrow valley through the Tiger Hills to join the Assiniboine. The downstream extension of the old valley is drained by the Pembina River. The grassy plains drained by the Souris River were once the grazing grounds of countless buffalo herds. They form an important part of the hard red spring-wheat belt of central North America. Near Estevan and Bienfait, Sask., there is extensive strip mining of lignite coal. Productive oil wells are near Midale, Sask., Virden, Man., and other points. There are coal mines also near Minot, N.D., the largest city on the Souris River. (D. F. Pu.)

SOUSA, JOHN PHILIP (1854–1932), U.S. bandmaster and composer of military marches, called the "March King." Son of a Portuguese father and a German mother, he was born on Nov. 6, 1854, at Washington, D.C., where he studied the violin and harmony with J. Esputa and G. F. Benkert. In 1867 he began to follow the career of his father as trombonist, but later took engagements as an orchestral violinist, playing at Philadelphia under Offenbach in 1877, and as a conductor. He brought the U.S. Marine Band, which he led from 1880 to 1892, to a higher standard of performance, and in September 1892 gave the first concert with his own band with which he toured in the U.S. and Europe (1900–05), and which he eventually took on a world tour (1910–11). Sousa wrote about 140 military marches, remarkable for their rhythmic and instrumental effects, the most celebrated of which is "The Stars and Stripes Forever" (1897). Others are "Semper Fidelis" (1888), "The Washington Post" (1889), and "The Liberty Bell" (1893). Between 1879 and 1915 he wrote 11 operettas, of which the most successful was *El Capitan* (Boston, 1896). During World War I he was adviser to the U.S. government on military music, but later continued his tours. He compiled for the U.S.

Naval Department *National, Patriotic and Typical Airs of All Lands* (1890); and wrote three novels, among them *The Fifth String* (1902), and treatises on the trumpet and violin. He died at Reading, Pa., on March 6, 1932.

See J. P. Sousa, *Marching Along* (1928).

SOUSAPHONE: *see* TUBA.

SOUTAR, WILLIAM (1898–1943), Scottish poet, was second in importance only to Hugh Macdiarmid among the writers of the Scottish Renaissance movement. He was born in Perth on April 28, 1898, the son of a master-joiner, and educated at Perth Academy and Edinburgh University. During World War I he served for two years in the navy, where he contracted the osteoarthritis from which he suffered thereafter. From October 1923 he was a semi-invalid, and after the failure of an operation in May 1930 he was bedridden. He was saved from apathy and despair by his delight in the variety of nature, his love of mankind, and his devotion to the craft of letters. His "bairn-rhymes" in Scots, *Seeds in the Wind* (1933), are beast-fables which express a mature insight into the life of things in terms of the "innocent eye" vision of childhood. In *Poems in Scots* (1935) he developed the ballad style toward the objective expression of individual lyricism. During his last ten years his principal output in Scots consisted of "whigmaleeries," humorous poems full of comic exaggeration, interweaving the fantastic and the familiar. He was fond of miniatures, publishing *Riddles in Scots* (1937), while as a poet in English he was at his best in the pointed epigrams of *Brief Words* (1935) and the short nature-lyrics of *The Expectant Silence* (1944). He died in Perth on Oct. 15, 1943, leaving a mass of unpublished verse and prose.

BIBLIOGRAPHY.—*Collected Poems*, ed. by H. Macdiarmid (1948); *Diaries of a Dying Man*, ed. by A. Scott (1954); A. Scott, *Still Life: William Soutar, 1898–1943* (1958); *Poems in Scots and English*, ed. by W. R. Aitken (1961). (A. M. S.)

SOUTH, THE, an extensive region covering approximately the southeastern quarter of the United States. There is no clear line of demarcation between it and the North but the Mason and Dixon Line (*q.v.*) and the Ohio River are popularly regarded as dividing the two regions. The South is generally considered to include the states of Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Tennessee, and Arkansas, as well as parts of the states of Texas, Oklahoma, Missouri, Kentucky, West Virginia, Maryland, and Delaware. Since the end of the colonial period the South has formed a major section of the United States, substantially distinct from other sections in agriculture, economic and social structure, and political attitudes.

Geography.—In its physical geography, the South does not possess any natural unity. In fact, it includes a number of quite diverse geographical districts. Along the coast, from Maryland to the Rio Grande, there is a low coastal plain, marked by a series of alluvial rivers flowing eastward to the Atlantic or southward to the Gulf of Mexico. On the Atlantic shore, this coastal plain rises gradually to the "fall line" where the streams cease to be navigable. There begins an area known as the piedmont, where the land is rolling or hilly, erosion has taken place, and reddish-yellow clay soils predominate in contrast to the sandy or alluvial soils of the coast. At its western edge, the piedmont rises into the Appalachian mountain system, the southern part of which extends from Maryland to northern Georgia and Alabama. The easternmost part of this system consists of the Blue Ridge Mountains, behind which lies the main Appalachian ridge. Beyond the Appalachian ridge is the high and mountainous country of the Cumberland Plateau. This mountainous terrain dominates the western parts of Maryland, Virginia, North Carolina, the northern part of Georgia, the eastern parts of Tennessee and Kentucky, and most of West Virginia.

Just as the Appalachian mountain system is the key to the geography of the eastern or Atlantic coastal part of the South, so the Mississippi River Valley is the key to the western or Gulf coastal part. Although portions of Alabama, Mississippi, Louisiana, and Texas drain directly into the Gulf of Mexico, other parts of these states and all of Kentucky, Tennessee, Missouri, and Arkansas drain into the Mississippi. On its west bank, such major rivers as the Missouri, the White, the Arkansas, the Ouachita, and the Red flow eastward through fertile alluvial valleys directly into the Mississippi. On the eastern side of the valley, the principal southern rivers are the Cumberland and the Tennessee, but instead of flowing westward from the Appalachian ridge into the Mississippi, these streams describe great semicircular arcs, first swinging southward and then northward across Tennessee and Kentucky into the Ohio River and thence into the Mississippi. Their courses run through rich limestone valleys that give great and distinctive beauty to the Nashville Basin in Tennessee and the bluegrass region of Kentucky.

Population.—This geographical variety has produced a diversity rather than a uniformity of Southern folk—a diversity ranging from the isolated mountaineers of Appalachian valleys to the semi-amphibious bayou dwellers of southern Louisiana, and including the "red necks" of the southern hill country, the poor whites of the piney woods, the cattle-herding plainsmen of Texas, the peasantlike Negroes who form more than 80% of the population of some counties along the Mississippi, and the cotton-planting farmers, both great and small, who, more than any others, have constituted a "Southern" type.

In terms of the origins of its population, the South presents a pattern unlike that of the rest of the country, but which, again, shows lack of uniformity within the region itself. The contrast with the non-South lies in the fact that the South received only a small proportion of the immigrants who came to the United States after 1800, and hence it has a relatively small number of people of southern, central, and eastern European antecedents. Most of the population is descended from the settlers who arrived during the colonial period. In the 17th century, English settlers established themselves in the coastal areas of Virginia, Maryland, and the Carolinas. In the 18th century, a heavy influx of Scots-Irish (Presbyterians of lowland Scotland who had migrated first to northern Ireland and then to America) and German settlers moved southward from Pennsylvania into the southern piedmont. During the same century, the French settled in Louisiana, where their language, religion, and culture are still in evidence, and at Mobile, in southern Alabama. The Spanish at the same time occupied Florida and Texas but did not settle in sufficient force to leave an important Hispanic tradition. The most extensive influx of the 18th century was the involuntary immigration of Negroes brought as chattel slaves directly or indirectly from the west coast of Africa. The importation of Negroes had begun as early as 1619, but did not attain large numbers until after 1700. By the year 1790, however, Negroes formed more than one-third of the population in the states of Maryland, Virginia, South Carolina, and Georgia, and in 1960 they continued to form more than 20% of the population in Virginia, North Carolina, South Carolina, Georgia, Alabama, Mississippi, Louisiana, and Arkansas. Florida, Maryland, Tennessee, Delaware, and Texas had Negro populations of between 10 and 20%. Illinois, with 10.3%, was the only state outside the South where Negroes made up as much as 10% of the population.

Climate and Soil.—In terms of climate and soil, the South shows features of uniformity, somewhat more than it does in geography and in the ethnic origins of its population. Almost the entire region lies in the same latitudes as North Africa, and the climate is, in general, mild to hot. In the summers, temperatures rise, day after day, above 90° F (32° C) in the shade. But the maximum heat is less significant than the long growing season—a period of 200 to 290 frost-free days during which slow-growing crops such as cotton, rice, and sugarcane may be brought to maturity. There is also abundant rainfall, averaging 40 to 50 in. (1,000 to 1,300 mm.) per annum in the upper South and 50 to 60 in. in the Gulf Coast region.

Because of the warmth of the region, it was not glaciated during the Ice Age, and this lack of glaciation gives a distinctive quality to the soil. Except in the limestone basins of Kentucky and Tennessee, the black belt across parts of central Georgia, Alabama, and Mississippi, and the river valleys with their rich alluvial deposit, the soil is predominantly of coarse texture and of mediocre fertility, with high proportions of clay in the piedmont region,

with its "red hills," and of sand in the coastal districts. Soil exhaustion has occurred widely in these areas, both because of their natural lack of fertility and because of the soil-depleting effects of the cotton and tobacco cultures. As a result, the South regularly consumes about half of the nation's output of commercial fertilizer. Also, erosion has occurred very widely in the piedmont area because of the torrential rains, coupled with the exposure of the soil in plowed hills and furrows which are used in the row cropping of cotton and tobacco. This erosion had been so severe that by 1930 the South had suffered a total loss of fertility in 22,500,000 ac. of its land—an area larger than the entire state of South Carolina.

Crops.—The one feature above all others that has made the South historically a distinctive and somewhat homogeneous region is the fact that, for more than three centuries, it relied almost wholly upon an agricultural economy, centred upon staple crops that were cultivated for sale on the world market. Even as late as 1948, the South still derived twice as much of its income from agriculture (16%) as did the non-South (8%). The first of these crops was tobacco, first cultivated in Virginia about 1612, just five years after the founding of the colony. Tobacco culture spread into Maryland and northern North Carolina, and for nearly 200 years it provided the chief export product of North America. Meanwhile, rice about 1698 and indigo about 1744 took hold in the Carolinas and quickly grew to dominate the economy of the coastal part of that colony. The culture of sugarcane was successfully launched in French Louisiana in 1794, and it also expanded to become the chief economic activity of the region. But none of these crops rivaled cotton, which created Southern wealth, shaped Southern institutions, and, as "king cotton," grew to symbolize the whole South at the height of its power.

The long-fibred, or black-seed, cotton was profitably cultivated on the coastal islands of Georgia and Carolina before the American Revolution, but the only strain that would grow in most areas on the mainland was the short-fibred, green-seed plant. The latter seemed commercially worthless because the lint clung so firmly to the seed that the separation of the fibres was impracticable. This situation was suddenly changed in 1793, however, when Eli Whitney, a Connecticut Yankee teaching school in Georgia, invented a device, known as a cotton gin, to separate the fibres. Whitney's gin made it possible for the South to supply cotton for a world textile market in which the new technology of mechanical spinning and weaving had created an insatiable demand for raw material. The production of Southern cotton doubled every decade between 1800 and 1861, until it reached 4,400,000 bales in 1861. As this growth took place, the staple-crop economy spread into the Gulf Coast region and within a single generation the South expanded from the Oconee River in Georgia to the banks of the Brazos in Texas.

The Plantation System.—The staple-crop economy of tobacco, rice, indigo (declining sharply after the Revolution), sugarcane, and cotton shaped the character of the South. Since production on a big scale was profitable, large agricultural units known as plantations grew up; although vast numbers of small, owner-cultivated farms remained, the plantation became a dominant factor. Since the crops were used for commercial purposes rather than for subsistence (such as prevailed in the backwoods), it paid the proprietor to employ labour and thus increase production. The slow-maturing crops needed attention most of the year and thus required a permanent labour force. Because the agricultural tasks were simple enough to be performed by unskilled labour working in gangs, owners resorted to the use of Negro slaves. Thus the one-crop system, the plantation unit, and the presence of a biracial population, with the Negroes in a subordinate position, were all developed in the South early in the 18th century, and their grip was tightened by the overpowering rise of the cotton economy. By 1850 there were 3,204,000 slaves in the South and it was estimated that 1,815,000 of these were connected with the cultivation of cotton.

Ultimately, the complex of staple crops, plantations, and Negro cultivators (whether slave or free) gave to the South qualities that made it a distinctive region and set it in an adverse sectional relationship to other parts of the United States. But until the 19th century, the sectional alignment was far from clear-cut. Historians now reject a once-prevalent interpretation which explained the differences between North and South in terms of the dissimilarities between Puritans in colonial New England and cavaliers in colonial Virginia. The British colonists who settled the two regions were not so sharply differentiated as the tradition has claimed. Moreover, in the colonial South, the plantation economy and slavery were rather narrowly confined to limited areas around Chesapeake Bay and in the coastal district of the Carolinas and Georgia. The democratic, Presbyterian, Scots-Irish who populated the vast back country of the piedmont did not share in this regime. They practised a subsistence economy, cultivated grain crops, had no use for slaves, and were frequently at odds with the aristocratic, Anglican planters. These subsistence farmers of the interior seemed to have more in common with the similarly circumstanced settlers in the interior of Massachusetts and the backwoods of Pennsylvania than either group had with the grandees of the coastal districts—whether merchant princes in Boston, Mass., or planter-aristocrats in Charleston, S.C., and along the James River. Correspondingly, the latter groups showed a greater affinity with one another than with their neighbours to the west. The antagonism between the coastal and the interior districts was chronic, and repeatedly manifested itself in incidents of friction such as Shays's Rebellion in Massachusetts and the uprising of the Regulators in the Carolinas. At the time of the American Revolution this east-west division seemed likely to remain the primary line of sectional cleavage.

North-South Differences.—At that time, the divisions between North and South still remained somewhat tenuous. The leaders of both sections agreed in regarding slavery as an evil. At the same time, both sections practised it, and it was sanctioned by law in every state. The actual concentration of slaves was overwhelmingly in the South—94% of all slaves in 1790 were held south of Pennsylvania. But many districts in the South had almost no slaves, and slavery was widely regarded as a localized and declining institution.

Between 1775 and 1830, however, both North and South experienced transformations that heightened their dissimilarities and generated deep antagonisms between them. The Northern transformation came as the region slowly began to industrialize. This process committed it politically to the idea of economic self-sufficiency within a national market protected by tariff walls, where agricultural areas and manufacturing areas would complement one another by producing to meet each other's needs and serving as markets for each other's products. Industrialization generated rapid economic change in the North, and this gave sanction to the idea of progress and to the idea of a free society in which the individual can alter his occupation and status to meet changing conditions. Moreover, the use of machines freed the worker from many brute tasks and, by increasing productivity, raised the standard of living. Both of these developments tended to enhance the dignity of the worker.

Meanwhile, in the South the advent of the cotton economy made it possible to convert the whole region to staple-crop agriculture, as had never been possible with tobacco, rice, and sugar. Cotton would grow almost everywhere in the South and it was more flexible than other staple crops in lending itself to production on any scale, large or small. The cotton economy moved quickly into the up-country, thus obliterating the former divisions between the coastal and the piedmont areas. By carrying the plantation and slavery with it, it increased the value of slaves and reinvigorated the declining slave system.

Economically, the South looked to the British textile industry for its market, and it therefore opposed the growing economic nationalism of the North and West. Socially, the plantation system discounted the commercial values of thrift, prudence, enterprise, and progress; it exalted qualities of magnanimity, command, manly prowess, and physical courage; and it adopted a cult of chivalry, with a code duello, to enshrine these virtues. The Southern social philosophy, holding to a country-gentry ideal, presented a sharp contrast with that of the North, for it stressed the con-

servative values of status in a fixed social order rather than of freedom, of stability rather than of progress, and of a way of life rather than accumulation of wealth. Such an emphasis was almost inevitable in a region where one-third of the population occupied a position of fixed legal subordination.

The Slavery Controversy.—The growing divergence between North and South, as well as the dissimilarities in their values, found a focus in the question of slavery. In the North, the antislavery sentiment of the Revolutionary era continued to grow. Between 1777 and 1804, every state north of Maryland either abolished slavery outright or provided for gradual abolition. The early antislavery movement was moderate in tone and sought to bring about emancipation gradually by persuading slaveowners to free their slaves voluntarily. But about 1830 it entered a new and more militant phase in which immediate abolition was demanded and slaveowners were denounced in most abusive terms. The militant abolitionists were always a minority, and historians disagree as to the extent to which the North became "abolitionized," but it is clear that antislavery sentiment became very widespread. The immense popularity of Harriet Beecher Stowe's *Uncle Tom's Cabin* (1852) and the election to the presidency of Abraham Lincoln, who had said that "if slavery is not wrong, then nothing is wrong," serve as indexes to this feeling.

While the North passed from mild opposition to strong condemnation of slavery, the South passed from mild opposition to an unqualified defense of the "peculiar institution." Various Southern spokesmen defended slavery on the ground that it was sanctioned by the Bible, that the economic exploitation of wage earners was worse than the exploitation of slaves, that the Negro was biologically inferior, and that in a well-ordered society it was better for one class to be set apart for the menial duties. Belief that slavery was a positive good became a test of Southern orthodoxy.

With all the cultural dissimilarities and economic conflicts of interest between North and South converted into a dispute on the slavery question, sectional antagonisms were chronically acute between 1830 and 1860. The South perceived that it was not growing as rapidly as the industrial North in wealth, power, or population. A minority psychology became fixed upon the whole region, and a number of its political leaders, of whom John C. Calhoun was foremost, developed elaborate constitutional arguments, proclaiming the ultimate sovereignty of the states and their right to secede from the union. The full explosive effect of the slavery question was to some extent kept out of politics by the fact that no one contested the right of individual states to permit or prohibit slavery within their own borders. But whenever new lands were acquired, either by the Louisiana Purchase (1803) or by the War with Mexico (1846–48), bitter contests took place over the status of slavery in these areas prior to statehood. Thus there was a series of crises: in 1820 over slavery in the Louisiana territory; in 1832–33 over South Carolina's effort to nullify the tariff act; and in 1848–50 over slavery in the territory acquired from Mexico. In each of these crises a clash was averted by the adoption of measures of compromise, which have been compared to treaties between two nations. But despite these efforts to save the Union, relations between the sections became increasingly strained and were brought almost to the breaking point by a series of violent episodes in the 1850s—the repeal of the Missouri Compromise by the Kansas-Nebraska Act in 1854, open combat between antislavery and proslavery factions in "bleeding Kansas," the decision of the Supreme Court in 1857 in the Dred Scott case that Congress could not prohibit slavery in the territories, and John Brown's raid at Harpers Ferry, in 1859. Finally, in 1860, Abraham Lincoln was elected president on a platform pledging opposition to slavery in the territories, and by a purely sectional majority. The states of the lower South took his election as a signal to act. Seven of them (South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Texas) adopted ordinances of secession, and proceeded (1861) to form a new union, the Confederate States of America (*q.v.*).

Lincoln denied the legality of their act and sought to maintain the authority of the Union. As a result, fighting began at Ft. Sumter, in Charleston Harbour (April 1861). Soon thereafter, four other states (Virginia, North Carolina, Tennessee, Arkansas) also seceded and joined the Confederacy. For four years, the United States and the Confederate states fought each other fiercely in the American Civil War (*q.v.*).

Southerners in the heat of secession and war believed that the "Southland" or "Dixie," as they now called it, had achieved a full and separate nationality, not only in the political but also in the cultural sense. Historians since then, with an eye to features of cultural distinctiveness, have also recognized an important degree of "Southern nationalism." It is a serious question, however, whether Southern separatism resulted primarily from deep cultural differentiation, or whether it was the reaction of a group who were still an integral part of the American people, but who had been psychologically (rather than culturally) alienated from their fellow Americans by a long period of minority status, conflicting interests, and defensive guardianship of the institution of slavery. The readiness with which the South returned to the Union after the war suggests the latter.

The Reconstruction Era.—At any event, the South lost the war, and this loss settled two questions. First, it killed the idea of state sovereignty and secession. Second, it ended the institution of slavery. But there were other problems which the war did not solve, and which continued to make for Southern distinctiveness for another century. For one thing, the South continued to rely on a one-crop economy, and it continued to cultivate this crop with the labour of Negroes, who discovered that as long as a man went on making his living by hoeing cotton, the transition from slavery to freedom did not altogether revolutionize his life. For another, the South continued to insist upon the inferiority and the subordination of the Negro. Slavery had been partly an economic system for the ownership of labour and partly a social system of racial control, and though emancipation ended the one, it did not end the other.

During the so-called Reconstruction period (1865–77), the victorious Republican Party made more or less earnest efforts to assure Negro equality, but those who wanted the party to help the Negroes were probably never as numerous as those who wanted the Negroes to help the party. After the latter perceived that the party might fare better without Negro help, the crusade for equality was generally abandoned, and the South was left to work out its own arrangements. As a consequence, the institution of slavery was replaced by three institutions: the economic system of sharecropping, the political system of one-party politics, and the social system of segregation, supported both by law and by custom.

Sharecropping.—The system of sharecropping or tenancy resulted from need for a new link between the soil and the cultivator. The former slaves lacked funds to buy land; the landowners lacked funds to pay wages. Hence, an agreement was made that the landlord would furnish the land, the freedman the labour, and each would receive a share of the proceeds for his contribution. The cultivator, however, necessarily received the necessities of life on credit during the annual periods between harvests, and the cotton economy was so unprofitable that the tenant's share seldom left a decent surplus after his debts had been settled. Thus Southern tenant farmers were chronically in debt; their indebtedness limited freedom to move or change occupation; and they had a lower standard of living than any other large group in the American population. Moreover, the chronic overproduction of cotton, with its attendant low prices, forced more and more small landowning farmers into tenancy. Between 1880 and 1930 the proportion of Southern farms operated by the tenants increased from 36 to 55%. During the last 30 years of this span, white tenants increased by 400,000 and Negro tenants by 147,000, so that what had begun as an arrangement for the freedmen became more and more a biracial institution.

One-Party System.—The one-party system was an arrangement to neutralize the political power of the Negroes, who had legally been enfranchised by the 15th Amendment (1870). After a brief period of Southern support for Populism in the 1890s, Southerners conducted their political activity solely within the Democratic Party. Contending that the Democratic Party was a private and not a public instrumentality, they excluded Negroes

from participation in the primary elections in which party nominees for office were chosen. Moreover, between 1890 and 1910, most Southern states placed in their constitutions literacy and property qualifications from which large groups of whites, at least in some states, were exempted by clauses allowing them to qualify if their forebears had voted before 1860 ("grandfather clauses"). By these devices, by inequitable enforcement procedures, and by a relatively costly poll tax which most Negroes found difficult to pay, Negroes were effectively excluded from political life. At the same time, the white "solid South" remained Democratic: 11 states from Virginia to Texas elected only four Republican governors between 1876 and 1960. Because of this concentration in one party, Southerners gained a disproportionate share of committee chairmanships and tended to dominate the U.S. Congress during Democratic administrations. After the 1964 presidential election, when the Republican nominee carried five Southern states, some political observers predicted the eventual demise of one-party rule in the South.

Segregation.—The third of the institutions succeeding to slavery was legalized segregation. The practice of segregation, it may be noted, was far older than the segregation laws and it extended far beyond the South. Throughout the United States, Negroes were generally denied access to places of public resort. They were excluded from trade unions, forced to live in segregated residential areas, prohibited from marrying whites, and prevented from mingling freely with the rest of the population. In most areas, this was done informally or by extralegal devices such as restrictive housing covenants, but in the South it was done by law. Among Southern segregation laws, the statutes requiring segregated education were important. Laws requiring Negroes to use segregated railroad facilities were upheld as constitutional by the Supreme Court in the case of *Plessy* v. *Ferguson* (1896).

Economic Trends.—Until 1932 the South remained an impoverished and undiversified region. The growth of a textile industry in the Carolinas, and the movement to develop a "new South" after the Civil War had not seriously qualified the region's commitment to cotton, to agriculture, and to a rural way of life. The Negroes remained a kind of peasantry and the income of the South stood at only $372 per capita in 1929, while income outside the South was $797 per capita. But after 1933 the South experienced an economic revolution, resulting partly from the New Deal and the collapse of cotton tenancy and partly from rapid economic diversification and industrialization.

During the depression that began in 1929, cotton prices sank to a point that bankrupted the cotton economy. The plight of the sharecropper was desperate and cultivation continued only because of the lack of alternatives. When this situation was at its nadir, the government stepped in with two devices that exercised an effect more revolutionary than was intended. Agricultural benefit payments for taking acreage out of production led landlords to curtail sharecropping operations. Partly because of this and partly because of the nationwide shift of population to the cities, the number of tenant farms in the South declined from 1,791,000 in 1930 to 366,000 in 1959, while the number of Southern nonwhite tenants declined from 600,900 to 123,000. The second New Deal device was the system of unemployment relief, which drew Negroes temporarily to Southern towns where relief payments were issued. During and after World War II, however, there was heavy Negro migration to Northern industrial centres. By 1960 the proportion of Negroes who lived in the South was only 60%, whereas in 1920 it had been 85%. By 1960 the average Negro was no longer a rural-dwelling, sharecropping, cotton cultivator but was a city-dwelling, wage-earning industrial worker.

With World War II and its aftermath, industrial developments and diversification revolutionized the economy within the South. Military installations were heavily concentrated in the South during the war, and after the war several new industries enjoyed spectacular growth. In 1948 the value of cotton produced was $2,189,000,000, but this amount was surpassed both by the value of livestock ($3,072,000,000) and of textiles ($2,832,000,000), and was rivaled by the value of chemical products ($1,082,000,000), lumber products ($889,000,000), petroleum products ($642,000,000), and paper products ($539,000,000), all of which had increased by more than 247% in the preceding eight years. Cotton was no longer king.

As a result of this economic change, the percentage of workers in the South employed in agriculture declined sharply. Also, the proportion of Southerners living in cities rose from 15% in 1900 to more than 55% in 1960 (compared with a national increase from 39 to 70%). But perhaps the most drastic change of all was reflected in the greater increase of per capita income in the South. While such income was rising from about $700 to $2,600 for the U.S. as a whole between 1929 and 1964, it rose from about $350 to more than $1,900 for the Southern states.

Desegregation.—At the middle of the 20th century, the steady growth of industrialism, urbanization, and wealth were constantly diminishing those factors of poverty, agrarianism, and ruralism that had operated to preserve the South's historical traditions and its distinctive identity. But although the distinctive elements in the Southern region had been much reduced, the Southern adherence to a biracial system remained strong enough to assert itself with powerful force when challenged in 1954.

From about 1905 (when a protest was framed by Negro leaders at Niagara, N.Y.), there had been a rising demand in liberal circles for the elimination of discriminatory practices toward Negroes, and after about 1927 (when the white primary was ruled invalid), the Supreme Court began to interpret the guarantees of the 14th Amendment as prohibiting various forms of discrimination. For many years it did not question the legality of segregation in cases where facilities were "separate and equal," but in 1954 in the case of *Brown* v. *Board of Education*, it reversed the *Plessy* v. *Ferguson* decision and declared that segregation in publicly supported schools was unconstitutional.

Sectional antagonisms flared again as the Southern states denounced "judicial usurpation," and invoked the principle of local self-government, while liberals criticized the South for racial prejudice and for the defiance of federal authority. Eleven states (Virginia, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Louisiana, Texas, Arkansas, and Tennessee) adopted laws to avoid, directly or indirectly, the integration of the schools, and "citizens' councils" were organized to resist it. In the border states, where resistance was not so strong, desegregation achieved several relatively quick and easy triumphs, as in Washington, Baltimore, St. Louis, and notably Louisville. But in the lower South, resistance was determined, federal court orders were resisted, and some states adopted policies of "massive resistance." By 1965, partly because of continued Southern segregation and partly because of generalized race discrimination throughout the country, Negro unrest had become intense and mass demonstrations were taking place on a wide scale. A century after the South had made its supreme assertion of a separate identity, and after many decades of steady decline in the significance of the differentials that distinguish the South, the historical heritage of the South's biracial system was still strong enough to create a major sectional issue and to induce a high degree of Southern solidarity in asserting local autonomy as opposed to uniform national policy. In the presence of this immediate issue, the less tangible aspects of Southernism—its emphasis upon the virtues of the agrarian way of life, its conservative concept of the social order, its ideal of social standards set by the leaders rather than by the masses—were so much overshadowed that it is difficult to estimate the extent to which these features, by themselves, would have continued to make the South a significant and distinctive entity.

See also references under "South, The" in the Index.

Bibliography.—W. H. Stephenson and E. M. Coulter (eds.), *A History of the South*, 10 vol. (1947 *et seq.*); Almon E. Parkins, *The South: Its Economic-Geographic Development* (1938); Rupert B. Vance, *Human Geography of the South* (1932); Howard W. Odum *et al.*, *Southern Regions of the United States* (1936); Calvin B. Hoover and Benjamin U. Ratchford, *Economic Resources and Policies of the South* (1951); V. O. Key and A. Heard, *Southern Politics in State and Nation* (1949); Jay B. Hubbell, *The South in American Literature, 1607–1900* (1954); Lewis C. Gray, *History of Agriculture in the Southern United States to 1860*, 2 vol. (1933); Ulrich B. Phillips, *Life and Labor in the Old South* (1929) and *American Negro Slavery* (1918); Kenneth Stampp, *The Peculiar Institution* (1956) and *The Era of Re-*

construction, 1865–1877 (1965); C. Vann Woodward, *Origins of the New South, 1877–1913* (1951); Wilbur J. Cash, *The Mind of the South* (1941); Twelve Southerners, *I'll Take My Stand: the South and the Agrarian Tradition* (1930); Southern Education Reporting Service, *Southern Schools: Progress and Problems*, ed. by P. McCauley and E. D. Ball (1959); Harry S. Ashmore, *An Epitaph for Dixie* (1958); William B. Hesseltine and David L. Smiley, *The South in American History* (1960). (D. Pr.)

SOUTH AFRICA, REPUBLIC OF (Republiek van Suid-Afrika), a country occupying the southernmost part of the African continent between the Atlantic and Indian oceans. From 1910 to 1961 it was known as the Union of South Africa, a constituent part of the Commonwealth of Nations, comprising the former British colonies of Cape of Good Hope, Natal, Transvaal, and Orange River Colony, which became the four provinces of the Union, the last named being known since union as Orange Free State.

The southern, western, and eastern limits of the republic are formed by the coastline. Landward it is bounded on the northwest by South West Africa, on the north by Bechuanaland (which achieved independence in 1966 under the name Botswana) and Rhodesia, and on the northeast by Mozambique (Portuguese East Africa) and Swaziland. The area of the republic is 471,445 sq.mi. (1,221,044 sq.km.). The administrative capital is Pretoria; Cape Town is the seat of the legislature.

South West Africa (*q.v.;* 317,827 sq.mi.), formerly a German possession, was mandated to the Union after World War I and is administered as part of the republic. Within the confines of the republic, but not part of it, lies Basutoland which became independent in 1966 and took the name Lesotho.

In December 1947 South Africa formally annexed two small subantarctic islands, Prince Edward and Marion, which lie in the Indian Ocean approximately 1,200 mi. (1,900 km.) SE of Cape Town.

This article is divided into the following sections and subsections:

I. PHYSICAL GEOGRAPHY

1. Geology and Structure.—Geological formations of South Africa range from some of the oldest rocks of the earth's crust to the most recent formations. The oldest rocks include metamorphosed sediments and lavas of the Swaziland or Kheis system which, together with other more restricted formations, form the Archean complex. Gold and other minerals are associated with these rocks. Next in age are the Proterozoic formations, the oldest of which is the Dominion Reef system of lavas and sediments. This is followed by the Witwatersrand system, 25,000 ft. (7,600 m.) thick, consisting mainly of quartzites and shales with thin reefs of "banket" (quartzitic conglomerates) from which most of the gold of Transvaal and Orange Free State is derived. Next come the Ventersdorp system of lavas and sediments, and the Transvaal system consisting of Black Reef quartzite at the base followed by dolomite and a succession of quartzites, shales, and lavas of the Pretoria series, the system attaining a thickness of about 16,000 ft. (4,900 m.). Toward the close of the deposition of the Pretoria beds vast intrusions of basic and acid magma in central Transvaal formed a lopolith (*i.e.*, a saucer-shaped mass) of norite and allied basic plutonic rocks and the red granite of the Bushveld Igneous Complex. Now exposed by the removal of the overlying beds, this lopolith is revealed as the largest feature of its kind in the world. In these rocks are many minerals, including vast quantities of chrome, platinum, and iron. After the intrusion of the Bushveld magmatic rocks the deposition of the sandstones and conglomerates of the Loskop, Matsap, and Waterberg systems took place in the late Precambrian.

In the Paleozoic, deposition continued with the formation of the Nama system, of quartzite, limestone, and shale, now found in western Cape Province and in southern South West Africa. The Cape system followed; at its base the Table Mountain Sandstone series, 5,000 ft. (1,500 m.) thick, forms most of the prominent mountain features of the western Cape, including Table Mountain itself. The Bokkeveld series of sandstones and shales, 2,500 ft. thick, was next deposited and the third series, the Witteberge, of whitish sandstone, terminated the deposition of the Cape system. In Permian times the whole system was strongly folded. In the Carboniferous the deposition of the Karroo system, which now covers a greater area of the republic than any other formation, began. Having a maximum thickness of about 25,000 ft., the system is divided into the Dwyka, Ecca, Beaufort, and Stormberg series. At the base of the Dwyka series is the Dwyka tillite, which was followed by sandstones and shales with numerous doleritic intrusions and crowned in Basutoland with lava sheets now about 4,000 ft. thick—the Stormberg (or Drakensberg) lavas, probably of early Jurassic age. Except in the Cape folded zone and the eastern area affected by the Natal monocline, the Karroo sediments have remained almost horizontal.

During the Cretaceous there was intense erosion and deposition in the valleys of the Cape folded zone and marine deposition along the south and east coasts; these movements brought into being the valley conglomerates and the estuarine sandstones and clays of the Uitenhage series, and the coastal limestones of Zululand. Deposition continued into the Tertiary in the coastal areas, forming the limestones of the Alexandria and Bredasdorp formations in the south and southwest, and in the east the Zululand Tertiary limestones.

In the plateau basin of the interior (the Kalahari Basin) to the north of the Orange River, the Kalahari system has been deposited, at first apparently under fluviatile conditions and later by eolian action. The basal gravels and partly cemented sands are covered by a mantle of loose, unconsolidated Kalahari sand, generally reddish in colour but grayish in damp localities, the whole formation having a maximum thickness of about 1,000 ft. The sand mantle extends continuously northward into the southern part of the Democratic Republic of the Congo.

2. Physiography.—The surface of South Africa may be considered as having two generally distinct physiographic elements: the plateau surface and the surface between the plateau edge and the coast, which may be called the marginal areas (*see* also *Geographical Regions* below). Separating these two divisions is the plateau edge, or the Great Escarpment, which is the feature formed by the headward erosion of the drainage of the marginal area. This escarpment is sharply defined or indistinct depending on whether the plateau rocks are hard-overlying-soft or generally of undifferentiated hardness. Thus the boldest part of the escarpment lies along the border of Lesotho and Natal where the Stormberg lavas overlie the softer Stormberg sandstones. There is formed the Natal Drakensberg, with spectacular peaks such as the Mont-aux Sources (10,822 ft. [3,299 m.]) in the north and Champagne Castle (11,072 ft. [3,375 m.]) and Giant's Castle (10,868 ft. [3,313 m.]) farther south; the highest point so far surveyed near the plateau edge, and the highest point in South Africa, is Thadentsonyane (Thabana-Ntlenyana) (11,425 ft. [3,482 m.]), about 15 mi. (24 km.) SW of Giant's Castle. The height of the scarp above the comparatively flat surface of the Natal high veld is about 6,000 to 7,000 ft. (1,800 to 2,100 m.). To the north of the Natal Drakensberg the Stormberg lava does not cap the Karroo sediments and the plateau edge is less well defined; but in eastern Transvaal another structural element, the Black Reef quartzite of the Transvaal system, is responsible for the boldness of the escarpment, which is there known as the Transvaal Drakensberg, its highest point, Mt. Anderson, a little behind the Black Reef escarpment itself, being 7,498 ft. (2,285 m.). To the north of the Black Reef quartzite the escarpment is cut in granite and is generally not well defined.

The southern edge of the plateau is given prominence by the presence of strong dolerite sheets which sharply define the Sneeuwberg (Snow Mountains) to the north of Graaff-Reinet, where the dolerite Compass Berg has an altitude of 8,215 ft. The Nieuwveld escarpment to the north and west of Beaufort West is of similar formation. The western edge of the plateau is cut in the south in Karroo beds (the Roggeveld—*i.e.*, Rye Veld—Mountains), in the middle in the Table Mountain Sandstone (the Bokkeveld—*i.e.*, Goat Veld—Mountains), and in the north in granite, the Nama beds and the rocks of the Richtersveld complex. It is continued in South West Africa.

Within the Great Escarpment the plateau surface ranges in altitude from about 3,000 ft. in the west to nearly 11,500 ft. in the east. Two types of plateau surface are distinguishable: the Karroo surface and the pre-Karroo surface, now being exposed by erosion. On account of the predominantly horizontal lie of the sediments and their association with dikes and sills of hard dolerite, the Karroo surface is tabular in character. Great expanses of flat surface are broken by table mountains and koppies (hillocks), generally formed by dolerite capping soft shales or sandstones. In the eastern part of the plateau the Stormberg lavas and the vigorous erosion of the Orange River headwaters have given the Basuto Highlands a rugged, dissected character. The pre-Karroo surface has a more diversified topography chiefly because of the greater variety of the rocks. Where the rocks are hard and disturbed the surface is generally rugged, with prominent ridges of quartzite or ironstone and deep valleys formed by the softer rocks. The quartzite ridges of the Witwatersrand are of this character, as are also the Magaliesberg group of ridges which form much of the margin of the Transvaal bushveld basin. In northern Transvaal the Zoutpansberg ridges are of a similar type, the quartzites dipping toward the Limpopo Trough. Where, however, the pre-Karroo rocks are more nearly horizontal, local plateau surfaces are formed such as the Waterberg Plateau in western Transvaal and the Kaap Plateau in northwestern Cape Province. At the northern boundary of Transvaal the low floor of the middle Limpopo Basin is the result of large-scale trough faulting.

In the marginal zone below the plateau edge prominent features are largely caused by the vigorous erosion of steeply graded streams. Thus hard rocks, such as the quartzites and ironstones of the Swaziland system, form the prominent ridges of northern Swaziland and adjacent Transvaal. The Lebombo Mountains, along the eastern border of Transvaal, are low ridges, about 1,500 to 2,500 ft. above sea level, formed of Karroo lavas dipping steeply eastward. In the midlands of Natal, monoclinal folding has given the Karroo beds a steep eastward dip.

The main ridges in this region

PROVINCES, MAJOR CITIES, AND TOWNS IN THE REPUBLIC OF SOUTH AFRICA

are gigantic spurs jutting out from the Drakensberg between the marginal stream valleys. In the south the marginal zone contains the "Cape folds" of Permian age. In the western part of the folded zone the anticlinal ridges of Table Mountain Sandstone run nearly north-south, slightly arcuate and concave to the coast. Two main lines of ranges are distinguishable: to the west the Hottentots Holland Mountains, stretching northward to the Olifants River Mountains, and to the east the Cedarberg line trending from just south of Vanrhynsdorp toward Ceres. The highest point in this western zone is the Great Winterhoek Mountain, 6,815 ft. In the eastern part of the folded zone two main ranges trend eastward, slightly concave to the south: the northern range, known by various names, may be called the Swartberg; the southern range is the Langeberg line. Between these two is a third rather fragmented series of ridges which may be called the Kammanassie line. The highest peak in this eastern zone is the Seven Weeks Poort Mountain, 7,632 ft. Between the folded zone and the coast is the coastal foreland and between the Langeberg line and the Swartberg lies the Little Karroo Basin. Between the folded belt and the Great Escarpment is the erosional basin known as the Great Karroo (*see* also KARROO).

Hydrography.—The main physiographic division between the interior plateau surface and the marginal area is reflected in the drainage systems of South Africa. The plateau river systems are much larger than the marginal systems; the two most important rivers draining the plateau are the Orange (with its affluent the Vaal), which flows into the Atlantic Ocean, and the Limpopo, which discharges into the Indian Ocean through Portuguese East Africa. (*See* further ORANGE RIVER; VAAL; LIMPOPO.)

There are very few lakes in South Africa, and although there are natural hollows on the surface of the land, most of these are filled with water for only part of the year. These accumulations of water are called pans (Afrikaans *panne*) and are found especially on the flatter outcrops of the Dwyka and Ecca beds of the Karroo system where the surface has been weathered into shallow hollows. The sandveld of Orange Free State has many pans (of sizes ranging up to about 12 sq.mi. [31 sq.km.]) because percolation is hindered by the presence of hardpan two or three feet below the surface. The water in pans may be saline, brackish, or almost fresh.

In contrast to these flat areas, where occasional heavy rain results in sheets of standing water, are sloping surfaces which are vulnerable to erosion by swift-moving water. If the protective vegetation cover is removed by burning or close grazing, soil is exposed to the attack of heavy rain and inches of topsoil may be washed downhill in a single storm. Runoff may concentrate in tracks made by animals going regularly to water holes, and as these bare paths offer no check to running water they are quickly attacked and dongas (gullies) are formed. Cultivated land in parts of the eastern region of heavy summer rainfall has been gashed by dongas and lesser runnels called rills.

Coasts.—The comparative straightness of the northern part of the west coast is probably due to the granite character of this area and the lack of large rivers. The southern part has the large inlet of Saldanha Bay. Lack of fresh water in the vicinity led to its abandonment as a harbour in favour of the more exposed Table Bay, where abundant fresh water was available. Table Bay is backed by the spectacular Table Mountain (*q.v.*) and by the Cape Peninsula which ends southward in Cape Point, the "Cape of Good Hope." On the south coast the chief features are False Bay, Cape Agulhas (Portuguese "needles"; the southernmost point of Africa), and several other capes. At the eastern end Algoa Bay is typical of bays formed by erosion of the softer rocks between Table Mountain Sandstone outcrops. The east coast is remarkable for its straightness, which is thought to be caused by crustal bending and faulting. Earthquakes off the coast of Zululand indicate present crustal instability.

Off the south coast lies the Agulhas Bank extending about 130 mi. (210 km.) SSE of Cape Agulhas, where its depth is about 130 fathoms. Certain parts of the bank are favourable for fishing. Off the west coast, about 7 mi. (11 km.) from Cape Town, lies Robben Island, formerly noted for its seals (Dutch *robben*). The island, about 2¼ mi. long and 1¼ mi. wide, has been the site of a prison settlement, a leper asylum, a defense base, and again a prison station. The seals have entirely disappeared.

3. Climate.—South Africa is dominated by the tropical belt of high atmospheric pressure which encircles the globe between the latitudes of about 25° and 30° S. In the winter months this belt is continuous over land and sea; in summer it is broken by the occurrence of low pressure systems over the land surfaces. In summer, therefore, a monsoon effect is produced by the occurrence of the "low" over South Africa and the "highs" which move eastward and northeastward along the south and east coasts. The interplay of the wind systems resulting from this general relationship affects both the temperature and the rainfall.

Actual temperatures are strongly affected by altitude, which offsets to some extent the normal increase of temperature resulting from decrease of latitude. From the south coast northward the temperatures increase little except in the lowland areas. At the coast Port Elizabeth has a mean annual temperature of 17.8° C (64° F); Graaff-Reinet, at an altitude of 2,440 ft. in the Great Karroo Basin, 17° C (63° F); Johannesburg (5,750 ft., on the high veld) 16° C (61° F); Pietersburg (4,244 ft., northern high veld) 17° C; Messina (1,800 ft. in the Limpopo Trough), 23° C (73° F). The monthly range of temperature is least at the coast and greatest in the interior. Thus the range between the mean monthly temperatures of January and July is for Port Elizabeth 6.7° C (12° F), Graaff-Reinet 11.7° C (21° F); Johannesburg 11.1° C (20° F); Pietersburg 10.6° C (19° F); Messina 9.4° C (17° F).

There is also a marked effect of the coastal currents; the cold Benguela Current flowing northward along the west coast gives low temperatures to the west coastal areas, while the warm southwestward-flowing Mozambique Current on the eastern side of the subcontinent is responsible for the higher temperatures of the eastern coastal areas. Thus on the west coast Port Nolloth has a mean annual temperature of 13.9° C (57° F); the corresponding temperature of Durban on the east coast is 21.1° C (70° F).

Frost occurs over most of the interior plateau and in the higher parts of the marginal zone. The mean duration of the frost period is highest in the Basuto Highlands (more than 180 days) and lowest in the low veld and coastal areas where frost rarely occurs except occasionally in the valley bottoms. The duration of sunshine varies from more than 80% of the possible duration in northern Cape Province to 70% over most of the remainder of the interior, and to less than 60% in the coastal areas.

Three fairly distinct rainfall regions can be recognized according to seasonal occurrence: the summer, the winter, and the all-season regions. In the centre and east of the republic the rainfall has a predominantly summer incidence, being mainly monsoon in character and often associated with heat thunderstorms. The moisture is brought into the interior by winds with an easterly component blowing toward the continental low pressure area developed over the southern Kalahari. There is thus a gradual progression of the summer rainfall from the east coast toward the northwestern part of Cape Province from August to January, and a recession eastward from February to May. Except along the east coastal zone little rain falls in this area in the winter. The highest rainfall in this eastern area occurs on the Great Escarpment, where the average annual precipitation is in places more than 60 in. (1,525 mm.), on other high mountain features, and on parts of the eastern coastal belt. The summer rainfall is mostly of high intensity, which, coupled with high evaporation, reduces its value for agriculture and increases its soil-eroding tendencies.

In the southwest of the republic is the winter rainfall area in which the rain occurs mainly from May to October. This rainfall is mainly cyclonic and orographic in type, occurring most heavily on the mountain ranges where annual falls of 200 in. (5,080 mm.) are not uncommon. In the coastal lowlands the highest annual averages are about 25 in. (635 mm.); to leeward of the mountain ranges the averages fall sharply to about 10 in. (255 mm.) or less. In the third region rain may fall in any season. Thus in the south coastal area rainfall is almost evenly distributed through the year and varies between 20–30 in. (510–760 mm.), while in the narrow

corridor in the interior, between the margins of the summer and winter rainfall areas, average annual rainfall is generally less than 10 in.

For the republic as a whole, it is estimated that about 30% of the surface receives on an average less than 10 in. of rain a year; about 34% receives between 10 and 20 in., 25% between 20 and 30 in., and the remainder more than 30 in. The reliability of the annual rainfall varies from 70% in northwestern Cape Province to about 85% in the southern and eastern coastal areas and to more than 90% along the western part of the south coast. It is estimated that about 51% of South Africa is in drought once every five years, 38% once in three years, and 24% once in two years, drought being defined here as the occurrence of less than 85% of the average rainfall for the year.

4. Vegetation.—The vegetation of South Africa may be divided into five main types: forest, savanna, grasslands, semidesert and desert, and sclerophyll or Cape maquis.

Forest.—Two general types of forest can be recognized, the temperate and montane and the eastern subtropical and tropical. The largest area of temperate forest is situated in the eastern part of the south coastal zone where mean monthly temperatures range from about 12° C (53° F) to about 22° C (71° F) and the mean monthly rainfall from about 1½ to 4½ in. (40–115 mm.). The best-known part of this association is the Knysna Forest in which the main trees include the tall yellowwoods (*Podocarpus* species), growing to a height of 150 ft. (45 m.), the black ironwood (*Olea laurifolia*), the Cape beech (*Myrsine melanophloeos*), the stinkwood (*Ocotea bullata*), and the sneezewood (*Ptaeroxylon obliquum*). Similar associations occur in smaller areas on the south-facing slopes of the Cape mountains, especially on the Hogsback and Amatola escarpments, and in the Natal midlands where, however, only small remnants of the former great forests are left. In eastern Transvaal large forests formerly covered the escarpment of the Transvaal Drakensberg, but these have been exploited for mining operations and the largest natural forest in this area is now conserved in the Woodbush Mountains. The subtropical and tropical forest formerly stretched along a narrow zone behind the east coast where the rainfall is 30–50 in. a year.

J. H. WELLINGTON

NATURAL VEGETATION REGIONS OF SOUTH AFRICA

The dense bush and evergreen forest includes trees of the genera *Sclerocarya, Protorhus, Albizzia, Trichilia,* and *Ficus.* Formerly there were valuable trees such as the Cape mahogany (*Trichilia emetica*), the Cape ebony (*Heywoodia lucens*), and the Cape box (*Notobuxus macowani*), but these have almost all been heavily exploited. (*See* also *The Economy: Production: Forestry* below.)

Savanna.—The bushveld, as it is generally called in South Africa, is generally the dominant association where the rainfall is between about 15 and 30 in. a year and frosts and cold are not severe enough to kill the young trees. The bush may be open, with sparse tree growth, or closed, with dense tree growth. Where no species is dominant the bush is termed mixed. In the republic mixed bush is the most common type, occupying most of the Middle Veld (between about 2,000 and 4,000 ft.) in eastern Cape Province and Natal, and in Transvaal covering the whole of the plateau area below the altitude of about 4,000 ft. In the drier areas the most common trees are acacias, with aloes, euphorbias, and baobabs forming a distinctive feature in northern Transvaal. In the moister parts of Transvaal deciduous broad-leaved trees dominate, among which are the maroola (*Sclerocarya caffra*), various sterculias, the South African mahogany (*Afzelia quanzensis*), and in the lower valleys various palms. The commonest grasses include the genera *Themeda, Digitaria,* and *Panicum.*

Grasslands.—These occupy the higher, cooler areas where the annual rainfall exceeds about 15 in. It is usual in South Africa to distinguish between "sweet" veld grasses, which occur mainly in the western, drier part of the grassland area, and which are nutritious and palatable to stock throughout the year, and "sour" veld, consisting of grass types that during the dry season lose their nutritious character and become woody and indigestible. On the plateau the climax type over most of the area is the red grass (*Themeda triandra*). Retrogression to the more xerophytic love grass (*Eragrostis*) follows overgrazing and soil erosion and, later, to bristle grass (*Aristida*) or Bermuda grass (*Cynodon*). In the higher elevations the grassland is temperate in character with the fescues *Festuca costata* and *F. caprina,* and *Bromus firmior* being common.

Semidesert.—Vegetation is comprised mainly of scattered Karroo shrubs and succulents, which occupy most of the area having an annual rainfall of between 5 and 10 in. The shrubs are mostly perennial woody bushes about 3 to 12 in. high and spaced from about one foot to several feet apart. The leaves are generally very small and the root systems large. The commonest include various species of *Pentzia,* the bitter karroo (*Chrysocoma tenuifolia*), and, in brack soil, various gannas (*Salsola* species) and the saltbushes (*Atriplex* species). Some of these bushes can survive a drought of nine or ten months and they form excellent subsistence feed for small stock. The succulents are of two types: low bushes, rarely more than eight inches high, with ericoid shrublets and many species of *Mesembryanthemum,* and a taller type consisting mainly of euphorbias, aloes, and *Portulacaria.* In the lower Orange River Valley the rocky slopes harbour many desert succulents.

Cape Maquis.—Cape maquis is confined to southwestern and southern Cape Province where the plants are protected against the hot, dry summers. The

shrubs are found from sea level to about 3,000 ft. With annual rainfalls of about 20–30 in. they grow to a height of about 15 ft.; near the lower rainfall limit of about 15 in. the shrubs are 3–5 ft. high, the leaves often very small and leathery. The families typical of the region are Proteaceae, Restiaceae, Ericaceae, Peneaceae, Grubbiaceae, Bruniaceae, Rutaceae, and Iridaceae. A remarkable difference between this region and most other regions of Mediterranean vegetation is the general absence of trees.

5. Animal Life.—Zoologically the republic is recognized as forming the southern part of the southern African subregion of the Ethiopian faunal region.

Mammals.—South African mammals are all placental mammals. The insectivores include little elephant shrews and golden moles. There are more than 20 genera of bats.

The primates are represented by the bush baby or night ape (*Galago*) and by the Cercopithecidae, which include the vervet monkey, the samango monkey, and the chacma baboon. The order Pholidota is represented by the Cape pangolin or scaly anteater.

The carnivores include Canidae (foxes and jackals), Mustelidae (polecat, weasel, ratel or honey badger, otter), Viverridae (civet, genet, mongoose, meerkat), Protelidae (aardwolf, hyena), and Felidae (wildcat, serval, lynx, leopard, lion, cheetah). The Pinnipedia are represented by the Cape sea lion, or fur seal, and the elephant seal, which occasionally visits the south coast. The Tubulidentata has one representative, the aardvark or antbear; the Proboscidea also one, the African elephant. Of the Hyracoidea there are two genera, the hyrax and the dassie, the former being one of the commonest mammals in the republic.

Among the odd-toed hooved mammals are the black and the white rhinoceroses and the mountain and Burchell's zebras. The even-toed ungulates include the bush pig, warthog, hippopotamus, giraffe, Cape buffalo, and about 20 genera of antelope (or buck, as they are generally called in South Africa), including duiker, grysbok, klipspringer, reedbuck, waterbuck, impala, springbok (or springbuck), gemsbok, sable antelope, bontebok, wildebeest, bushbuck, kudu, and eland. There are two genera of hares.

In point of numbers the largest order is probably that of the rodents. Six families are represented, the animals included in them being the mole rat, cane rat, squirrel, spring hare, and dormouse, but the largest family is that of the Muridae, which includes rats, mice, and gerbils. Five genera of murids are peculiar and endemic. In the South African seas are about 20 genera of whales and dolphins.

The present state of the fauna in the republic may be summarized by saying that most of the small mammals are still found wild. The larger animals, especially the carnivores and antelopes, survive only in protected areas. Lions and elephants are found now only in or near the game reserves. The larger buck are preserved on some farms and in the reserves. Smaller buck still roam wild, and the leopard and baboon, the jackal and hyena are still found in the less frequented areas. The blue antelope (bloubok) and the quagga of the Cape have become extinct. In the early 1960s the elephants of the Knysna Forest—the tallest of all African elephants—existed in a herd of fewer than ten beasts and also the Addo elephant herd had no adult bull.

Birds.—Most of the bird families are of the Ethiopian type but of the groups of families one is tropical in character, another Antarctic or southern in its origin, and yet another consists of migratory types. The tropical group consists of well-known types from the central African forests such as the parrots and lovebirds, the touracos, the rollers, and many others. These are found mainly in the eastern part of the republic, having come southward along the East African savanna corridor. The southern group includes most of the sea birds: the jackass penguin (*Spheniscus demersus*), the albatrosses and petrels, including the Cape pigeon (*Daption capensis*), the Cape hen (*Procellaria aequinoctialis*), and several shearwaters.

Of the large group of migratory birds from the Palearctic region are such well-known species as the European swallow, the martin, swift, cuckoo, willow warbler, garden warbler, yellow wagtail, tree pipit, golden oriole, and many others. Of the endemic Ethiopian families only the Promeropidae, the long-tailed sugarbirds, are peculiar to South Africa, the Cape sugarbird and Gurney's sugarbird being confined respectively to the Cape and the eastern areas. Among the other interesting species are the rock jumper (*Chaetops*), the ground woodpecker (*Geocolaptes*), the bald ibis (*Geronticus*) of the eastern mountains, the blue crane (*Tetrapteryx*), and the blue korhaan (*Eupodotis*), all of which are peculiar to South Africa. The largest number of families are in the orders Passeriformes (including, among others, the larks, wagtails, babblers, bulbuls, thrushes, warblers, shrikes, sunbirds, weavers, sparrows, and canaries); the Charadriiformes (including the plovers, phalaropes, skuas, gulls, and terns); and the Gruiformes (including the cranes, bustards, curlews, jaçanas, crakes, and coots). The family Struthionidae is represented by the southern ostrich, which is still wild in certain areas; the Sagittariidae by the snake-killing secretary bird, which is widely dispersed over the savanna and grassveld. An interesting recent newcomer is the Indian myna, which was introduced at first into Natal and now is found in southern Transvaal.

Reptiles and Amphibia.—Turtles, both aquatic and terrestrial, are well represented. The lizards, chameleons, leguans, and geckos include numerous species and subspecies native to the subcontinent. The representative of the Crocodilia is the Nile crocodile, whose southern limit on the west side of the subcontinent is the Cunene River and on the east the Tugela River.

Of the more than 100 species of snakes about 25% are poisonous. The nonpoisonous snakes include the useful mole snake and the curious egg-eaters (*Dasypeltis scabra*). The back-fanged snakes, which are moderately poisonous, include such well-known species as the skaapstekers or "sheep-strikers" and the boomslang or tree snake. Among the highly poisonous front-fanged snakes are the deadly mamba, three species of cobra, and the closely allied ringhals; *i.e.*, "ringed-neck," a spitting snake peculiar to South Africa. The vipers are represented by the puff adder and the horned adder. South Africa also possesses one species of python, which may attain a length of 20 ft., and also one venomous sea snake (*Hydrus platurus*).

The amphibians of South Africa are chiefly frogs and toads. There are no tailed amphibians, and there is only one genus of limbless amphibians. Frogs vary in size from the minute *Arthrolepis*, which may be less than half an inch, to the great bullfrog *Rana adspersa*, or *Pyxicephalus adspersus*, which sometimes attains a length of more than a foot. Among the toads one of the largest, the leopard toad, (*Bufo regularis*), is four to five inches long and is widespread in South Africa. The smallest is the tiny striped mountain toad, *Bufo rosei*, about an inch long. From the commercial point of view the most important amphibian is the aquatic, tongueless, clawed frog called the platanna (*Xenopus laevis*); it provides a dependable test for human pregnancy in the early stages.

Fresh-Water Fishes.—Of about 200 species of fresh-water fishes nearly half belong to the carp family, of which the best known is the barbel. Kurpers comprise about one-quarter, and of the remaining quarter the most numerous are the catfish and tiger fish. Trout have been successfully introduced.

Insects.—All the important orders of insects are represented in South Africa. The orthopterans are numerous and diverse, simulating leaves, flowers, sticks, and other forms. All are carnivorous. The brown locusts, which formerly swarmed over South Africa, are now destroyed in the immature stages. Termites or "white ants" are abundant in species and numbers. Many families of bees, wasps, and ants also are represented. True bumblebees are absent, but the carpenter bee (*Xylocopa*) is common and there is an immense variety of solitary wasps. There are many endemic beetles; especially abundant are the scarabs, or dung beetles, and the chafers. A few types of butterflies and moths are found in the southwest but many occur in the east, some of the genera being peculiar to South Africa. The Diptera include the culicine and anopheline mosquitoes, and the tsetse fly, which was formerly widespread in Zululand.

South African arachnids are abundantly represented by scorpions and solifuges, or hunting spiders, which flourish especially

in the hot, arid conditions of the Karroo and Namaqualand. Ground-living spiders are abundant. *Peripatus*, the wormlike "missing link" in the ancestry of insects, is endemic but is fast disappearing as the natural forest is destroyed.

Game Reserves.—In South Africa these serve the important purpose of conserving the game, much of which is in danger of becoming extinct. There are five national parks and a large number of provincial reserves. The Kruger National Park, the largest, extends for about 200 mi. along the Transvaal-Mozambique border following the Crocodile-Limpopo River, with an average width of about 40 mi. Nearly all the larger game animals (the notable exceptions being the gemsbok and the springbok) and the carnivora are found in the park, as are also the elephant, hippopotamus, and giraffe. The eland and the inyala frequent only the northern parts of the park. The Kalahari National Park is situated between the Nossob and Auob rivers in the extreme north of Cape Province. This is also known as the Gemsbok reserve on account of the large numbers of these animals in the park, but springbok also are very numerous and there are also hartebeest, eland, and kudu, and a fair number of carnivores including lion, leopard, cheetah, and wild dog. The Addo Elephant National Park, about 35 mi. N of Port Elizabeth, is notable for its small herd of elephants which, with the Knysna herd, are the southernmost of the African elephants. The Cape buffalo also thrives there, the only place now in Cape Province where it is found.

The Bontebok National Park (about 1,800 ac. [730 ha.]), about 17 mi. S of Bredasdorp in Cape Province, was formed to preserve the last remaining bontebok, of which a herd of about 150 has been established. The Mountain Zebra National Park, near Cradock in Cape Province, was formed to preserve the few remaining mountain zebras. In the 1960s these were increasing slowly. Of the provincial reserves the most notable are those of Natal. The Hluhluwe and Umfolozi reserves (in Zululand) have preserved the white and the black rhinoceros.

See also WILDLIFE CONSERVATION: *International Co-operation.*

II. GEOGRAPHICAL REGIONS

In South Africa physiography offers the best single basis for regional division. As described above (see *Physiography*) the two primary divisions are the interior plateau and the marginal areas.

1. The Interior Plateau.—Five major regions may be recognized: the High Veld, the Middle Veld, the Transvaal Plateau Basin, the Limpopo Trough, and the Kalahari Basin.

The High Veld.—Lying between the altitudes of about 4,000 and 11,500 ft. (1,200–3,500 m.) above sea level, the southern part of the surface is formed by the beds of the Karroo system. Eroded by the Orange-Vaal headwaters, this structure gives rise to a tabular type of physiography consisting of flat plains and scattered relict mountains and koppies, generally capped by a sill of hard dolerite. The surface is ideal for farming of a pastoral type, being covered with grass suitable for cattle in the eastern part and with grass and karroo shrubs suitable for sheep and goats in the western part. Soils, however, are rather infertile except on dolerite outcrops. The summer rainfall in the eastern part (25–35 in.) is adequate for maize (corn), but not in the western part (15–20 in.), where kafir corn (a sorghum) is more suited to the drier conditions. In the east of the region, the Basuto Highlands, the high plateau surfaces provide summer pastures only; in the valleys are narrow alluvial terraces which form sheltered habitable areas in winter and provide limited cultivation of maize and kafir corn. The northern part of the High Veld region has a different type of surface. The Karroo beds have been removed by long-continued erosion exposing the pre-Karroo surface which was planed by the Permo-Carboniferous ice sheet. This surface, now being eroded by the headwaters of the Limpopo and Vaal rivers, has become more uneven than the Karroo surface as the Proterozoic quartzites, shales, and lavas have been thrown into relief. The great economic factor is the mineral content of the Witwatersrand beds, the "banket" or quartz-conglomerates of which contain the gold particles that have made the Rand the world's greatest gold field. The development of the Witwatersrand has been facilitated by the abundant coal supplies from the neighbouring Karroo beds.

The Cape Middle Veld.—This lies to the west of the High Veld at an altitude of about 2,500–4,000 ft. There is no natural boundary between these two regions, but the lower altitude of the Middle Veld is responsible for its higher summer temperatures and its greater distance from the east coast for its lower rainfall (5–15 in.). The region is thus arid to semiarid and where there is no irrigation is suitable only for small stock farming of an extensive type. Intensive irrigation is practised along the Orange River in this region.

The Transvaal Plateau Basin.—This basin, to the north of the High Veld, comprises the great Transvaal bushveld basin and its highland rims. The floor of the basin is formed mainly of norite and red granite. The norite soils are fairly fertile and repay cultivation, but the granite soils are sandy and infertile. Overlying the plutonic rocks in the northern part of the basin is the Stormberg lava which has decomposed into very fertile black clay soil; the rainfall, however, is low (25 in.) and rather unreliable. The Springbok Flats, as this part of the basin floor is called, is thus an area of mixed farming rather than of intensive cultivation. The highland rim of the basin is formed by the High Veld in the southeast, by the *bankeveld* (or cuesta) topography in the southwest and east, and mainly by the Waterberg Plateau in the north. The Pietersburg Plain and the Zoutpansberg Ranges form the extreme northern part of the region. The highest economic potentialities of the region are associated with the irrigation areas along the Crocodile, Olifants, and Marico rivers and with the smaller schemes in the valleys of the *bankeveld* marginal areas. Minerals such as chromite and platinum are present in large quantities in the plutonic complex of the basin floor and low-grade iron ore in almost inexhaustible quantities in the marginal quartzites.

The Limpopo Trough.—This is the middle portion of the Limpopo Basin. Formed mainly by large-scale faulting, the valley has an altitude of between about 1,500 ft. at its eastern end (Pafuri) and 3,500 ft. at its southwestern end (Lobatsi). With low rainfall (15–25 in.) and high temperatures this area is of little agricultural importance, being mainly suited to extensive ranching. Copper in the Messina area is the only important mineral.

The Kalahari Basin.—Only the extreme southern part of the Kalahari falls within the republic. The sand surface is more or less covered with grass and sparse tree growth and water is generally absent except in scattered pans and along the dry watercourses. The area is thus scarcely inhabited away from the river courses such as the Molopo in the east and the Nossob and Auob in the west. (*See* KALAHARI DESERT.)

2. The Marginal Areas.—In the area marginal to the plateau the eastern, southern, and western areas form fairly distinct types of country.

The Eastern Marginal Area.—The fall from the plateau edge to the coast is in many parts taken in three steps or surfaces, forming a high veld surface above about 4,000 ft., a midland surface between about 2,000 and 4,000 ft., and a coastal zone below about 2,000 ft. In Natal and the Transkei these are all areas of fairly high summer rainfall (25–40 in.) so that cultivation is generally possible, the crops being determined mainly by the temperature. On the eastern high veld surface the dominant occupation is pastoral farming, in the midlands mixed farming, and in the coastal zone of Natal the cultivation of sugarcane.

The Southern Marginal Area.—The surface is dominated by the Cape folded ranges. Agriculture is confined mainly to the well-watered coastal belt and to the irrigable valleys of the Little and Great Karroos. The Karroo vegetation is well suited to sheep and goat farming and the irrigated lands to the cultivation of lucerne and, in some areas, of fruit.

The Western Marginal Area.—The surface falls rather irregularly and the climate is too dry for any sustained agriculture to the north of about Vanrhynsdorp. Southward the area between the folded mountains and the coast is well suited to winter cereals, especially wheat, and to summer fruits such as grapes, peaches, plums, and apricots. (J. H. WN.)

III. THE PEOPLE

1. Ethnic Groups.—South Africa contains peoples of many diverse ethnic origins and differing cultures, languages, and forms of social organization. This complex situation may best be described historically, in terms of immigrations into the region.

Bushmen and Hottentots.—These, the surviving representatives of the original inhabitants of South Africa, are of non-Negro physical type, with languages which, though related to one another, are unrelated to any other language families. They are together known as Khoisan (*q.v.*). The Bushmen are, and would seem always to have been, hunters and gatherers, living a wandering life and divided into small isolated bands lacking any form of centralized chieftainship. Traces of their inhabitance, mainly in the form of rock engravings, show that they were once widespread

CONSTANCE STUART FROM BLACK STAR

BUSHMAN HUNTERS OF THE KALAHARI DESERT

throughout southern and central Africa; but they are now virtually confined to the more arid areas of South West Africa and Botswana. The Bushmen of what is now the Republic of South Africa were mainly killed, driven out, or absorbed by Negroes and whites. The Hottentots were pastoralists, keeping cattle and sheep, and occupied most of the southwestern corner of Africa at the time of the Dutch arrival at the Cape. Only the Nama Hottentots of South West Africa now form a distinct Hottentot-speaking group. Others have become part of various mixed-blood communities, described below. In the 1960s Bushmen numbered about 55,000 and the Nama about 25,000. (*See* further BUSHMAN; HOTTENTOT.)

Nguni.—The main element of the South African population consists of various Negro peoples speaking southern Bantu languages. They are, indeed, conventionally classified according to their linguistic affiliation and collectively referred to as Bantu for statistical purposes. The first Negro group to enter South Africa would seem to have been the cluster of tribes speaking Nguni dialects. They occupied what is now Natal and had covered much of the eastern Cape before the arrival of Vasco da Gama, and held the same areas when met by the Dutch and English in the 18th century. Nguni languages contain three of the so-called clicks (*c*, *q*, and *x*) of the Bushmen and Hottentots, and the Nguni tribes show some Hottentot physical characteristics; presumably they drove the Hottentots before them in their southern migrations and absorbed many of the Hottentot tribes.

The Nguni are usually divided into the southern Nguni of the Cape and the northern Nguni of Natal. The southern or Cape Nguni consist of several tribes, mostly living in the Transkei area: the Pondo, Thembu (Tembu), Xhosa, Bomvana, Mpondomise, and others, in all numbering well over 2,500,000. The northern Nguni include the Zulu, Swazi, Bhaca, Fingo, and others, and also total over 2,500,000, in Natal and the surrounding areas. As a consequence of the political upheavals of the early 19th century associated with the rise of the Zulu king Shaka (Chaka), several offshoots of northern Nguni moved northward: the Ndebele of the Transvaal, the Ndebele (Matabele) of Rhodesia, and the Ngoni of Malawi, Zambia, and Tanganyika.

All the Nguni of South Africa were traditionally organized into chiefdoms, reckoning descent patrilineally and observing strict patriarchy, and having economies of mixed agriculture and cattle raising. Some of them, particularly the Zulu and Swazi, established kingships and became important military nations in the 19th century. (*See* further NDEBELE; NGONI; NGUNI; SWAZI; ZULU.)

Sotho.—Behind the Nguni came the various Negro peoples speaking Sotho dialects, which are closely related to Nguni but lack the Hottentot-derived clicks. The Sotho-speakers consist of the southern Sotho, or Suto, of Lesotho; the western Sotho, or Tswana, of Botswana; and the northern Sotho of the Transvaal (the Pedi, Lovedu, and others). In the northern Transvaal are also the Venda, sometimes classified with the Sotho but usually counted as a distinct group. In their economy the Sotho peoples resemble the Nguni tribes, but they differ in certain other respects: some of them live in large towns rather than in the dispersed settlements of the Nguni, and they approve of marriage between cousins, whereas the Nguni prohibit it. They reckon descent patrilineally, like the Nguni, and are also organized into chiefdoms, often with rulers of marked sacred or even divine attributes. Some rulers, such as those of the Lovedu in the Transvaal, are women; they play a male role and "marry" "wives," whose children may inherit whereas any children of the ruler are counted as illegitimate. The rulers of the western and southern Sotho chiefdoms have greater secular and military power and resemble those of the Zulu and Swazi. (*See* further SOTHO; SUTO; TSWANA.)

With the growth of the diamond and gold industries and the development of cities such as Johannesburg and Kimberley there developed the system of migrant Bantu labour to the towns, mines, and white farming areas outside the Cape (in the Cape itself Coloured labour has always been predominant). This has led to an increasing Negro population in the towns, both purely migrant (*i.e.*, temporary) with its roots in the tribal reserves, and also fully urbanized and "detribalized," living permanently in the locations of the cities and towns. A good proportion of the temporary labour, especially that employed in mines, comes from outside South Africa, from Lesotho, Botswana, Portuguese East Africa, Rhodesia, Zambia, and Malawi. The permanent African urban population has tended to drop tribal customs and even tribal languages and to become a new amalgam with an urban, characteristically Western way of life, although economically it is an extremely depressed one.

See also AFRICA: *Ethnography (Anthropology): Southern Africa.*

European Intrusion.—The Europeans have lived in South Africa since the founding of Cape Town in 1652. Originally Dutch, the white population was first concentrated in the western Cape. The Dutch made early contact with the Hottentots and seem at first to have treated them with kindness and to have regarded them as of more or less equal standing (there was occasional intermarriage and much concubinage). With the extension of white farms along the Cape seaboard, whites came into direct competition with Hottentots over water and grazing; they drove some of them into the more arid regions to the north and made others into little more than serfs. The Dutch imported slaves, mainly from the East Indies and Madagascar. There was much miscegenation between whites (both Dutch and English) and slaves, Hottentots, and Bushmen, and the descendants of these various unions later became the Cape Coloured people. The Dutch, to show their allegiance to their new country, took the name Afrikaner.

The whites came into contact, and also into conflict, with Negroes on the eastern borders of the Cape during the 18th century. The ensuing "Kaffir Wars," fought between white and Negro farmers who desired the same lands and water supplies since both groups were mixed agriculturalists and cattle keepers, led to considerable tribal disruption and movement. This resulted in the founding of the Basuto nation, which was largely made up of refugees from elsewhere, and the appearance of the Fingo of the Ciskei from the remnants of broken Nguni tribes, and later of many small, racially mixed communities of white-Hottentot-Negro origin (the Griquas and others).

Asians.—The last main element in the South African population is Indian. Indians were brought to Natal in 1860 mainly as indentured labourers on the sugar estates in the humid coastal zone.

A century later they numbered nearly 400,000 in Natal, two-thirds of whom lived in Durban, with smaller communities elsewhere, mainly as traders. There are also a few hundred Chinese and smaller numbers of Mauritians, St. Helenians, and others, almost all in the towns.

2. Religions.—The traditional religious beliefs and practices of the peoples of South Africa vary widely. Those of the Bushmen and Hottentots were based upon worship of the moon and other heavenly bodies, although both peoples believed in an afterlife also. Among the Bantu-speaking groups the principal indigenous cults are those of the ancestors, sacrifice to them being made by the heads of lineages and clans. A high god was recognized but sacrifice was not made to him. Beliefs in witches and sorcerers were and are prevalent, witches being thought to be women throughout this region. In most of the states and large chiefdoms the ruler has important ritual duties at first-fruit ceremonies and other rites, which have both religious and political aspects. In some states the ruler has great sanctity: the king of the Venda is regarded as divine before his death, for example, and the queen of the Lovedu is recognized widely for her near divinity and power to control rain and fertility.

Christianity has become widespread among all the Bantu groups, the Methodist Church having the largest number of adherents. In most areas there are, besides orthodox mission churches, separatist churches led by Bantu priests, bishops, and archbishops (the latter often self-consecrated), who resent white control of most mission churches and regard themselves as prophets or messiahs. The separatist churches consist mainly of various "Zionist" churches, in which speaking with tongues and "rolling" are practised, and "Ethiopian" churches in which antiwhite, tribalistic, and nationalistic emphasis is strong.

Most Indians follow Hindu or Muslim rites. The majority of the Cape Coloureds are Christian, only the Cape Malays being Muslims. Among Europeans the most powerful churches are the Dutch Reformed churches. These churches, which have a mainly Afrikaner membership, accept racial segregation as divinely ordained, and as a result they have come under much criticism from Protestant churches outside South Africa. (*See* REFORMED CHURCHES: *History: The Netherlands: Former Dutch Colonial Areas.*) The majority of other Europeans belong to the Anglican, Methodist, Roman Catholic, or Presbyterian churches, and there is also a substantial number of Jews. None of these bodies accepts the views of the Dutch Reformed churches on racial segregation, and various individual members have expressed opposition to the South African government's racial policies.

3. Languages.—The two principal official languages are Afrikaans and English. At the 1960 census Afrikaans was given as first among 58% of the white population and English among 37.3%, but with instruction in both languages in schools, an increasing number of people are bilingual. (*See* also NETHERLANDIC LANGUAGE; SOUTH AFRICAN LITERATURE: *Afrikaans.*)

Africans speak various Bantu dialects; some, such as Xhosa (in the Transkei Bantustan), are accepted as official languages within the homelands and are used as vernaculars in education. A large number of Africans also speak English or Afrikaans, and a lingua franca known as Fanagolo is also used in the mining areas. (*See* AFRICAN LANGUAGES; BANTU LANGUAGES; BUSHMAN LANGUAGES.) (J. F. M. M.)

IV. HISTORY

Whether or not the Phoenicians sailed around Africa as claimed by Herodotus, southern Africa was unknown to the Greeks and the Romans and to Europeans before the 15th century. Then, largely as a result of the patronage of Prince Henry the Navigator, Portuguese mariners steadily pushed down the West African coast until in 1488 Bartolomeu Dias de Novais rounded the Cape of Good Hope and in 1498 Vasco da Gama, in a flagship of 120 tons, 80 ft. long, reached the objective, India. Thereafter the Portuguese gained maritime supremacy in the Indian Ocean and used it to exploit the Asiatic spice trade and, less effectively, the East African gold trade. But, though they colonized Angola and Mozambique, they made no settlements as far south as what is now the Republic of South Africa, regarding its resources as negligible and its inhabitants as too primitive for commercial or evangelical effort. Throughout the 16th century, therefore, though South African waters were often traversed by Portuguese fleets, the only civilized people who spent much time ashore were the survivors of the many ships that were wrecked on the coast.

A. DUTCH FOUNDATIONS

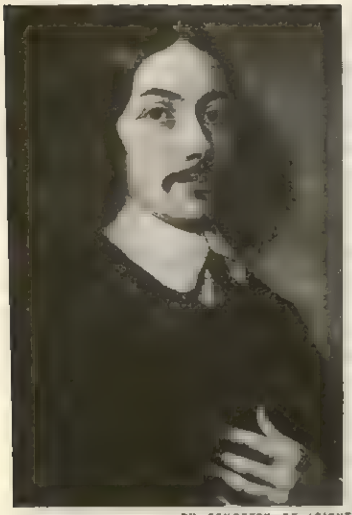
1. Trading Companies.—By the end of the 16th century Dutch and English seamen had also begun to trade with Asia by the Cape route. They found it desirable to pause for refreshment (the death rate on the long voyage being often as high as 50%), and during the first half of the 17th century they usually stopped at St. Helena, the Cape, or Mauritius. In 1615 the English tried vainly to form a settlement at the Cape with ten felons reprieved from the hangman's noose, and five years later two English captains made the empty gesture of annexing all Africa for James I; but it was St. Helena that eventually became the main port of call for the English company.

Meanwhile the Dutch East India Company (*q.v.*) outstripped all rivals and obtained a virtual monopoly of the East Indian spice trade. For several decades Dutch ships made a practice of putting into Table Bay to take in fresh water, to barter with the Hottentots for meat, and to find the latest information about the affairs of Europe or Asia in dispatches they left one another under inscribed stones. Then in 1647 Leendert Janssen was shipwrecked in Table Bay with the crew of the "Noord Haarlem," and after his return to the Netherlands he presented a report advising the company to found a permanent refreshment station at the Cape. Largely because the Anglo-Dutch War of 1652–54 was imminent, the directors eventually agreed; three ships were to take out building materials, seeds, and implements, and 90 men were to build a fort and develop a vegetable garden.

2. Jan van Riebeeck.—To command the expedition the directors chose Jan van Riebeeck (*q.v.*), a doctor by training, who had already had six years of commercial experience with the company. He landed at the Cape on April 7, 1652, and remained ten years, during which time droughts, floods, and pests, indiscipline among his subordinates, and thefts by the Hottentots often brought the settlement to the brink of disaster. But it survived; a timber-and-sod fort was built, the coastline was charted, the interior was explored, and a belt of land extending about eight miles from Table Bay to Wynberg was brought under cultivation.

However, the Cape did not become the simple outpost the directors had had in mind. At that time Hottentots, pastoral nomads loosely organized in small tribes, occupied most of the coastal regions of South Africa as far east as the Great Fish River. The Cape Peninsula had been used by three such tribes, who regarded the Dutch as intruders in their domain, especially when it became apparent that they were not birds of passage but were building fortifications and extending their area of cultivation. The Dutch bartered with the Hottentots for sheep and cattle and from time to time the Hottentots stole them back again. Van Riebeeck calculated whether he was strong enough to use force, but when the directors cautioned him to do no such thing he fell back on the idea of confining the settlement within a palisade. Yet even at this early stage in South African history it was not possible to keep the races completely apart, because the Dutch continued to barter with the natives.

Van Riebeeck also initiated two experiments which had far-reaching consequences. At first all the work of the station was done by the company's servants, but in 1657 a few slaves were imported and more came in 1658 and thereafter. The second experiment started in 1657 when the company, hoping it would save money if grain were grown and stock bred by unpaid freemen instead of by paid employees, freed nine married men from their contracts and gave them 30-ac. farms along the Liesbeeck Valley at Rondebosch. These people, who were known as burghers, were to produce grain and stock and sell them to the company at agreed prices. So Van Riebeeck left behind him an embryonic colony of settlement which used slave labour and was confronted with a "native problem."



BY COURTESY OF (RIGHT) THE SOUTH AFRICAN INFORMATION SERVICE, (LEFT) THE GRANGER COLLECTION

(LEFT) JAN VAN RIEBEECK, LEADER OF THE FIRST WHITE SETTLEMENT IN SOUTH AFRICA; (RIGHT) AN ASPECT OF CAPE TOWN AS PAINTED ABOUT 1720 BY AN UNKNOWN ARTIST

who recalled Van der Stel in 1707. But the underlying cause of the discontent—the commercial policy of the company—was not remedied.

4. Evolution of the Afrikaner People.—The company, probably convinced that white settlement led to political agitation, had no vision of a New Netherlands in South Africa and sponsored no immigration schemes after the time of the Van der Stels. Nevertheless, during the 18th century a New Netherlands of a sort was arising from the materials present. The white population grew, mainly by natural increase, to about 15,000 in 1795, by which time they regarded South Africa as their only home and spoke a language, Afrikaans, which deviated from the language of the Netherlands. Those who lived in Cape Town—the only port and the only real town in the colony—were mostly traders and innkeepers. In the southwestern part of the colony, near Cape Town, were the grain and wine farmers, some of them proprietors of large estates run by slave and Hottentot labour. Beyond them were the trekboers, about half the total white population.

The trekboers were seminomadic, many of them completely nomadic, pastoral farmers. Most of the land they occupied was unsuitable for cultivation; moreover, the market for agricultural produce was glutted and slaves did the skilled work of the colony. There was, therefore, nothing to keep the increase of the white population in the southwestern Cape and much to encourage them to disperse and become self-sufficient pastoral farmers.

The way was paved by W. A. van der Stel, who, unlike his father, allowed burghers to concentrate on stock farming. Within a few years it became possible for any burgher to take occupation of 6,000 ac. or more of land and treat it as his property, paying the company a small annual licence. The result was a continuous expansion of the frontiers of white settlement. In the north the trekboers were slowed up by aridity and finally stopped near the Buffels River about 1760. In the east they advanced along either side of the arid Great Karroo until in the 1770s the two streams of pioneers began to mingle beyond it in the vicinity of what became Graaff-Reinet, about 400 mi. (640 km.) from Cape Town.

Thus a few thousand white persons scattered over an area about twice as large as England and about the size of Colorado. Their simple needs were met by their stock and by hunting and by traveling traders who bought their sheep and cattle for the Cape Town market and sold them a few imported goods. The company did little for them beyond establishing courts of landdrosts and heemraden at Swellendam (1745) and Graaff-Reinet (1785). From time to time officials expressed concern that the isolation of the trekboers was causing them to degenerate. But although all the trekboers were extreme individualists and some were ruffians, most of them contrived to preserve a sound domestic morality, a smattering of literacy, and a respect for their religion, a rather strait Calvinism.

5. Slaves, Hottentots, and Bushmen.—During the 18th century the slaves, the Hottentots, and some of the Bushmen were beginning to interbreed with one another and with Europeans to produce the community now known as the Cape Coloured people. Although few slaves were imported, the slave population outstripped the white to number nearly 17,000 in 1795. Negroid types from the west and east coasts of Africa predominated, with Malays from the East Indies forming an important minority. There was a considerable amount of extramarital miscegenation between white men and slave women. Most of the slaves were the property of the agricultural farmers, for whom they formed the main labour supply. Others were owned by the company and the

3. Immigration and Expansion.—In 1662 the Cape settlement numbered about 250 white persons, of whom nearly half were company servants, and in the next few years there was little increase. But during the governorships of Simon van der Stel (1679–99) and his son W. A. van der Stel (1699–1707) the company made its one serious effort to encourage immigration. In 1707 a census showed a burgher population of 1,779 men, women, and children, owning 1,107 slaves. This was a slow rate of progress when compared with that of the English colonies in North America.

Nevertheless, the modern Afrikaner people are descended for the most part from those who were enumerated in 1707. These were company servants who had taken their discharge at the Cape and people who were sent out to settle. Most of them were of Dutch and German stock. There was also a small but significant group of fewer than 200 French Huguenots who, after the revocation of the Edict of Nantes in 1685, went to the Netherlands and then were sent to the Cape. As a matter of policy the company scattered the Huguenots among the other settlers, and within little more than a generation they became fully assimilated, discarding the French language and even pronouncing their names in the Dutch manner.

The area of settlement expanded with the population—to Stellenbosch beyond the sandy Cape Flats (1679) and then to the Berg River Valley. To meet the needs of local administration a general purposes official (landdrost) was appointed to Stellenbosch in 1685; sitting with four local burghers (heemraden), he was given jurisdiction in minor civil cases. Nearly all the company servants, however, lived in Cape Town, where the governor and his chief subordinates formed a Council of Policy which issued regulations for the colony, subject to the overriding authority of the governor general and council of the Indies in Batavia and the directors in the Netherlands.

From the first the burghers had fretted at many of the regulations, which imposed restrictions and taxes on internal trade; for example, the rights to deal in much local produce, including cattle, were auctioned annually to contractors, as well as the rights to retail imported goods. Furthermore, by the end of the century the Cape was producing more wine, brandy, and grain than was consumed locally and by passing ships, and there was little opportunity for export.

What was particularly galling to the burghers was the tendency of the officials to farm and trade on their own account, using their authority to make sure that their own produce did not remain unsold. This led to an outcry against the governor, W. A. van der Stel, who embarked on large-scale farming operations and cornered the market so successfully that some malcontents, including the Huguenot diarist Adam Tas, sent a petition to the directors,

Cape Town burghers, a few by the trekboers. The treatment of the slaves, which had at first been fairly mild, tended to deteriorate, and the 1754 slave code was a stiff one.

The nearest approach to a Hottentot war had taken place in the 1670s. But it was smallpox that took the heaviest toll of the Hottentots, especially a shattering epidemic in 1713. After that, as the trekboers advanced into their territories, the surviving Hottentots lost their land and their cattle—their only means of an independent existence. Many were reduced to a status not unlike serfdom, forming the main labour supply of the trekboers, whose simple demands they were able to satisfy. Others withdrew beyond the frontiers of white settlement to maintain a more or less independent existence until the next century.

THE BETTMANN ARCHIVE

HOTTENTOTS: A LATE 18TH-CENTURY ENGRAVING

There was a certain amount of miscegenation between white men and Hottentot women, although intermarriages had been unlawful since 1685. The offspring were usually disowned by their fathers and became known, bluntly, as Bastaards; some entered into service, others retreated before the advancing trekboers, leading trekboer lives themselves.

The small Bushmen were still more primitive than the Hottentots. Before white settlement began they had been driven into the more arid and mountainous parts of South Africa by the Hottentot and other African tribes. But, though every man's hand was against them, they contrived to make hit-and-run raids to get sheep and cattle for food, and, if cornered, they were able to kill human beings with their poisoned arrows. As the trekboers in turn began to enter Bushman country and to suffer from their raids they organized Boer commandos to hunt them. Toward the end of the 18th century several Bushmen bands made a prolonged stand, operating from mountain fastnesses along the northern fringes of the area of white settlement; but the commandos gradually cleared them out, killing 2,500 Bushmen and capturing 650 in the decade ending in 1795. Those captured were children, who were taken into service.

Their experience with slaves, Hottentots, and Bushmen had important and enduring effects on the racial attitudes of the Afrikaner people. Most of the nonwhites they came into contact with were at so low a cultural level that the Afrikaners tended to think of themselves as innately superior. Their response was not merely hostile, however, because they all—Capetonians, agriculturalists, and trekboers—came to depend on the use of nonwhite servants.

6. African Opposition and Economic Discontent.—Before the last quarter of the 18th century the trekboers had not been seriously impeded by human obstacles. Then, however, they came up against formidable opposition.

Bantu-speaking tribes had occupied the eastern half of South Africa for several centuries. They were more numerous than the Hottentots. They had a more complex social and political structure. They cultivated the soil as well as hunting game and herding cattle. They made iron weapons and tools. South of the Limpopo River were two main groups of Bantu-speaking peoples: the Sotho, who lived on the plateau; and the Nguni, in the coastal belt. The westernmost Nguni tribes slowly expanded westward as population increased and tribes split, but the trekboers moved toward them more rapidly. Early in the 18th century Boer hunters began to make contact with the Nguni and bartered with them for cattle and ivory. The company vainly tried to prevent such contacts and delay the convergence of the Boers and the Nguni by issuing proclamations prohibiting barter and defining limits beyond which the Boers were not to go. When the governor, J. van Plettenburg, toured the frontier in 1778 he found that Boers were already intermingled with Africans in the vicinity of the Great Fish River.

A long series of frontier "wars" followed—basically struggles for possession of the land. The first fighting took place in 1779–81 and was sufficient to show the Boers that they faced tough opposition. They therefore appealed to the company for aid; although a landdrost, M. Woeke, was appointed to Graaff-Reinet in 1785, no troops were sent. Woeke's life was soon made intolerable by the Boers, and in 1793 he was superseded by H. C. D. Maynier. Further fighting broke out in that year and again the result was inconclusive. The Boers were disgruntled. For nearly a century they had been left by the company to make their own terms with man and nature, and when at length they had asked for help all they received was a government official who not only failed to defeat the Nguni but also allowed Hottentots to come to his court with complaints against them. In 1795, on the eve of the British conquest of the Cape, they drove out Maynier and declared Graaff-Reinet to be a republic. But the name "republic" was a euphemism: in fact there was anarchy at Graaff-Reinet. Similar events took place at Swellendam, where another "republic" came into being.

Meanwhile there was trouble in the southwestern part of the colony also. The company had maintained rigorous commercial restrictions, together with a system of government that gave the colonists practically no voice in the determination of policy. The trekboers had managed to escape most of the effects, but the restrictions had always borne heavily on the agricultural farmer who had produce to sell and on the Cape Town man who lived by trade. In the 1770s, which were lean years, the discontent came to a head with the organization of the Cape Patriots movement, which had some connections with the anti-Orange Patriot Party in the Netherlands. A petition was drawn up criticizing many of the officials and asking for full burgher representation on the Council of Policy and the right to trade freely throughout the company's possessions. When it was laid before the directors in Amsterdam, their only response was to dismiss the official who was most roundly criticized—the fiscal, W. Boers.

In 1784 a second petition was taken above the heads of the directors to the States-General of the Netherlands. This time the burghers were given equal representation with officials on the Court of Justice (though not on the Council of Policy) and the right to trade with the company's possessions—a right that was nullified by the conditions that were attached. By then, however, the disaffection had subsided because the American Revolutionary War had brought a large French garrison to Cape Town to help the Dutch to defend it against the British, and the garrison provided an abnormally large market for Cape produce and brisk business for the traders and innkeepers. In the early 1790s bad times came again as the company, trying desperately to stave off bankruptcy, made a final effort to make the colony pay its way. The garrison was reduced, public works were suspended, and—the last straw—new taxes were imposed. So, when the British invading force arrived in 1795, it found depression in the west and rebellion in the east. The company had never done much for the white colonists and still less for the nonwhite peoples.

B. South Africa Divides (1795–1870)

1. British Conquest.—In the course of the Napoleonic Wars, the Cape was captured by a British naval and military force in September 1795, nominally on behalf of the prince of Orange, who had taken refuge in England from the Dutch republicans. In their struggle with France the British considered it essential not to let the Cape fall into enemy hands, and until the Treaty of Amiens a hold was kept on it for strategic reasons. The British tried to conciliate the Afrikaners, and by abolishing the restrictions on internal trade and maintaining a large garrison they caused an economic revival in the southwestern part of the colony. On the eastern frontier, on the other hand, the British were no more successful than their predecessors had been. Although the republics of Swellendam and Graaff-Reinet soon submitted to the new government, anarchy developed in Graaff-Reinet again in 1799, after the arrest of a prominent trekboer, A. van Jaarsveld, on a charge of forgery. Not only the Xhosa but also many of the Hottentot servants of the Boers made the most of the opportunity

created by the divisions among the whites and caused havoc as far west as Knysna. The British general F. Dundas patched up a truce; but H. C. D. Maynier, appointed resident commissioner of the two eastern districts, failed to satisfy the trekboers, who rebelled again in 1801, when once again their Hottentot servants joined with Xhosa tribesmen to harry the farms over a wide area.

The eastern districts were therefore in turmoil when Britain gave up the Cape under the terms of the Treaty of Amiens (1802). By this time the company was defunct, and the Dutch government —the Batavian Republic organized by the French in 1795 after their conquest of the country—succeeded to its charge. J. A. de Mist, the commissioner-general, and Gen. J. W. Janssens, the governor, tried to restore the morale of the Boers and showed a considerable understanding of the problems of the country; but these men did not stay long enough to consolidate their reforms, for in 1806 the British sent out another force to recapture the Cape from Napoleon's Dutch allies. Janssens lacked the means to put up much resistance and once more the colony passed into the hands of the British, whose title was confirmed in 1814 as part of the general peace settlement.

2. British Policy, 1806–36.—Until 1823 British policy was cautious and conservative. The governor had autocratic powers, but the Afrikaners were used to that. Moreover, their Roman-Dutch law was retained, and local administration remained as of old in the hands of landdrosts and heemraden, most of them Afrikaners. These were comparatively prosperous years. The southwestern farmers profited from their access to British markets, where their wine sold well thanks to a substantial tariff preference. Elsewhere, however, extensive pastoral farming remained the norm. Nevertheless, the British did introduce some changes in these years. They appointed more landdrosts. In 1809 they promulgated a code for the treatment of Hottentot servants, who were to be employed only under written contracts registered at a landdrost's office and were given legal protection against ill treatment; on the other hand, they were not to be allowed to leave their employer's farm without a pass signed by the employer. From 1811 onward high court judges, who had previously sat only in Cape Town, made annual circuits to hear cases in all the districts.

Mild though these reforms were, they caused trouble. There was much grumbling in 1812 when the judges on circuit in the eastern districts tried a number of Afrikaners on charges of ill treating their Hottentot servants. Three years later a trekboer, F. C. Bezuidenhout, opened fire on a party that had come to arrest him for repeatedly ignoring his landdrost's orders to answer a charge of cruelty; in the skirmish Bezuidenhout was killed, whereupon his brother proceeded to organize opposition to the government; but the Slachter's Nek Rebellion soon ended when about 60 rebels were rounded up and tried, and five ringleaders were hanged.

The British also brought superior force to bear against the Xhosa on the eastern frontier and thus started the process of the conquest of the African tribes. In 1811 British regular troops and Boer commandos drove the Xhosa back to the eastern side of the Fish River, and in 1819 to the eastern side of the Keiskama.

In 1820, 5,000 British settlers were brought out and placed on 100-ac. lots in the Zuurveld, on the western side of the Great Fish River. The 1820 settlers did not, however, form the human barrier that was intended, because the land was not suited to agriculture; within a few years most of them had abandoned their lots and become townsmen in Port Elizabeth and Grahamstown. Others became traders among the African tribes to the east. Those who stayed on the land eventually made good by producing wool for export to England, especially when large farms came on to the market at the time of the Great Trek. This was the first subsidized immigration scheme since the 17th century, and the 1820 settlers were the first white immigrants who were not assimilated by the Afrikaner people.

When they had got their bearings the 1820 settlers were dismayed to find that the legal system of the colony was alien to them, the judges were officials without any particular legal training, and Lord Charles Somerset, the governor from 1814 to 1826, was an autocrat who made it impossible for them to publish a newspaper. Their agitation for British liberties and institutions evoked some response. The Charter of Justice of 1827 created a Supreme Court of qualified judges with security of tenure, abolished the heemraden, and introduced English rules of procedure and evidence, including the jury system, while it left untouched the Roman-Dutch substance of the law. In 1834 an Executive Council of officials and a Legislative Council of officials and nominated nonofficials were established. By that time the freedom of the press had been won, English had become the official language of the courts, and British teachers had been imported to start village schools in which English was the medium of instruction.

During these years South Africa was attracting the attention of the British humanitarians or evangelicals who were striving to free the nonwhite populations of the colonies from their legal disabilities. They were supplied with severe criticisms of South African conditions by John Philip, South African superintendent of the London Missionary Society, which had established a number of mission stations up and down the country. Philip reached the conclusion that the Hottentot code of 1809, with its stringent pass regulations, was unjust; in 1826 he returned to England to appeal to the government and the British public, and in 1828 the House of Commons passed a resolution for the emancipation of the Hottentots. Seeing the trend of affairs, Gen. (afterward Sir) Richard Bourke, the acting governor, had already passed an ordinance, no. 50 of 1828, which abolished the pass system and gave "free persons of colour" complete liberty of movement. Meanwhile, with their eyes focused primarily on the West Indies, the British government had been insisting on improvements in the treatment of slaves; and in 1833 the reformed Parliament passed the Emancipation Act under which all slaves in the colonies became free after a period of apprenticeship, and £20,000,000 was voted as compensation to their owners. When their period of apprenticeship ended in 1838 the former slaves in the Cape Colony stepped into the position of other "free persons of colour." Thereafter the emancipated slaves and Hottentots—the Cape Coloured people—became a rural and urban working class in the Cape Colony.

Meanwhile frontier trouble had flared up again. Unrest had developed among the border Nguni tribes for various reasons, including the frequency of military raids on their villages made in reprisal for the theft of stock from white farmers. In 1834 Gen. Sir Benjamin D'Urban (*q.v.*) arrived as governor with instructions to devise a more equitable policy, but he did nothing about it until large bodies of Africans invaded the colony and did serious damage. D'Urban then reasserted British military superiority and annexed the land up to the Great Kei River, from which most of the Africans were to be expelled (May 1835). But he found it impossible to enforce the expulsion, and the colonial secretary, Lord Glenelg, disapproved; consequently, in 1836 the territory was disannexed and an attempt was made to pacify the tribes by making treaties with their chiefs. By that time, however, the frontier was no longer a clear line of division between the races; white traders and missionaries were working among the tribes to the east of it, and Africans were working on white farms to the west.

3. The Great Trek.—For several generations the trekboers had been left more or less to their own devices. Now, under an alien government actuated by liberal ideals which they did not comprehend, many of their institutions were being transformed. They were short of labour and they found it difficult to comply with the regulations and get the compensation money for their slaves; their property was being pilfered by wandering Hottentots and they suffered further losses during the African invasion in 1834. Behind all these things they saw the hand of missionaries, who, they considered, misrepresented them in England. The reversal of D'Urban's frontier policy was the last straw that made many of them decide to throw off British dust and find a new home where they could, in the words of Piet Retief (*q.v.*), their ablest leader, "preserve proper relations between master and servant." Later, Anna Steenkamp recorded that the British had placed their slaves "on an equal footing with Christians, contrary to the laws of God and the natural distinctions of race and religion, so that it was intolerable for any decent Christian to bow down beneath such a

yoke; wherefore we withdrew in order thus to preserve our doctrines in purity." Imbued with this spirit the *voortrekkers*, about 12,000 Afrikaners, men, women, and children, left the colony with their sheep, their cattle, their ox wagons, and their Coloured servants between 1835 and 1843.

In the previous two decades there had been widespread destruction and dislocation among the African tribes, caused by the Zulu under Shaka, and their offshoot, the Ndebele or Matabele, under Mzilikazi (Moselekatse). (*See* ZULULAND.) By the time of the Great Trek two areas seemed suitable for white settlement—Natal, south of the Tugela River, and the High Veld, on either side of the Vaal. But appearances were deceptive, for these two areas were kept denuded by Zulu and Ndebele impis (regiments). The first *voortrekkers* made for the High Veld, where they were attacked by the Ndebele, whom they routed in 1837. Most of them then decided to settle in Natal. Piet Retief and a pioneer group went to the kraal of Dingane (Dingaan), Shaka's successor, to negotiate for a cession of land; and there, in February 1838, he and 70 Boers and 30 Coloured servants were treacherously massacred, while another 300 Boers, mainly women and children, and 200 Coloured servants, were killed in upper Natal. On Dec. 16, 1838, however, Andries Pretorius gained a decisive victory over the Zulu at Blood River. Most of the *voortrekkers* then proceeded to stake out farms in Natal. Border troubles ensued, and when in 1840 a Boer commando expedition made a sortie to the south against the kraals of a chief named Ncaphayi (Ncapaai), his neighbour Faku, who had a treaty with the British, appealed for aid.

Previously the British had been undecided what to do about the *voortrekkers*; the Cape of Good Hope Punishment Act of 1836 had made them liable, as British subjects, to trial in British courts for crimes committed south of latitude 25° S, but little had been done to make it effective. Then, however, anxious lest the activities of Natal commandos lead to disturbances on the Cape frontier, the governor, Sir George Napier, sent a small force to Natal, which, after some fighting, became British territory in 1843. Most of the *voortrekkers* who had settled in Natal then inspanned their oxen again and trekked back to the High Veld, north of the Vaal River, where they proved to be immune from further British attention for a generation.

Between the Orange and the Vaal, however, the British were drawn in for a while. There was a racial medley in that region: Griquas (Hottentot-European half-castes), who had lived north of the Orange River for several decades; Africans, survivors of the Ndebele devastation of the High Veld, who had been rallied by Moshesh (*q.v.*, or Moshweshwe), a Suto, or Sotho chief, near the Caledon River; and Afrikaners. The British had made a treaty with the West Griqua chief, Andries Waterboer, in 1834, and they made treaties with the East Griqua chief, Adam Kok III, and with Moshesh in 1843, hoping to use them to keep the peace along the Orange River. Trouble followed when Kok, in accordance with his treaty, tried to hand an Afrikaner over to the Cape authorities for trial under the Punishment Act. The Afrikaners took up arms and were dispersed by troops from the Cape at Zwartkoppies (1845). Two years later Gen. Sir Harry Smith became governor of the Cape Colony and British high commissioner in South Africa. An optimist, he believed that the South African imbroglio would be pacified if the area of British authority was extended. Accordingly, after overcoming further African resistance on the eastern frontier of the Cape, he annexed the land up to the Great Kei River as the province of British Kaffraria (1847) and the land between the Orange and the Vaal as the Orange River Sovereignty (1848). The colonial secretary, Earl Grey, accepted these annexations most grudgingly.

Humanitarian influence was waning in Britain and the main concern of the Colonial Office was to cut down expenses. Therefore, when further fighting broke out on the eastern frontier and similar difficulties developed with Moshesh in the Orange River Sovereignty, the British government decided to reduce its commitments in South Africa by recognizing the independence of the *voortrekkers*. This was done in two stages. In 1852 the Transvaalers were granted "the right to manage their own affairs without any interference on the part of the British government" in a convention signed at the Sand River; and in 1854 similar arrangements were made with the Afrikaners of the Vaal-Orange area at Bloemfontein.

South Africa was thus divided into three camps: Afrikaner republics; British colonies; and still-independent African societies.

4. Afrikaner Republics.—At the time of the Bloemfontein convention there were about 15,000 white persons engaged in pastoral farming in what then became the Orange Free State. They soon agreed on a rigid constitution, which vested legislative power in an elected council (*Volksraad*), executive power in an elected president—the voters being the white male adults—and local administration in landdrosts and heemraden on the old Cape model. The East Griqua question was settled in 1861 when Adam Kok sold out his remaining land rights in the Free State and trekked with his people to found a new Griqualand East nearer the coast (*see* GRIQUALAND EAST AND GRIQUALAND WEST). The Suto question was more difficult. The British had abandoned the sovereignty because they shirked the expense of controlling the brisk rivalry between the Afrikaners and the Suto for the Caledon River valley, which included some of the best arable land in southern Africa.

When fighting broke out in 1858 the Suto outmaneuvered the Free Staters, who were glad to accept the mediation of the high commissioner Sir George Grey (first convention of Aliwal North). The Free Staters, conscious of their weakness, then tried to unite with the Transvaal but were stopped by Grey, who threatened to withdraw the recognition of their independence and advised them to federate with the Cape Colony; yet when they expressed a desire to do so, the British government refused. Grey considered that a federation, under the crown, of all the colonies and republics in South Africa would be in the general interest. The British government, however, was averse to any extension of its area of responsibility. Further fighting started in 1864; this time the Free Staters were more successful and were able to dictate stringent terms, by which Moshesh lost most of his arable land (Treaty of Thaba Bosiu, 1866). But the peace was not kept and Moshesh asked Britain to annex Basutoland. This was done in 1868 and the next year the frontier line between the Free State and Basutoland was agreed upon (second convention of Aliwal North; *see* also LESOTHO).

The intervention of the British preserved for the Suto people more land than had been left to them in 1866; but it gave serious offense to the Free Staters, depriving them of the fruits of their victory. Thereafter the Free State, its frontiers pacified and absolved from the responsibility of administering a large number of Africans, became what Lord Bryce once described as a model republic, under the wise leadership of J. H. Brand, president from 1864 to 1888 (*see* BRAND, SIR JOHANNES HENRICUS). The Cape Colony took over Basutoland in 1871, but later got into difficulties in trying to force the Suto to forfeit their arms and was glad to hand it back to Great Britain in 1884.

Meanwhile the *voortrekkers* north of the Vaal River, numbering perhaps 22,000 at the time of the Sand River Convention, preserved their old trekboer mode of life. It was not until 1857 that most of them overcame their factional disputes and united to form the South African Republic under the presidency of M. W. Pretorius, son of the victor of Blood River (*see* PRETORIUS). Their constitution, which created institutions similar to those in the Free State, frankly stated that "The people desire to permit no equality between coloured people and the white inhabitants, either in church or state." Otherwise it was an ambiguous document, and from the first the *Volksraad* ignored any restraint but that of force. In 1863 there was further factional fighting; and Boer expansion frequently led to disputes with African tribes, the weaker of which were broken up, the remnants becoming squatters and servants under white landowners.

5. British Colonies After the Great Trek.—Not more than a quarter of the Afrikaners had left the Cape Colony during the Great Trek, though others went north in later years. Those who remained gradually became accustomed to the more equal order, and in the course of time a liberal tradition developed in the colony, in contrast with the illiberal tradition of the republics.

When the colony received representative government in 1853, the only franchise qualifications were low economic ones, irrespective of race or colour; but most of the Coloured people were unable to qualify and the vast majority of the electorate were white men. By that time the loss of population had been offset by further British immigration; and the decline in viticulture caused by the removal of British tariff preferences was remedied by an increase in the production of wool for export.

In British Kaffraria the high commissioner, Sir George Grey, tried to achieve quick results from a policy of civilizing the Africans by encouraging trade, mission schools, medical services, and public works, by bringing in white settlers, and by using white magistrates to undermine the authority of the chiefs. There was a reaction in 1857, when the Africans destroyed their own stock and crops in the superstitious belief that the white man would then vanish from the country. The result was catastrophic: thousands starved to death and emaciated survivors poured into Cape Colony in search of work. In 1865 British Kaffraria was incorporated in the colony. Soon afterward the British government urged Cape Colony to assume full responsibility for its own affairs, and in 1872 a responsible government bill was carried and J. C. Molteno became first prime minister.

British immigrants had settled in Natal, so that its white population became predominantly British and crept up to 9,000 in 1856, when representative institutions were established, and to 50,000 in 1893, when responsible government was obtained. Mixed farming was practised inland and sugar was grown in the subtropical coastal belt.

One result of the defeat of Dingane was an influx of Africans from Zululand into Natal, many of them returning to the region from which they had formerly been evicted by Shaka. Thus the African population rose steeply from about 20,000 in 1840 to nearly 500,000 in 1893. They were handled by Sir Theophilus Shepstone (*q.v.*), the son of a missionary who had come to South Africa with the settlers in 1820. He placed most of the Africans in a number of reserves or locations, comprising about one-seventh of the total area of the colony. Their tribal laws and customs were maintained and they were able to preserve their tribal economy for another generation, when they began to suffer from land shortage. Africans subject to tribal law were not eligible for the franchise, but in 1864–65 laws were passed which enabled Africans to apply to be exempted from tribal law and, seven years later, for the franchise. However, the regulations were such that few became exempted and hardly any became voters.

The colonists were critical of Shepstone's policy, mainly because it did not make the Africans work for them. The coastal sugar planters, in particular, were short of labour, and it was to help them that Natal began to import Indian labourers in 1860. By 1866, 6,000 had arrived, and when their five-year contracts expired most of them chose to stay in Natal, as they were entitled to do. In 1872 the Indian government prohibited further emigration because some of the Indians had complained of ill treatment; but the migration was resumed on a large scale, with a heavy government subsidy, in 1874, and by the end of the century the Indians outnumbered the whites in Natal.

C. Through War to Union (1870–1910)

1. Diamonds and Gold.—After 1870 the South African economy was rapidly transformed. Previously South Africa had been an economic backwater, with most of its inhabitants, whites and nonwhites, engaged in inefficient near-subsistence farming; and it had lacked the means to attract the capital and the skilled personnel to create the facilities of a modern country. In 1867 alluvial diamonds were found along the Orange and the Vaal rivers; in 1870 the dry diamond diggings began to be worked at Kimberley; and in 1886 the gold rush to the Witwatersrand began to eclipse all previous gold rushes. The effects were remarkable. The sum of South Africa's imports and exports rose from £49,000,000 in the decade 1861–70 to £128,000,000 in 1871–80, to £179,000,000 in 1881–90, to £357,000,000 in 1891–1900, and to £700,000,000 in 1901–10. In the last of these decades gold (59%) and diamonds (19%) accounted for 78% of South Africa's exports; and by 1910, £121,000,000 had been invested in the Witwatersrand gold-mining industry and dividends totaling £77,000,000 had been paid by it.

Immigration raised the white population from about 328,000 in 1875 to 1,117,000 in 1904. Railway construction was even more striking: in 1870 there were only 69 mi. (111 km.) of railway track in all South Africa; by 1886 there were 1,800 mi. (2,900 km.) and the line from Cape Town had reached Kimberley; by 1895 there were 3,600 mi. (5,800 km.) and Johannesburg was connected with five ports—Cape Town, Port Elizabeth, East London, Durban, and Lourenço Marques (Portuguese East Africa).

Thus South Africa at last entered the world economy and acquired a special place in it as the source of about a third of the world's annual supply of gold and of more than half its diamonds.

The control of these great industries became concentrated in a few hands. The digger phase on the diamond fields was short-lived; companies forced out individuals and then the companies amalgamated until in 1891 De Beers Consolidated Mines controlled the entire South African diamond industry. The digger phase on the gold fields was even shorter, but there the process of concentration never reached quite the same finality; companies became organized into groups or corporations, and the groups became associated in the Chamber of Mines for certain purposes only, notably for the recruitment of African labour. Thus great financial power came to be exercised by a few men, above all by Cecil Rhodes, who went to South Africa for his health in 1870 and became the most successful of all the mining financiers, controlling both De Beers and the Consolidated Gold Fields, one of the strongest gold corporations (*see* also Rhodes, Cecil John).

The skilled work in the mining industries was done by whites, mainly immigrants at first, who drew high wages. The unskilled work was done by low-paid Africans who went to the mines from all over southern Africa, especially from Portuguese East Africa, to work for limited periods during which they lived in closed compounds, separated from the ordinary life of the community. In 1898, for example, the Witwatersrand gold-mining industry employed 9,476 whites at an average monthly wage of £26 and 67,797 Africans at an average monthly wage of £2 9*s*. Previously many Africans had been employed as farm labourers in return for squatting rights or low wages. Now the agrarian pattern of race relationships was being adapted to the mining industries, which, having a voracious appetite for labour, were giving a distinct impetus to the long-drawn-out process of the disintegration of tribalism.

The discovery of diamonds led to a serious controversy. The Transvaal claimed part of the diamondiferous area and the Orange Free State claimed the rest of it, while David Arnot, a white attorney, claimed it all on behalf of his client Nicholas Waterboer, the chief of the West Griquas. Arnot managed to persuade the British Colonial Office that the Griquas had the best case; therefore when a new high commissioner, Sir Henry Barkly, reached South Africa in January 1871 he had instructions to take a strong line. Barkly persuaded M. W. Pretorius, the Transvaal president, to submit his case to arbitration by R. W. Keate, the governor of Natal. When Keate discovered that Pretorius had unwittingly placed great reliance on a forged document, he decided against him. Although this award did not directly affect the Free State claims, which were the strongest, Barkly, failing to understand their merits, annexed to the crown the entire area claimed for Waterboer in the same month (October 1871). Five years later a land court decision exposed the fallacies in Waterboer's claim to sovereignty over the diamond fields; whereupon J. H. Brand, the Free State president, went to London and exacted £90,000 compensation from the imperial government. In 1880 Griqualand West was incorporated in the Cape Colony.

2. Annexation and Retrocession of the Transvaal.—Lord Carnarvon, colonial secretary in Disraeli's 1874 ministry, hoped that the South African states and colonies might unite in a self-governing federation under the British crown. The following year he summoned a conference to discuss the project in London, but it was a failure. The Free State, smarting under the recent annexations, would have nothing to do with the project; nor would

THE BETTMANN ARCHIVE

ZULU WARRIORS IN BATTLE DRESS (1888)

the Cape, whose ministers' dignity had been ruffled by the conduct of the historian J. A. Froude, who had toured the colony on behalf of Carnarvon urging federation. Federation from the south having failed, Carnarvon turned his attention to the north and found that the South African Republic was ripe for plucking.

That republic, unlike the Free State, had never acquired stability. Pretorius was bundled out of the presidency after the annexation of Griqualand West, and the Transvaalers chose as his successor Thomas François Burgers (*q.v.*), a Cape colonial preacher of the Dutch Reformed Church. But in Burgers they had more than they had reckoned for; he was critical of Transvaal backwardness and proposed to remedy it by social, educational, and religious reforms. He also borrowed money, intending to build a railway to Delagoa Bay to free the Transvaal from its dependence upon British trade routes; but not a single track of the railway was laid, whereas the loan strained the meagre resources of the republic and alarmed the Cape and Natal merchants whose interests it threatened. When in 1876 Burgers personally waged an unsuccessful campaign against the Pedi (one of the Sotho peoples) in the northern Transvaal, he was totally discredited. The Afrikaners, including Paul Kruger, were plotting against him, and the traders were conniving with the Cape and Natal merchants to bring in the British. Carnarvon heard their cries and commissioned Shepstone to annex the Transvaal. Entering the republic in January 1877 with 25 police and a staff of seven, Shepstone annexed it in April in a proclamation which promised "the fullest legislative privileges compatible with the circumstances of the country and the intelligence of the people."

But British rule in the Transvaal was neither efficient nor tactful and the elected legislature did not materialize. Then the Zulu War revealed the weakness of British arms: the Zulu, their martial spirit revived by Cetshwayo (*q.v.*), routed a large British force at Isandhlwana in 1879 before they were vanquished in the following year (*see* ZULULAND). By that time the Transvaalers, who had passively acquiesced in the annexation, were moving toward open rebellion under the leadership of Paul Kruger, Petrus Jacobus Joubert (*qq.v.*), and M. W. Pretorius, and in 1881 they, too, wiped out a British force at Majuba. Gladstone's ministry, which assumed office in Britain in 1880 but failed to reform the Transvaal administration in time to avert the rebellion, then gave the Transvaalers a qualified independence (Pretoria Convention, 1881). Three years later they allowed the republic full internal autonomy, while maintaining British control over its relations with foreign states, other than the Free State (London Convention, 1884). Carnarvon's project had completely miscarried, and his permissive act for the unification of South Africa (1877) was stillborn.

3. The Rhodes-Hofmeyr Alliance.—In the Cape, which remained the most civilized of the South African states, the advent of responsible government and the series of British annexations prompted a group of Afrikaners to form a political organization, the Afrikaner Bond, in 1879. At first the Bond's program was crudely sectional, but three years later Jan Hendrik Hofmeyr (*q.v.*) gained control and turned it in the direction of Anglo-Afrikaner cooperation, with a British South African federation, which should include the republics, as the ultimate goal. Meanwhile Cecil Rhodes had entered the Cape Parliament and, as his vision of British expansion in Africa was reconcilable with Hofmeyr's, he was able to form a ministry with Bond support in 1890.

Meanwhile, although it had washed its hands of Basutoland in 1884, the Cape Colony by 1894 had annexed all the land below the mountain escarpment as far east as the Natal border at the Umzimkulu River. The Nguni inhabitants of these Transkeian territories remained in occupation of most of their lands after their political autonomy had been extinguished; but steps were taken to prevent them exercising much influence over the colonial Parliament. Laws were passed providing that land occupied on tribal tenure should not satisfy the economic requirement for the franchise (1887) and raising the economic qualification and introducing a simple educational test (1892). A start was also made in encouraging African communities to adopt individual land tenure and to elect local administrative councils (1894). Thus the Cape government drew a line between "tribal" and "civilized" Africans, treating tribal Africans as minors and giving civilized Africans the same rights as white men in theory, though in practice not many Africans got the vote. By 1895, therefore, the Rhodes-Hofmeyr alliance was bidding fair to heal Anglo-Afrikaner tensions in the Cape Colony, without destroying the prospect of a constructive relationship with Africans.

4. British Clash with the Transvaal.—During the late 19th century the Bantu-speaking tribes of South Africa which had previously maintained their autonomy north of the Vaal River came under white control as a result of Boer expansion and the intervention of Germany and Britain. In 1884 Germany took over South West Africa and Transvaalers began to encroach westward toward it, setting up the republics of Stellaland and Goshen. This expansion, which seemed likely to cut off the British from access to central Africa via the Cape, was checked, largely on the insistence of Rhodes, in 1885, when Great Britain, after a display of force, proclaimed a protectorate over northern Bechuanaland and annexed southern Bechuanaland as a crown colony. Transvaal efforts to keep their north clear of the British were also foiled by Rhodes, who secured a prospecting concession from the Ndebele chief Lobengula in 1888, obtained a royal charter incorporating the British South Africa Company to exploit the concession in 1889, and dispatched a pioneer expedition to occupy what became known as Rhodesia in 1890 (*see* RHODESIA: *History*). Finally, the Transvaalers tried to push toward the east in order to get a port on the Indian Ocean, but, although they incorporated fresh territory, including Swaziland, they did not reach the coast, as Britain annexed Zululand in 1887 and Tongaland, the last gap, in 1895. The Afrikaner republics were thus encircled by the colonies of Great Britain and its oldest ally, Portugal—hemmed in a kraal, as Paul Kruger complained. Southern Bechuanaland was taken in by the Cape Colony in 1895; Zululand and Tongaland by Natal in 1897; but northern Bechuanaland remained a British protectorate.

In 1883 Paul Kruger was elected president of the Transvaal. His character had been molded in the hard school of the Great Trek and of commando fighting against Afrikaner factions and African tribes; and recently he had led the opposition to President Burgers and the British. His policy was to regain complete independence for the republic and to preserve Afrikaner control of it. In 1885 he was constrained to suggest a customs union with the Cape, whose government rejected the offer. Soon afterward his financial difficulties were solved by the gold discoveries and he proceeded to use the mining industry as a milch cow, imposing heavy taxation and granting monopolies of essential materials, such as dynamite.

The growth of such an industry in the heart of his republic, however, posed serious problems. The Afrikaners became outnumbered by the newcomers who poured in from all parts of the world, most of them British subjects. To give these *uitlanders* ("outlanders"), as they were called, an effective vote seemed like

suicide; to withhold such a vote was bound to cause trouble with the British. Kruger and the *Volksraad* never hesitated: they progressively raised the franchise qualifications until by 1894 no *uitlander* could vote in presidential elections, and only those who were 40 years old and had lived 14 years in the Transvaal, during 12 of which they had been subject to an oath of allegiance, could vote in *Volksraad* elections. In short, on the one hand the government got most of its revenue from the *uitlanders* and denied them political rights; on the other hand the *uitlanders* continued to swarm into the republic where many of them prospered exceedingly. Such a situation required patience. Although Kruger was reelected for a third term in 1893, his majority over the more moderate P. J. Joubert was small and a change seemed likely in 1898.

But Rhodes was not patient. Finding that Kruger stood between him and a British South African federation, he planned to overthrow him. Kruger nearly played into Rhodes's hands in 1895 when, contrary to the London Convention, he tried to injure Cape trade with the Transvaal and the Cape railways that were competing with the Delagoa Bay line by making the rates over the Transvaal sector of the Cape line prohibitive and by closing the Vaal drifts to prevent the carriage of goods by wagon. But he gave way when the colonial secretary, Joseph Chamberlain, agreed to support the Cape ministry's protests, by force if necessary.

Rhodes then pushed ahead with his conspiracy: the *uitlanders* were to rise, an armed force under Leander (afterward Sir Leander) Starr Jameson (*q.v.*) was to go to their assistance, and the high commissioner was to hurry to Pretoria to "restore order." Chamberlain knew of Rhodes's plans when he provided a suitable jumping-off place for Jameson's force by ceding to the British South Africa Company a strip of land along the Transvaal border of the Bechuanaland Protectorate, but he later withdrew his support when he learned that the *uitlander* rising would not take place to give the invasion an air of respectability. Therefore, when Jameson, contrary to Rhodes's last-minute instructions, rode into the Transvaal from Pitsani near Mafeking on Dec. 29, 1895, he was disowned by the colonial secretary and the high commissioner. Four days later he surrendered to Transvaal commandos at Doornkop.

5. Outbreak of the South African War.—The raid fiasco cleared the decks for the South African War. Rhodes had to resign the premiership of the Cape Colony, where, in 1898, W. P. Schreiner became head of a ministry which relied upon the support of the Afrikaner Bond, now deeply suspicious of British designs. In the Orange Free State, which had opposed the Transvaal over the drifts crisis, M. T. Steyn was elected president and formed a military alliance with the Transvaal. While Steyn and the Bond were not uncritical of the Kruger regime, Kruger's Anglophobia seemed to have been vindicated in the Transvaal, where he won the 1898 presidential election by a large majority. In England, on the other hand, a parliamentary committee failed to probe Chamberlain's dealings with the raiders, a telegram from the German emperor congratulating Kruger turned Jameson and Rhodes into popular heroes, and the press paid much attention to *uitlander* grievances.

As the tension mounted, Chamberlain, taking the high ground of British paramountcy in South Africa, claimed to be competent not merely to control the foreign relations of the Transvaal (a right expressly granted by the London Convention) but also to intervene on behalf of the *uitlanders* (which was not). If the right to intervene existed, there were grounds enough for exercising it. The tendency of the *Volksraad* to give sweeping powers to the president was exemplified in the judges' crisis: in 1897 Chief Justice John (afterward Sir John) Gilbert Kotze delivered a judgment which meant that the greater part of the laws of the Transvaal had been enacted unconstitutionally and were null and void, whereupon the *Volksraad* deprived the courts of the testing power and gave Kruger the power to dismiss any judge who disagreed; and the next year Kruger dismissed Kotze. In 1898 an Anglo-German agreement removed the likelihood of German intervention in the event of war in South Africa, and Chamberlain and the high commissioner, Sir Alfred (afterward Viscount) Milner, proceeded to apply the screw. In so doing they brushed aside several attempts by the Cape government and the Afrikaner Bond to avert a catastrophe.

The central issue was the Transvaal franchise. In June 1899, in conference at Bloemfontein, Milner proposed a simple five years' franchise, Kruger refused, and the negotiations collapsed. The last chance of peace was lost in August when Jan Christiaan Smuts (*q.v.*), Kruger's young state attorney, offered a five years' franchise and Chamberlain rejected the conditions that were attached to the offer. Milner had long since made up his mind that war was "inevitable," and in September Chamberlain arranged for 10,000 British troops to be sent to augment the meagre British forces in South Africa. The republicans, who had been arming furiously for several years, replied by issuing an ultimatum which expired on Oct. 11, 1899. (For events of the war and for the Treaty of Vereeniging that ended it, *see* SOUTH AFRICAN WAR.)

6. Reconstruction and Union.—After the peace of Vereeniging, Milner, high commissioner and governor of the crown colonies —the Transvaal and the Orange River Colony—concentrated on their material reconstruction from the ravages of war. The gold mines resumed large-scale operations, and the Afrikaners were returned to their land from the prison and concentration camps and supplied with food, stock, seed, and implements, no less than £10,000,000 being spent by Great Britain on their rehabilitation. Milner also prepared the ground for political union by bringing all the colonies into a South African customs union and by amalgamating the railways of the Transvaal and the Orange River Colony.

All this work was well done by Milner and his band of able young men, who were dubbed the "kindergarten." But the opportunity was not used to emancipate the nonwhite inhabitants of the crown colonies from their legal disabilities, which remained much as they had been during the republican regime. Nor did Milner gain the confidence of the Afrikaners: they disliked him as "the man who made the war," who openly supported an agitation for the suspension of the Cape constitution in 1902, who imposed an English educational system on the crown colonies, and who was responsible for bringing Chinese labourers to the Witwatersrand gold mines in 1904. For the importation of Chinese labourers there were, indeed, strong economic reasons, because the supply of African mine labour had temporarily fallen off; but in Britain the

(ABOVE) BROWN BROTHERS; (RIGHT) THE BETTMANN ARCHIVE

The South African War: (Above) Boer combatants in the field. (Right) British artillery contingent outfitted with an observation balloon en route to Johannesburg, 1901

decision played a large part in the fall of A. J. Balfour's Conservative government in December 1905, and the change of government was followed by a radical change in the distribution of power in South Africa.

Sir Henry Campbell-Bannerman's Liberal ministry, anxious to make amends to the Afrikaners for a war which had been brought about by its predecessors, decided to give the new colonies responsible government, with exclusively white electorates. The Liberals defended this decision by referring to the Treaty of Vereeniging, but it did not, in fact, debar them from allowing Coloured and Asian people to vote. Consequently in March 1907 Louis Botha (*q.v.*), former commandant general of the forces of the South African Republic, became premier of the Transvaal, with Smuts as his right-hand man, and in December 1907 A. Fischer became premier of the Orange River Colony. The swing of the pendulum was finally completed in February 1908, when J. X. Merriman's South African Party, consisting largely of members of the Afrikaner Bond, ousted Jameson's Progressives from office in the Cape Colony.

The high commissioner, Lord Selborne, who succeeded Milner in 1905, argued the case for union in a memorandum published in July 1907, and, once in power, the three Afrikaner parties took it up enthusiastically, with the support of their oppositions. White public opinion in South Africa moved steadily in favour of union, partly out of an idealistic feeling that it was best to bury the hatchet and make a new start on a basis of white equality, and partly because an African rebellion in Natal in 1906 led many to think union essential for white security; some, moreover, saw in union the best way of guarding against any further British interference in South African affairs. There were also compelling economic reasons for union: the four colonies were interdependent and yet, without political union, their material interests were so divergent that the customs union seemed almost certain to collapse.

Accordingly a national convention, comprising 30 members appointed by the four colonial parliaments and 3 nonvoting members from Rhodesia, met in 1908 and 1909 under the chairmanship of Sir (afterward Baron) John Henry de Villiers, chief justice of the Cape. A constitution was drafted and unanimously approved by the convention and it was carried with scarcely any dissentients in the parliaments of the Cape Colony, the Transvaal, and the Orange River Colony, and by a three-to-one majority of the electorate in a referendum in Natal. It was enacted by the British Parliament in September 1909, substantially as it had been submitted to the British government by delegates from South Africa. The South Africa Act came into force on May 31, 1910, and the four colonies became the provinces of the Union of South Africa (*see* also *Constitution*, below).

The convention, except for the Natal members, wanted a close union and complete flexibility, and they were largely successful; the principal feature of the South African constitution was the grant of power to Parliament to legislate on practically every subject by simple majorities in each house. Nevertheless, on two subjects they did not grant Parliament such extensive powers. First, the Afrikaner members wanted special protection for the section which gave Dutch (to which Afrikaans was added in 1925) equal status with English as an official language of the Union. Second, the convention could not agree on a uniform franchise for the Union, the Cape delegates favouring a colour-blind franchise and the others a rigid colour bar. Eventually a compromise was reached whereby the franchise qualifications were to remain as they were in each province and the established franchise rights of the Cape nonwhites were specially protected. Accordingly, the sections on language equality and the Cape nonwhite franchise were to be amended only with the approval of two-thirds of the members of both houses of Parliament in a joint sitting. The act also made it possible for Rhodesia to join the Union on terms to be approved by the Privy Council and for the High Commission Territories of Basutoland, the Bechuanaland Protectorate, and Swaziland to join on terms laid down in a schedule, but only with the consent of the imperial government, which had special commitments to their African inhabitants.

D. The Union and the Republic

1. The Race Issue.—The dominating factor in the history of South Africa has been the exceptionally complex character of the population. In 1910 there were nearly 6,000,000 inhabitants, of whom 21.5% were whites, 67% Africans, 9% Coloured people, and 2.5% Asians. By 1960 there were 16,000,000, of whom 19.3% were whites, 68.3% Africans, 9.4% Coloured people, and 3% Asians. In 1910 the whites included an Afrikaner majority, mainly rural, and a British minority, mainly urban. Between them they owned most of the land and the capital, did most of the skilled work, and possessed 93% of the votes.

The Africans were still predominantly tribal rather than modern in culture, but, as a result of their conquest and of white missionary activity and economic enterprise, there was already a distinct trend toward the disintegration of tribalism. Although most Africans still had homes of sorts in the reserves, those were scattered lands which amounted to only one-fourteenth of the area of the country and by no means provided them all with a livelihood; consequently, many Africans went out to earn wages on white farms and in the towns. Moreover about 1,000,000 Africans had long been established as squatters on white farms and a few were already completely urbanized. The Africans were subject to pass laws which restricted their movements outside the reserves, to Masters and Servants laws which made breach of contract a criminal offense in many types of employment, to special taxation, and to other special laws. These laws varied from province to province, and in the Cape about 7,000 African voters were exempted from some of them.

The Coloured people were Western in culture. Most of them lived in the Cape Province, where their legal status was that of the whites, and about 14,000 were voters, whereas in the other provinces they were subject to discrimination.

The Asians were mainly the product of the indentured Indian immigration scheme to Natal, where they outnumbered the whites and were subject to special laws, including a £3 tax. They were excluded from the Orange Free State, but about 11,000 had settled in the Transvaal, where, too, they came under special laws.

The policies of South African governments differed on many important issues; but all, being responsible to a predominantly white electorate, stood more or less explicitly for "the maintenance of white supremacy," a task which became increasingly difficult as a result of the rapid tempo of economic development in South Africa and the change in the balance of power and the climate of public opinion elsewhere.

2. The Botha-Smuts Regime, 1910–24.—In 1911 the Afrikaner parties of the former colonies merged to form the South African Party, which, under Botha and Smuts, governed the country until 1924. Botha and Smuts had been zealous fighters for republican independence so long as there was a chance of success, but they were realistic enough to accept the inevitable at Vereeniging and imaginative enough to see great prospects for South Africa when the hot mood of British imperialism had spent itself. By 1910 they hoped for the coalescing of the British and Afrikaner elements into a white South African nation which would freely cooperate with the British Commonwealth in peace and in war. The ideal went too far, however, for the more conservative Afrikaners, who were concerned with preserving their group identity, and not far enough for the more self-conscious nonwhites, for whom it implied a permanently inferior status.

The government wished to keep most of the Africans in the reserves to prevent the whites from being swamped, and to use African manpower as the unskilled base of all forms of economic enterprise—objectives which involved certain contradictions. As the Africans became more accustomed to a money economy, some of them were liable, unless checked by law, to compete successfully with the less efficient whites. This was already happening in the gold-mining industry, and laws were passed in 1911 to preserve the racial hierarchy in that industry. Such competition was also happening in some rural areas, where Africans were pooling their resources to buy more land; a Natives Land Act was passed in 1913 to limit the areas within which such purchases could be made. Neither of these acts was wholly effective. In 1916 a

commission reported that if the policy of territorial segregation was to be carried out the reserves should be substantially increased in size, but little was done about it because the whites were not prepared to make the necessary sacrifices.

The Asian question had reached a more crucial phase. Already before union Mohandas Gandhi (*q.v.*) had organized nonviolent resistance against the Transvaal government, and after union he resumed the struggle against the £3 tax in Natal. As a result the Indian government stopped the importation of indentured Indian labourers in 1911 and the £3 tax was removed in 1914. By that time the pattern of later disputes had developed: the Union government regarded the Indians as temporary visitors and tried to persuade them to return to India; and most of the Indians regarded themselves as permanent residents and demanded full rights of citizenship.

The conservative Afrikaners had found their leader in J. B. M. Hertzog (*q.v.*). Although he had joined the Botha cabinet in 1910, Hertzog regarded his colleagues' policy as liable to cause the Anglicization of the Afrikaner people, and he publicly advocated a different, "twin-stream" policy, which led to his exclusion from a reconstituted cabinet in 1912 and to the foundation of the Nationalist Party. When in 1914 the government unhesitatingly took the part of Great Britain in World War I and Parliament voted funds for the conquest of German South West Africa, a number of former republican generals, some of whom held appointments in the South African defense force, came out in rebellion to avenge Vereeniging. After mastering the revolt, Botha took command of the South West African expedition and forced the Germans to capitulate on July 9, 1915. South African volunteer contingents also fought in East Africa and on the Western Front in Europe and many individuals joined British units. Smuts, after serving as commander in chief in East Africa, did notable work in the British war cabinet. In July 1919 he and Botha, on behalf of South Africa, signed the Treaty of Versailles and the covenant of the League of Nations, under which South West Africa became a Union mandate. Back in South Africa Botha died before August was out, and Smuts, succeeding to the premiership, faced the discontents caused by thwarted Afrikaner nationalism, a steep rise in the cost of living, and industrial troubles.

As early as 1907 Smuts had intervened on the side of the employers in a strike of white workers on the Witwatersrand, and in 1914 he had ended another strike by declaring martial law and illegally deporting nine strike leaders. After the war there was a serious crisis. In December 1921 the Chamber of Mines, faced with rising costs and a fall in the price of gold, decided to organize the industry more rationally by using Africans for semiskilled work. There was a violent reaction by white labour, which Smuts suppressed at a cost of 230 lives. The result was that, although the threat to relax the colour bar was not carried out, the Labour Party, representing the aggrieved white workers, made an electoral pact with the Nationalists. In the hope of redressing the political balance Smuts wooed the Southern Rhodesian electorate to accept incorporation in the Union, but when a referendum was held on the issue in 1922 they preferred to remain separate. Two years later the South African Party was heavily defeated at the polls and Hertzog became premier of a Nationalist-Labour coalition.

3. Hertzog's Nationalist Policy, 1924–33.—Hertzog's main objectives were to complete the emancipation of South Africa from imperial control and to provide greater protection for the whites from the Africans and for the Afrikaners from the British. He played a notable part in the events leading to the Balfour Report (1926) and the Statute of Westminster (1931), which gave statutory definition to the established convention that the British government could not exert authority over a dominion. South Africa's sovereign status was also asserted by the adoption, after a long and bitter controversy, of a distinctive national flag (1927), by the appointment of ambassadors to Italy, the United States, and the Netherlands (1929), and by the separation of the office of governor-general, the head of the South African government, from that of high commissioner, the representative of the British government in South Africa (1931).

Economic nationalism was fostered by the foundation of a state-controlled iron and steel works at Pretoria, by increased protection for agriculture and industry, by a reduction in imperial preferences, and by a commercial treaty with Germany. White supremacy was bolstered by the provision of sheltered employment for "poor whites" in state enterprises; by a Mines and Works Amendment Act (1926), which was more effective than its predecessor in shutting Africans out from skilled mining trades; by a Native Administration Act (1927) and a Riotous Assemblies Act (1930), which gave the executive wide powers over individuals; and by franchise acts (1930–31), which extended the vote to all white men and women, while they left the Cape nonwhite vote as before, restricted to men who possessed property and educational qualifications. Hertzog's legislative program was still incomplete when the world depression undermined the prosperity of the country and the popularity of the government, which continued to cling to the gold standard after Great Britain had abandoned it in 1931. After the government, yielding to pressure from some of its own supporters, left the gold standard in December 1932, there was a fairly rapid economic recovery, but political confusion continued until, in 1933, Hertzog and Smuts formed a coalition government which secured overwhelming support from the electorate.

4. Hertzog-Smuts Regime, 1933–39.—The coalition was based on a great deal of common ground between Hertzog and Smuts. Hertzog no longer feared political cooperation with Smuts's British supporters, believing that the Afrikaners had been raised to the point where they were immune from the danger of Anglicization, and Smuts was reconciled to most of Hertzog's achievements and was prepared to collaborate with him in rounding them off. Consequently, in 1934 the two major parties fused to form the United Party; but Hertzog failed to carry with him a small group of Afrikaner irreconcilables, led by D. F. Malan (*q.v.*), who formed the Purified Nationalist Party, while C. F. Stallard at the other extreme dissociated himself from Smuts and formed the Dominion Party. By then the constitutional settlement had been completed by the Status Act and the Seals Act (1934) which secured, so far as words could do, that South Africa was master in its own house. The government then proceeded to complete Hertzog's segregation legislation. A Representation of Natives Act (1936) removed the Cape Province African voters from the common roll and gave them the right to elect three white members to represent them in the lower house, gave the Africans throughout the Union the right to elect four white senators, and created a Natives' Representative Council with advisory powers. A Native Trust and Land Act (1936) provided for a considerable increase in the size of the reserves. A Native Laws Amendment Act (1937) authorized the executive to prevent more Africans from entering the towns and to compel municipalities to segregate African from white residents. It was Hertzog's hope that the enlarged reserves would become capable of maintaining almost all the African people, so that those who worked for whites could be regarded as temporary visitors from the reserves. But it is notable that he saw the need for consultation with an African council, and that he did not try to discriminate against the Coloured people. Moreover, his government expanded the social services not only for the whites but also for the nonwhites; there was, for example, a considerable increase in government grants for education.

Nevertheless, Africans, Asians, and Coloured people were becoming disturbed by the great contrast between their living standards and those of the whites and by an accumulation of laws which caused 500,000 Africans to be convicted of statutory and municipal offenses in a year and many more to come into unpleasant contact with a police force recruited from among the poorer Afrikaners, who were imbued with the strongest race prejudices. As tension began to mount the Nationalists claimed that "white South Africa" would not be safe unless the restrictions on Africans were increased and the Coloured people were also segregated from the whites. Though Hertzog strongly disagreed, the Nationalists were able to profit from the celebrations held to mark the centenary of the Great Trek in 1938, when they appealed to all Afrikaners to remain true to the principles of the *voortrekkers* and of Paul Kruger, placing special emphasis on their race attitudes.

5. World War II.—The outbreak of World War II caused a

crisis in South Africa. Although Hertzog and Smuts had cooperated successfully on many issues, they differed widely on foreign affairs. Hertzog took an indulgent view of Nazi Germany, believed that its expansion was no concern of South Africa, and moved in the House of Assembly, on Sept. 4, 1939, that South Africa should remain neutral. Smuts, keenly aware of the wider implications of Naziism, maintained that it was the interest and the duty of South Africa to support Great Britain, and he won the crucial division by 80 votes to 67. On Sept. 5 Smuts formed a ministry with the support of the Labour Party, the Dominion Party, and the majority of the United Party, and war was declared on Germany.

The Smuts government concentrated its energies on the war. About 200,000 white men (more than half of them Afrikaners) and 125,000 nonwhites (mainly Africans and Coloured men) joined the forces, and many of them served with distinction in the Ethiopian, Mediterranean, and Madagascan theatres of war. The nonwhites were not allowed to bear arms but were distributed among the combatant units, for whom they performed vital work as stretcher-bearers, labourers, etc. Industry was efficiently switched to the production of munitions and clothing for military purposes. Smuts himself remained in close contact with Winston Churchill and watched carefully the strategy of the war, often leaving his able lieutenant, J.H. Hofmeyr, Jr. (1894–1948), to carry a lion's share of the burden of administration.

South Africa's achievements were impressive, considering the strength of the opposition to the war. Fortunately for the government, the opposition splintered into fragments during the most critical period: J. F. van Rensburg's Ossewabrandwag ("Guard of the Ox Wagon") and Oswald Pirow's New Order, accepting the racial doctrines of the Nazis and their contempt for parliamentary government, patently hoped to profit from a Nazi victory; the Malanites expounded the ideal of an Afrikaner republic in which, as in Kruger's republic, the British would not necessarily have political rights; while the Hertzogites stood firm by their principle of equality between the whites. In the first flush of his disappointment, indeed, Hertzog led his defeated minority from the United Party to join hands with Malan and form the Reunited Nationalist Party, but the reunion was not a happy one and before the end of 1940 Hertzog had been driven out for his tenderness toward the British. His colleague, N. C. Havenga, subsequently founded the Afrikaner Party to maintain his ideals.

The general election of 1943 was a victory for Smuts in that his prowar coalition secured a majority of 67 seats in a lower house of 153; but it was also in a sense a victory for Malan's Nationalists, who won all the 43 opposition seats. Thereafter, as the prospect of a Nazi victory faded, the Nationalists consolidated their position as the political instrument of self-conscious Afrikanerdom, absorbing elements from the Ossewabrandwag, the New Order, and the Afrikaner Party, and drawing support from a variety of cultural and economic organizations, including the Broederbond, a secret group of Afrikaner elite.

As the war drew to a close Smuts drafted the preamble to the United Nations Charter signed at San Francisco in 1945 and returned to mold postwar South Africa along the lines of a generous demobilization scheme, an expansion in the social services for all races, a planned development of agricultural, mineral, and industrial resources, and large-scale white immigration.

6. Industrial Expansion in South Africa.—Subsequent events can be understood only in relation to the fact that from about 1938 onward South Africa was experiencing an industrial expansion as intense as the British industrial revolution of the early 19th century, with the added complication of racial differences between the unskilled workers on the one hand and the skilled workers, the capitalists, and the majority of the electorate on the other. The way had already been paved by the rise of the mining industries and by the foundation of many manufacturing industries during World War I and the 1930s; the rate of expansion increased during and after World War II. In the 24 years between 1936 and 1960 the total population increased by 67%. In the 28 years between 1935 and 1963 the geographical national income at current prices increased more than eightfold (the value of mining production fivefold, of agricultural production sixfold, and of manufactures thirteenfold). In the same period the price index increased by two-and-a-half times; so that even allowing for the change in the value of money, the growth of manufacturing industries was remarkable.

The gold-mining industry continued to play an important role in the South African economy, especially in easing the balance of payments problem. Fears that the supply of gold would soon be exhausted were removed by the opening up of new mines on the Far West Rand and in the Orange Free State, where production started in 1951. Nevertheless, gold production did not increase as rapidly as other industries and the relative importance of goldmining decreased as the economy became more diversified. The mining industries continued to be organized on racial lines and the manufacturing industries adopted a similar basis. Skilled work in South Africa remained virtually a white monopoly and was well paid; most of the unskilled work continued to be performed by nonwhites for low wages.

As a result of this expansion more than four-fifths of the white population were townsmen by 1960, the "poor whites" were absorbed by industry, and the earlier economic distinctions between the Afrikaners and the British inhabitants became blurred. Industry also attracted many nonwhites. More than 3,500,000 Africans (nearly a third of the total African population) were in the towns at the time of the 1960 census. Into some of these towns the influx was so rapid that housing and other amenities were grossly inadequate and there was much crime. But in spite of low wages and squalid living conditions, the urban Africans earned far higher incomes than the rural Africans did.

7. The 1948 Election.—As the 1948 election approached, disturbing facts were becoming known in South Africa. Although the reserves had been enlarged under the 1936 act, only 40% of the African population was to be found in them at the time of the 1946 census. The Social and Economic Planning Council reported that the quality of the reserve lands was deteriorating; and early in 1948 a commission revealed that many Africans were becoming permanent inhabitants of the towns. These facts did not square with the argument that the Africans had adequate homes in the reserves and could therefore be treated as inferiors elsewhere. Moreover the Natives' Representative Council was frustrated because the government ignored its advice and the African National Congress, which had been cautious since its foundation in 1912, was now demanding the removal of discriminatory laws in terms of the ideals of the United Nations. This raised the question whether white supremacy, which had developed in the simple pastoral economy of the 18th century, could be maintained in the industrial economy of the 20th century.

The Smuts government continued to profess adherence to the segregation policy but tried to deal with practical grievances in a conciliatory spirit, while Hofmeyr went farther than his colleagues on occasion, feeling toward a policy which might be acceptable to all races. But Asians were protesting vigorously against legislation that prevented their buying or occupying new premises in Durban, and the Indian government attacked South Africa in the United Nations General Assembly. Consequently, when Smuts tried to persuade the assembly in 1946 to allow the Union to incorporate South West Africa, he met with a rebuff, and the assembly passed the first of a series of resolutions condemning racial discrimination in South Africa.

The Nationalist leaders wished to reduce South Africa's links with Britain, the Commonwealth, and the United Nations, to advance the power of the Afrikaner people, and, above all, to preserve white supremacy by every possible means. They expressed horror at the "liberalism" of the United Party, declared it would cause the "suicide" of white South Africa, and proclaimed a policy of apartheid (*q.v.*), according to which whites, Coloureds, Asians, and Africans would be separated from each other and each race would be able to "develop along its own lines in its own area." Though as far as the Africans were concerned this was little more than Hertzog's segregation policy in a new dress, it was presented to the electorate as something new and dynamic. The Nationalists also exploited the discontents caused by wartime controls and were

assisted by the fact that the electoral system favoured the rural voters. The result was that in May 1948 Malan, with the assistance of Havenga's small Afrikaner Party, was able to form a government, with a majority of five in the House of Assembly. This result, which was a surprise to Malan as well as to Smuts, was decisive for South Africa.

8. Nationalist Party Policy from 1948.—After the 1948 election the Nationalist Party consolidated its power, absorbing the Afrikaner Party and gaining strength in the House of Assembly in each election—from 73 seats in 1948, to 94 in 1953, to 103 in 1958, to 105 in 1961, and to 126 (in a larger House) in 1966. The process was promoted by legislation and administrative action: Smuts's immigration scheme was scrapped; British immigrants, like others, were required to wait several years before they could vote; the white people of South West Africa were given six seats in the House of Assembly; the voting age was lowered to 18; and the Coloured voters were given separate and limited parliamentary representation and the representatives of the Africans were removed from Parliament. Nationalist power was also promoted by the strong leadership of the prime ministers Malan (1948–54), J. G. Strijdom (*q.v.;* 1954–58), and H. F. Verwoerd (*q.v.;* 1958–66). The United Party, led by J. G. Strauss after the death of Smuts in 1950 and by Sir D. P. de Villiers Graaff from 1956, remained the main opposition party, but it lost seats at each election and suffered from secessions on both flanks. Several new parties were founded, including a Liberal Party and a Progressive Party, but they had little success at the polls.

The Nationalist governments enacted a mass of racial legislation to preserve white supremacy. They created a population register to fix the racial category of every South African. They made marriages and unions out of wedlock between whites and nonwhites unlawful. They systematically divided the towns as well as the rural areas into zones in which members of only one race could own or occupy property or conduct business. They assumed control over African and Coloured school education, eliminating the mission schools. They excluded nonwhites from the established universities and founded separate colleges—one each for Coloureds, for Asians, for Xhosa, for Zulu, and for Sotho. They intervened in the labour-union movement to separate white from nonwhite. They gave officials sweeping powers to remove "undesirable" Africans from towns. They extended the practice of reserving particular types of jobs for whites. They enforced segregation where it did not previously apply, as in buses, trains, post offices, libraries, motion-picture houses, and theatres in the Cape Peninsula. White and nonwhite South Africans could thus rarely meet, except as masters and servants or rulers and subjects.

Nonwhites were almost totally excluded from the authoritative political system. After a long struggle the Cape Coloured voters were removed from the common voters' roll and given the right to elect four whites to represent them in the House of Assembly. At first the government lacked the two-thirds majority laid down in the South Africa Act for such legislation. An act removing the Cape Coloured voters from the common roll was passed through Parliament by the ordinary simple-majority procedure in 1951, to be declared invalid by the Supreme Court. Under a further act the Nationalist members of Parliament then sat as a high court and reversed the judgment, but the Supreme Court held that the "high court of Parliament" was itself invalid. Eventually the Senate was enlarged and its composition changed to give the Nationalists 77 seats out of 89 in that house (1955). The Coloured voters were then removed from the common roll by the necessary two-thirds majority in a joint sitting of both houses, and the Supreme Court acquiesced (1956).

In 1959 the representatives of the Africans were removed from both houses of Parliament and the Cape Provincial Council. The House of Assembly then contained 156 whites elected by the white voters of the Union, later Republic, and South West Africa and four whites elected by the Cape Coloured voters. In 1966 the House of Assembly was enlarged to 166. The Senate, as again reconstituted in 1960, contained 43 whites elected by white electoral colleges and 11 nominated by the government. All provincial councillors were white people elected by white voters, except two in Cape Province who were elected by Coloureds. Thus the compromise embodied in Hertzog's 1936 legislation was destroyed. The only nonwhites who had any say in the composition of the Parliament, which was supreme over all South Africa, and in the principal subordinate bodies, the Provincial Councils, were the Coloured people of Cape Province, and their say was meagre.

Instead, the nonwhite peoples were given various types of segregated subordinate institutions. For the Africans there was a system of "Bantu authorities" in the Bantu areas, consisting mainly of chiefs, councillors, and their nominees. The chiefs were on the government's payroll and were dismissible by the government. In the Transkei, the largest Bantu area, a Legislative Assembly, with a majority of chiefs ex officio and a minority of members directly elected by Africans, and a cabinet and prime minister responsible to the assembly, were established in 1963. Such bodies had powers in defined fields, but subject to the overriding authority of the government of South Africa. The first prime minister of the Transkei was Chief Kaiser Matanzima, who had the support of the chiefs in the assembly but not of the elected members, and who endorsed the principle of racial separation. In the towns some Africans could elect members of Bantu urban councils. A Coloured Persons Representative Council with an elected majority was established in 1964; and a similar body was envisaged for Asians. The Bantu urban councils and the Coloured Persons Representative Council had no more than advisory powers.

The other major achievement of the Nationalists was to change South Africa into a republic. In a referendum in 1960 a small majority of the white voters expressed a preference for a republic. In March 1961 Verwoerd asked a Commonwealth prime ministers' conference whether South Africa might remain a member of the Commonwealth when the change took place; but he withdrew his request when other members of the conference criticized his government's racial policies. Consequently when South Africa became a republic on May 31, 1961, it left the Commonwealth. The republican constitution was substantially the same as the constitution of the Union, except that an indirectly elected state president replaced the queen, and her representative, the governor-general, as head of state. C. R. Swart was elected state president.

9. Internal Opposition to Nationalist Rule.—In spite of all the separatist laws the peoples of South Africa were becoming more and more interdependent as a result of the continuous growth of the modern sector of the economy; and many South Africans, white as well as nonwhite, opposed the government's racial policies. White people of British descent resented the near-monopoly of senior posts in the civil service by Afrikaners, the introduction of the republic, and secession from the Commonwealth. Businessmen, experiencing a shortage of skilled labour, concluded that the colour bar was an impediment to sustained economic growth. Clergy regarded the government's policies as un-Christian and intellectuals regarded them as irrational. Several groups, including the Black Sash, a women's organization, were formed to protest against the most blatant abuses. However, these white critics did not shake the confidence of the government, for the election results showed that it was continuously increasing its electoral support.

African opposition was more serious. The leaders of the ANC organized a series of campaigns to elicit the support of the African masses and to intimidate the government into making concessions; and they had the cooperation of Coloured and Asian congresses and small white groups. In 1952–53 several thousand people, mainly Africans, deliberately courted arrest by "passive resistance to unjust laws." In 1955, 3,000 delegates attended a "Congress of the People" and adopted a "Freedom Charter." In 1957, 45,000 Africans walked many miles from their locations to their places of work in Johannesburg and Pretoria for ten weeks, rather than pay increased bus fares. There were also intermittent riots and clashes with the police in many parts of the country. These reached a high point in 1960. The Pan-Africanist Congress, a radical offshoot of the ANC, organized demonstrations against the pass laws; at Sharpeville, near Johannesburg, the police opened fire against demonstrators, killing 69 Africans and wounding 178; and blood was also shed in Cape Town and elsewhere. Between 1959 and 1961 there was a revolt against the Bantu authorities

IAN BERRY—MAGNUM

The struggle against apartheid: (Left) Africans burning pass books; (above) police breaking up an African rally, 1960

system in the Transkei; and between 1961 and 1964 there were several hundred acts of sabotage, in the form of time-bomb attacks upon public buildings and railway and electrical installations in the main industrial centres.

The government responded in an uncompromising manner. Parliament enacted many laws to help the executive break the revolutionary opposition. The nonviolent resistance campaign of 1952–53 was called off in the face of harsh new penalties for protest actions. The Congress of the People movement of 1955–56 collapsed when the government arrested 156 persons and charged them with high treason. Order was restored in 1960 by the declaration of a state of emergency and resort to government by proclamation, and by the outlawing of the ANC and the PAC. The Transkei revolt was handled under proclamations giving extraordinary powers to chiefs and white officials. The sabotage movement of 1961–64 was countered by General Law Amendment acts, which allowed the executive to outlaw any organization, to ban any person from attending meetings or publishing statements, and to arrest any person and hold him incommunicado without trial. The result was that by the end of 1964 all known revolutionary leaders were in prison or exile, all known revolutionary organizations had been broken up, and the radical internal opposition seemed to have been crushed.

10. External Opposition.—There was strong foreign opposition to South Africa's racial policies. The governments of the newly independent African states wished to complete their revolution by destroying the last strongholds of white supremacy in Africa; but although they had the diplomatic support of the Asian and Communist countries, they could not muster enough resources to coerce South Africa without the cooperation of South Africa's principal trading partners—Britain, Western Europe, and the United States. They therefore tried to use the United Nations as an instrument of coercion. The General Assembly had condemned South Africa's racial policies since 1948; the Security Council since 1960. In 1962 the assembly, by majority vote, called on member states to apply economic sanctions against South Africa and to break off diplomatic relations; and in 1964 the Security Council set up an expert committee to consider the feasibility of sanctions. But although Britain, France, and the United States subscribed to many resolutions criticizing apartheid, they did not commit themselves to coercion. In the mid-1960s, when the UN was weakened by its financial problems and the major powers were distracted by the conflict in Vietnam, the prospect of a UN blockade of South Africa seemed to be diminishing. By then, moreover, South Africa had well-equipped and well-trained military forces, consisting of about 125,000 white men, who could be rapidly mobilized and who were capable of overcoming any but a very large-scale attack on South Africa.

Nevertheless, there were still doubts about the future of the regime. Though the radical internal opposition had been crushed, the conditions which had created it continued; and the situations in contiguous territories were fluid. During 1966 Lesotho (Basutoland) and Botswana (Bechuanaland) became independent; Rhodesia's future remained uncertain; Portugal's hold on Angola and Mozambique was being challenged; and, although in July 1966 the International Court of Justice rejected an application brought by Ethiopia and Liberia in regard to South West Africa, in October the UN General Assembly voted to revoke South Africa's mandate over that territory (*see* SOUTH WEST AFRICA: *History*).

Verwoerd, the prime minister, was assassinated in September 1966 and was succeeded by B. J. Vorster, a member of the extreme right wing of the Nationalist Party. (L. M. T.; X.)

V. POPULATION

At the first census (1911) after union the population totaled 5,972,757; by 1960 it had nearly trebled, reaching 16,002,797. The increase is due only slightly to immigration and overwhelmingly to natural increase, which in 1960 was 16.4 per 1,000 for whites, 28.5 for Asians, and 32.9 for Coloureds. No reliable statistics are available for the Bantu (a term that has officially replaced "native" or "African").

From Tables I and II the numerical preponderance of the Bantu in the population is evident. Since 1911 they have maintained their rather more than three to one lead over the whites. About 32% of the Bantu population is urban, the majority living in municipal townships with a considerably smaller number in white areas as servants. Industrial compounds house about 146,000 and about 474,000 live in mine compounds. Of the Bantu rural population (68% of their total numbers) more than half live in the Bantu homelands; most of the remainder are labourers or servants on white farms. Table II shows that the Bantu are most numerous in Transvaal, which is partly because of their extensive employment in the industries of the southern Transvaal industrial and mining area and partly because of the numerous homelands in the

TABLE I.—*Population of South Africa, 1911-60*
(in 000)

Census year	All races		Whites		Bantu		Asians		Coloureds	
	Total	Urban	Total	Urban	Total	Urban	Total	Urban	Total	Urban
1911	5,973	*	1,276	*	4,019	*	152	*	525	*
1936	9,588	3,204	2,003	1,361	6,596	1,246	220	146	769	415
1946	11,416	4,393	2,372	1,769	7,831	1,856	285	200	928	540
1951	12,671	5,397	2,642	2,071	8,560	2,329	367	285	1,103	713
1960	16,003	7,481	3,088	2,582	10,928	3,471	477	397	1,509	1,031

*Not issued.

TABLE II.—*Provincial Areas and Populations, 1960 (Final Census)**

Provinces and capitals	Area (sq mi)	All races	Whites	Bantu	Asians	Coloureds	Density per sq.mi
Cape Province (Cape Town)	278,380	5,362,853	1,003,207	3,011,080	18,477	1,330,089	19 3
Natal (Pietermaritzburg)	33,578	2,979,920	340,235	2,199,578	394,854	45,253	83 7
Transvaal (Pretoria)	109,621	6,273,477	1,468,305	4,633,378	63,787	108,007	57.2
Orange Free State (Bloemfontein)	49,866	1,386,547	276,745	1,083,886	7	25,909	27.8
Total	471,445	16,002,797	3,088,492	10,927,922	477,125	1,509,258	33.9

*Data exclude Walvis Bay (374 sq.mi. [969 sq.km.]).

province. The second highest number is in Cape Province because of the population of about 1,500,000 in the homelands of the Transkei and Ciskei. Of the total number in Natal (including Zululand) about 900,000 are in the homelands, the rest being in the white area.

The white population has increased by more than 1,000,000 since 1936. From Table I it is clear that this section of the population is predominantly urban. From Table III it is seen that the majority of urban whites is concentrated in the southern Transvaal industrial area, which includes the Witwatersrand (chiefly Johannesburg, Germiston, Benoni, and Springs), Pretoria, Vereeniging, and Vanderbijlpark. The four ports of Cape Town, Durban, Port Elizabeth, and East London together contain about 646,000 whites. With the provincial administrative capitals Bloemfontein and Pietermaritzburg the white population in these "conurbations" is rather more than three-fifths of the total white population. Of the rural white population about 400,000 are on the 106,000 white farms, which cover an area of about 226,000,000 ac. (91,500,000 ha.).

TABLE III.—*Population, Metropolitan Areas, 1960*

Metropolitan area	All races	Whites	Bantu	Asians	Coloureds
Johannesburg	1,152,525	413,153	650,912	28,993	59,467
Cape Town	807,211	305,155	75,200	8,975	417,881
Durban	681,492	196,398	221,535	236,477	27,082
Pretoria	422,590	207,202	199,890	8,046	7,452
Port Elizabeth	290,693	94,931	123,183	4,247	68,332
Germiston	214,393	86,314	121,496	2,389	4,194
Bloemfontein	145,273	63,046	75,944	2	6,281
Springs	141,943	38,217	100,797	1,384	1,545
Benoni	140,790	41,992	90,236	2,996	5,566
Pietermaritzburg	128,598	40,065	55,991	26,827	5,715
East London	116,056	49,295	56,603	1,727	8,431
Welkom	97,614	27,096	70,230	nil	288
Kimberley	79,031	24,739	36,134	1,080	17,078
Virginia	41,057	11,934	29,020	nil	103
Odendaalsrus	21,268	6,102	15,050	nil	116

The Coloured population, deriving from a mixture of various racial groups, is concentrated mainly in Cape Province, where it supplies most of the labour for the farming and manufacturing industries. Included for statistical purposes in the Coloured total are 63,000 Malays, of whom 55,000 are urban in Cape Province.

The Asians have increased more rapidly than any other racial group. They consist almost entirely of Indians who were brought over from India in the second half of the 19th century for the Natal sugar plantations. Of the total 477,125 in 1960, 394,854 were in Natal, about 307,000 being urban, and of the urban Asian population 236,477 were in the Durban metropolitan area, exceeding the whites in this area by more than 40,000. (J. H. WN.)

VI. ADMINISTRATION AND SOCIAL CONDITIONS

1. Constitution.—Following the passage of the South Africa Act, 1909, by the British Parliament, the Union of South Africa was formed on May 31, 1910, out of the colonies of the Cape of Good Hope, Natal, Orange River Colony (Orange Free State), and Transvaal, which became the subordinate provinces of the new state. The Union was a member of the British Empire, and its sovereign independence was later guaranteed by the Statute of Westminster (1931) and the Status of the Union Act, 1934. On May 31, 1961, after a popular referendum held among white voters in October 1960, which favoured by a narrow margin the establishment of a republic, the Union became the Republic of South Africa under a constitution adopted by the Union Parliament. On becoming a republic South Africa ceased to be a member of the Commonwealth of Nations (see *History: Nationalist Party Policy from 1948*, above).

The new constitution (further amended in later years) is virtually the same as that of the Union enacted in 1910, with certain changes such as effected in the article dealing with the office of president and substitution of "the State" for "Crown" or "the King or Queen." It provides for a state president (who is head of the republic), an Executive Council of ministers of state appointed by him, a Parliament comprising a Senate and a House of Assembly, and, in each province, an administrator appointed by the state president-in-council for five years, an Executive Committee of four members under the administrator's presidency, and an elected Provincial Council which elects the Executive Committee and, like it, serves for five years. In 1920 the former German territory of South West Africa was assigned to South Africa under mandate. The people of this territory are regarded as South African subjects and are represented in the Parliament of the republic.

The Executive.—The state president is elected for a term of seven years by secret ballot of an electoral college consisting of members of both houses of Parliament. He is a constitutional ruler acting on the advice of his ministers, whom he appoints from among sitting and prospective members of Parliament having the confidence of the majority in the House of Assembly, and they together form a cabinet which in practice is the Executive Council. Ministers in charge of departments may not exceed 18 in number, but the state president may also appoint not more than six deputy ministers.

The Legislature.—Parliament has unrestricted legislative power "to make laws for the peace, order, and good government of the republic." It consists of an upper house, the Senate, and a lower house, the House of Assembly. Ordinarily, it legislates by means of bare majorities in both houses sitting separately, but in the case of the entrenched clause of the Republic of South Africa Constitution Act providing for the equality between the two official languages, English and Afrikaans, it must legislate by means of a two-thirds majority of all members of both houses sitting together.

Although both houses are original legislative bodies, in practice the Senate is in the main a revising chamber with delaying powers in respect of bills introduced in the lower house. In 1960 the Senate's constitution was changed to form a chamber consisting of 54 members, of whom 14 were elected in the Transvaal, 11 in Cape Province, 8 each in Natal and the Orange Free State, and 2 in South West Africa. In addition, eight were nominated for South Africa (four of them for their knowledge of the Coloured peoples) and two for South West Africa, one for similar knowledge. The Coloured population in the Cape had one representative. Qualifications for senators were the same as for members of the lower house, except for a minimum age limit of 30 years. Four seats formerly held by senators elected by Bantu voters were abolished in 1960.

In 1966 the House of Assembly comprised 160 members directly elected by white voters in the republic, 6 members directly elected by white voters in South West Africa, and 4 white members elected by Coloured voters in Cape Province. Three representatives previously elected by Bantu in the Cape ceased to be members of the house in 1960. Members must be adult South African citizens of white descent, and the representatives of white voters are elected from single-member constituencies delimited by a judicial commission at intervals from five to ten years. Seats are allotted to the provinces (and Provincial Councils) in proportion to their adult white population. Under the 1965 delimitation the Cape returned 54 members, Natal 18, the Orange Free State 15, and the Transvaal 73. Electoral divisions within each province have an approximately equal number of voters but may exceed or fall short of the average by between 15% and 30% under a so-called loading system.

With the abolition of Bantu representation and the lowering of the age qualification for white voters in 1959, the parliamentary electorate comprises white South African citizens of both sexes who are 18 years and over and a separate roll of Coloured males in the Cape possessing prescribed literacy and property qualifications.

2. Government.—There are 40 departments of state, staffed by about 260,000 public or civil servants. Departmental responsibility is vested in individual members of the Cabinet, which is presided over by the prime minister. Statutory bodies have been created to supervise broadcasting, electricity supply, and the development of residential group areas according to race; to produce iron, steel, phosphates, and other fertilizers; to extract oil from coal; to finance the development of industries and fisheries; to prospect for petroleum; to develop the Orange River Basin; and to promote the economic development of the Bantu and Coloured peoples in separate areas. The railways, harbours, and air-

ways are operated by an administration responsible to the minister of transport. More than 80 major public bodies have been set up for regulatory and advisory purposes, including an economic advisory council, a national education advisory council, a hotels board, a tourist corporation, and an immigration council.

Provincial Administration.—The legislative powers of the Provincial Councils are defined and limited by the Republic of South Africa Constitution Act of 1963 and by the Financial Relations Consolidation and Amendment Act, 1945, and their ordinances may not be repugnant to an act of Parliament. The constitution guarantees the provincial boundaries. The provincial administrations comprise the administrator; the Executive Committee, which under an act of 1962 is elected by the majority parties in the councils instead of by proportional representation as in the past; a provincial secretary; and various departments and branches. They deal with education, hospitals, roads, shop hours, public entertainments, horse racing, libraries, new townships, town planning, and motor vehicle licensing. They also control local government authorities. Under the constitution, Parliament may not abolish a Provincial Council or abridge its powers except on a petition by the council. The South West Africa administration is similarly constituted with a Legislative Assembly instead of a Provincial Council.

Local Government.—Each province has its own system of local government. In the Cape there are municipalities with city and town councils, village management boards, periurban (suburban) local boards, and divisional councils (rural). They are wholly elected except for local boards. In Natal there are elected city and town councils and town boards, and elected or appointed health committees. Periurban development is controlled by a local health commission appointed by the administrator. In the Orange Free State there are municipalities with city and town councils and village management boards, all elected. Elected small holdings committees of management in periurban fringes or settlements are controlled by an appointed small holdings area control board. In the Transvaal, there are elected city and town councils and village councils, and elected and appointed health committees. Periurban development is controlled by a special health board appointed by the administrator. South West Africa has institutions similar to those of the Cape.

Nonwhite Administration.—Local authorities control Bantu populations within their borders and establish elected Bantu advisory boards in their Bantu townships. In 1961 provision was made for the establishment of partially elected Bantu urban councils in the townships. In the homelands, a reorganization of local government was undertaken into tribal, regional, and territorial authorities under tribal chiefs and the supervision of the Department of Bantu Administration and Development. In 1963 the first Legislative Assembly of 109 members, with a chief minister and cabinet and limited powers (*i.e.*, excluding foreign affairs, defense, etc.) was established in the Transkei homeland, the first of the projected "Bantustans." Local councils for Coloured people are being created in the Cape under the supervision of the Department of Coloured Affairs and the Coloured Advisory Council. A National Advisory Council was set up for the Indian population by the Department of Indian Affairs.

Political Parties.—There are two main political parties in South Africa and one other with representatives in Parliament. They are the Nationalist Party (106 seats in the House of Assembly in 1965), the United Party (49 seats), and the Progressive Party (1 seat). Smaller groups unrepresented in Parliament include the Liberal Party and the Republican Party; Coloured voters are represented by four white members (1 United Party and 3 Independents). The Nationalist Party came into office in May 1948.

3. Taxation.—Three-quarters of national taxation comprises income tax, customs and excise duties, and death duties. There is a growing reliance on the income tax (and supertax) levied progressively, which accounts for more than 50% of total revenue, but incidence is not excessive and is reduced by family and sickness allowances. In 1963 the pay-as-you-earn system of income taxation was introduced. From time to time special repayable tax levies are imposed. The provinces rely mainly on personal and income taxes (18% of their income), motor vehicle licence revenue, and racing and entertainment taxation; but they are state aided to the extent of at least 50% of their income. In 1957 the subsidy basis was revised. Local authorities receive only 7% in grants, and 20% of their income derives from rates on property, generally considered regressive in incidence. Charges for goods and services account for more than 30% and trading profits for more than 10%.

While such taxation facilitates central budgetary control in the interests of economic stability, increasing urbanization throws a growing financial burden on local authorities, and an inquiry into the redistribution of revenue resources began in 1956. By the 1960s approximately 80% of all taxation was levied centrally, 13% provincially, and 7% locally.

4. Living Conditions.—The standard of living is high for the white population; for most Bantu it remains low but is nevertheless higher than in most African countries.

Employment and Wages.—A maximum 46-hour week was introduced in 1941 and the tendency has been to reduce working hours, in some cases to a 40-hour or five-day week. The Industrial Conciliation Act, 1956, provides for the registration of employers' organizations and labour unions, the voluntary settlement of disputes by industrial councils or conciliation boards and, finally, arbitration (which is compulsory in essential services where strikes and lockouts are prohibited). Bantu are excluded from this act and cannot become members of registered unions, but the Native Labour (Settlement of Disputes) Act, 1953, created a central native labour board and subordinate regional committees and labour officers. It also prohibited strikes by, and lockouts of, Bantu. Outside this legislation, wage determinations are made by a wage board established under the Wage Act, 1937, which also determines Bantu wages where the minister refers disputes to it from the Central Bantu Labour Board.

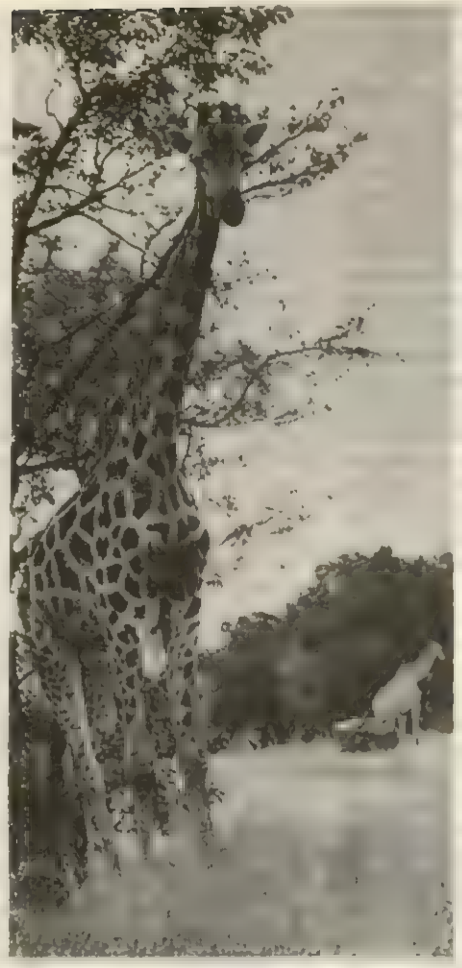
Other legislation includes the Registration for Employment Act, 1945, which created a national employment service (including employment exchanges and extending to sheltered employment and unemployment relief), and the Workmen's Compensation and Unemployment Insurance Acts, 1941. Bantu labour bureaus and unemployment are dealt with by the Department of Bantu Administration and Development, either regionally or through local authorities.

Labour Unions.—Only a few trade unions, properly so-called, survive; *i.e.*, unions that represent the workers in a specific trade. Most unions have become industrial labour organizations, representing all workers in a particular industry. Under the Industrial Conciliation Acts, as amended in 1956, unions are prohibited from affiliation with political parties. In the 1960s the total number of industrial organizations and trade unions exceeded 170, over half of which were for white workers, less than one-quarter were Coloured, and more than one-quarter were racially mixed. The major organizations include the South African Trade Union Council, the South African Federation of Trade Unions, the Koordinerende Raad van Suid-Afrikaanse Vakverenigings, the South African Congress of Trade Unions, the National Union of Liquor and Catering Trades, and the Mechanics Unions' Joint Executives.

Bantu unions do not exist. The rights of Bantu workers are defined by the Native Labour (Settlement of Disputes) Act of 1953. This act provides for a Central Bantu Labour Board, Bantu labour committees and works committees, and Bantu labour officers. Disputes which cannot be settled by the labour officers in collaboration with the regional labour committees and the divisional inspector of labour are referred to the central board. Under the Industrial Conciliation Act (1956) certain occupations are reserved for certain racial groups in order to protect the racial groups from unfair competition.

Housing.—At the 1951 census 564,277 non-Bantu residential units were enumerated, of which 443,616 were occupied by whites. In 1964 new dwellings completed in 18 principal municipalities and 45 smaller towns averaged more than 9,200 a month. Residences for whites are usually built privately, but public housing projects were stimulated by wartime measures that set up a national housing and planning commission with the power to erect houses, to assist local authorities and utility companies, and to make loans to individuals. Most projects are economic or "no-

loss," but assisted housing is provided for lower income groups in which the central and local authorities share the loss. A Ministry of Community Development and Housing was established in 1961 to advance housing programs for all races.



(LEFT) ERNST HAAS—MAGNUM; (TOP) AUTHENTICATED NEWS; (BOTTOM) MARC AND EVELYNE BERNHEIM FROM RAPHO GUILLUMETTE

Some aspects of the land: (Left) Giraffes and kudu in Kruger National Park; (above) Bantu students at the University College of the North, Pietersburg; (below) modern Cape Town with Table Mountain at left

Bantu housing is assisted by a Bantu Housing Board and undertaken in Bantu townships by urban local authorities with public and private funds. A massive project to rehouse urban Bantu was undertaken from 1951 onward. The "squatters' camps," or slum shacks, which had accumulated around towns to which Bantu had been attracted by industrialization, were removed and new townships built. During ten years more than R 200,000,000 (£100,000,000 or $280,000,000) was spent in providing these families with new houses, services, and transport. The townships are planned on modern lines and built by modern methods; each house has its own garden, and there are community amenities, schools, halls, playgrounds, churches, etc. Ethnic groups are housed in their own separate quarters. An example is Daveyton, outside Benoni. Houses are leased, and rents are government-subsidized where incomes fall below R 360 (£180 [$504]) a year (or R 300 in some cases). Bantu have no freehold property rights in townships but may occupy their houses on a long lease. The cost of services to urban Bantu townships is defrayed from funds accruing under the Native Services Levy Act, 1952, which obliges employers to pay a maximum of *2s.9d.* (39 cents) a week for each Bantu employed, part of which is reserved for transport to the townships. Outside these areas Bantu (and also Coloured and Asian) housing is provided by the Group Areas Development Board. Very little nonwhite housing is privately built.

5. Welfare Services.—Welfare services are run by central, provincial, and local authorities, and by private organizations. The Department of Social Welfare administers legislation dealing with the protection of children, adoptions, family allowances, probation services, juvenile delinquency, the registration of welfare organizations, assistance to the blind, pensions, and poor relief. It subsidizes the voluntary work of about 1,800 registered organizations, including national bodies for cripple care, child welfare, the blind, the deaf, first aid, mental health, and maternal and family welfare. Church organizations also undertake relief and family welfare work.

In the provinces and municipalities, other services are financially assisted or provided in conjunction with hospitals, public health, education, and recreation (all races). Coordination of welfare services is effected by a statutory nonofficial national board and local boards.

Public health and medical services fall under a central Ministry of Health, the four Provincial Councils, and more than 700 local authorities. The Ministry of Health, created under the Public Health Act of 1919, has advisory and executive functions, chiefly in regard to the prevention and control of infectious and communicable diseases and the promotion of health. It has in its employ numerous full-time or part-time district surgeons in urban and rural areas. Health centres and clinics serve mainly the indigent. Larger municipalities have health departments and salaried medical officers. The provinces maintain public general hospitals, including homes for the chronically ill, for whites and nonwhites. These hospitals are financed from taxation, and in necessitous cases treatment is free or partly free. In addition there are many private, mission, mining, and industrial companies' hospitals. Medical-aid societies operate among sections of the public in a number of areas. In 1962 a central authority to coordinate this movement on a nationwide voluntary basis was established. By the 1960s South Africa had about 8,500 medical practitioners, mostly white but including a number of Bantu, Coloured, and Asian doctors.

6. Justice.—*Courts.*—The common law of the republic is based on Roman-Dutch law (*q.v.*), the uncodified law of Holland that was retained after the Cape's cession to Britain in 1806. The judiciary comprises a Supreme Court with local, provincial, and appellate divisions, the High Court, and circuit courts of South West Africa, and inferior courts. The judges, headed by a chief justice, are appointed by the state president-in-council; they cannot be removed from office except by an address of both houses of Parliament, and their salaries cannot be reduced. Appeals from the divisional courts and the high and circuit courts of South West Africa lie to the appellate division of 11 judges, with a quorum of 5 except where the validity of an act of Parliament is in question, when the quorum is 11.

The inferior courts comprise magistrates' courts, Bantu courts, and water courts (dealing with the use of water resources). Most civil and criminal litigation is a matter for the magistrates' courts, each responsible for a magisterial district. Regional magistrates are appointed in populous areas with criminal jurisdiction over a number of districts. The Bantu courts do not have jurisdiction over Bantu in criminal cases or where whites are involved.

Police.—The police are organized nationally into 11 divisions. The force comprises about 15,000 whites and a similar number of nonwhites, and it extends to South West Africa. Each division contains a number of districts and police-station areas. There are

also the security police, directly responsible to head office, and the traffic police of provincial and local authorities.

7. **Education.**—The Ministry of Education, Arts, and Science is responsible for primary and secondary education for white children who are physically handicapped or in need of care and for vocational schools, including technical high schools and colleges. It is also responsible for the autonomous universities of Cape Town (founded 1873), Natal (1909), Rhodes (1904), and the Witwatersrand (1922) (all English speaking), the Orange Free State (1855), Potchefstroom (1869), Pretoria (1930; from 1908 it was the Transvaal University College), Stellenbosch (1918) (Afrikaans speaking), Port Elizabeth (1965) (bilingual), and South Africa (a correspondence university). In 1965 it was decided to establish a new Afrikaans-speaking university on the Witwatersrand. In addition there are three university colleges for Bantu—Fort Hare, eastern Cape, for Xhosa (founded in 1916 for nonwhites; Bantu only since 1959); Zululand for Zulu only, at Ngoye (1959); and the University College of the North for Tsonga, Sotho, and Venda, at Turfloop, Pietersburg, Transvaal (1959); and one for Indians at Durban (1961). In 1959 the Department of Bantu Education assumed responsibility for all education of Bantu, whether primary, secondary, technical, teaching, or university, other than education in certain mission schools (which are subsidized and partly controlled except in the case of Roman Catholic schools). In the 1960s the Department of Coloured Affairs took over all Coloured education and the Department of Indian Affairs all Indian education.

The provinces administer all other primary and secondary education and teachers' training colleges (private schools excepted). Natal apart, local school committees are elected by parents, and elected and nominated district boards undertake general administration and (with the committees) select teachers. In Natal education is directly administered by the province. In all provinces schools for whites are generally single medium (*i.e.*, either English speaking or Afrikaans speaking), and school selection depends on the language in which the child is more fluent. Attendance is compulsory and free for white children between 7 and 16 years of age. For Coloureds and Asians, attendance is not compulsory but is free up to certain secondary standards, varying with the province. There are extensive bursary facilities for higher education. In 1963 the Ministry of Education established a National Advisory Education Council to foster a uniform national education policy.

8. **Defense.**—The South African Defense (Further Amendment) Act, 1961, provides that every white citizen at age 18 is subject (under an annual ballot) to nine months' compulsory military training for one year and another period of three weeks during the following two years.

The South African defense force comprises the permanent force, active Citizen Force, rifle commandos, reserves, school cadets, and auxiliary services (nonmilitary and nonwhite personnel). The permanent force consists of the Army, Air Force, and Navy organized in ten army commands, three air force groups, and two naval bases. It is staffed by regular personnel primarily concerned with training and administration. The active Citizen Force comprises personnel undergoing peacetime training, and the school cadets consist of boys between 12 and 17 years of age. (L. P. Gr.; L. H.; X.)

VII. THE ECONOMY

South Africa has passed through three phases in economic development. The first phase was the pioneer period of subsistence agriculture which lasted until about 1870, when the diamond fields were opened up. Gold discoveries followed, culminating in the development of the Witwatersrand gold field and giving a stimulus to agriculture and foreign exchange for the purchase of manufactured goods. With the outbreak of World War I and the inability to get supplies from overseas came the stimulus to manufacturing, ushering in the third phase, in which secondary industry became a greater producer of wealth than either agriculture or mining. This is reflected in the national income; between 1914 and 1953 manufacturing industries passed from eighth to first place among contributors to the national income.

The government's policy is to support private enterprise, with the minimum of interference. Governmental assistance is given in the fields of development, finance, and export promotion. The National Resources Development Council, set up in 1947, advises the minister of economic affairs in order to promote general development. There are a large number of regional development associations formed to guide local development. Foreign investments were heavy, especially after World War II, although in 1960 and 1961, because of racial unrest, the economy suffered from the outflow of foreign capital, which was later stemmed.

A. Production

1. **Agriculture.**—Agriculture has developed on a natural basis of climate, topography, and soil. Cultivation has flourished where ground is moderately flat, rainfall adequate or irrigation practicable, and the soils sufficiently fertile. Two parts of the country fulfill these requirements—the lowland surface of southwestern Cape Province, with winter rainfall exceeding 10 in. (255 mm.), and the eastern part, with summer rainfall exceeding 25 in. (635 mm.). In southwestern Cape wheat and grapes, with deciduous fruit trees, are important; in the eastern part maize (corn), sugar, and subtropical fruits are grown.

Livestock raising, while common in all crop-growing areas, is found mainly in the moister parts too high or too steep for cultivation or in areas of grassland too dry for cultivation. Many of these stock farms rely on underground sources of water for their animals, and through boreholes tap supplies in a variety of water-bearing rocks. Cattle are largely confined to the eastern summer-rainfall areas with long grass, sheep and goats to the drier areas where Karroo shrubs can withstand long droughts and provide adequate sustenance for small stock. Field crops and horticultural products are together of greater value than livestock products, though the value and volume of both sides has been steadily increasing. In the 1960s about 1,700,000 people were engaged in farming, including 120,000 whites.

Crops.—Maize (corn), the staple diet of the African, is the most valuable of the republic's crops. Two kinds are generally grown: white dents and yellow flints. About 6,000,000 morgen (12,700,000 ac.) are sown to maize, the production fluctuating between about 40,000,000 bags (of 200 lb.) and 50,000,000. Excess of production over local requirements has necessitated exports involving considerable annual losses. This is one of the republic's problems, since production is likely to increase as more hybrid seed becomes available. Wheat is grown mainly in southwestern Cape, and also in eastern Orange Free State and, under irrigation, in southern and central Transvaal. Annual production has varied from about 5,000,000 to 11,700,000 bags. As the Bantu are buying more wheat bread, an annual import of wheat is generally necessary. Sugarcane growing is limited mainly to the east coastal area of Natal (including Zululand). The subtropical climate necessitates a 16- to 20-month reaping cycle. About 550,000 ac. are under cane, of which about 290,000 ac. are harvested each season. Production is about 1,400,000 short tons of sugar from 12,000,000 tons of cane, of which about 800,000 tons are consumed locally. Exports, which vary greatly with world demand, average about 600,000 short tons. Other important crops include tobacco, peanuts, potatoes, and kafir (sorghum).

Fruit.—The gross value of the annual production of fresh and dried fruit is about R. 99,000,000 (£49,500,000 or $138,600,000). Deciduous tree fruits and grapes flourish best in southwestern Cape, citrus and subtropical fruits in the eastern part of the republic. The main deciduous tree fruits grown are (in order of quantity exported) apples, pears, plums, peaches, and apricots. Exported grapes total approximately 32,000 short tons, and 4,000,000 gal. of wine and a smaller quantity of brandy are also exported. More than 80% of the citrus grown consists of Valencia and navel oranges; grapefruit, lemons, and naartjies (tangerines) are grown in much smaller quantities.

Livestock.—The livestock population comprises mainly cattle, sheep, and goats, with comparatively small numbers of pigs and donkeys and much poultry. Cattle number about 12,000,000; approximately 4,000,000 are in Cape Province (mainly in the eastern part), 4,000,000 in Transvaal, 2,500,000 in Natal, and nearly 2,000,-

000 in Orange Free State. Of these about 7,000,000 belong to whites, 1,200,000 to Bantu on white farms, and 3,800,000 to Bantu in the homelands. Sheep number approximately 38,000,000, of which nearly 25,000,000 are in Cape Province, nearly 9,000,000 in the Free State, nearly 4,000,000 in Transvaal, and over 1,000,000 in Natal. About 28,000,000 of these are merinos, 1,200,000 karakuls, and 4,000,000 nonwooled sheep. About 3,500,000 of the nonmerino sheep belong to Bantu. Goats number nearly 5,500,000, more than 3,500,000 being in Cape Province, 900,000 in Natal, 650,000 in Transvaal, and 85,000 in the Free State. Of these about 870,000 are angora goats. About 3,000,000 goats are Bantu owned.

Cattle products include meat, fresh milk, butterfat, cheese milk, and milk for condensing. Products from sheep and goats include not only meat and milk (from goats) but wool and mohair. Miscellaneous livestock products include ostrich feathers.

2. Forestry.—Out of the republic's total land area of slightly over 302,000,000 ac. (122,200,000 ha.), natural forest covers less than 925,000 ac. (374,000 ha.). There are more than 3,000 timber and wattle plantations varying from 25 to over 10,000 ac. (4,000 ha.). Of the large plantations (5,000 ac. or more) most are in eastern Transvaal, Natal, southern Cape Province, and Zululand. In the natural forests the most valued hardwoods are stinkwood and sneezewood. The softwoods include the yellowwoods, the Cape box, and many others. The natural forest is carefully preserved by the state Forestry Department, and only small quantities of timber are cut each year. The chief broadleaf species include eucalyptus, matchwood poplars, and the tan wattles. An important factor in the development of exotic softwood plantations is the high natural wood increment, one of the highest in the world (the average annual increment is of the order of 150 cu.ft. per acre, which may be compared with European and North American areas where it is between 17 and 35 cu.ft.). In some of the wetter afforested areas in South Africa the average increment is 250 cu.ft. per acre. South Africa's afforestation potential is estimated to be about 2,000,000 ac.

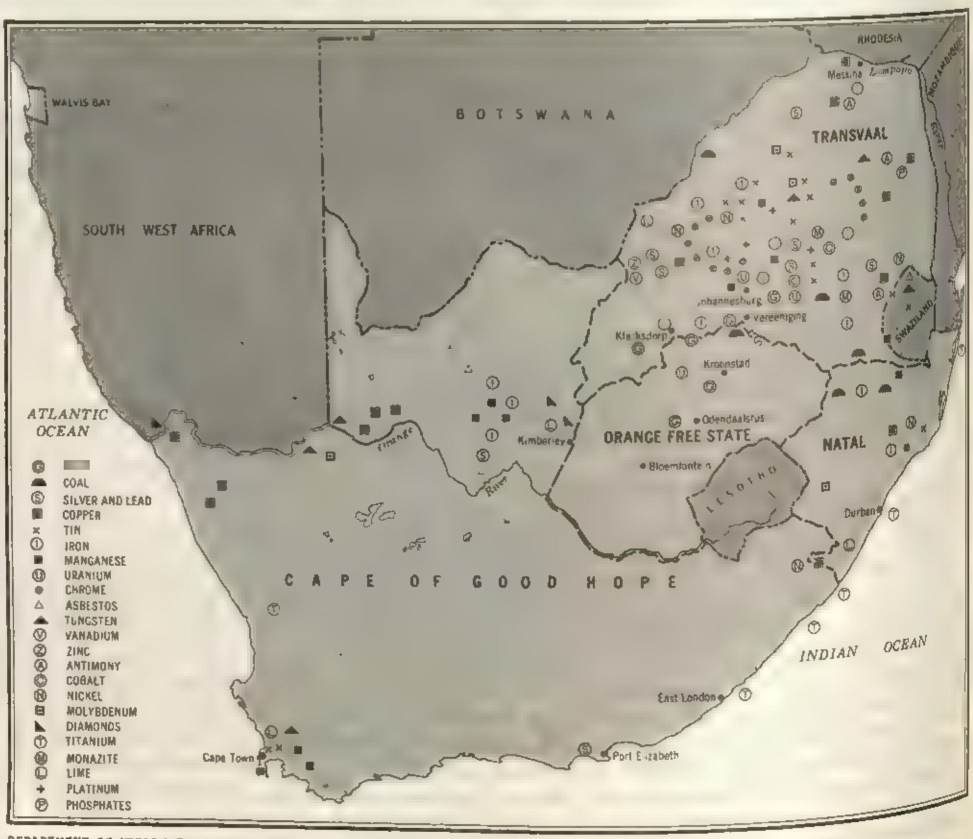
The annual national consumption of softwoods is about 130,000,000 cu.ft. of logs. Of this about one-quarter is imported while three-quarters is produced locally. Since, however, pines take 30–40 years to mature, and well over half of the plantations have been planted since 1950, 75% of the logs delivered to the sawmills are immature. The local woods are therefore generally marketed at prices 15%–25% below those of imported timber, but mature South African timber compares favourably with the imported wood. The export of forest products includes wattle bark and bark extract. (*See* also *Physical Geography: Vegetation*, above.)

3. Fisheries.—Marine fishing is carried on mainly off the west and south coasts. The west coastal waters with their low temperatures and abundant marine plankton abound in fish life and provide the largest catches. Off the west coast the most profitable types of fishing are inshore shoal fishing and offshore trawling. The two principal shoal fish are the pilchard and the maasbanker, with mackerel varying considerably in quantity. Trawling is the main source of fresh fish consumed locally. Stockfish, or Cape hake, is by far the most important item in the trawled catch, though much smaller catches are taken of the 14 or 15 other varieties, of which the most valuable are kingklip, the Agulhas sole, panga, and kabeljou. Most of the catch of line fisheries is of snoek, or sea pike, the remainder consisting mainly of geelbek (Cape salmon), kabeljou, steenbras, redfish, and sharks. Vitamin A is derived from the shark livers and there is an export of dried and salted shark meat to the Far East. An important item in the inshore catch is the crayfish, commercially termed rock lobster, which is found mainly off the west coast and is exported largely to the United States in the form of frozen tails and canned lobster. There is a whaling industry based on Durban and Saldanha Bay. Fin, sperm, sei, humpback, and blue whales are caught for whale oil, sperm oil, whale meat and meat extract, and vitamin oil.

4. Minerals.—Mining provided in the 1960s the second largest contribution (after manufactures) to the national income. Production of gold bullion exceeded 29,000,000 fine oz. annually and provided a revenue of more than R 730,000,000 (£365,000,000 or $1,022,000,000). Most of the output came from the 55 producing mines controlled by the Transvaal and Orange Free State Chamber of Mines. In decreasing order of production the seven main gold fields are Orange Free State, Far West Rand, Klerksdorp, East Rand, Central Rand, Evander, and West Rand. All of these mine the "reefs" or gold-bearing conglomerates of the Witwatersrand geological system, which has surface outcrops on the Witwatersrand but is found mostly at depth in the other gold fields. On the Rand (*i.e.*, Witwatersrand) the main reef dips gently southward and is mined to a depth of over 11,000 ft. (3,350 m.).

Since 1952 uranium, which in the form of uraninite is associated with the gold in the reefs, has been produced in ever increasing quantities. The republic's uranium ore reserves are estimated at 1,100,000,000 tons, of which the uranium content appears to be greater than the reserves of Canada or the United States. Nineteen of the older mines, though producing gold at a loss, make a substantial profit on uranium.

Employment on the gold mines includes about 51,000 whites and 400,000 Bantu drawn from many parts of southern Africa. The place of gold in the economy of the republic is well established, but perhaps its most important effect is as a stabilizer in the fluctuations of foreign trade (see *Foreign Trade*, below).



DEPARTMENT OF INFORMATION, SOUTH AFRICAN EMBASSY, LONDON

PRINCIPAL MINERAL DEPOSITS IN THE REPUBLIC OF SOUTH AFRICA

Of other precious minerals, platinum is produced as the market demands; the annual output is valued at about R 16,000,000. Small quantities of silver and osmiridium are found in the gold ore. Diamonds to the value of about R 44,000,000 are derived from mines and alluvial deposits. In the base mineral group coal from the Ecca beds of the Karroo system is mined in southern Transvaal and Natal; in the 1960s the output exceeded 49,000,000 tons annually, of which nearly 1,000,000 tons were exported. Iron ore production, mainly from Thabazimbi in Transvaal, Postmasburg in northern Cape Province, and from northern Natal, fluctuated between 5,000,000 and 6,000,000 tons, the bulk of the ore being a hard, high-grade hematite with low phosphorus and sulfur content. Other minerals include asbestos, copper, manganese, chrome ore, and tin.

5. Power.—In 1922 the South African government established the Electricity Supply Commission (ESCOM), a nonprofit-earning body, to control and supply power for the industrial development of South Africa. The commission operates in eight licensed areas of supply: Rand and Orange Free State, Cape Western, Cape Eastern, Natal Southern, Natal Central, Border, Eastern Transvaal, and Cape Northern areas. The total power output of the commission in the 1960s was in excess of 22,000,000,000 kw.hr., representing about 70% of the country's output, the rest being provided by municipalities and private undertakings (mines and industries). The supply to rural settlements and farms has steadily increased. The power stations, with two small exceptions, are all thermal, coal-burning types.

6. Industries.—Government-sponsored industrial activities include the South African Coal, Oil, and Gas Corporation (SASOL), brought into being for the extraction of oil from coal; the Iron and Steel Industrial Corporation, Ltd. (ISCOR), which owns furnaces and mills and provides about two-thirds of the country's steel requirements; Klipfontein Organic Products (KOP), which produces chemicals and insecticides; and the Phosphate Development Corporation, producing fertilizers. Industrial research is sponsored by the South African Council for Scientific and Industrial Research (CSIR), a corporate body set up to advise the minister of economic affairs on all questions of scientific and technological methods affecting the utilization of the natural resources of the republic. The CSIR controls the Bureau of Standards and a number of national laboratories including those for research in chemistry, physics, telecommunications, building, mechanical engineering, road construction, personnel administration, nutrition, and industries. An important function of the council is the promotion of cooperative research by the formation of industrial research associations which are independent bodies subsidized by the council.

Manufacturing is by far the largest item (about 85%) of all private secondary industry, while building construction, electricity, gas and steam, and services make up the remainder. In the manufacturing sector four areas are outstanding in importance: southern Transvaal (including Pretoria, Vereeniging and Vanderbijlpark), western Cape, Durban-Pinetown, and Port Elizabeth-Uitenhage.

Of the total gross output of private manufacturing, the groups producing most were: food, metal products, clothing and footwear, basic metal industries, chemicals and chemical products, and nonmetallic mineral products. The total number of employees in manufacturing industries in the 1960s exceeded 830,000, comprising about 26% whites, 57% Bantu, and 17% Asians and Coloureds.

B. Trade and Finance

1. Foreign Trade.—A characteristic feature of South Africa's trade has been a constant disequilibrium between the import and export of merchandise, excluding gold. In 1925, for example, the excess of imports over exports, excluding gold, was £12,800,000; in 1930, £17,100,000; in 1935, £44,300,000. In each year the deficit was changed into a surplus by exports of gold bullion amounting respectively to £34,300,000, £35,900,000, and £71,400,000. Gold is therefore of increasing importance in the balance of payments.

Although South Africa is no longer a member of the Commonwealth of Nations, preferential tariff rates are maintained with Great Britain under a bilateral agreement and in the framework of the General Agreement on Tariffs and Trade (GATT).

Imports.—In the 1960s by far the most valuable item was the class "metals, metal manufactures, machinery, and vehicles." In this class the chief commodities were automobiles and trucks, railway locomotives and rolling stock, tractors and parts, and airplanes. These goods were supplied by the U.K. and other Commonwealth countries; of the other suppliers the U.S. and West Germany far exceeded the other trade partners (Japan, Italy, the Netherlands, Rhodesia, France, Belgium, and Ceylon). Second in value was the class "fibres, yarns, textiles, and apparel." Third in value came "oils, waxes, resins, paints, and varnishes," with crude oil, paraffin, and motor spirit (gasoline) from Iran as the main items. Next in value were "drugs, chemicals, and fertilizers," "foodstuffs," "minerals, earthenware, glassware, and cement," and "wood, cane, and wicker manufactures."

Exports.—In the 1960s the class with the highest value (excluding gold) was "metals, metal manufactures, etc." Atomic energy material covered almost half of this amount. Other minerals were of comparatively small value: copper, lead ore concentrates, manganese ore, chrome ore. Manufactured metal goods include mining machinery, ferroalloys, wrought iron, and manufactured iron and steel. Small items deriving from gold mining were silver and osmiridium. Foodstuffs came second, the chief items being fresh and preserved fruits (citrus exports go to more than 35 countries), maize (corn) and maize meal, fish (fresh, dried, frozen, and fish meal), and sugar. The class "animals, agricultural and pastoral products" was next in value. The most important commodities in this class were wool, hides and skins, fish meal, wattle and wattle bark extract, and mohair. Minerals, earthenware, and glassware came next (diamonds, asbestos, etc.), followed by jewelry (cut and polished diamonds), then by oils, waxes, etc., and other commodities of small value.

The chief trading partners receiving exports and reexports were the United Kingdom, Rhodesia, the United States, Japan, West Germany, Italy, Israel, Belgium, France, and the Netherlands.

2. Banking and Currency.—In South Africa the Ministry of Finance determines the country's general financial policy, which is administered by the central bank and operated mainly by the commercial banks and other specialized financial institutions. The central bank is the South African Reserve Bank, which came into existence in 1921 during the monetary chaos that followed World War I. The bank is a privately owned joint-stock company with limited liability; its governor and deputy governor and three other directors on the board of 11 are appointed by the government. The bank has the sole right of note issue in the republic, but its notes are not redeemable in gold.

Under the terms of the South African Reserve Bank Act of 1944 the bank was required to maintain in gold coin and bullion a reserve of at least 30% of the aggregate amount of its note issue and of its other liabilities to the public, but by an amendment in 1948 the minimum reserve ratio of the bank was reduced to 25%.

The commercial banks are required to hold on deposit with the Reserve Bank balances amounting to not less than 10% of their demand liabilities payable in the republic and 3% of their time liabilities. The Reserve Bank, however, is empowered to require the commercial banks to maintain, when the national interest demands it, supplementary reserves not exceeding 10% of their total liabilities to the public. The Reserve Bank found it necessary in the national interest to exercise these powers in 1958. The Reserve Bank may alter the supplementary reserves to 6% to ease a strain on the banking system. Registered banking institutions include commercial banks, people's banks, loan banks, and deposit-receiving institutions.

To promote agricultural development the Land and Agricultural Bank of South Africa was established in 1912 as a nonprofit-making institution. To promote an active money market the National Finance Corporation of South Africa was established in 1949. A close cooperation with the Reserve Bank is provided for, and the corporation may not be placed in liquidation except by an act of Parliament.

Currency.—The South African coinage was decimalized in 1961 when the rand (issued as two gold coins of R 1 and R 2) became the new monetary unit equal to the former half-pound (10 shillings or $1.40). The new silver coins were 50 cents (crown), 20 cents (florin), 10 cents (shilling), and 5 cents (sixpence); and the bronze coins comprising the cent (twelve-tenths of a penny), the two-cent, and half-cent. Although no longer a member of the Commonwealth, South Africa remains within the sterling area.

3. National Finance.—Exchequer receipts fluctuated in South Africa from £147,400,000 (equivalent to R 294,800,000) in 1950 to R 1,088,000,000 by the mid-1960s; in the same period exchequer issues increased from R 293,600,000 to more than R 1,018,000,000. On the revenue account (in excess of R 600,000,000) the receipts were derived from customs and excise, post office, and inland revenue. Of these the inland revenue amount was by far the largest, with income tax the highest source of revenue, followed by departmental receipts, interest and dividends, transfer duty, and nonresident shareholders' tax.

Parallel expenditure from the revenue account included, as its main items, provincial administration, social welfare and pensions, posts and telegraphs, defense, public debt, and agriculture.

In the decade following 1954 the net national income increased from R 3,025,000,000 to around R 5,600,000,000. The main sources were manufacturing, mining, agriculture, forestry and fishing, and transport.

The Bantu homelands five-year plan (1961–66) called for a total estimated expenditure of R 114,000,000.

C. Transport and Communications

1. Roads.—In South Africa the construction and maintenance of roads is the responsibility of provincial administrations. In each of the provinces four classes of roads are recognized: national, main, district, and minor roads. Provincial main roads are 80 ft. (24 m.) wide and, in general, are of the all-weather type, with gravel or bituminous surface. District roads, 60 ft. (18 m.) wide, are usually of the fair-weather type, with gravel surface. Minor roads vary from almost veld tracks to ordinary farm roads. The extent of such road types is of the order of 13,000 mi. of blacktop (bituminous, asphalt, tarmac, etc.), 39,000 mi. of gravel-surfaced, and 150,000 mi. of earth roads.

In 1935 the National Roads Act created a National Road Board for the administration of national roads under the responsibility of the central government. The board's program, announced in 1936, provided for the construction, reconstruction, and repair and maintenance of about 5,120 mi. of roads, declared "national." These were specified as bituminous surfaced with a minimum width of 100 ft. Fifteen routes were designated, their function being to link together the larger centres and to facilitate communication with adjoining territories. Thus route no. 1 links Beitbridge (over the Limpopo River on the Rhodesian border) with George on the south coast, passing through Pretoria, Johannesburg, and Bloemfontein; route no. 2 links Cape Town with Durban, passing through Mossel Bay, Port Elizabeth, and East London; route no. 3 links Johannesburg with Durban; route no. 4 links Pretoria with Komatipoort on the Mozambique border. The remaining routes interconnect these four routes or provide branch roads or extensions to other important centres.

2. Railways.—With few exceptions the railways and harbours of South Africa and South West Africa are the property of the government, the control and management being under the direction of the South African Railways and Harbours Administration, responsible to the minister of transport. Of the republic's 12,000 mi. of railway lines, 5,000 mi. are main lines of 3-ft.-6-in. gauge, and nearly 7,000 mi. are branch lines, of which less than 500 mi. are of 2-ft. gauge. Connected with the railways are the South African Railways' auxiliary road motor services, which cover almost 32,000 route miles.

Electrification of South Africa's railways has taken place mainly in industrial areas as, for example, southern Transvaal, where, in addition to electrification on the Witwatersrand and between Johannesburg and Pretoria, the lines from the coal mines of the Witbank area are connected by electrified lines with Kroonstad and the Free State gold field. In Natal the main line has been completely electrified from Durban to the Transvaal border, and in southwestern Cape the line to the interior has been electrified for 330 route mi. (530 km.) to Beaufort West. In the republic 1,684 route mi. were electrified by the mid-1960s. Diesel locomotives are replacing steam chiefly for traction in the drier parts of the country and in South West Africa.

3. Shipping.—There were about 470 ships registered in the mercantile marine in the 1960s with a registered gross tonnage of about 285,000. The total number of ships arriving at South and South West African ports in a year was nearly 15,000, including coasters. In respect of cargo Durban is the first port in the republic and in Africa, followed in South Africa by Cape Town, Port Elizabeth, and East London. Ship arrival and passenger transits are also greater at Durban than at the other South African ports.

A marked feature of shipping in South Africa is the large number of foreign cargo- and passenger-carrying liners calling regularly at ports either as destination or in transit to other parts of the world. There is thus generally ample transport for South Africa's overseas imports and exports. Since, however, precooling is essential for the country's important fruit exports, the large precooling stores at Cape Town and Port Elizabeth have special significance. An important feature of Cape Town harbour is the Sturrock graving dock, which is able to accommodate even the largest oil tankers.

4. Air Transport.—South African Airways is included in the South African Railways and Harbours Organization and is the only operator with regular services. There are more than 270 licensed civic airports in South Africa. Of these, however, Jan Smuts Airport (Johannesburg) is the only one of international standard; four others are of national standard. Services are classified as domestic, regional (Johannesburg-Rhodesia-Mozambique), and international (Johannesburg-London, Johannesburg-Australia).

5. Postal Services and Telecommunications.—These are maintained at about 3,150 post offices, 1,580 of which are also money-order offices. The post office has a savings bank and also issues listeners' radio licences.

Included in telephone services are more than 90,000 farm connections. Besides internal telephones there are radiotelephone services to foreign countries and ships at sea. A radiotelegraph and telephone system between South Africa and South West Africa was introduced in 1960. There are more than 1,070 departmental teleprinters in use, and country offices equipped for teleprinter working total over 240. Twenty-two radioteleprinter subscriber circuits are in operation to overseas countries, while international telex service is available via London to 59 overseas countries.

Domestic radio transmission is controlled by the South African Broadcasting Corporation, a public utility established in 1936, which is responsible to the Minister of Post and Telegraphs and derives its revenue from listeners' licences and from its commercial service (Springbok Radio). There are 28 broadcasting stations transmitting three national programs, English, Afrikaans, and commercial. Bantu services are broadcast on the first two of these. There is no television service.

See Cape of Good Hope; Natal; Orange Free State; Transvaal; *see* also references under "South Africa, Republic of" in the Index. (J. H. Wn.; X.)

Bibliography.—*Physical Geography:* Weather Bureau, *Climate of South Africa,* parts i–iv (1957); *South African Geographical Journal* (annual, 1917 *et seq.*); publications of the Republic Geological Survey; L. C. King, *South African Scenery,* 3rd ed. (1964); J. H. Wellington, *Southern Africa,* vol. i (1955); R. S. Adamson, *The Vegetation of South Africa* (1938); E. Palmer and N. Pitman, *Trees of South Africa* (1961); J. R. Ellerman *et al., South African Mammals* (1953); A. Roberts, *Birds of South Africa,* 2nd rev. ed. (1958); W. Rose, *Reptiles and Amphibians of Southern Africa* (1950); J. L. B. Smith, *The Sea Fishes of Southern Africa* (1950); S. H. Skaife, *African Insect Life.* (J. H. Wn.)

The People: I. Schapera (ed.), *Western Civilization and the Natives of South Africa* (1934), *Government and Politics in Tribal Societies* (1956), *The Khoisan Peoples of South Africa* (1930), *The Tswana* (1953), (ed.) *The Bantu-Speaking Tribes of South Africa* (1937); L. Marquard, *The Peoples and Policies of South Africa,* 3rd ed. (1962);

N. J. van Warmelo, *Preliminary Survey of the Bantu Tribes of South Africa* (1935); E. H. Ashton, *The Basuto* (1952); A. T. Bryant, *Olden Times in Zululand and Natal* (1929); M. Hunter, *Reaction to Conquest* (1936); E. J. Krige, *The Social System of the Zulus* (1936); E. J. and J. D. Krige, *The Realm of a Rain-Queen* (1943); H. Kuper, *An African Aristocracy* (1947), *The Swazi* (1952), *Indian People in Natal* (1960); V. G. J. Sheddick, *The Southern Sotho* (1953); J. H. Soga, *The Ama-Xosa* (1932); H. A. Stayt, *The BaVenda* (1931); B. G. M. Sundkler, *Bantu Prophets in South Africa* (1948); J. S. Marais, *The Cape Coloured People 1652–1937* (1939); S. C. Patterson, *Colour and Culture in South Africa* (1953); W. D. Hammond-Tooke, *Bhaca Society* (1962); A. T. Bryant, *A History of the Zulu and Neighbouring Tribes* (1964); L. Kuper, *An African Bourgeoisie* (1965). (J. F. M. M.)

General Historical Surveys: C. W. de Kiewiet, *History of South Africa, Social and Economic* (1941); *The Cambridge History of the British Empire*, vol. viii, 2nd ed. (1963); A. M. Keppel-Jones, *South Africa* (1949); A. J. H. van der Walt *et al.* (eds.), *Geskiedenis van Suid Afrika*, 2 vol. (1951); J. H. Hofmeyr, *South Africa*, 2nd ed. by J. P. Cope (1952); L. Marquard, *The Story of South Africa* (1955); E. A. Walker, *History of Southern Africa*, 3rd ed. (1957), *Historical Atlas of South Africa* (1922).

Special Subjects: W. M. Macmillan, *The Cape Colour Question* (1927), *Bantu, Boer, and Briton* (1929); C. W. de Kiewiet, *British Colonial Policy and the South African Republics, 1848–72* (1929), *The Imperial Factor in South Africa* (1937); R. I. Lovell, *The Struggle for South Africa, 1875–99* (1934); I. D. MacCrone, *Race Attitudes in South Africa* (1937); S. H. Frankel, *Capital Investment in Africa* (1938); P. J. van der Merwe, *Die Trekboer in die Geskiedenis van die Kaapkolonie* (1938); J. S. Marais, *Maynier and the First Boer Republic* (1944), *The Fall of Kruger's Republic* (1961); S. T. van der Horst, *Native Labour in South Africa* (1942), *African Workers in Towns* (1964); M. Roberts and A. E. G. Trollip, *The South African Opposition, 1939–45* (1947); E. A. Walker, *The Great Trek*, 3rd ed. (1948); M. W. Spilhaus, *The First South Africans* (1949); J. van der Poel, *The Jameson Raid* (1952); S. C. Patterson, *The Last Trek: a Study of the Boer People and the Afrikaner Nation* (1957); Gwendolen M. Carter, *The Politics of Inequality* (1958); L. M. Thompson, *The Unification of South Africa, 1902–10* (1960), *Politics in the Republic of South Africa* (1966); F. A. van Jaarsveld, *The Awakening of Afrikaner Nationalism* (1961). C. R. Hill, *Bantustans: the Fragmentation of South Africa* (1964); W. H. Vatcher, *The Rise of Afrikaner Nationalism* (1965); J. E. Spence, *Republic Under Pressure: a Study of South African Foreign Policy* (1965); J. S. Galbraith, *Reluctant Empire: British Policy on the South African Frontier, 1834–1854* (1963); G. H. Le May, *British Supremacy in South Africa, 1899–1907* (1965). Monographs and articles are published in the *Archives Yearbook for South African History* (1938–), the *South African Journal of Economics* (1927), and *The Round Table* (1910 *et seq.*) and in the publications of the South African Institute of Race Relations and the South African Bureau of Racial Affairs.

Biographies: B. Williams, *Cecil Rhodes*, new ed. (1938); I. D. Colvin, *The Life of Jameson*, 2 vol. (1922); E. A. Walker, *Lord de Villiers and His Times* (1925), *W. P. Schreiner* (1937); F. V. Engelenburg, *General Louis Botha* (1929); C. M. van den Heever, *General J. B. M. Hertzog* (1946); J. Rose-Innes, *Autobiography* (1949); E. W. Smith, *The Life and Times of Daniel Lindley* (1949); J. C. Smuts, *Jan Christiaan Smuts* (1952); W. K. Hancock, *Smuts: the Sanguine Years, 1870–1919* (1962); A. J. Luthuli, *Let My People Go* (1962); J. G. Lockhart and C. M. Woodhouse, *Rhodes* (1963); Alan Paton, *Hofmeyr* (1965).

Collected Historical Records: J. Bird, *The Annals of Natal, 1495–1845*, 2 vol. (1888); G. M. Theal, *Basutoland Records, 1833–68*, 3 vol. (1883), *Records of the Cape Colony, 1793–1831*, 36 vol. (1897–1905), *Records of South-Eastern Africa*, 9 vol. (1898–1903); G. W. Eybers, *Select Constitutional Documents Illustrating South African History, 1795–1910* (1918); A. P. Newton, *Select Documents Relating to the Unification of South Africa*, 2 vol. (1924); C. Headlam, *The Milner Papers, South Africa, 1897–1905*, 2 vol. (1931–33). The publications of the Van Riebeeck Society (1918 *et seq.*) comprise original journals, reports, etc. (L. M. T.)

Economy: Government publications: *Monthly Bulletin of Statistics; Official Yearbook of the Republic of South Africa; Farming in South Africa* (monthly); *Commerce and Industry* (monthly); *Statistics of Production: Monthly Abstract of Trade Statistics; Trade and Shipping of South Africa and the Territory of South-West Africa* (annual); M. Cole, *South Africa* (1961); J. H. Wellington, *Southern Africa*, vol. ii (1960); L. P. Green and T. J. D. Fair, *Development in Africa: a Study in Regional Analysis* (1962); D. H. Houghton, *The South African Economy* (1964); L. Katzen, *Gold Mining and the South African Economy* (1965); R. A. Pelletier, *Mineral Resources of South and Central Africa* (1965). (J. H. Wn.)

Official publications and reports of government departments on special topics are published at Pretoria by the government. *See* also the government publication *Union Statistics for Fifty Years, 1910–1960*, and later volumes.

Bibliographies: The most accessible detailed bibliography is in *The Cambridge History of the British Empire*, vol. viii, rev. ed. (1963). *See* also S Mendelssohn, *South African Bibliography*, 2 vol. (1910); G. M. Theal, *Catalogue of Books and Pamphlets Relating to Africa South of the Zambesi* (1912); I. Schapera, *Select Bibliography of South African Native Life and Problems* (1941). For recent publications, *see* the *Quarterly Bulletin of the South African Public Library* (1946–). (L. M. T.)

SOUTH AFRICAN LITERATURE. The earliest writings about South Africa, the country, and its inhabitants, date back several centuries, but South African literature—the work of authors who were born in South Africa, or have lived there and identified themselves with South African thought and life—is of more recent origin because the turmoil of South African affairs prevented any serious creative writing before the 19th century. After a difficult start in its development from the Dutch language, Afrikaans also produced a strong literature of its own, and a review of South African literature must be divided into that written in English and that in Afrikaans.

IN ENGLISH

Some critics hold that English literature in South Africa has followed the main lines of development in other English-speaking countries. This view should be accepted with reservations. It is true that many early writers were British, and that many writers in the 20th century left South Africa, so that their work belongs to other traditions; nevertheless it is also true that South African literature in English has a character of its own.

Poetry.—In poetry this individuality is revealed by an indigenous subject matter and the use of imagery and diction that are characteristically South African. The first poet of note was a Scot, Thomas Pringle (1789–1834), who spent six years in South Africa and later published *Ephemerides* (1828) and *African Sketches* (1834). After a considerable interval there were William Charles Scully (*The Wreck of the Grosvenor*, 1886; *Poems*, 1892), Francis Carey Slater (*Footpaths Thro' the Veld*, 1905; *Settlers and Sunbirds*, 1919; *The Karroo*, 1924), the Rev. Arthur Shearly Cripps (*Lyra Evangelistica*, 1911; *Pilgrimage of Grace*, 1912; *Lake and War*, 1917), and another Scot, Charles Murray (*Hamewith*, 1909; *A Sough o' War*, 1917; *In the Country Places*, 1920). These are poets of the older school, and their verse is of no particular interest except for its "South African" quality. Cullen Gouldsbury (*Songs out of Exile*, 1912; *From the Outposts*, 1914), a poet and novelist of great promise, died at 35, before he had fulfilled it.

The first poet to combine this distinctive "South African" quality with an outstanding poetic gift was Roy Campbell (*q.v.*; 1901–57). William Plomer (1903–) also became prominent with *The Family Tree* (1929) and *Selected Poems* (1940). Later poets include Frank Prince, Rico Martin Titlestad, Anthony Delius, and Guy Butler.

Prose.—Travel books and "letters home" were the earliest prose works written about South Africa, and the works of Thomas Baines, John Barrow, William John Burchell, and David Livingstone (*q.v.*) became well known in the 19th century. The letters of Lady Anne Barnard (*q.v.*), edited as *South Africa a Century Ago* (1901), give a vivid account of the years 1797–1801 at the Cape.

The South African novel has characteristics that differentiate it from other novels in English. Almost all novels deal with topics peculiar to South Africa, although notable exceptions are the first and most famous South African novel, *The Story of an African Farm* (1883), by Olive Schreiner (*q.v.*); *Mary Glenn* (1925), *Sons of Mrs. Aab* (1931), and *What Hath a Man?* (1938) by Sarah Gertrude Millin (*q.v.*); Ethelreda Lewis' *The Harp* (1924); and Elizabeth Charlotte Webster's *Expiring Frog* (1946), all of which deal with problems of general, rather than particular concern. The most popular subject is the trials of the "mixed," or in official definition, Coloured, people, who have been isolated from both black and white persons (*see* further SOUTH AFRICA, REPUBLIC OF: *History*). Sarah Gertrude Millin has become the major writer of this theme. The subject is also treated by William Plomer in his short stories *I Speak of Africa* (1927); and by Laurens van der Post (1906–), in *In a Province* (1934). Peter Abrahams (1919–), a Coloured writer, produced impressive work in *Dark Testament* (1942), *Song of the City* (1945), *Mine Boy* (1946), and *Return to Goli* (1953).

The African has formed the subject of hundreds of novels, from

adventure stories to studies of race relations. Powerful work has been produced by William Plomer (*Turbott Wolfe*, 1925), Sarah Gertrude Millin (*The Coming of the Lord*, 1928; *The Herr Witchdoctor*, 1941), Alan Paton (*Cry, the Beloved Country*, 1948; *Too Late the Phalarope*, 1953), and Dan Jacobson (*The Trap*, 1955; *A Dance in the Sun*, 1956; *The Price of Diamonds*, 1957; *The Evidence of Love*, 1960), who all studied race relations as they affect the individual; and Harry Bloom (*Episode*, 1956; *Whittaker's Wife*, 1962), Nadine Gordimer (*The Lying Days*, 1953), Stuart Cloete (*The Mask*, 1958; *Rags of Glory*, 1963), and Eugene O'Donnell (pseudonym of Laurence James McCauley; *The Night Cometh*, 1960), who all wrote of the political implications of racial problems. Other novels include historical works by Stuart Cloete (*Turning Wheels*, 1937; *Watch for the Dawn*, 1939; *The Hill of Doves*, 1942) and Dorothea Fairbridge (*That Which Hath Been*, 1910); A. A. Smith's *Trader Horn* (1927–28); Percy Fitzpatrick's charming dog story *Jock of the Bushveld* (1907); Horace Rose's *Golden Glory* (1915); and Daphne Muir's *A Virtuous Woman* (1929).

Pauline Smith's *The Little Karroo* (1925) must head any list of South African short stories. William Westrup (d. 1943) and Ernest Glanville (1856–1925) are noteworthy for their tales of animal and veld. Later, Herman Charles Bosman (*Mafeking Road*, 1947) and Nadine Gordimer (*The Soft Voice of the Serpent*, 1953; *Six Feet of the Country*, 1956; *Friday's Footprint*, 1960) became leading exponents of this genre. Other authors who have written short stories are Sarah Gertrude Millin (*Two Bucks Without Hair*, 1957), Stuart Cloete (*The Soldiers' Peaches*, 1959; *The Looking Glass*, 1963), and Jack Cope (*The Tame Ox*, 1960).

No significant drama in English had been produced in South Africa by the mid-20th century. (J. P. L. S.)

IN AFRIKAANS

Afrikaans, one of the youngest members of the Germanic group of languages, and a derivative of 17th-century Netherlandic (*see* NETHERLANDIC LANGUAGE), is the language used by the descendants of Dutch, German, and French colonists who settled in South Africa after 1652 and before the British occupation in 1806. Although the Afrikaans vocabulary is almost identical with the original Netherlandic stock, the grammatical structure has been much simplified. In Afrikaans the noun has no grammatical gender and has lost all inflectional forms except the singular and plural; and the verb has one form for all persons, expresses the tenses by means of modal words, and has lost the difference between strong and weak conjugations. The main traits of Afrikaans were probably present by *c.* 1750.

Toward mid-19th century Afrikaans was used in letters, in some newspapers, and in a few pamphlets, but until the 1920s Dutch (Netherlandic) was still used for literary and official purposes. In 1875 a society, Die Genootskap van Regte Afrikaners, was founded to translate the Bible into Afrikaans and to foster official use of the language. The society founded a monthly magazine, *Die Afrikaanse Patriot*, and started a printing press; but of the literary production of the period 1860–1900 only the Rev. J. L. Cachet's didactic stories, collected as *Sewe Duiwels* (1907), the Rev. S. J. du Toit's historical novel *Di Koningin fan Skeba* (1898), and F. W. Reitz's adaptation of Burns's "Tam o' Shanter" (*"Klaas Geswint en sy Perd,"* 1870) are still read.

The South African War (1899–1902) saw the end of Die Genootskap van Regte Afrikaners, but also the birth of a new spirit in the movement for the recognition of Afrikaans. Phonetic orthography was discarded in favour of a spelling more closely related to Dutch; new writers and scholars took the lead in promoting the use of the language; new journals were founded. Die Afrikaanse Akademie was founded in 1909. As a result of these activities Afrikaans replaced Dutch in schools (1914), university professorships in Afrikaans were instituted, and by 1924 the churches had adopted Afrikaans. In 1925, under the premiership of J. B. M. Hertzog (*q.v.*), Afrikaans was made an official language. The Afrikaans version of the complete Bible was published in 1933, and the first volume of the Afrikaans standard dictionary, begun in 1925, was completed in 1950 (by 1961 vol. iv [H.-I.] had appeared). After 1935 an increasing number of Afrikaans journals, magazines, and newspapers was established.

A literature distinctively Afrikaans in outlook and sentiment grew out of the experience of the South African War. The first poets of real merit whose poems arose from this experience were Eugène Nielen Marais (1872–1936), Johannes François Elias Celliers (1865–1940), Jakob Daniel du Toit (*q.v.;* 1877–1953), and C. F. L. Leipoldt (*q.v.;* 1880–1947). The poem that brought fame to Celliers was *"Die Vlakte"* ("The Plain"; 1908), while Marais made his mark as a poet with outstanding nature-studies. Largely because of the national spirit of this poetry, Afrikaans very soon came to be recognized as the cultural language of South Africa.

This led to a renewed vigour in prose writing, and, although Dutch authors still influenced many novelists, a sound contribution to the beginning of a mature Afrikaans prose style was made by Cornelis Jakob Langenhoven (1873–1932), Gustav Schoeman Preller (1875–1943), Daniel François Malherbe (1881–), and Jochem van Bruggen (1881–1957). While still largely concerned with the national experience of the war, these writers were turning their attention to the realistic and psychological novel. The best examples are Malherbe's *Die Meulenaar* (1926) and van Bruggen's *Ampie* (1924–42), a trilogy on the poor-White problem. Van Bruggen's work as a whole—and especially *Die Sprinkaanbeampte van Sluis* (1933), which was less popular, but of greater literary merit, than *Ampie*—profoundly influenced contemporary novelists. A. A. Pienaar's short stories on wild life in the African jungle, *Uit Oerwoud en Vlakte* (1921), also belong to this period.

Meanwhile younger poets, primarily concerned with the impact of modern science and technology on a mainly agrarian and patriarchal culture, began to write, and the stress shifted from national to personal experience. F. P. (Toon) van den Heever (1894–1956) especially in his *Gedigte* (1919), embodied his doubt, and his rebellion against the religion of his people. Although characterized by differences of outlook, style, and interests, the following verse also belongs to this period: T. J. Haarhoff, *Die Liefde van Catullus* (1933); S. J. du Toit, *Tussen die Dae* (1947); A. G. Visser, *Gedigte* (1925), and *Die Purper Iris* (1930).

About 1925 a new group emerged to dominate the literary scene for the next decade. The most marked advance was in the novel and short story, the realistic and the social novel being particularly favoured. The outstanding figure was Christiaan Maurits van den Heever (1902–57; *Op die Plaas*, 1927; *Droogte*, 1930; *Somer*, 1935; *Laat Vrugte*, 1940; *Die Held*, 1948; and a biography, *Generaal J. B. M. Hertzog*, 1943). Other prominent authors of this group were C. H. Kühn, also known as Mikro (*Toiings*, a trilogy, 1934; *Klaradynstraat*, 1947), Johannes van Melle (*En Ek is Nog Hy*, 1942; *Verspeelde Lente*, 1961), the brothers G. C. and S. B. Hobson (*Kees van die Kalahari*, 1929; *Buks*, 1933), Holmer Johanssen (*Die Onterfdes*, 1944), P. J. Schoeman (*Op die Groot Spore*, 1942), and G. H. Franz (*Moeder Poulin*, 1946). Those of this group who were also poets—van den Heever (*Deining*, 1932; *Aardse Vlam*, 1938) and I. D. du Plessis (*Stryd*, 1935; *Kwatryne*, 1941)—might be regarded as the transition to the next generation, the poets of 1935.

In the poetry of this latter group national experience is at last ignored, and the poems centre on personal feeling. Rhyming verse is discarded in favour of modern experiments. The new form of poetry passed its zenith soon after 1950, but until then it was predominant, the outstanding writers being N. P. van Wyk Louw (*Alleenspraak*, 1935; *Die Dieper Reg*, 1938; *Raka*, 1941), D. J. Opperman (*Heilige Beeste*, 1945; *Negester oor Ninevé*, 1947; *Engel uit die Klip*, 1951), and W. E. G. Louw (*Die Ryke Dwaas*, 1934), one of the founders of the important periodical *Standpunte* (1945–). At the same time, more lyrical poetry was being written by Uys Krige (*Oorlogsgedigte*, 1942; *Hart Sonder Hawe*, 1949), Elisabeth Eybers (*Belydenis in die Skemering*, 1936; *Die Stil Avontuur*, 1939), S. J. Pretorius (*Vonke*, 1944; *Inkeer*, 1948), and Olga Kirsch (*Mure van die Hart*, 1948).

As yet Afrikaans has few dramatists of merit, the most noteworthy being C. F. L. Leipoldt (*Die Heks*, 1923), J. F. W. Grosskopf (*Die Klipdolk, en ander Kortspele*, 1941), Henry Allan Fagan (*Die Ouderling*, 1934), and Uys Krige. Younger dramatists are

W. A. de Klerk (*Die Jaar van die vuur-os*, 1952) and G. J. Beukes (*As ons twee Eers getroud is!*, 1952). An interesting interlude came with two verse-dramas in the classical style: D. J. Opperman's *Periandros van Korinthe* (1954) and N. P. van Wyk Louw's *Germanicus* (1956).

Quite a stir has been caused by avant-garde novels, stimulated by English and French models. Of these, *Lobola vir die Lewe* (1962) by André Brink, and *Sewe Dae by die Silbersteins* (1962) by Étienne Leroux, are worthy of note. (A. J. Ce.)

Bibliography.—*Anthologies:* E. H. Crouch (ed.), *A Treasury of South African Poetry and Verse* (1907; 2nd ed., 1909); F. C. Slater (ed.), *The Centenary Book of South African Verse, 1820 to 1925* (1925; revised, 1945); E. R. Seary (ed.), *South African Short Stories* (1947); M. A. Bax-Botha and D. Bax (eds.), *Afrikaans verhalend Proza* (1953); R. Macnab (ed.), *Poets in South Africa* (1958).

English Literature: M. Nathan, *South African Literature: a General Survey* (1925); G. M. Miller and H. Sergeant, *A Critical Survey of South African Poetry in English* (1957). *See* also S. Mendelssohn, *Mendelssohn's South African Bibliography*, 2 vol. (1910; 2nd ed., 1957).

Afrikaans Literature: L. van Niekerk, *De eerste afrikaanse Taalbeweging en letterkundige voortbrengselen* (1920); E. C. Pienaar, *Taal en poësie van die twede afrikaanse taalbeweging*, 4th ed. (1931); G. Dekker, *Afrikaanse Literatuurgeskiedenis* (1935; 5th ed., 1960); P. C. Schoones, *Die Prosa van die tweede afrikaanse beweging*, 3rd ed. (1939); C. M. van den Heever and P. de V. Pienaar (eds.), *Kultuurgeskiedenis van die Afrikaner*, 3 vol. (1945–50); D. J. Opperman, *Digters van Dertig* (1953). *See* also T. J. Haarhoff, *Afrikaans, its Origins and Development* (1936).

SOUTH AFRICAN WAR, also known as the Anglo-Boer War or Boer War, was fought between Great Britain and the two Boer republics, the South African Republic (Transvaal) and the Orange Free State, between October 1899 and May 1902.

Causes.—The war had its origins in the relations that had developed between Britain and the South African Republic. Britain, especially after 1895, when Joseph Chamberlain became colonial secretary, aimed at establishing its supremacy over southern Africa and at preventing the encroachment of its colonial rivals, France and Germany, while Paul Kruger (*q.v.*), president of the South African Republic, was determined to maintain the political and economic independence of his Boer republic. Several factors helped build up the tension that ultimately led to war. The first of these was the issue of British suzerainty. In 1877 Britain annexed the Transvaal, but in consequence of a Boer rebellion (December 1880) and the defeat of British troops at Majuba (February 1881), by the Pretoria Convention (August 1881) the Transvaal was granted self-government subject to the suzerainty of Queen Victoria; three years later, by the London Convention (February 1884), it received a fuller measure of independence (*see* Transvaal: *History*). But Britain's retention of control over the republic's treaties, other than with the Orange Free State, and its claim at a later stage that it had not abandoned suzerainty were a source of resentment to Kruger and his burghers (enfranchised citizens).

The second factor was the influx of large numbers of immigrants (known as *uitlanders*), who came into the republic from Britain and elsewhere following the discovery of gold on the Witwatersrand in 1886. Although they paid the bulk of the republic's taxation, they were denied any effective share in the government; the English language was banned in the law courts and schools; and juries consisted only of burghers. The *uitlanders* also complained of the government's grant of monopolies in dynamite and other commodities vital to the mining industry and of corruption on the part of the officials. But Kruger, fearing that if given political rights the *uitlanders* might gain control of the Boer republic, rejected their demands for the franchise and for redress of their grievances.

The third factor was the conflict between Kruger's ideal of independence and isolation and the attempts of Cecil John Rhodes (*q.v.*), head of the British South Africa Company (which had control over the country north of the Limpopo River) and prime minister of the Cape Colony, to bring about an economic federation of the whole of southern Africa that would lead to a political federation under the British flag. The British occupation between 1885 and 1895 of territory to the north, west, and east of the republic, which encircled it and cut it off from access to the sea, aroused the fears of Kruger. He resisted Rhodes's attempts to include the republic in a customs union, and to ensure economic independence from the British colonies to the south he gave Hollander interests a concession to build a railway line from Pretoria to Delagoa Bay in Portuguese East Africa. A fierce rivalry in railway rates developed between the Netherlands Railway Company, which was responsible for this line, and the Cape authorities. War seemed imminent when in October 1895 Kruger closed the Vaal River drifts (fords) that the Cape authorities had been using to send goods into the Transvaal by wagon to avoid the high rates imposed on the Transvaal section of their Cape to Johannesburg line; but under pressure from Chamberlain, Kruger reopened the drifts on Nov. 7.

The Jameson Raid.—Hostility between the Boers and the British was further increased by the fiasco of the Jameson Raid (December 1895). Rhodes arranged that L. S. Jameson should lead an armed force into the republic from Pitsani, in Bechuanaland, to support a rising of the *uitlanders* against the Kruger government. The rising failed to materialize, and Jameson's raid ended in his capture at Doornkop (Jan. 2, 1896), but it aroused among the Boers deep suspicion of Britain's intentions. (*See* Jameson, Sir Leander Starr.) Kruger, convinced that war with Britain could be only a matter of time and, by receiving a congratulatory telegram from the German emperor, William II, encouraged to hope for German assistance in the event of war, began a systematic program of rearmament. Moreover, Kruger's prestige was increased not only among his own people but also among the Cape Dutch and the Free Staters, and in 1897 he formed an alliance with the Orange Free State.

Relations between Britain and the South African Republic steadily deteriorated after the raid. Sir Alfred (later Viscount) Milner, who was appointed high commissioner in 1897, was determined to uphold British supremacy in South Africa and was prepared to "work up a crisis" by demanding the franchise and other civic rights for the *uitlanders*. Kruger, on the other hand, regarded firmness on the franchise question as essential for the maintenance of Boer independence. At the Bloemfontein Conference (June 1899) both the high commissioner and Kruger adopted an attitude of intransigence, and the negotiations broke down over the question of the *uitlander* franchise. Thereafter Chamberlain openly adopted the *uitlanders'* cause; for him, too, it was part of the more important issue of British paramountcy. In August, J. C. Smuts, the state attorney of the South African Republic, offered a franchise for the *uitlanders* after five years' residence, but Chamberlain rejected the conditions attached, which included abandonment of the British claim to suzerainty over the republic. Both sides then drifted rapidly into war. On Sept. 22 the British government drew up a drastic program of demands and decided to mobilize an army corps. Mobilization also proceeded both in the South African Republic and in the Orange Free State, which, although having endeavoured throughout to avert war, had resolved to support its ally. On Oct. 9 the republics presented an ultimatum, demanding the withdrawal of British troops from the frontiers, which only narrowly forestalled the British ultimatum.

THE COURSE OF THE WAR

On the expiry of the Boer ultimatum (Oct. 11, 1899) there were only 27,000 British troops in South Africa, while the republics had about 48,000 men under arms. The British Army Corps that was later dispatched to South Africa included highly trained field artillery and cavalry, as well as some mounted infantry, which were at that time unique in Europe. But it was hampered by incompetence in the high command and by the difficulty of the terrain—vast expanses of open country with poor communications. The Boers had an initial advantage in their mobility and their experience of the country. They were mounted riflemen, loosely organized in district commandos (*see* Commando) under elected officers, each army being supported by a small corps of state artillery. But in the early stages of the war the lack of discipline and the weakness in the command of the Boer forces offset these advantages.

The war can be divided into two stages. The first was the period

of regular warfare, which ended with the annexation of the Transvaal in September 1900; the second was the period of prolonged skirmishing, which continued until May 1902, when peace was signed.

The war opened with the investment of Mafeking by a Transvaal force under Gen. P. A. Cronje and the envelopment of Kimberley by Free State commandos under C. J. Wessels (October 1899). Meanwhile, a large Boer army under Gen. P. J. Joubert, commander of the Transvaal forces, had invaded the northern triangle of Natal, where Lieut. Gen. Sir George White was commander in chief of 16,000 British forces, most of which were divided between Ladysmith and Dundee. The Boers advanced on these centres from the north, east, and west. Although one Boer force under Lukas Meyer was repulsed at Talana Hill, outside Dundee, on Oct. 20, and another under Gen. J. H. M. Kock was defeated at Elandslaagte on the following day, the Boers forced the British to fall back on Ladysmith. White's attempt to break the Boer lines resulted in the Battle of Ladysmith (Oct. 30), in which the British losses totaled more than 1,200 (69 killed, 249 wounded, 954 missing) against the Boers' 200. Thereupon the Boers started the long siege of Ladysmith and seemed about to overrun the whole of Natal.

Buller's Offensive.—Gen. Sir Redvers Buller, who had been appointed to the supreme command in South Africa, arrived in Cape Town (Oct. 31, 1899) followed by his army corps. The situation with which he had to deal was serious. On Nov. 1, M. T. Steyn, president of the Orange Free State, sent his commandos into the Cape Colony, where they found support among the colonists, despite the imposition of martial law by the Cape government in the invaded districts. On the western front Mafeking and Kimberley were in a state of siege, and in Natal the British position was critical. Buller split his army corps into three unequal parts. One, under the command of Lord Methuen, was to relieve Kimberley and Mafeking in turn; the second, under the command of Lieut. Gen. Sir William Gatacre, was to contain the commandos invading the Cape Colony; the third, under Buller's own command, was to relieve Ladysmith. But Buller's whole scheme broke down in the series of disasters that the British sustained during "Black Week" in the middle of December. On Dec. 10 Gatacre essayed a night march and attack on the Boer position near Stormberg Junction, but, having been misled by his guides, he was himself surprised and forced to return to Molteno with a total loss of about 700 men. On the following day Methuen, who had succeeded in crossing the Modder River on his march to the relief of Kimberley, was defeated in the Magersfontein Hills, just south of the besieged town, by the forces of the Boer generals J. H. de la Rey and Cronje. The British lost about 1,000 men, including Maj. Gen. A. J. Wauchope. The Boers' victory was largely due to De la Rey's original and effective measure of digging and defending long narrow trenches along the foot of the hills. Then on Dec. 15 Buller was decisively defeated at Colenso, on the Tugela River, by the able Boer general Louis Botha. Botha routed the British forces and lost 29 men against the British loss of 1,139. Buller immediately sent orders to White, who was still besieged in Ladysmith, to destroy his guns and surrender the town. But White refused to do so, and the siege continued.

In England "Black Week" brought a reaction of deep shock. The Cabinet appointed (Dec. 16) Field Marshal Lord Roberts to the supreme command, with Maj. Gen. Lord Kitchener as his chief of staff. Volunteer forces, both mounted and infantry, were raised, and the militia were invited to come forward. The empire—Australia, New Zealand, Canada, and others—contributed contingents (which amounted in the end to a total of 40,000 volunteers), and at the same time the Boer forces were augmented by volunteers from many countries of Europe. In the early days of 1900 the struggle swayed to and fro. On the western front, in the far north Cape, the settlement of Kuruman surrendered (January 1900) to the Boers' besieging force; in the Cape midlands Maj. Gen. John French (later 1st earl of Ypres) was pressing on Colesberg; while in Natal the fate of Ladysmith hung in the balance. On Jan. 6 the Boers made an attempt to overwhelm the besieged town but without success. This was followed by two attempts on the part of Buller to relieve it. In the first of these he was forced to withdraw with very heavy losses in the Battle of Spion Kop (Jan. 24). On Feb. 5 he again crossed the Tugela River but failed through indecision to press the advantage he had attained and was defeated at Vaal Krantz.

Roberts' Advance.—Roberts and Kitchener arrived in Cape Town on Jan. 10. Roberts' plan was to assemble his army on the Modder River in the west and to carry out an eastward march to strike at Bloemfontein, the capital of the Orange Free State. He kept his plan secret and created the impression that he was concentrating his forces in the Cape midlands area, near Colesberg. On Feb. 11 Roberts began his march from between the Modder and Orange rivers. He sent French from Ramdam (near Graspan) on a flanking movement northeast toward Kimberley, where Rhodes's vociferous demands and the serious health situation in the town, whose populace was near starvation, created an urgent need of relief. Sweeping round Cronje's main force in the Magersfontein Hills, French put an end to the siege of Kimberley on Feb. 15. Thereupon Cronje abandoned his Magersfontein defenses and escaped eastward through the British lines, which were also moving east. But, hampered by his long train of wagons, he was surrounded at Paardeberg (Perdeberg), southeast of Kimberley, and after a long and stubborn resistance was forced to surrender to Roberts with about 4,000 men on Feb. 27. The next day Buller in Natal at last succeeded in forcing the Tugela defense line and relieved Ladysmith, which by this time was in a desperate state. The Boers' besieging force retreated in disorder. The turning point in the first stage of the war had been reached. Kimberley and Ladysmith had been relieved, the Boer Army in the west had been captured almost entire, that in Natal was retreating in confusion, and in the Cape the invaders De la Rey and H. Schoeman had been forced to flee to Bloemfontein. On March 5 Presidents Kruger and Steyn offered peace proposals to the British government on the basis of continued independence for the republics, but after the fall of Bloemfontein these were rejected by the British prime minister, the marquess of Salisbury.

Capture of Bloemfontein.—Roberts, meeting with little resistance from the Boers, continued his advance on Bloemfontein, which he entered and occupied on March 13. He considered that the war in the Free State was over and on March 15 issued a proclamation allowing those burghers who laid down their arms to return to their homes. But the Boers gained a valuable breathing space by the delay of Roberts' army in Bloemfontein, where an outbreak of enteric fever took a heavy toll of lives. At the same time the death of Joubert (March 27) brought the more vigorous Louis Botha into command of the Transvaal forces. Further strengthened by a decision to reduce all transport to a minimum and to impose stricter discipline on their forces, the Boer leaders again took the initiative. Christiaan de Wet captured the waterworks at Sannah's Post outside Bloemfontein after inflicting a defeat on Gen. R. G. Broadwood (March 31) and followed this up by securing the surrender of a British detachment near Dewetsdorp, 40 mi. SE of Bloemfontein, on April 4.

Relief of Mafeking.—Early in May, Roberts set out from Bloemfontein on his march to the Boer capital of Pretoria. On May 17 a flying column, from near Kimberley, relieved Mafeking, where Col R. Baden-Powell had maintained morale in increasingly difficult circumstances. Roberts, having reached the Vaal River and announced the annexation of the Orange Free State as the Orange River Colony (May 24), continued his successful advance, invaded the Transvaal, and on May 31 entered Johannesburg. With practically no resistance he was able on June 5 to occupy Pretoria, from which Kruger had already fled east along the Delagoa Bay railway line to Middelburg. The occupation of Pretoria was secured by Roberts' displacement of Botha's force at Diamond Hill (June 11–12), east of the capital. But the end of hostilities was not yet in sight. As Roberts concentrated on his eastward offensive against Botha, his rear was threatened by De la Rey, who harried the British forces in the western Transvaal. At the same time De Wet was engaged in disrupting railway communications in the Orange Free State. However, on July 30, a party of more than 4,000 Boers under Marthinus Prinsloo was forced to surren-

der after being trapped in the Brandwater Basin, south of Bethlehem, in the Orange Free State, but De Wet himself had already escaped and made his way northward.

Roberts' main army was joined (Aug. 14) by the forces of Buller, who had marched through northern Natal. On Aug. 27 they defeated the Boer commandos at Bergendal, near Dalmanutha, 45 mi. E of Middelburg, and caused them to disperse. Kruger, who had been fleeing east in advance of the British forces, left the republic (Sept. 11) and entered Portuguese East Africa. On Oct. 19 he set sail for Europe, where it was felt he would be safe and might obtain aid for the Boer cause. S. W. Burger was appointed to act as president in his absence.

The Guerrilla War.—Roberts had already proclaimed the annexation of the Transvaal (Sept. 1). Then, believing that the war was over, he laid down his command in South Africa in favour of Kitchener (November 1900) and in January returned to England. But a new war had already begun, in which the Boers adopted guerrilla tactics. They concentrated on disrupting railway communications, capturing isolated posts, and avoiding contact with the columns sent in pursuit of them. In the Orange Free State, De Wet roused the burghers to the new tactics and raised local commandos. He was unable at the beginning of December 1900 to cross the swollen Orange River and carry out his proposed invasion of the Cape, but two of his generals, P. H. Kritzinger and J. B. M. Hertzog, entered the colony (Dec. 16) and made their way southward, evading the British columns sent in pursuit of them. The new tactics were also adopted in the Transvaal, De la Rey operating in the west and Botha in the east.

Concentration Camps.—Kitchener, on taking over command in November 1900, was faced with the problem of dealing with the new war. Having appealed to Britain for more manpower, he extended the "scorched earth" policy that Roberts had already begun. Boer farms were devastated, and the women and children of the men on commando were interned in concentration camps, where for a time the death rate from disease was very high. From the British point of view Kitchener's internment policy was not a success; it added an extra burden to the fighting forces, and it further stimulated the spirit of the Boers. At the end of January 1901 Kitchener launched two large-scale offensives, the first against De Wet, who on Feb. 10 crossed the Orange River to invade the Cape Colony. But closely pursued by Col. H. C. O. (later Viscount) Plumer, he gave up his attempt and retired (Feb. 28) across the river. A second offensive, under French, was launched (Jan. 28) against Botha in the eastern Transvaal, from where he was planning an invasion of Natal. Botha managed to escape with his forces from Ermelo (Feb. 6), whereupon he was asked by Kitchener to discuss peace terms. A meeting took place at Middelburg on Feb. 28, 1901. Kitchener proposed peace terms, which included a general amnesty for Boers and for Cape and Natal rebels and the grant of self-government to the northern colonies as soon as possible. But, largely because the British government was not in favour of an amnesty for the rebels, the negotiations broke down (March 16).

Blockhouses.—As the Boers continued their skirmishing tactics, Kitchener adopted a policy of building corrugated iron blockhouses along the railway lines. When these proved effective, he extended them away from the railways to form fortification lines across the Free State and later across the Transvaal, so that eventually the whole country was split up into small areas, within which the Boer forces could be localized and overwhelmed by British troops. Meanwhile, he continued his policy of interning women and children and on Aug. 7 issued a proclamation to the effect that all Boers who had not surrendered by Sept. 15 would be punished and the leaders permanently banished from South Africa. But the Boers, who had been exhorted by the exiled President Kruger to continue fighting, were not daunted. Botha made an unsuccessful attempt at invading Natal early in September. On Sept. 3 Smuts invaded the Cape Colony and by the end of October was within 50 mi. of Cape Town. During the next six months Kitchener concentrated his efforts on the three able Boer generals who were still harassing the British forces: Botha in the eastern Transvaal, De la Rey in the western Transvaal, and De Wet in the eastern Free State. The last notable Boer victory was gained by De Wet at Tweefontein, 50 mi. W of Harrismith, on Christmas Day. Thereafter Kitchener brought his blockhouse system to completion and by a series of raids forced Botha to retire from the eastern Transvaal in February 1902. In the same month he launched intensive drives across the Free State, carefully planned to comb the countryside, and in this way he succeeded in rounding up large numbers of De Wet's men. In the western Transvaal, De la Rey's forces were gradually worn down during March, while in the Cape Smuts continued his skirmishing. The Boers' resistance steadily weakened. On April 12 their leaders met Kitchener in Pretoria and agreed to discuss peace proposals.

Treaty of Vereeniging.—On May 15 representatives of the Boer governments and 60 delegates elected by the Boer commandos of the two republics met at Vereeniging to discuss peace. A deputation, consisting of Botha and four other Boer leaders, was chosen to negotiate with Kitchener and Lord Milner, the high commissioner, for limited independence. The British, however, stood firmly by their Middelburg proposals (*see* above). A draft treaty, which consisted essentially of these proposals, was drawn up and, in the form approved by the British government, was submitted to the convention. On May 31, 1902, the treaty was agreed to and signed.

The terms of the treaty were as follows: (1) the burghers to lay down their arms and acknowledge King Edward VII as their lawful sovereign; (2) all prisoners outside South Africa and internees on acknowledging themselves subjects of the king to be repatriated; (3) no burghers who should surrender to be deprived of their personal liberty or property; (4) no proceedings to be taken against burghers for any legitimate acts of war; (5) the Dutch language to be taught in public schools where the parents of the children should desire it, and to be allowed in courts of law when necessary for the better administration of justice; (6) the possession of rifles to be allowed on the taking out of licences; (7) military administration to be succeeded by civil government as soon as possible, and representative institutions leading to self-government to be introduced as soon as circumstances permitted; (8) the question of granting the franchise to "natives" not to be decided until after the introduction of self-government; (9) no special tax to be imposed on landed property to defray the expenses of war; (10) district commissions to be set up to facilitate resettlement. For this purpose the British government would be willing to grant a sum of £3,000,000 and also offer loans on liberal repayment terms.

The war, which had lasted two and a half years, had cost Britain about 22,000 men and the Boers not far short of 6,000, as well as 4,000 women and 16,000 children who had died in the concentration camps. At the end of the war there were almost 26,000 Boers in prison camps and more than 116,000 were confined in concentration camps.

RESULTS OF THE WAR

As a result of the war the two Boer republics ceased to exist, and crown colony rule was established in the Transvaal and in the Orange River Colony. The whole of South Africa was brought under the British crown, and thus the subsequent political union of the former republics with the existing colonies of the Cape and Natal was made possible. The Treaty of Vereeniging contained a promise of self-government for the new colonies, and this was granted to the Transvaal in December 1906 and to the Orange River Colony in June 1907. For both political and economic reasons a movement for the closer union of the four self-governing colonies developed, and this led to the establishment of the Union of South Africa in 1910.

The war had the effect of arousing a new national consciousness among the Afrikaner people throughout South Africa. This gave rise first to a cultural and then to a political revival. In the crown colonies, in opposition to Lord Milner's policy of anglicization, a "Christian National Education" movement developed. Private schools were established, and these fostered Afrikaner national sentiment among the rising generation. National consciousness expressed itself also in a language movement. The Taalbond (lan-

guage union), revived in 1903, encouraged the use of High Dutch to combat English. Attempts were also made to promote Afrikaans, the spoken language of the Afrikaner people, as a written language and to obtain official recognition for it. The Afrikaans poetry and prose that the language campaign produced gave rise to a virile national literature. At the same time the Afrikaners in the new colonies were beginning to form political organizations. In the Transvaal the Het Volk ("The People") Party was formed in January 1905, followed in May 1906 in the Orange River Colony by the Orangia Unie ("Orange Union"), both pledged to the ideal of self-government.

For Britain the war brought to an end the period of late-19th-century imperialism. Moreover, it revealed to Britain many defects in its military organization and taught it tactical lessons, such as the importance of accurate rifle fire and of taking cover, which were to prove valuable in World War I.

See also references under "South African War" in the Index.

BIBLIOGRAPHY.—L. S. Amery (ed.), *The Times History of the South African War, 1899–1900*, 7 vol. (1900–09); Sir J. F. Maurice and staff, *History of the War in South Africa* (comp. by direction of His Majesty's Government), 4 vol. (1906–10); C. Headlam (ed.), *The Milner Papers: South Africa, 1897–1899, 1899–1905*, 2 vol. (1931–33); D. Reitz, *Commando*, 2nd ed. (1953); C. R. de Wet, *The Three Years' War* (1902); A. C. Martin, *The Concentration Camps, 1900–1902* (1957); R. Kruger, *Good-bye Dolly Gray* (1959). *See* J. D. Kestell and D. E. van Velden, *The Peace Negotiations Between the Governments of the South African Republic and the Orange Free State, and the Representatives of the British Government* (1912). For detailed bibliography (official publications, contemporary works and later works) *see* bibliography section xii in *Cambridge History of the British Empire*, vol. viii, 2nd ed. (1963). (C. E. M. O'D.)

SOUTH AMERICA, the fourth largest of the continents, is located between longitudes 34° and 82° W and latitudes 13° N and 55° S. It has an approximate area of 6,879,000 sq.mi., about 13% of the land surface of the earth. Its estimated population (1963–64) was 160,000,000, equal to approximately 23 persons per sq.mi. The continent, with adjacent islands, is made up of 11 sovereign republics—Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Guyana (until 1966 the colony of British Guiana), Paraguay, Peru, Uruguay, Venezuela—and 2 European colonies—French Guiana, and Surinam, or Dutch Guiana. South America is joined to Central America by the Isthmus of Panama.

Geographically, the continent is dominated by the Andes Mountain chain, which extends in a generally north-south direction for 4,500 mi. The Magdalena (Colombia), Orinoco (Venezuela), São Francisco (Brazil), and Río de la Plata (Argentina, Brazil, Paraguay, and Uruguay) are major river systems, but each is dwarfed by the Amazon, the largest river in the world in terms of area drained and volume of flow. The barren, saline Atacama Desert of the west coast stands in marked contrast with the tropical jungles of Bolivia, Brazil, Ecuador, Peru, and the Guianas on the one hand and temperate Valdivia, Chile, with an average annual rainfall in excess of 100 in., on the other.

This article is organized into sections as described in the following outline:

I. PHYSICAL GEOGRAPHY

A. Geology

The most ancient core areas of South America lie in the Brazilian Shield and the Guiana Highlands. Analogous to the Laurentian Shield of North America, they consist largely of a basement complex of pre-Cambrian gneisses and schists, exposed in some areas and overlain in others by sedimentaries mostly of Paleozoic Age. Many of the older formations have their counterparts in age, composition, and attitude in southern and western Africa, and the belief is widely held, although not universally accepted, that the two continents were joined until Jurassic time. Much of the rest of the South American continent must ultimately owe its existence to deposition in marginal seas of the products of erosion from the core areas. Extensive and repeated periods of uplift and subsidence brought about major transgressions and retreats of the sea over these marginal areas, with consequent deposition of massive beds of marine sediments and a gradual increase in the size of the continent.

A vast geosyncline on the western and northern margins of the continent became the repository in the Mesozoic Era of accumulations of sediments many thousands of feet thick. In the Upper Cretaceous the great weight of these deposits gave rise to crustal instability with crumpling and uplift of the strata, a process that continued throughout the Tertiary and resulted in the formation of the Andes. Intrusions of magma accompanied this activity, and volcanism was widespread. Volcanic activity is still a feature of the Andes, although it has entirely ceased in such older lands to the east as the Paraná Basin of southern Brazil and Paraguay, where successive and widespread outflows of basaltic lava engulfed the land in the Jurassic period.

Extensive glaciation has occurred in South America in at least two distinct geologic epochs, the Carboniferous and the Pleistocene. In the former case the main centre appears to have been in southern Brazil; in the latter, in the southern Andes. A few small relics of a formerly extensive continental ice sheet still survive in southern Patagonia.

B. Physiography

1. The Coasts.—The South American continent rises abruptly from the ocean floor along much of its coast, but the steepness of the continental margin is more marked on the western side than on the eastern. The west coast between latitudes 5° and 35° S displays little development of a continental shelf and drops off sharply to a series of deep submarine trenches parallel to the shore. In the neighbourhood of 26° S the Richards Deep exceeds 25,000 ft. in depth within 50 mi. of the coast. In general, the Pacific coast north of the Patagonian archipelago is but little articulated and is completely lacking in good natural harbours.

From 41° S to Cape Horn the Pacific coastal zone is a series of islands, channels, and fjords, results of subsidence and glaciation. North of 41° S the land between the Andes and the sea averages 40–50 mi. in width. In southern and central Chile a longitudinal interior valley is bordered on the west by a coast range whose elevation increases toward the north. Northern Chile has a high cliff coast situated in front of a series of interior basins. In Peru, Ecuador, and Colombia the coast range is largely absent.

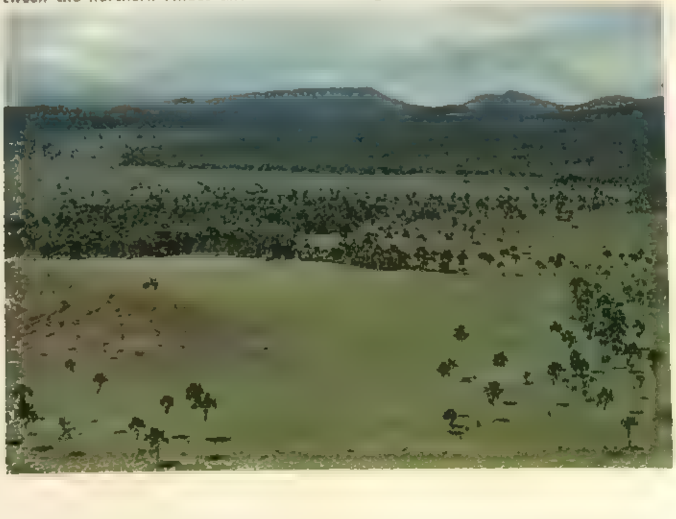
The east coast is in strong contrast with the west. Between Cape São Roque and Rio Grande do Sul the coast generally follows the sinuous eastern edge of the Brazilian Plateau. Along this coast lie some of the world's most beautiful harbours, notably those of Recife, Salvador, Vitória, Rio de Janeiro, and Florianópolis. Farther south the estuary of the Río de la Plata is the most marked indentation of the entire South American coast. It is the outlet for the combined waters of the Paraná and Uruguay river systems. The Patagonian coast is poorly endowed with natural harbours. Although many embayments occur, their utility is reduced by shallow water, high tidal ranges, and coastal cliffs.

The east coast of South America has few islands, and these, except Trinidad, off the coast of Venezuela, and Marajó, at the mouth of the Amazon, are mostly small. Trinidad (area 1,980 sq.mi.) is separated from the continent by the Gulf of Paria. Along the island's northern edge is a range of mountains with a

Lake Maracaibo, Venezuela, centre of the petroleum industry of Latin America. The lake receives the drainage of the slopes of the eastern cordillera of the Andes Mountains. Although it is called a lake, it is, in fact, a 133-mi.-long bay of the Gulf of Venezuela

Magdalena River, Colombia. A major commercial artery for several centuries, the Magdalena drains the central and eastern slopes of the Colombian Andes and empties into the Caribbean near Barranquilla

Plains near Esmeralda, Venezuela, part of the Orinoco watershed which lies between the northern Andes and the Guiana Highlands (in background)



PHOTOGRAPHS, (TOP) J. D. BARNELL—IMAGE INTERNATIONAL, (BOTTOM LEFT) THOMAS B. HOLLYMAN—PHOTO RESEARCHERS, INC., (BOTTOM RIGHT) HECTOR R. ACEBES—PHOTO RESEARCHERS, INC.

SOUTH AMERICA

Machu Picchu ruins, Peru. Ancient Inca village perched high in the Andes overlooking Urubamba canyon 70 mi. from Cuzco, capital of the Inca empire. Behind the ruins is the peak of Machu Picchu from which the village takes its name

Range in the Andes Mountains, central Chile, part of a s... mountain ranges stretching the entire length of the Americas and attaining its highest ...ion along the Chile-Argentina border

Volcanic peaks, Imbabura province, northern Ecuador. The eastern cordillera of the Andes in Ecuador is one of the world's major areas of volcanic activity

PHOTOGRAPHS, (TOP LEFT) MARC AND EVELYNE BERNHEIM—RAPHO GUILLUMETTE, (TOP RIGHT) RONNY JAQUES—PHOTO RESEARCHERS, INC., (BOTTOM) THOMAS B. HOLLYMAN—PHOTO RESEARCHERS INC

Brasília. A [illegible] the Plaza of the Three Powers in the public-building section of Brazil's modern capital

Tropical vegetation, Amazon basin, Brazil. The wet, warm climate of the lowland equatorial portion of South America fosters luxuriant plant and animal growth. The tribal Indian culture to which the hunter belongs is undergoing modification

Some aspects of land usage and topography in South America: (left) a fertile mountain valley under cultivation near La Paz, Bolivia; (right) cattle in Santa Cruz district of eastern Bolivia, a developing farming area; (below) Marajó Island at the mouth of the Amazon River, Brazil

PHOTOGRAPHS, (TOP) J. D. BARNELL—IMAGE INTERNATIONAL, (LEFT) KARL WEIDMANN FROM NATIONAL AUDUBON SOCIETY, (CENTRE LEFT) THOMAS [illegible] HOLLYMAN—PHOTO RESEARCHERS, INC., (CENTRE RIGHT) RENE BURRI—MAGNUM, (BOTTOM RIGHT) SIMON—PHOTO RESEARCHERS, INC

SOUTH AMERICA

Boat dock on the Paraná River north of [illegible] ;s Aires. Larger of the two river systems that empty into the Río de [illegible] ;ta, the Paraná rises in Brazil and drains large areas of Brazil, the Parag[illegible] ;asin, and the pampas

Iguaçu Falls. Along part of Brazil-Argentina border, the 2½-mi.-wide Iguaçu River drops about 270 ft. over a series of about 200 falls separated by islands and grouped roughly in a horseshoe pattern

Copper mine, Chuquicamata, Chile. [illegible] ;veral mines situated in the Atacama Desert, it works the largest p[illegible] ;r deposits in the world

Tierra del Fuego. Glacial lakes, deciduous forests, and rugged, and, in places, almost barren terrain are typical of this southernmost part of South America, separated from the mainland by the Strait of Magellan

PHOTOGRAPHS (TOP LEFT) ED DREWS—PHOTO RESEARCHERS, INC., (TOP RIGHT) ALPHA, (CENTRE RIGHT) FPG, (BOTTOM) ROGER TORY PETERSON—PHOTO RESEARCHERS, INC

FIG 1.—COUNTRIES AND MAJOR CITIES AND SURFACE FEATURES OF SOUTH AMERICA

mainland of Peru, the Galápagos Islands on the Equator, and the islands of the Guayas River delta near Guayaquil.

2. Landforms.—The continent of South America is composed of the following major surface forms: the cordilleras of the Andes in the west and north; the Guiana, Brazilian, and Patagonian uplands in the east and south; and the lowlands that divide the highlands from one another, namely, from north to south, the Orinoco depression, the Amazon Basin, the Paraguay Basin, and the pampas.

The Andes, one of the greatest mountain systems in the world, are more than 4,500 mi. long and are exceeded in average height only by the Himalayas. The southern Andes are a single chain, but north of latitude 28° S they form a complex system. Widening to a broad plateau in Bolivia, they narrow to a series of parallel ranges in Peru, continue in two parallel ranges in Ecuador, and then fan out in Colombia into three distinct ranges, two of which end near the Caribbean shore while the easternmost continues through northern Venezuela into Trinidad. Volcanic cones are present in three separate areas (for further details, *see* ANDES).

The Guiana Highlands and the Brazilian Shield together represent the ancient core area of the South American continent, divided into two unequal sections by the Amazon geosyncline. Waterfalls and rapids occur along all the streams passing over the edge of these uplands. Angel Fall in the Guiana Highlands is the highest uninterrupted waterfall in the world (3,212 ft.). The Paulo Afonso Falls on the São Francisco River in northeastern Brazil were harnessed in the late 1940s for a major hydroelectric installation. Crystalline rocks, in the areas of high rainfall, weather into characteristic dome-like shapes, as exemplified by the celebrated Sugar Loaf overlooking the bay at Rio de Janeiro. The highest point in the Guiana Highlands is Mt. Roraima (9,094 ft.), and on the Brazilian Shield, Pico da Bandeira (9,495 ft.). Patagonia is mostly a succession of arid plateaus, characterized by horizontal beds of sandstone and widespread lava caps.

In South America the development of lowlands is far greater in the interior of the continent than near the periphery. Only on the Atlantic coast, around the mouth of the Amazon and near the Río de la Plata, do extensive lowlands occur next to the sea. Typically, the lowland areas are pouch shaped, with their greatest width inland. The Llanos del Orinoco occupy a structural depression between the northern Andes and the Guiana Highlands. The Orinoco River does not flow along the central axis of the llanos but at the base of the Guiana Highlands, in part a consequence of the great deposition of alluvium by tributary streams issuing from

maximum height of 3,085 ft., which is geologically a continuation of the Cumaná Range of Venezuela. On the south side of this island is the famous asphalt lake—the largest known deposit of its kind. West and north of Trinidad and lying farther off the coast are several small islands of historical interest and considerable commercial importance: Tobago, Margarita, Tortuga, and the Curaçao group. The main island of the Fernando de Noronha group of volcanic islands, 230 mi. off Cape São Roque, has an area of only 7 sq.mi. Although separated from the mainland by a channel 13,000 ft. deep, it really stands upon the submerged corner of the continent. The Falkland Islands in latitude 51° S also stand upon the submerged edge of the continent. Their fauna and flora indicate that they were once a part of the mainland, from which they are now separated by shallow water. On the west coast north of the Patagonian archipelago are the Juan Fernández Islands west of Valparaíso, the famous guano islands close to the

the rapidly eroding Andes. The Amazon Basin has an exceedingly low gradient, Manaus, 800 air miles from the Atlantic, being only 130 ft. above sea level, and Iquitos, 1,700 air miles inland, only 350 ft. In spite of this low gradient, far less of the Amazon lowland is subject to inundation than is commonly supposed, the great bulk of the land lying above the level of highest floodwater on the major streams. In its widest section the upper Amazon Basin affords access over low divides (about 1,000 ft. elevation) into the Orinoco depression to the north and the Paraguay Basin to the south. The Paraguay-Paraná geosyncline begins in the north with the great region of swamp (Pantanais grandes) in western Mato Grosso and with the Gran Chaco of Bolivia, Paraguay, and Argentina. Slopes are so gentle that drainage is slow and imperfect. Toward the south the Chaco grades into the pampas, while the swamps of Mato Grosso find their general counterpart in the adjoining Mesopotamia section of Argentina.

3. Drainage.—The most important river systems of South America—the Amazon, Orinoco, and the Paraná-Río de la Plata—have the greater part of their drainage basins in the lowland belt. The chief headwater tributaries of the Amazon and the Orinoco have their sources in the Andean Highlands, where they have cut deep valleys. The principal sources of the Paraná-Río de la Plata system are in the Brazilian Plateau. These three river systems together drain an area of about 3,700,000 sq.mi. The Amazon is the largest river in the world, though not the longest. At the narrows of Óbidos, where it is constricted by spurs of highland projecting from north and south, it has a width of $1\frac{1}{4}$ mi., a depth of 270 ft., and an average current of 2.7 mph. During flood stages it rises about 25 ft., and the current in the narrows increases to 4.2 mph. The main stream is navigable for ocean steamers as far as Iquitos, about 2,300 mi. from the sea. In time of high water the main stream of the Orinoco is navigable for 1,000 mi. or more.

Other important rivers are the Magdalena in Colombia, the Essequibo in Guyana, and the São Francisco in eastern Brazil. The Magdalena is navigable in two sections separated by the rapids at Honda for river steamers as far as Girardot, the river port of Bogotá. The São Francisco, which rises in the highlands of Minas Gerais, is navigable nearly to the falls of Paulo Afonso, 140 mi. from its mouth, and for 850 mi. above the falls.

Most of the lakes of South America are mountain lakes in the Andes or along their base. Of the Andean lakes, Titicaca (*q.v.*) and Poopó are the largest. The glacial lakes along the eastern base of the Andes in southern Argentina and those of the Llanquihue District of southern Chile are described in the articles ANDES, ARGENTINA, CHILE, and PATAGONIA. Along the east coast there are occasional lakes of brackish water, such as the Lagoa dos Patos of southern Brazil, produced by depression of the coast and the closing of the mouths of estuaries formed thereby or by barrier beaches thrown up by the sea. Lake Maracaibo in northern Venezuela is a large, narrow-necked bay of the Caribbean rather than a true lake, but its waters are brackish.

C. CLIMATE

South America, the only southern continent to extend far into temperate latitudes, tapers toward the South Pole and hence does not experience the extremes of temperature of the northern continents. The winters in temperate latitudes are warmer and the summers cooler than in North America or Asia. The continent is broadest in the equatorial section and has larger areas of hot tropical lowlands than any other continent.

Average annual temperatures in the Amazon Basin are about 80° F. (27° C.), and seasonal variations are slight. Quito, near the Equator but at an elevation of 9,350 ft., averages about 55° F. (13° C.) in every month. Seasonal temperature ranges increase with latitude and reach their maximum in northwestern Patagonia, where July averages 32° F. colder than January. Winter freezing at low elevations is common in Tierra del Fuego and Patagonia, is by no means rare on the Argentine pampas or in south-central Chile, and occurs in southern Brazil at elevations above 1,000–2,000 ft.

The northward-flowing Humboldt (Peru) Current and associated upwelling water keep temperatures on the Pacific coast cool

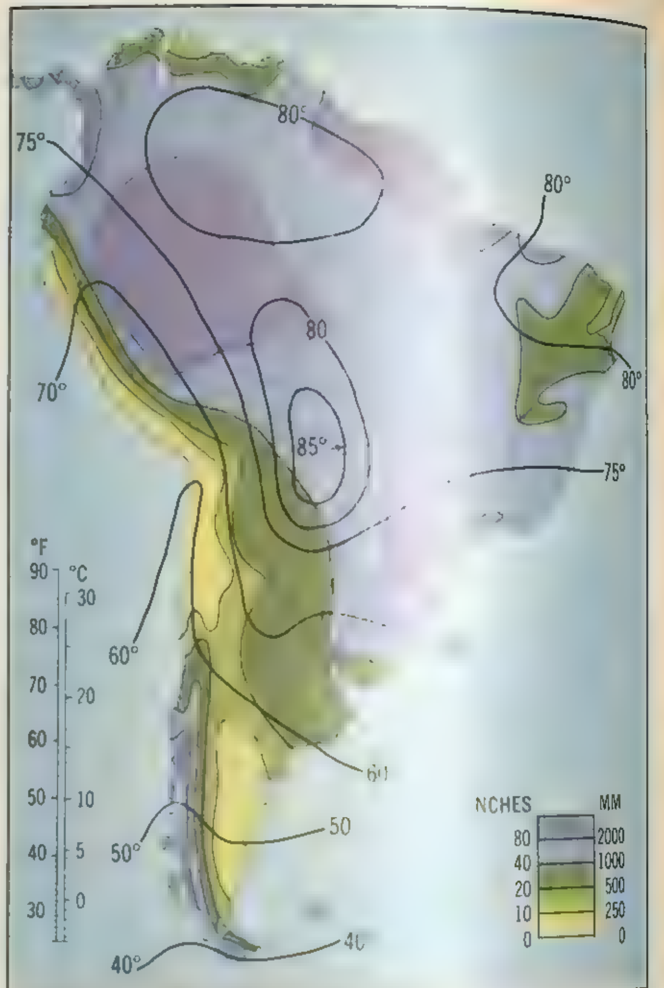


FIG. 2.—AVERAGE ANNUAL RAINFALL AND TEMPERATURE (IN °F)OF SOUTH AMERICA

as far north as the Equator. The coast of Argentina is cooled by the Falkland Current. Elsewhere around the continent warm water washes the shores: the South Atlantic equatorial current flowing west encounters the bulge of Brazil and divides into the south-flowing Brazil Current and a coastal current that flows northwest into the Caribbean; on the northern Pacific coast the equatorial countercurrent sends a warm-water eddy to wash the shores of western Colombia and Ecuador.

The rainfall pattern of the continent is closely connected with the distribution of winds and air masses. Most of tropical South America east of the Andes has prevailing winds from the northeast, east, or southeast. They transport moisture from the warm Atlantic Ocean and give copious rainfall by convection. Rainfall totals average 90–120 in. per year along the lower Amazon and the coast of the Guianas. The middle Amazon Basin has somewhat less rain (60–70 in.), and the upper basin has more than 90 in., possibly exceeding 150 in. in exposed portions of the Andean foothills. The interior of the Brazilian bulge is generally dry and periodically subject to extreme drought. The Llanos del Orinoco and the Guiana Highlands have moderate to high rainfall, with a strong maximum in summer. Much of northwestern South America has two rainfall maxima associated with the two periods of high sun. The Pacific shores of Colombia and northern Ecuador are extremely rainy, while the Gulf of Guayaquil marks the transition to the dry west-coast conditions of Peru and northern Chile. The Atacama Desert along this part of the coast is one of the driest areas in the world. South of 33° S rainfall is adequate west of the Andes, and south of 38° S it becomes excessive. Cyclonic precipitation is characteristic of middle and southern Chile, associated with the westward movement of storms from the Pacific. Valdivia (latitude 40° S) averages 105 in. of rain per year, and the fjorded coast farther south receives even more. Argentine Patagonia lies

in the rain shadow of the Andes and is mostly steppe and desert. The grasslands of Argentina, Uruguay, and southern Brazil have moderate rainfall (25–50 in.) well distributed throughout the year. Eastern Paraguay is moist (50–75 in.), while the Chaco is moderate to dry and has a dry winter.

D. Vegetation

The pattern of vegetation for the continent is complex, especially in the highlands and in areas of critical rainfall differences. The actual vegetation cover deviates most from the natural forms in areas of dense human settlement and in regions where the land has been converted into permanent farms or artificial pasture. A question still remains as to how much influence the pre-Columbian Indian exerted on the natural vegetation cover. It is possible that the greater extent of tropical grassland north of the Amazon lowland and in the northern fringes of it as compared with areas south of the Amazon is in part a reflection of greater aboriginal population density in the northern areas and of a greater use of fire for making clearings and driving game. The comparable areas south of the Amazon have a cover of semideciduous woodland that has been much less affected by fire.

The largest more or less uniform vegetation region is that of the tropical lowland rain forest in the Amazon Basin and the Guianas, with an outlier along the east-central Brazilian coast. It is the largest area of its kind in the world and has several distinct subtypes. The number of plant species growing together in this forest is greater than in any other plant association in South America. Additional characteristics are dense growth, generally slender trees, storied vegetation, lianas, parasites, and epiphytes, and, at times, buttressed trunks.

A semideciduous phase of the tropical forest is found in the Venezuelan part of the Guiana Highlands and in some inland sections of central and southern Brazil. It is somewhat less luxuriant than the true lowland rain forest, and some of the plants that compose it drop their leaves seasonally because of dryness or low temperatures. Tropical highland forest, in part cloud forest, is found on the eastern flanks of the Andes from latitude 20° S northward. Mangrove forests fringe most of the warm-water coasts of the continent from the Gulf of Guayaquil clockwise almost to Rio de Janeiro. A palm forest, also tropical, is characteristic of a portion of the state of Maranhão.

Temperate or mid-latitude forests are of two main types: the araucaria evergreen forest of southern Brazil and the more extensive deciduous forests with small araucaria enclaves in southern Chile. In southern Brazil the araucaria, a coniferous softwood, is the basis of a significant lumber and paper industry.

One step below the forests in magnitude are the woodlands of the continent, composed chiefly of lower trees, more widely spaced, with an admixture of shrubs and grass. As compared with the forests, they are found in areas of less total and less uniform rainfall. Semideciduous woodlands known as *campos cerrados* are characteristic of central Brazil. Especially thorny woods are found in northeastern Brazil, in much of the Chaco, and in northernmost Colombia and Venezuela. Park landscapes occur along the Paraguay River and in the hill lands west of the Chaco. They are dotted with well-spaced trees, the intervening areas being covered with grass and shrubs, and represent a most attractive plant association.

Shrub lands display a further diminution in scale, and the typical plants are bushy rather than treelike. They are characteristic of generally dry areas and of those with pronounced seasonal drought. Various forms are found in the Andean foothills behind the tropical coastal desert of Peru, in central Chile, in the lands west of the Argentine pampas, and in eastern Patagonia.

Tropical grasslands are characteristic of the llanos, of the coastal fringe of the Guianas behind the mangrove forest, of parts of the Guiana Highlands in northernmost Brazil, and of certain sections of the Amazon, including the northern and interior parts of Marajó Island. South of the Amazon forest, tropical grasslands are rare, but two large areas of palm-savanna swampland are found.

Temperate grasslands are characteristic of the pampas of Argentina and of adjacent areas in Uruguay and southernmost Brazil. At intermediate elevations they occur near the araucaria forests of southern Brazil, as well as in the Peruvian Andes. At high Andean levels moist alpine grassland is found in equatorial and mid-latitude areas, whereas a dry phase (*puna*) covers most of the intervening section in the Peruvian and Bolivian altiplanos and in parts of northern Chile and Argentina. Extreme aridity in the Atacama Desert of coastal Peru and northern Chile causes an almost complete absence of vegetation.

FIG. 3.—MAJOR VEGETATIONAL ZONES OF SOUTH AMERICA

E. Animal Life

The native animals of South America belong largely to the so-called neotropic realm of animal geography. For obvious reasons there is a fair correspondence between the occurrence of specific animals and the distribution of native vegetation.

The tropical forests contain a fauna characterized by several kinds of monkeys, the tapir, rodents (including the capybara, the largest rodent in the world), birds (parrots, egrets, hummingbirds, etc.), and reptiles (especially crocodiles). Large members of the cat family are relatively rare, as are the large herbivores. The insect world is represented in great numbers. A most characteristic, though not common, member of the Amazonian fauna is the manatee, a vegetarian, aquatic mammal. Just as characteris-

tic, and unfortunately all too common, is the piranha, a small, fierce freshwater fish that attacks other animals that venture into its waters. A school of piranhas has been known to kill a cow and to consume it completely, except for the skeleton, in less than five minutes.

The greatest contrast to this tropical lowland development of animal life is found in the high Andes and in the cold steppes and deserts of Patagonia. Among the important animals native to the central Andes are the four members of the American branch of the camel family: the guanaco, llama, alpaca, and vicuña. The guanaco and vicuña have not been domesticated, whereas the llama and alpaca are found only in the care of man. The llama is the common beast of burden of the Andean Highland Indian, the alpaca is a prolific bearer of fine wool, and the vicuña has a fleece of unbelievable softness and thickness. A rodent found in parts of the high Andes and famous for its magnificent fur is the chinchilla.

The guanaco ranges beyond the Andes into Patagonia and at one time was fairly common on the Argentine pampas. Even wider in range is the rhea, the three-toed South American ostrich, which extends from Tierra del Fuego into central Brazil.

The shrub lands and woodlands south of the Amazon have a fauna that well expresses their transitional character between the Amazon on the one hand and the Chaco and pampas on the other. Truly tropical forms occur there, such as monkeys, anteaters, sloths, tapirs, and parrots, but animals from the colder south are also found, including the rhea, the armadillo, and the raccoon. Large members of the cat family, including jaguars, leopards, and pumas, occur in the Brazilian state of Mato Grosso and other inland areas. Some of these animals are called *tigres* by the South Americans, but true tigers in the wild state are confined to the Eastern Hemisphere.

II. NATURAL RESOURCES

1. Water.—Because of diverse topography and rainfall conditions, the water resources of the continent vary enormously in different areas. In the Andes, navigation possibilities are slight except on the Magdalena River, Lake Titicaca, and the Chilean-Argentine lakes in the south. Irrigation for agriculture is of importance from northwestern Peru to Patagonia. Less than 10% of the Andes' estimated hydroelectric power potential had been harnessed by the mid-1960s.

The Brazilian Shield has better hydroelectric possibilities than the Andes because of the large size of many rivers, the harder rocks, and the occurrence of cliff margins. The harnessing of the Paulo Afonso Falls provides electricity for a large part of northeastern Brazil. Waters impounded near São Paulo are made to fall over the Serra do Mar to provide cheap power and a reliable municipal water supply.

The Amazon River system has more than 8,000 mi. of navigable waterways but presents few possibilities for power development because of the gentle slopes throughout the region.

2. Soils.—Thousands of square miles of dark prairie soil of aeolian and alluvial origin occur in the pampas of Argentina and Uruguay. They are among the best farming soils in the world. Smaller, disconnected areas of good soils are located in various Andean and west-coast valleys, notably in the Valle Longitudinal of Chile, the Guayas lowland of Ecuador and the Cauca Valley of Colombia.

The Paraná Basin is deeply covered with a rich, purplish-red soil called *terra roxa*, derived by weathering from basaltic outflows. This soil is especially suited to coffee growing, as is the volcanic soil of the Colombian Andes.

In the Amazon Basin the soils are generally poor. Small areas of bottomlands have fertile soils, but these are subject to flooding. Infertility and high acidity make most tropical lowland soils submarginal for agriculture. To improve them with lime and fertilizer is rarely feasible.

FIG. 4.—LAND USE AND MAJOR FUEL AND MINERAL AREAS OF SOUTH AMERICA

3. Minerals.—South America is well endowed with petroleum. Roughly one-fourth of the area contains sedimentary basins that might, other conditions being favourable, permit the occurrence of oil-bearing strata. As of 1965 only minor strikes had been made at distances greater than 300 mi. from the Andes, most of the operating oil fields being located in structural basins flanked by Andean ranges, especially in the vicinity of the Caribbean.

Venezuela, with 17,000,000,000 bbl. of proved reserves as of the end of 1963, had by far the greatest wealth of petroleum in South America. It ranked second in the Western Hemisphere and seventh in the world, being exceeded in addition to the U.S. by four countries in the Middle East and by the Soviet Union. Other South American countries with large amounts of proved petroleum reserves include Argentina, Colombia, Peru, and Bolivia.

Natural gas is commonly associated with petroleum, but its commercial utilization has been limited by the remoteness of fields from major population centres. Gas is beginning to be used com-

mercially in Venezuela, but its greatest use occurs in Argentina, where a 1,000-mi. pipeline extends northward from the Comodoro Rivadavia field in Patagonia to supply greater Buenos Aires.

The dearth of good coal in South America represents one of the greatest resource contrasts between the two American continents. Although scattered deposits are found in many parts of the Andes, as well as in southern Brazil, only three countries, Colombia, Chile, and Brazil, have a modest production. Brazil's coal is excessively high in ash content and must be mixed with higher-grade imported coal before it can be utilized industrially. The paucity of coal is a serious disadvantage to the growth of heavy industry on the continent, notwithstanding the plentiful supply of iron ore.

South America is extraordinarily well endowed with deposits of iron ore. The deposits are associated mainly with the ancient core areas of the Guiana Highlands and the Brazilian Shield. Very extensive, high-grade ore bodies have been discovered at or immediately below the surface, and they can be mined directly and cheaply in open pits through the use of power shovels of large size. As a result of the investment in Venezuela of hundreds of millions of dollars by United States corporations, the production of iron ore in that country increased from 2,500,000 tons in 1953 to 12,500,000 tons in 1960, followed by a sizable decline in the early 1960s because of political unrest in the producing area. The main iron districts lie near the Caroní River, a southern tributary of the Orinoco, and their economic development has been greatly facilitated by the availability of cheap water transport.

In Brazil the situation is less favourable, the large ore bodies lying mostly in the interior of the state of Minas Gerais, far from navigable waterways. Although the deposits are rich and more extensive than those in Venezuela, large-scale utilization awaits the improvement of land transport and coastal docking facilities.

Copper occurs mainly in the geologically young lands of the central Andes, in Chile and Peru. Only the great copper mines of North America and central Africa have greater world significance than those of Chile. Commercial tin ores in the Western Hemisphere are confined mainly to Bolivia. The mines are located in the high Andes, some of them being earlier silver mines that were converted to tin mining when the richer silver ores were exhausted. Total tin resources in the Bolivian Andes are estimated to rank third in the world, following southeast Asia and equatorial Africa.

Brazil ranks high in manganese reserves, one of the world's great deposits being located in southwestern Mato Grosso. A smaller but more easily accessible deposit in Amapá lies immediately north of the mouth of the Amazon. Brazil ranks with the Soviet Union and India as a leading producer of manganese.

Other significant minerals in South America are bauxite (aluminum ore), which is especially plentiful in the Guianas and adjacent parts of Brazil; platinum, in Colombia; silver, in Peru and Bolivia; and nitrate, in Chile.

4. Land Use.—Most South Americans live within a few hundred miles of the coasts, and most regions of intensive land use are confined to the same peripheral locations. For the continent as a whole, about 5% of the land is cultivated, whereas about 19% is permanent grazing land and fully 47% is forest. In the early 1960s the proportion of cultivated land by countries varied from a high of 12% for Uruguay to a low of 1% for Paraguay and one-thirtieth of 1% for French Guiana.

Regions of commercial agriculture are widely scattered. The irrigated coastal valleys of Peru are noted for their long-staple cotton, and the Guayas lowland of Ecuador has become the world's leading area of banana production for export. In Brazil more than 2,500,000 ac. are devoted to sugarcane, mainly along the south-central coast and in nearby interior valleys, and tobacco, bananas, and cacao occupy other low-lying districts not far from the sea. The world's greatest coffee-growing district is located on the upland *terra roxa* soils of the Brazilian states of São Paulo and Paraná. The other major coffee area of South America lies in the Colombian Andes, where higher-quality mild coffees are grown.

Mid-latitude agriculture is dominant in the southern part of the continent. The greatest agricultural region of South America is the pampas of Argentina and Uruguay. Wheat is the chief crop, but corn and flax are also significant exports.

Notwithstanding the importance of South America's trade in agricultural commodities or its significance as a source of foreign exchange for the South American countries, it must be emphasized that the vast bulk of the continent's agriculture is devoted to supplying local needs. For subsistence and local commercial markets corn (maize) is grown over much of the continent and leads in total acreage planted. In the tropical lowlands cassava (manioc) and sweet potatoes are widely planted for local consumption, and in the Andes potatoes and other cold-country root and tuber crops are cultivated. Rice and wheat have long been staple food crops and are widely grown wherever conditions are favourable.

Pastoral industries with emphasis on cattle are dominant on the Argentine pampas and adjacent grasslands of Uruguay and southern Brazil, as well as on the llanos of Colombia and Venezuela. Sheep raising for wool is commercially significant in Uruguay, Patagonia, and Tierra del Fuego.

Forest industries based on planted eucalyptus and native araucaria have been organized in southern Brazil; south-central Chile appears to have possibilities of similar development.

Extensive manufacturing districts are not yet common on the continent. Their greatest development has taken place in the environs of São Paulo and Buenos Aires. Examples of individual industrial centres are numerous, however, and include such settlements as Volta Redonda (Brazil), Concepción (Chile), Córdoba (Argentina), Medellín (Colombia), and San Félix (Venezuela).

The great Amazon Basin, the largest undifferentiated region on the continent, is still mainly restricted to a collecting economy, with emphasis on Brazil nuts. (H. J. BN.)

III. ANTHROPOLOGY

Recent archaeological discoveries indicate that man first arrived in South America more than 20,000 years ago. The predominance of large stone projectile points in some Paleo-Indian sites, and of chopping tools in others, suggests that some of the early migrants were primarily hunters, while others specialized in the gathering of wild plant foods.

The introduction of agriculture, which dates from at least 2600 B.C., initiated a relatively rapid cultural evolution that reached different levels in different parts of South America. Some areas remained at the hunting-fishing-gathering level of subsistence. Others adopted agriculture and made certain other cultural advances but continued to live in relatively small and independent communities. Still others developed further, both technologically and socially, and attained the level of chiefdoms. The greatest cultural development of all, which occurred only in the central Andes, was the rise of a full-scale civilization. In the *Handbook of South American Indians*, Julian Steward called these four culture types Marginal, Tropical Forest, Circum-Caribbean, and Andean, respectively, and this usage will be employed below.

1. Racial Type.—The Indians of South America all belong to the Mongoloid racial stock and came originally from Asia by way of North America. The distribution of various morphological characteristics among them has been studied, but as yet no proposed classification of physical types has been generally accepted. The study of blood groups has revealed a high predominance of O, a small amount of A, and practically no B or AB.

2. Linguistic Groups.—The number of language families recognized for South America has varied. A. F. Chamberlain distinguished 83 families in 1913, Paul Rivet 77 in 1924, Cestmir Loukotka 94 in 1935, and J. Alden Mason 91 in 1950. It was generally conceded, however, that more detailed study would decrease the number of independent linguistic groups recognized. In 1956 Joseph Greenberg proposed a tentative classification that reduced the languages of South America to three phyla or superfamilies: Macro-Chibchan (including Chibchan); Andean-Equatorial (including Araucanian, Arawak, Aymara, Chon, Quechua, and Tupí-Guaraní); and Ge-Pano-Carib (including Carib, Ge, Guaycuruan, and Panoan).

3. Marginal Culture.—Societies subsisting entirely or largely on wild food sources were found over large areas of South America in aboriginal times. These groups occupied habitats unsuitable for agriculture. From the Chaco south, it was generally too dry

or too cold for cultivated plants; in eastern Brazil the savanna could not be readily farmed with the tools available. The differences in culture level among Marginal peoples resulted largely from differences in the richness of the wild foods in their respective environments.

The relative importance of hunting, fishing, and gathering varied with the habitat. Among the Tehuelche of Patagonia the hunting of guanacos and rheas provided almost the only source of food. Other tribes, especially those of the Chaco, were more generalized in their subsistence activities and consumed a wide variety of vegetable products along with meat and fish. North of the pampas the principal hunting weapon was the bow and arrow; from the pampas south to Tierra del Fuego it was the bolas (*q.v.*).

Community size was generally small among the Marginal societies. The Yahgan, living in a difficult environment, had local groups of about 15 or 20 persons. On the other hand, after the acquisition of the horse some of the tribes of the Chaco, pampas, and Patagonia had bands of 500 or more. The average size for a Marginal local group was probably about 50. The exploitative nature of their subsistence economy meant that Marginal peoples had to move their campsites periodically to stay abreast of their food supply. With almost no exception, each local group was economically and politically independent. Intertribal gatherings occurred for ceremonial occasions, but they were never permanent. Chieftainship might or might not be hereditary. The chief was usually weak except in societies where achievement in warfare served to increase his power.

Dwellings varied according to the materials available. The guanaco hunters of the south used tents made of animal skins, while the Amazonian Marginals had thatched huts.

Pottery was usually absent, and containers were leather bags, baskets, or gourds. Loom weaving occurred only in the Chaco, where it had diffused from the Andes. Land travel was by foot except among tribes that adopted horse riding from the Spaniards. Watercraft were absent among most Marginal peoples, but a few had bark canoes, rafts, and, in Patagonia, coracles. The band constituted a kinship as well as a local group. Sibs and moieties occurred only among the Ge groups of Brazil, and social classes only among a few horse-riding Indians of the Chaco. Marriage tended to be exogamous, and postmarital residence patrilocal. Matrilocality was found where women's share in subsistence was important, as in the Chaco. There were no specialists except the shaman and the chief, who were nevertheless not exempt from subsistence.

Available evidence indicates that warfare was sporadic and was waged for revenge, witchcraft, and trespass. It rarely involved the capture of prisoners, the taking of land, or the exacting of tribute. The basis of religion was a belief in a number of gods and spirits. Rituals were concerned principally with subsistence and life crises. The shaman was a diviner and curer.

4. Tropical Forest Culture.—The occurrence of Tropical Forest culture coincided closely with the distribution of the tropical rain forest in South America. In the absence of all but the crudest farming tools, the slash-and-burn cultivation characteristic of this culture could be carried on only in areas of loose forest soil. A plot of land was cleared during the dry season, burned just before the rains, and planted with a digging stick. Cultivated plants included corn, sweet potatoes, beans, and peanuts, but the staple crop was cassava. The hydrocyanic acid found in the poisonous variety of this tuber was removed with the help of a number of devices, the best known of which was a sleevelike press called a *tipiti*. Cakes, gruel, and a fermented drink were made from cassava meal.

Fishing was an important supplement to the diet and was carried on with bow and arrow, fish traps, and by drugging with poisonous vines. The hunting of large game was done chiefly with the bow and arrow and with traps; the blowgun, when present, was used mainly against birds and small mammals.

Tropical Forest villages were semipermanent and were located in or near forests and close to watercourses. Houses had a pole framework and were covered with thatch, generally of palm. Usually several families lived in a house, and in some areas the entire village might consist of a single large house. The hammock was used for sleeping by almost all tribes. Community size varied within fairly wide limits. In the Peruvian Montaña, Campa and Amahuaca settlements averaged about 15 persons; along the coast of Brazil, Tupinambá villages often had 600 or 800. The average for the entire area was perhaps 150 to 200. Each village had a chief and occasionally subchiefs or house chiefs. Political authority was not strong except in warlike societies. Chieftainship was sometimes hereditary but was generally only loosely so. Villages were autonomous, and confederacies between them were rare and never permanent.

Baskets were of many kinds and were made by a variety of techniques. Wood carving was common, the most representative item being the one-piece wooden stool. Featherwork was universal and was often beautiful and spectacular. Most tribes had some form of water transport. The dugout canoe was the most typical watercraft, but some tribes had only bark canoes or rafts.

Marriage was generally exogamous, although a number of groups permitted local endogamy. Polygyny occurred, especially for such persons as the chief or shaman. Clans were found among some tribes, such as the Mundurucú and Tucuna, but were not characteristic of the area. The practice of couvade was prevalent. According to this custom, a man took to his hammock or underwent other restrictions at the time his wife gave birth.

Warfare was common to most tribes of the region. Some, like the Mundurucú and the Island Carib, devoted a great deal of time to it. The securing of head trophies and the practice of cannibalism were frequent adjuncts of warfare. Offensive weapons consisted primarily of bows and arrows and clubs. For defense, villages were often palisaded and trails sometimes set with poisoned stakes. Characteristically, warfare was for revenge rather than for economic gain.

Ceremonialism and ritual were well developed. Puberty rites often included tests of courage for the initiates, such as the ant ordeal, whipping, and being bitten by an anaconda. Narcotics of many kinds were used, including tobacco, coca, and *paricá* snuff. Tobacco was particularly important because it was frequently smoked by the shaman in connection with divination and curing.

5. Circum-Caribbean Culture.—Around the Caribbean Sea from Cuba to Nicaragua and in the northern Andes there were societies that resembled Tropical Forest tribes in certain details, *e.g.*, use of the hammock, arm and leg ligatures, and drinking bouts, but were at a distinctly higher cultural level. Agriculture was carried out intensively. The digging stick was still the basic farming tool, but such techniques as planting in mounds and the use of urine fertilizer increased productivity. Irrigation canals and terracing were employed in some places. Corn was the most important crop in the mountain valleys, while cassava was the principal crop in the lowlands. Fruit orchards were also common. Fishing was of importance locally, but hunting was of little significance.

Communities of 2,000 or 3,000 were not uncommon. Towns were generally permanent and well laid out, usually around a large central plaza. For defensive purposes they were often palisaded. Flagstone-paved roads sometimes connected larger towns. Houses were of pole-and-thatch construction rather than of stone. Settlement was often dense: one valley occupied by the Tairona was reported to contain 250 towns. The Anserma had a total population of more than 40,000.

Confederacies and petty states were common in the area. The chief of such political units had considerable authority and prestige. He lived in a special house and was carried about in a litter or on the shoulders of retainers. Some of his many wives and concubines were buried with him after being stupefied with *chicha*, an alcoholic drink made from fermented corn. Society was stratified, with nobles, commoners, and slaves ranking below the chief. Slaves were prisoners of war or persons caught stealing. Commoners could become nobles by valorous exploits in war.

There were craft specialists in pottery making, weaving, goldworking, and stone carving. Metallurgy was particularly well developed. The Quimbaya of Colombia and the inhabitants of Coclé in Panama produced outstanding goldwork. *Tumbago*, an

alloy of gold and copper, was made into utilitarian and luxury goods. Trade was very important, and markets existed. The principal items of exchange were salt, fish, gold, and cotton. Water travel was in dugout canoes along the coasts and on the larger rivers. To cross narrow mountain valleys suspension bridges of vines were constructed. Attire varied considerably. Some tribes went naked or used a simple breechclout. Others wore elaborate cotton mantles or tunics. Class distinctions were reflected in differences in dress.

Marriage was customarily by purchase, and wealthier individuals had several wives. Ceremonies surrounded life crises, especially death. The dead were disposed of in various ways, depending upon locale and social status. Commoners were often cremated and buried in urns; chiefs were interred in stone-lined shaft graves. Funerals were often accompanied by drinking bouts in which huge quantities of *chicha* were consumed.

Warfare prevailed except among a few tribes, such as the Taino. It was waged for conquest of land, to exact tribute, and to capture slaves. Organized military tactics were employed, in some cases directed by special war chiefs. When confederacies fought each other, thousands of warriors might be involved. The favourite weapon was the *macana*, a wooden sword-club, but poisoned arrows and fire arrows were also used. Captured trophy heads were displayed on posts, and the bodies of notable enemies were dried over the fire and stuffed with ashes. Cannibalism was widespread.

The sun and moon were commonly regarded as important deities, and shrines dedicated to them were erected on mountain tops. In trying to influence the supernatural, priests made offerings and sacrifices. Idols, like the *zemis* of the Taino, were used to communicate with the spirit world.

6. Central Andean Culture.—The Inca Empire of the central Andes was the largest and best organized state in aboriginal America. It encompassed about 1,000,000 sq.mi. and had a population conservatively estimated at 6,000,000.

At the base of Inca civilization was a system of intensive agriculture. Cultivated land was extended to the fullest by terracing mountain sides and irrigating deserts. The *taclla*, or foot plow, was used to turn the soil. Llama dung and guano were employed as fertilizer. Able-bodied men farmed plots for the state and the church in addition to their own. Surpluses went into granaries to support priests and other specialists, to maintain the army, and to serve as emergency rations in times of famine. More than 40 different species of crop plants were grown, corn being the most important. In high altitudes, quinoa and potatoes were raised. The principal domestic animal, the llama, served as a beast of burden and as a source of meat, wool, hides, and fertilizer.

The household head, called a *puric*, was allotted a certain portion of land at marriage, which was added to at the birth of each child. The basic unit of settlement was the *ayllu*, an endogamous agricultural community of about 100 families. *Ayllus* were grouped into districts, districts into provinces, and provinces into quarters. The Inca Empire as a whole was called Tawantinsuyu, "four parts of the world." An official governed each administrative unit of the empire. The smaller administrative units were governed by *curacas*, or lower nobility, while the larger units were ruled by members of the higher nobility, called *orejones* ("big ears") by the Spaniards because of the earplugs they wore as symbols of their office. At the top of the hierarchy was the Sapa Inca, or "sole lord." The Sapa Inca was considered the descendant of the sun-god and, thus, divine himself. No one could look directly at the Sapa Inca, and he never drank from the same vessel or wore the same garment twice. He had hundreds of concubines, but his rightful heir was his son by his eldest sister.

There existed a system of forced labour known as the *mita*, by means of which men were drafted for public works. As many as 30,000 men were mobilized at one time to work on roads, fortresses, etc. The *mitimae* system involved forced resettlement and was employed to ease population pressure and to disperse rebellious peoples.

The army, the instrument for expanding the empire, was also recruited by draft. Draftees served either as bearers or as fighting men. On the march the army used the excellent highways and was quartered and provisioned in *tambos*, or fortified storehouses, at points along the road. Military formations and tactics were used during warfare. Soldiers were divided into slingsmen, bowmen, spearmen, etc.

Arts and crafts were practised by full-time specialists and showed great technical excellence. Pottery making, metallurgy, and weaving were especially well developed. Inca architects built solid and imposing structures of dry stone masonry. No system of writing was known, but the *quipu* permitted the keeping of detailed records.

The system of roads consisted of two arterial highways, one along the coast and the other through the highlands. Shorter roads connected the two at various points. Couriers running in relays carried messages or goods rapidly. The 1,400 mi. between Cuzco and Quito could be traveled in five days.

The Inca believed in a creator god, Viracocha, and in a number of other deities, including the sun. The assistance of the gods was invoked at public ceremonies connected with such events as the agricultural cycle and warfare. Many shrines and temples existed throughout the empire. *See* ANDEAN CIVILIZATION; INDIAN, LATIN-AMERICAN. (R. L. C.)

IV. POPULATION

The population of South America was estimated (1963–64) to be 160,000,000. Together with the population of Middle America this gave a total population for Latin America of approximately

FIG. 5.—POPULATION DENSITY OF SOUTH AMERICA

235,000,000, which was about 23,000,000 more than that of Anglo-America. Of the population of South America, almost exactly 50% is in Portuguese-speaking Brazil; Argentina, Colombia, and Peru rank next in order.

The population of South America is arranged in a pattern of more or less separate clusters. In the centre of each cluster there is a central city, of which eight in South America (and ten in Latin America as a whole) had passed 1,000,000 in 1963. Between these clusters there are only scanty populations. A large part of the continent has less than one person per square mile. Most of the clusters are located near a coast, and the vast interior, mostly in Brazil, is one of the world's larger areas of very thin population. Table I shows how this population is divided among the 11 countries and 2 colonies.

TABLE I.—*Area and Population of South America*

Country	Area (in sq.mi.)	Population (1963–64 est.)	Densities per sq.mi.
Total	6,879,122	159,942,730	23.4
Argentina	1,072,067	22,023,431	20.5
Bolivia	424,162	3,653,000	8.6
Brazil	3,286,470	79,937,000	24.3
Chile	292,256	8,369,000	28.6
Colombia	439,512	15,434,090	35.1
Ecuador	105,685	4,877,000	42.2
French Guiana	35,135	35,000	1.0
Guyana	83,000	621,390	7.5
Paraguay	157,047	1,936,000	12.3
Peru	496,222	11,854,000	23.9
Surinam	63,251	320,000	5.1
Uruguay	72,172	2,556,020	35.1
Venezuela	352,143	8,426,799	23.9

The population of Latin America is growing faster than that of any other part of the world. Between 1920 and 1960 the population increased 126.3%, as compared with 92.4% for second-place Australia and New Zealand. Table II shows the medium estimates made by the United Nations for the future population of Latin America by subregions. It is important to note that in all of Latin America the most rapid jump in the rate of increase took place between the pre-World War II period and the period after the war. The decline in the death rate that began generally in the late 1930s is the major cause of this accelerated gain in population. During the 1920s the death rate for 11 Latin American countries for which adequate and comparable data exist was lowered by 6%, but during the short period from 1945 to 1954 the death rate was lowered by 17.1%. With the continued high birth rates in these countries the result was a "population explosion."

TABLE II.—*Medium Estimates of Future Population by Subregions*
(in 000,000)

Region	1960	1975	2000
Low-latitude South America	107.0	163.0	339.0
Middle-latitude South America	32.9	41.7	55.8
Latin America as a whole	205.8	304.1	592.8

Source: Population Branch, United Nations.

The decrease in the death rate that has taken place in all the Latin American countries is essentially the result of the improvement of medical services and of public hygiene and has accompanied the decrease of illiteracy. The virtual elimination of malaria, the great decrease in smallpox cases, the effective attack on yellow fever and hookworm—all these medical gains have had the effect of saving the lives of children.

The population of South America is quite diverse in its racial composition. There are certain countries in which the proportion of Indians remains high. In Ecuador, Peru, Bolivia, Paraguay, and the southern part of Colombia, people of unmixed or almost unmixed Indian ancestry make up more than half of the population. In Argentina, Uruguay, and southern Brazil (from São Paulo state southward) the population is more than 95% pure European in origin. It is to this part of Brazil and to Argentina that the great stream of European immigration into South America has gone.

Since 1800 an estimated 12,000,000 immigrants have arrived in Latin America. Of these about 4,000,000 came from Spain, 4,000,000 from Italy, and 2,000,000 from Portugal. The remainder include a variety of peoples: Germans, Poles, and other Europeans; Syrians; Chinese; and Japanese. There are also parts of South America in which there is a high proportion of Negroes. The Negroes came first to the Portuguese colonies of Salvador and Recife in northeastern Brazil to work the sugarcane plantations established there during the 16th century. This area still has a large proportion of Negroes and mulattoes. After the beginning of the 18th century, when gold and diamonds were discovered in Minas Gerais, many Negroes were brought as slaves to that part of Brazil; this region thus contains a considerable proportion of Negroes in its present-day population. Similarly, deep in the interior, near Cuiabá, there is a concentration of Negroes around old colonial goldworkings. Negroes were also brought as slaves to sugarcane areas in Venezuela and in the Cauca Valley and Caribbean coastal lowlands of Colombia. Elsewhere, the population of South America is a mixture chiefly of Indians and Spaniards known in Spanish America as mestizo. Table III gives the best available estimates of racial proportions.

TABLE III.—*Estimated Racial Proportions*
(in %)

Country	European	Indian	Negro	Mestizo	Oriental
Argentina	97	3			
Bolivia	15	53		32	
Chile	30	5		65	
Colombia	20	7	5	68	
Ecuador	15	60	3	22	
Paraguay	(97% Indian and mestizo)				
Peru	15	46		38	1
Uruguay	90		2	8	
Venezuela	20	7	8	65	

Throughout Latin America there is a tendency for the densely

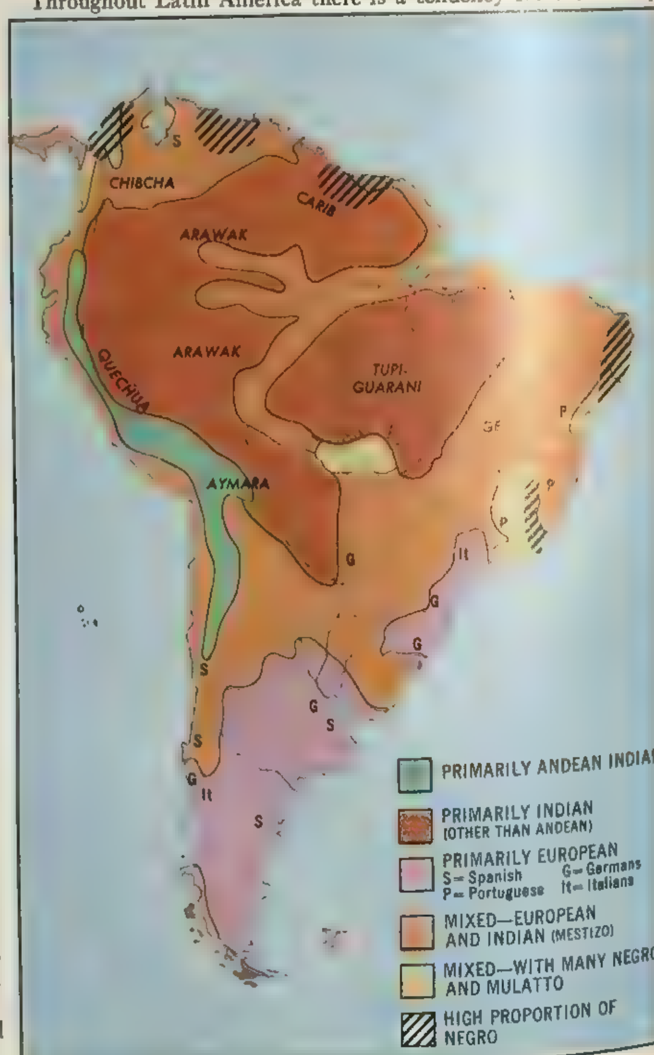


FIG. 6.—RACIAL COMPOSITION OF SOUTH AMERICA

populated areas to increase in population more rapidly than the average and for the thinly peopled areas to become even more sparsely populated. There are, to be sure, a few zones of pioneer settlement where new lands are being cleared and settled, but by far the greater part of the population movement is away from these frontiers rather than toward them. The movement generally is from the rural areas into the cities. Urban growth, especially after World War II, increased rapidly. The first city in South America to reach a population of 1,000,000 was Buenos Aires in 1906. Since then Rio de Janeiro, São Paulo, Santiago, Lima, Caracas, Bogotá, and Montevideo have all passed this mark. Not only have the cities increased in size, but also the proportion of the total population of each country living in the central cities has increased. This urban growth is a measure of the rapidity with which the Industrial Revolution is sweeping over Latin America. (P. E. J.)

V. HISTORY

When the mainland of South America was sighted by Columbus on his third voyage, in 1498, it was inhabited by numerous Indian tribes of widely varying cultures. By 1560 Spain had overrun much of the continent, conquered a large majority of the native inhabitants, discovered mines that gave the area a firm economic base, established the Roman Catholic Church, and evolved an administrative system that did not undergo any major modification until the 18th century, when threats from Great Britain forced a reorganization. Meanwhile, the Portuguese in Brazil developed a tropical-agriculture colony based upon Negro slavery. The British, French, and Dutch established their claims to certain of the "leftover" areas of the continent in the 17th and 18th centuries.

The Spanish colonies won their independence as a result of a costly 15-year struggle (1810–25), while Brazil separated from Portugal with a minimum of effort in 1822–23 and formed an empire, which lasted until 1889.

For approximately one century after political freedom had been won, the Spanish republics were ruled—often by force—in the interest of a narrow elite. After 1875, European immigrants and foreign capital and technicians contributed significantly to the material development of the nations.

Since World War I, several of the republics have undergone major transformations. New middle classes and working classes have emerged and have been incorporated into the political life. Economically, the republics have sought, with some success, to widen the range of their production and have made the modern factory the symbol of progress. For a full treatment of the history of South America, *see* LATIN AMERICA. *See* also references under "South America" in the Index. (J. J. J.)

SOUTH AMERICAN LANGUAGES. South America and the West Indies apparently had more than twice as many aboriginal languages as were found in Central and North America. It seems also that the South American languages were more diverse than those to the north. Linguists have enumerated no more than 40 distinct stocks from Central and North America, but the estimated number of stocks for South America range from 70 to more than 100. The aboriginal languages of the West Indies belonged to two stocks, Arawakan and Cariban, both of which became widespread in South America. (For a general discussion of the languages of the Americas and a brief description of the methods of classifying languages into stocks or families, *see* AMERICAN ABORIGINAL LANGUAGES; *see* also CENTRAL AND NORTH AMERICAN LANGUAGES.)

It is probable that the extreme linguistic diversity noted for South America rests in large part on two factors: (1) a lack of detailed knowledge of most South American languages; and (2) a method of classification based mainly on the comparison of brief vocabularies which is wholly inadequate for the establishment of more distant relationships. It seems that when more is known of the languages of South America and proper methods can be applied to their classification, the number of stocks will be considerably reduced.

The larger and the more widely distributed linguistic stocks of South America and the West Indies appear to fall into the following groups:

1. Chibchan. These languages, in the aboriginal period (*i.e.*, roughly A.D. 1500 and before), occupied most of Colombia, and extended southward to the latitude of Guayaquil in Ecuador and northward to include Panamá and most of Costa Rica and Nicaragua. One of the languages served a powerful Indian empire, with its centre near Bogotá. Later, but before the Spanish conquest in 1533, this empire was defeated by the Incas of Peru and the southernmost Chibchan languages were driven back or replaced by Quechuan languages (*see* no. 6, below). A number of Chibchan languages are still spoken, mainly by small Indian groups in Colombia and Central America, but these are gradually giving way to Spanish.

Some investigators (notably J. Alden Mason, 1950) proposed a Macro-Chibchan stock, composed of Chibchan plus the Misumalpan languages of Central America (*see* CENTRAL AND NORTH AMERICAN LANGUAGES). The evidence for this grouping is weak.

2. Cariban. This group spread, in preconquest times, from the Greater Antilles to central Mato Grosso in Brazil, and from eastern Peru to central Pará, Braz. Most of the languages, however, were found north of the Amazon River, in the territory now occupied by northern Brazil, the inland areas of the Guianas and Venezuela and lowland Colombia.

West Indian Cariban is extinct, but a considerable number of Cariban languages, perhaps as many as 200, have survived.

3. Arawakan. This is the largest and most widespread of all South American linguistic stocks. During preconquest times, Arawakan languages were found, in a number of disconnected areas, from Cuba and the Bahamas southward to the Gran Chaco and the sources of the Xingu River in southern Brazil, and from the mouth of the Amazon River to the eastern foothills of the Andes. Some of the languages of this group are extinct or nearly so (particularly in the West Indies), but more than 200 are still spoken in South America.

4. Ge. The languages of this group, in modern and aboriginal times, occupied a solid large area in eastern Brazil. About 50 languages survive, although most of these are slowly being replaced by Portuguese. Mason (1950) combined Ge and eight smaller stocks spoken near it into a Macro-Ge grouping, which, in his terms, was still "far from proven." The eight added stocks were Caingangan, Camacanian, Machacalian, Purian, Patacho, Malali, Caropa, and Botocudo.

5. Tupí-Guaraní is the second largest linguistic stock of South America. It has two major divisions: Tupian, spoken in eastern Brazil, mainly along the coast from Uruguay to the mouth of the Amazon, and inland along the Amazon to the foothills of the Andes; and Guaranian, spoken in Paraguay and the neighbouring region of Argentina.

Languages of both divisions were used by the first European traders and missionaries as contact languages in their dealings with the Indians. About 120 Tupí-Guaraní languages are still spoken, but these, as well as the two more widespread contact languages, are slowly giving way to Portuguese and Spanish.

Macro-Tupí-Guaraní, proposed very tentatively by Mason (1950), combines Tupí-Guaraní with two smaller stocks, Huitoto-Miranyan and Zaparoan in eastern Ecuador and Peru.

6. Quechuan. Archaeological and historical studies suggest that the Quechuan languages probably occupied a small area in the southern Peruvian highlands until, about 1450, they were spread over a far larger region by the rapid conquests of the Inca empire. When Pizarro conquered the Incas in 1533, Quechuan languages were spoken over most of the west coast of South America, from southern Colombia to central Chile, and eastward from the coast to the borders of the Amazonian basin. Quechuan languages are still spoken by several million people in roughly the same area, many of whom speak no other language. Nevertheless, the Quechuan languages are slowly losing ground to Spanish, which, by reason of its position as the language of government and education, plays an increasingly important role throughout the area. The modern Quechuan languages, about 28 in number, are said to be very much alike.

7. Aymaran. The languages of this group were spoken over a fairly large region in the southern Peruvian highlands and adjacent areas of Bolivia. Inca conquests probably reduced the Aymaran area somewhat, but there remain almost 500,000 speakers of the 14 Aymaran languages.

J. Alden Mason (1950) suggested that Quechuan and Aymaran might be combined into a single stock, which he called Quechumaran. The evidence for this relationship is not impressive.

Many of the remaining languages of South America are grouped into 45 stocks, which are more restricted in size and distribution than the ones listed above. These are:

In Venezuela: Chirianan, Cuica-Timote, Guahiban, Guarauan, Jirajaran, and Salivan.

In Colombia: Mocoa, Pamiguan, Puinavean, and Tucanoan.

In Brazil: Arauan, Borotuque, Carajan, Caririan, Catuquinan, Guaitacan, Huanhaman, Muran, Nambicuran, Panoan, Peba-Yaguan, and Zamucoan.

In Ecuador: Canelan, Jivaroan, and Omurano.

In Peru: Aguanoan, Cahupanan, Munichean, Sabelan, and Yunca-Puruhan.

In Bolivia: Chapucuran, Chiquitoan, Mosetenan, Tucanan, and Yurucarean.

In Paraguay (and adjacent parts of Brazil and Argentina): Guaycuruan, Mascoian, and Mataco-Maca.

In Chile: Alacalufan, Araucanian, and Ataguitan (also in adjacent parts of Argentina).

In Argentina: Chon (or Tehuelchean), Comechingonan, Huarpean, and Lule-Vilelan.

This listing by no means accounts for all the languages of South America. P. Rivet and C. Loukotka (1952) listed 108 stocks—the list above contains only *52*. But the additional "stocks" were actually very little known: McQuown estimated (1955) that there were over 600 languages in South America on which there were too little data to permit even a tentative classification. (*See* also INDIAN, LATIN AMERICAN.)

South America is unmatched in its linguistic diversity. The classification given above must therefore be regarded as highly tentative, and subject to drastic revision when better knowledge of the area becomes available.

BIBLIOGRAPHY.—J. Alden Mason, "The Languages of South American Indians," vol. 6, pp. 157–317 of *Handbook of South American Indians*, U.S. Bureau of American Ethnology *Bull. 143* (1950); N. A. McQuown, "The Indigenous Languages of Latin America," *Amer. Anthrop.*, 57: 501–570 (1955); P. Rivet, G. Stresser-Pean and C. Loukotka, "Langues de l'Amérique" in A. Meillet and M. Cohen, *Les Langues du monde* (1952). (H. HR.)

SOUTHAMPTON, HENRY WRIOTHESLEY, 3RD EARL OF (1573–1624), Shakespeare's patron, was born at Cowdray, Sussex, on Oct. 6, 1573. He succeeded his father in 1581 and became a royal ward under the care of Lord Burghley. Educated at St. John's College, Cambridge, and Gray's Inn, London, at 17 he was presented at court, where he was favoured by the queen and befriended by the earl of Essex. Southampton became a munificent patron of poets: Barnabe Barnes appended a sonnet in his honour to *Parthenophil and Parthenophe* (1593); Thomas Nashe dedicated *The Unfortunate Traveller, or, the Life of Jack Wilton* (1594) to him; and Gervase Markham his poem on Sir Richard Grenville's last fight (1595).


BY COURTESY OF THE NATIONAL PORTRAIT GALLERY, LONDON

HENRY WRIOTHESLEY, 3RD EARL OF SOUTHAMPTON. AFTER A PORTRAIT BY D. MYTENS, ABOUT 1618

Southampton is best known, however, as the patron of Shakespeare, who dedicated *Venus and Adonis* (1593) and *The Rape of Lucrece* (1594) to him. It has also been argued, inconclusively, that the sonnets were addressed to him (*see* SHAKESPEARE, WILLIAM: *The Poems: The Mystery of "Mr. W. H."*). If so, the earlier sonnets, urging marriage, must have been written before the beginning in 1595 of Southampton's intrigue with Elizabeth Vernon, one of the queen's waiting women, which culminated with their hasty marriage in 1598, incurring the queen's wrath and leading to their brief imprisonment.

Meanwhile in 1596 and 1597 Southampton had accompanied Essex on his expeditions to Cádiz and to the Azores. In 1599 he went to Ireland with Essex, who made him general of his horse, but the queen insisted that the appointment be canceled, and Southampton returned to London. He was deeply involved in the Essex Rebellion, on the eve of which he induced players at the Globe Theatre to revive *Richard II*, a play dealing with the deposition of a king, in order to stir up the populace. He was tried for treason on Feb. 19, 1601, and condemned to death, but his sentence was commuted to life imprisonment through the intervention of Sir Robert Cecil.

On the accession of James I, Southampton resumed his place at court. He was made a knight of the Garter and captain of the Isle of Wight in 1603, and was restored in blood by act of Parliament in 1604. In 1603 he entertained Queen Anne with a performance of *Love's Labour's Lost* by Richard Burbage and his company, to which Shakespeare belonged.

Southampton was an active member of the Virginia and East India companies. A born fighter, he was engaged in more than one serious quarrel at court and was imprisoned for a short time in 1603. He was a volunteer in support of German Protestants in 1614, and in 1617 proposed fitting out an expedition against the Barbary pirates. He became a privy councilor in 1619 but fell into disgrace through his determined opposition to the favourite, the duke of Buckingham. In 1624 he and his elder son volunteered to fight for the United Provinces against Spain, but on landing they were attacked with fever, and Southampton died at Bergen op Zoom on Nov. 10, 1624, a few days after his son.

See "Memoirs of Henry Wriothesley, the third Earl of Southampton," in James Boswell (ed.), *Shakespeare* xx, pp. 427 ff. (1821); A. L. Rowse, *Shakespeare's Southampton* (1965). (RD. C. G.)

SOUTHAMPTON, THOMAS WRIOTHESLEY, 1ST EARL OF (1505–1550), English civil servant and lord chancellor, was born in London on Dec. 21, 1505. The son of one herald, William Writh or Wriothesley, and nephew and cousin to two others, Thomas was well-placed for a career in the royal service. He was educated at Cambridge where he made the acquaintance of Stephen Gardiner, later master of Trinity Hall, bishop of Winchester, and a leading councilor to Henry VIII. Wriothesley subsequently married Gardiner's niece Jane Cheyne. Gardiner appointed him a clerk of the signet in 1530, but when Thomas Cromwell rose to power (1532–33) Wriothesley transferred to his service, finally becoming Cromwell's chief clerk and personal secretary. His promising connections not only got him possession of extensive monastic properties in Hampshire and elsewhere, but also moved him upward in the king's service. In 1538 he went on embassy to the Netherlands; in 1539 he sat in Parliament as one of the knights for Hampshire; in April 1540 he succeeded Cromwell as one of two joint principal secretaries of state. In the same month he was knighted. Cromwell's fall (June 1540) did not interrupt his career; indeed, even before the final crisis he had almost certainly reestablished contact with Cromwell's enemy Gardiner and had worked against his master. Wriothesley was a true Henrician who would have nothing to do with the pope and welcomed the dissolution of the monasteries, but remained a conservative in religion and viewed with apprehension Cromwell's negotiations with the Lutheran states. After Cromwell's fall the dearth of able administrators gave Wriothesley his chance. From 1542 he was one of Henry VIII's leading councilors, rewarded with a barony in January 1544 and (surprisingly, since he was neither a lawyer nor an ecclesiastic) with the lord chancellorship, the senior office of state, in April that year.

Boundlessly ambitious, he hoped to profit from the accession of a minor after Henry VIII's death (January 1547), but the ensuing political struggles proved that he should have remained the senior civil servant the discerning Cromwell had meant him to

be. In February 1547 Protector Somerset bought his support with the earldom of Southampton; a month later, ready to promote the Reformation, Somerset deprived him of the chancellorship. Naturally, therefore, Wriothesley supported the conspiracy which Somerset's rival, John Dudley, earl of Warwick (later duke of Northumberland), led against Somerset in October 1549. But once again he was outmaneuvered: so far from restoring the old religion and the fallen minister, Warwick proved more Protestant still and in February 1550 excluded Wriothesley from the Council. The earl died in London on July 30, 1550. An assiduous and efficient servant to others, he was negligible as a statesman, being gifted with neither vision nor originality nor even competence in the intrigues in which he engaged with such fervour.

See G. R. Elton, *The Tudor Revolution in Government* (1953).

(G. R. E.)

SOUTHAMPTON, a port, city (1964), county borough, and parliamentary borough of Hampshire, Eng., lies 76 mi. SW of London by road. Pop. (1961) 204,822. Area 15.1 sq.mi. The town stands near the head of Southampton Water, an inlet of the English Channel, on a peninsula formed by the Rivers Test and Itchen. It is the third largest in importance among English ports and the first in regard to passenger traffic. In A.D. 43 there was a Roman settlement, Clausentum, on the east bank of the Itchen, and inscribed stones, coins, pottery, etc., have been found. Southampton (Hamtune, Suhampton) superseded the Saxon Hamtune and was a royal borough before 1086. The earliest town charter was given by Henry II, probably in 1154–55, but the borough was not incorporated until 1445. Henry VI, in 1447, created Southampton a county in itself. A charter which remained in force until 1835 was granted by Charles I in 1640. It was from Southampton that the Mayflower sailed with the Pilgrim Fathers to America. The town has returned two members to Parliament since 1295. It is a quarter sessions borough. Southampton gives name to a bishopric suffragan to Winchester. Natives of the town include Isaac Watts (1674–1748), Sir John Millais (1829–96) and Earl Jellicoe (1859–1935).

An old house known as King John's Palace is one of the earliest 12th-century domestic buildings in England. The south (God's House Gate) and west town gates date from the early 14th century; Bar Gate (the north gate), as it stands, is mostly later and retains good Decorated work. St. Michael's Church has an 11th-century tower and arches and a font of Tournai marble (*c.* 1150). The priory of St. Denys, an Augustinian foundation of 1124, gives its name to a suburb by the Itchen. Netley (*q.v.*) Abbey is 3 mi. southeast.

The old guildhall, long used as a courthouse, is over the archway of the Bar Gate. Tudor House, a medieval merchant's residence, is the museum. King Edward VI Grammar School was founded in 1550, Taunton's School in 1760; the University College of Southampton, evolved from the former Hartley Institute, received its charter as a university in 1952. The town suffered severely from air raids in World War II, three-quarters of the old district in the centre having been devastated and important port installations destroyed. The civic centre, completed in 1939, comprises the new guildhall, municipal offices, police headquarters and law courts, and the public library and art gallery. The tower is a landmark for shipping and contains a clock and bells. There are a large sports ground (268 ac.), a common, and several parks.

The Port and Its Trade.—The modern importance of the port began with the creation of a pier and harbour commission in 1803 and the erection of the Royal pier in 1831. The prosperity of the town was enhanced by the opening of railway communication with London in 1840. The fine natural harbour has the advantage of a double tide. In 1950 the docks were taken over by the Docks and Inland Waterways Executive. The older docks lie at the confluence of the Test and Itchen rivers; large modern extensions lie to the northwest of these alongside the Test River. The old docks cover more than 200 ac. and in addition to the river quays there are three large tidal basins, the Ocean dock, Empress dock, and the Outer dock, leading to the Inner dock. The new docks, with 1½ mi. of deep-water quay facing the Test River, include the King George V dry dock, one of the largest in the world.

Southampton is England's chief port for express transatlantic services and is also used by ships bound for South America, East Africa, Australia, New Zealand, and the Far East. There are cross-channel services by British Railways ships and a ferry to the Isle of Wight. Southampton Airport is at Eastleigh, 5 mi. northeast.

Principal industries are shipbuilding and ship repairing, building and contracting, dock and harbour services, distribution services, and marine engineering. There is also a wide range of light industries. A large new oil refinery was opened at Fawley on Southampton Water in 1951.

SOUTH ARABIA, an area on the southern coast of the Arabian Peninsula, bounded on the west by Yemen, on the north by Saudi Arabia, on the east by Muscat and Oman, and on the south by the Gulf of Aden. The area was named the Aden Protectorate in 1937 and renamed the Protectorate of South Arabia in 1963; it is under the protection of Great Britain. Area (including the island of Socotra, lying at the mouth of the Gulf of Aden, opposite Cape Guardafui in Somalia) about 112,000 sq.mi. (290,000 sq.km.); pop. (1965 est.) 868,000. The protectorate is divided into Western (pop. 550,000) and Eastern (pop. 318,000) areas.

The Western protectorate includes 'Abdali (sultanate of Lahij), 'Alawi sheikhdom, 'Aqrabi sheikhdom, 'Audhali sultanate, Upper 'Aulaqi sultanate, Upper 'Aulaqi sheikhdom, Lower 'Aulaqi sultanate, amirate of Beihan, Dathinah state, amirate of Dhala ('Amiri); Fadhli sultanate, Haushabi sultanate, Lower Yafa'i sultanate, Upper Yafa'i sultanate, Sha'ibi sheikhdom, and Wahidi sultanate (Balhaf), together with five small sheikhdoms in the Upper Yafa'i area (Bu'si, Dhubai, Maflahi, Mausatta, and Hadhrami). The colony of Aden (*q.v.*) is on the coast of the Western Protectorate.

The Eastern Protectorate comprises the Qu'aiti sultanate of Shihr and Mukalla, the Kathiri sultanate, and the Mahra sultanate of Qishn and Socotra.

Physical Geography.—South Arabia extends from Ash Shaykh Sa'id, which marks its boundary with Yemen, eastward for about 740 mi. (1,190 km.) to Ra's ("cape") Darbat 'Ali, which marks that with Oman. The southern coastal strip varies in depth from 4 to 40 mi. (6.4 to 64 km.). There are some fertile parts, but in the main it is sand and lava. Behind this plain in the west are mountain ranges varying from 1,000 to 2,000 ft. (300 to 600 m.) and rising to a plateau of from 5,000 to 8,000 ft. (1,500 to 2,400 m.). This merges into the Yemen Highlands and falls away on the east to the sands of the Rub' al Khali. In the east the main feature is the extensive *jol* (plateau) of the Hadhramaut (*q.v.*), varying from 3,000 to 6,000 ft. (900 to 1,800 m.) which also falls away to the Rub' al Khali in a series of steppes. The mountain ranges of the west and the plateaus of the east are dissected by wadis (valleys) which bring down torrents from the Yemen and make limited agriculture possible.

The warm season lasts from May to September, with the southwest monsoon; the cool season from October to April, with the northeast monsoon. The average annual rainfall is about 1 in. (25 mm.) in the maritime region, and rather more inland.

The People.—Population estimates since 1934 have varied from 600,000 to 1,000,000. Almost all are Arabs and Sunni Muslims; ethnically those in the west are of Mediterranean stock, with pockets of Veddoids increasing eastward until in Mahraland virtually all are Veddoid. The Mahras speak a non-Arabic Semitic language and for that reason are not recognized as Arabs. In 1946 the population of Aden and the protectorate was estimated as 70% tribesmen, all armed, and 30% townsmen.

The social organization has been shaped by desert conditions and by Islam. Thus the descendants of the Prophet (mostly through his grandson Husain and called Sayyids) rank highest socially. The tribesmen are the most important single factor in society, for they are armed and in the majority. Townsmen, ranging from important merchants down to *dhafa* (mostly artisans), are unarmed. Truly urban conditions exist only in such towns as Lahij and Al Mukalla, where social conditions resemble those of Aden on a small scale. (*See* further ARAB.)

History.—The presence of Paleolithic and Neolithic artifacts is evidence that the country was populated in prehistoric times, but history really begins after 6000 B.C. when desiccation set in after

the retreat of the last Ice (here Pluvial) Age. The desert conditions turned the inhabitants, whatever their racial origin, into Arabs, attuned to a nomadic life. Tribal life began to evolve, and gradually four main regions developed: Saba, Qataban, the Minaean Kingdom, and the Hadhramaut (*see* further SABAEANS; ARABIA: *History: The South Arabian Kingdoms*).

Cities and the first civilization in Arabia developed in South Arabia from the trade in incense, the demand for which was great. It grew in the Hadhramaut and in regions east of it, and other peoples became rich on the overland carrying trade. The cities fell into ruins, however, and their citizens reverted to tribal life when the Romans ruined the overland trade by taking incense by sea. From that time onward conditions of tribal life continued in South Arabia, though from time to time the tribes fell under the control of dynasties based in the Yemen (*q.v.*).

The modern history of South Arabia can be said to date from 1839, when the British East India Company occupied Aden. This caused considerable uneasiness to the neighbouring tribes, which valued their independence but realized their weakness. However, the political agent of the East India Company and the Bombay government pursued a consistent policy of good relations, and to that end made agreements with six neighbouring chiefs in 1839. In the early 1880s the British and Indian governments agreed to the negotiation of protectorate treaties; the first were made in 1886. With confidence established, the number of treaties increased. All the protected territories were recognized by the Anglo-Turkish Convention in 1914 as within the British sphere. Finally a protectorate order-in-council in 1937 formally constituted the territories of the treaty chiefs into a "protectorate," and the Aden Protectorate acquired a legal existence and became the responsibility of the British Colonial Office rather than of the British Government of India.

Noninterference with the internal affairs of the tribes had over a long period developed into established policy, and it was continued after 1937; it had succeeded in making the protectorate a *cordon sanitaire* for Aden. In 1934 close contact had been made for the first time with the Hadhramaut, and some 1,400 tribes and tribal groups, constantly warring or marauding, were in 1937 persuaded to join in a general truce. In the following year the protectorate was divided administratively into the Eastern and Western Aden Protectorates. This was followed by the two principal chiefs signing advisory treaties (by which they undertook to accept the advice of a resident adviser). Thereafter formal government was organized without interfering with the internal affairs of some 2,000 tribal and other autonomous petty rulers. In 1944 advisory treaties in more restricted terms were introduced in the Western protectorate, and by the late 1940s a far more rigid regime became evident. The margin between advice and direct administration was often imperceptible, and military and air action to reduce some tribes to obedience increased.

In January 1954 federation was proposed to the chiefs but was not welcomed. However the British "forward policy" (the antithesis of the traditional "noninterference" policy) in regard to the protected states had been regarded by the imam of the Yemen as an infringement of the *status quo* agreed to in the Treaty of San'a' (1934), and he had begun what was virtually a war, supported by Egypt and some Communist countries. By 1958 the chiefs were so alarmed that six agreed to federate. In February 1959 Britain and the federation signed a treaty which recognized the latter's desire to develop into an independent state which Britain would assist it to achieve. It also extended protection and provided for the acceptance of advice. A new capital was founded on the Aden border in October and named Al Ittihad ("unity" or "concord"). By the end of 1962 five more chiefs had joined and in January 1963 Aden colony also joined, though remaining under British sovereignty. By 1965 all treaty chiefdoms except four had been induced to join and the total membership (including Aden) was 17 states.

In July 1964 the British government announced that the federation (including Aden) should become independent not later than 1968. Thereafter a succession of political and constitutional difficulties arose and little progress was made in negotiations. During 1964–66 there was also a considerable increase in Yemeni terrorist activity, largely inspired and assisted by Egypt.

Administration.—The transfer of responsibility for Aden and the protectorate to the Colonial Office in 1937 led on to the policy of federation, but deeply entrenched traditional forces still largely prevail internally. The constitution introduced in December 1962 gave legislative and executive authority to the federal government in a number of matters, particularly education and agriculture. In 1963 the governor became a high commissioner, and the order-in-council that gave legal force to the constitution made his authority prevail over the federation's in matters of defense, internal security, and external affairs. The legislature is the Federal Council; Aden has 24 members in it, and most of the other states 6 members each. The executive is the Supreme Council, consisting of those ministers prescribed by federal law and elected from the Federal Council and up to three other members appointed by the ministers and confirmed by the Federal Council. States represented in the Federal Council by at least six members can claim one minister for every six representatives. Ministers hold the chairmanship for one month in rotation.

Each state is self-governing, with such executive and legislative authorities as it may decide. One state, Dathinah, is a republic.

Social Conditions.—The conditions of society derive largely from the traditional social organization. Up to the late 1940s formal education other than religious virtually did not exist outside the Qu'aiti and 'Abdali states, but by the mid-1960s the Western Protectorate had about 7,500 places for boys and nearly 800 for girls at the primary level and for over 700 boys at intermediate level. Education for girls at that level did not begin until 1963. In the Qu'aiti and Kathiri (unfederated) states there were, in that year, 10,700 primary places for boys and girls and 11 intermediate schools. There is a secondary school at Al Mukalla and a teacher-training college at Ghayl ba Wazir which serves both states.

Outside Al Mukalla there were by the mid-1960s eight protectorate hospitals served by 25 doctors and linked to 76 rural health units. The health service was expanding.

The Economy.—Agriculture is the principal occupation, but a comparatively small area is cultivable. The most widely grown crops are millets and other small grains; some wheat and barley are grown at the higher levels. Dates are grown wherever possible, and in the Western Protectorate some coffee, particularly in the mountains of Yafa'. Herds are limited in number and quality by inadequate grazing, but there are more than 1,000,000 sheep and goats, about 70,000 cattle, 80,000 camels, and 13,000 donkeys.

Of cash crops tobacco from the Qu'aiti state was formerly the most important but has been superseded by cotton. By the mid-1960s cotton production from the Abyan Oasis, Lahij, and smaller areas was approaching 40,000 bales.

The growing of fruit and vegetables for the Aden market, first encouraged in Lahij in the 19th century, has greatly expanded, and there is a modern fruit and vegetable wholesale market at Khormaksar in Aden. Much has been done to develop the fishing industry, and many local craft are now power-driven. *See* also ADEN; ARABIA; HADHRAMAUT. (W. H. Is.)

SOUTH AUSTRALIA, a state of Australia in the central portion of the continent, occupying 380,070 sq.mi. (984,395 sq.km.), or about one-eighth of the total area. In the north the boundary runs successively with those of the Northern Territory and Queensland along latitude 26° S and in the east with those of Queensland, New South Wales, and Victoria along longitude 141° E. In the south the state has an irregular coastline (its only natural boundary), about 2,200 mi. (3500 km.) in length, and in the west the boundary marches with that of Western Australia along longitude 129° E. The state capital is Adelaide (*q.v.*).

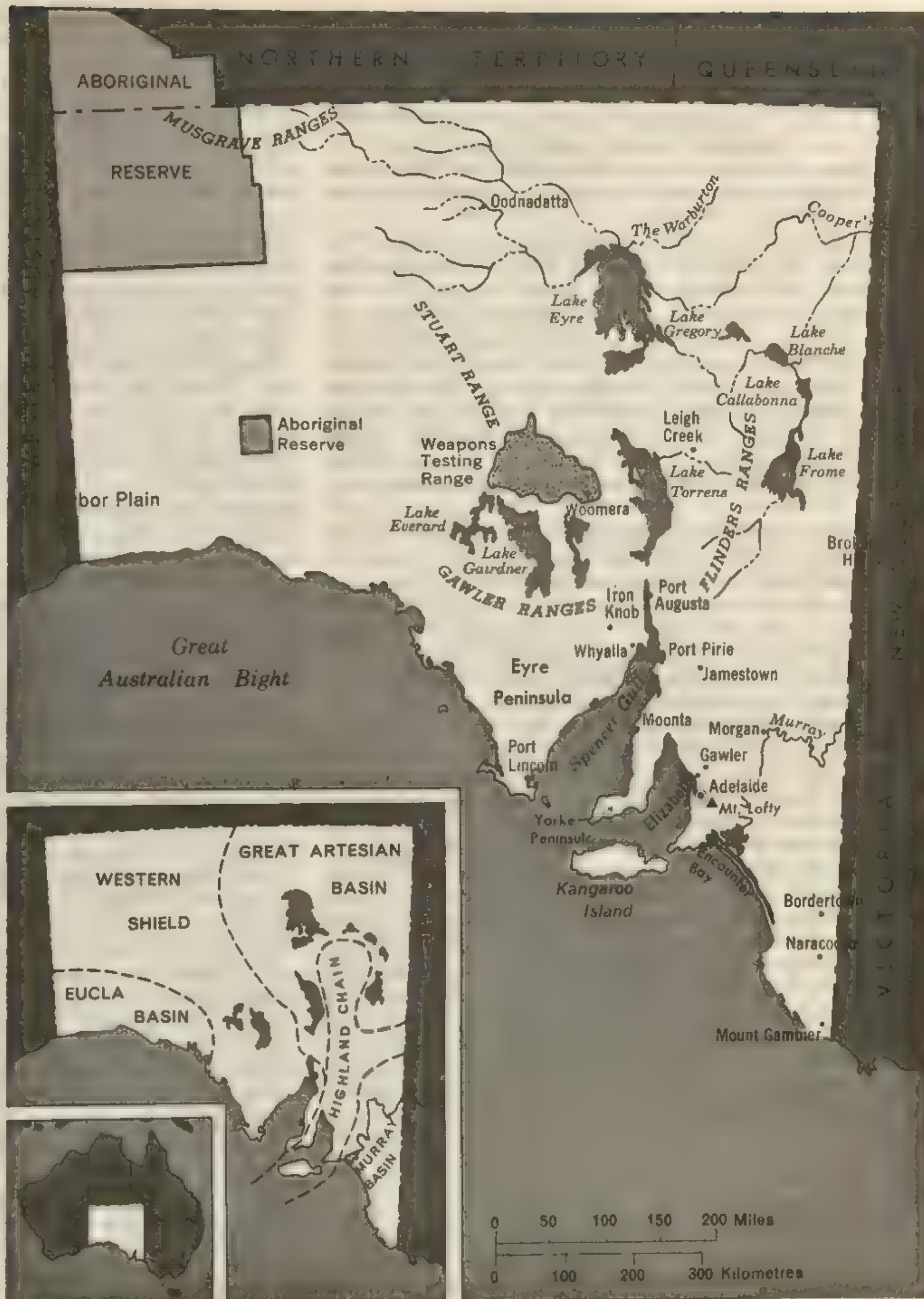
PHYSICAL GEOGRAPHY

Physical Features.—Situated athwart the junction line of the western plateau and the great central lowland of Australia, South Australia can be divided into a number of different physical regions. Its relief is nowhere so marked as that of the eastern states, and there is relatively little land above 2,000 ft. (600 m.). There are

five major physical regions: the western shield; the highland chain; and three sedimentary basins—the Eucla Basin in the far southwestern corner of the state, the Great Artesian Basin in the northeastern corner, and the Murray Basin in the southeast.

Most of the western part of the state is occupied by the eastern portion of the western shield. This extends from the Musgrave ranges in the northwest to the western portion of the Yorke Peninsula in the southeast and consists mainly of Archean igneous and metamorphic rocks. The highland chain lies between the western shield and the Murray Basin and is the most prominent physiographic feature in the state. It includes the Flinders ranges in the north, the spur of the Main Barrier Range between the Lake Frome depression and the northern part of the Murray Basin, and extends southward through the Mt. Lofty ranges to Kangaroo Island and the eastern part of Yorke Peninsula. Most of the rocks in this highland chain are sediments formed in a geosyncline in late Proterozoic and Cambrian time; *i.e.*, more than 500,000,000 years ago.

The depressed area southwest of the western shield is occupied by the Eucla Basin, which extends into Western Australia. In the northeastern part of South Australia lies the southern portion of the Great Artesian Basin, probably the largest of its kind in the world. Beds of freshwater sandstones are the main aquifers and are thickest in the northern part of the state, lying at depths averaging from 4,000 to 5,000 ft. (1,200 to 1,500 m.). The southern part of the basin is occupied by the Lake Eyre depression, part of which is 39 ft. (12 m.) below sea level. South and east of Lake Eyre there is a series of lakes —Frome, Callabonna, Blanche, Gregory, Torrens, and Gairdner—which occupy a number of mud- and marsh-filled depressions. The Murray Basin lies in the southeastern part of the state and extends into western Victoria. On the whole it is a very flat area covered with Recent sediments. A series of low ranges of sand dunes runs parallel to the coastline. Around Mt. Gambier a few craters, occupied by lakes, have become tourist attractions.



MAJOR PHYSICAL FEATURES, CITIES AND TOWNS OF SOUTH AUSTRALIA

Climate.—The state may be roughly divided into two climatic regions: a northern warm, dry region with high summer temperatures and less than 10 in. (250 mm.) of rainfall per annum, much of which is erratic and unreliable; and a southern region typically Mediterranean with moist winters and hot, dry summers. Very few areas in this latter region receive more than 35 in. (900 mm.) of rain each year. North of the 32nd parallel and in the eastern part of the state north of the Murray River, the mean annual rainfall is less than 10 in., but, as in most dry areas, distribution is very erratic. Some sections may have 10 in. in a few months followed by years of almost complete drought. The driest known area of the whole continent occurs around Lake Eyre (*q.v.*).

South Australia is warm and sunny. In June, July, and August most of the state experiences mean monthly temperatures between 10° and 15.6° C (50° and 60° F). In the summer, however, the southern part of the state has mean monthly temperatures around 18.3° C (65° F), while in the northern part of the state the monthly average is between 26.7° and 29.4° C (80° and 85° F).

Soils.—The podzols, or ash gray soil, are found in the wetter parts of the state where iron oxide and other soluble materials have been leached out leaving the surface soil ash gray in colour and poor in plant nutrients. Where the rainfall is slightly less and the evaporation higher the podzols give way to the red brown earths, which extend north of Adelaide beyond Jamestown and in the southeast around Naracoorte and Bordertown and on Eyre Peninsula. All these areas are important for agriculture and grazing. The mallee soils, rich in lime but poor in humus and phosphate,

are closely associated with the mallee type of vegetation. With the application of superphosphate and some of the important trace elements (copper, zinc, and molybdenum), many of these mallee soils, once thought to be of little economic value, are gradually being sown down with subterranean clover, and it is expected that they will add considerably to the pasture acreage of the state. Desert-steppe soils and desert sand hills extend over the northern half of the state and are used for extensive grazing where artesian water is available.

Vegetation.—The natural vegetation consists mainly of five major types. Eucalyptus woodland and forest is on the higher and wetter parts of the Mt. Lofty and Flinders ranges and in the southeastern part of the state. Mallee (dwarf eucalypt), she-oak, and tea tree once covered the remainder of the southern portion of the state, but large areas of mallee have been cleared for agriculture. In the drier north the mallee gives way to mulga scrub (acacia), and on the fringe of the Nullarbor Plain in the west saltbush is dominant. In the northwest and northeast spinifex grass is the chief plant of the very dry areas, but in places there is no vegetation at all but only sandy ridges or rough gibber (stony) plains.

The great need for forest reserves was recognized at the end of the 19th century, and, although the total area set aside for forestry purposes is only 0.1% of the area of the state, more than 150,000 ac. were under plantation in the early 1960s, 95% of which were planted with soft wood. These man-made forests have changed the landscape of the southeastern part of the state.

Water Supply.—Climatic conditions made the provision of an adequate water supply an early problem for settlers in South Australia. Throughout the arid areas they were forced to rely on surface catchments for which large tanks were often constructed below ground level. Artesian water later became available, and the artesian wells are still a valuable source of water for the drier parts of the state. In the settled areas local water supplies have not been able to keep pace with the industrial and commercial developments, and it has, therefore, been necessary to draw on the resources of the Murray River. In 1944 the Morgan-Whyalla pipeline was completed to convey water 223 mi. from the bend in the Murray River at Morgan to the industrial area and shipping port at Whyalla on Eyre Peninsula. A branch was subsequently constructed to take water to the new settlement at the Long-Range Weapons establishment at Woomera, 110 mi. NW of Port Augusta. In 1954 the Mannum-Adelaide pipeline, capable of delivering 14,600,000 gal. per yr., was completed. In the 1960s further schemes were under consideration. It is estimated that 90% of the population of the state is dependent upon the Murray for its water supply. (G. H. Ln.)

HISTORY

The First Settlement.—South Australia was officially settled as a new British province on Dec. 28, 1836. Thirty-four years earlier the South Australian coast had been explored by Matthew Flinders, but no organized settlement had been proposed until 1830, when plans made by the National Colonization Society were placed before the Colonial Office. Later the South Australian Land Company and the South Australian Association also put forward plans. Each proposed to apply the Edward Gibbon Wakefield (*q.v.*) theory of systematic colonization in a new territory where settlers might enjoy civil and religious liberties, but each in turn was rejected by the government as "too republican." These frustrating negotiations ended in 1834 when an appeal to Parliament brought the South Australian Act, which guaranteed a fixed price for land and ensured that no convicts would be sent to the colony. To make certain that no cost would be borne by the mother country, the Colonial Office was to have charge of all government affairs except land sales, emigration, and fund raising, which were to be controlled by an independent board of colonization commissioners.

When settlement began, this division of authority led to bitter factional strife, exacerbated by the ill-considered rules of the commissioners in London. A site for the capital, Adelaide, was chosen and soon divided into quickly sold lots, but the survey of the country sections was unduly delayed, and very little food was produced until 1840. Some settlers thrived on speculation in town land and imported supplies, while thousands of poor immigrant families, heedlessly sent out by the commissioners, remained idle. As the commission had ordained a free market for labour, the first governor, Capt. John Hindmarsh, was powerless; the second, Col. George Gawler, exceeded his instructions by employing immigrants on costly public works. Before his unauthorized bills reached London, the commission was already bankrupt, its funds frittered away in "puffing" the new province. However, a parliamentary committee that examined South Australian affairs in 1841 blamed Gawler for the province's disastrous financial state. During the next year, Parliament paid most of the debts, stopped emigration, removed the colonization commissioners, and placed the province wholly under Colonial Office control.

Progress Toward Self-Government.—The third governor, Capt. George Grey, was pledged to economize. He arrived to find land surveyed and settlers at work on their country sections. By 1844 the colony was paying its way. Although wool was the chief export, farmers were growing more wheat than the population of 17,000 could consume. In 1845 rich copper discoveries brought a spectacular mining boom that made many settlers independent of the absentee investors who had hitherto financed and directed them from England. Local politics thus gained new interest and became more controversial. The Colonial Office was defied over mining royalties and road taxes, and when it insisted that state aid be offered to all churches, only four denominations accepted it. The vigorous protests of Nonconformists brought an end to government grants in 1851, making South Australia the first part of the British Empire completely to separate church and state. At the same time, four years of heavy immigration increased the population to 67,000 and qualified South Australia for the grant of a partially elected legislative council with the right of drafting a constitution of self-government.

The gold rush to Victoria had drawn thousands of South Australians when the first constitution was prepared in 1853. To returning diggers and irate citizens alike the constitution was "a caricature of self-government." Their petitions led to a second constitution, which provided two elective chambers: a House of Assembly with triennial elections, manhood suffrage, secret ballot, and no property qualifications; and a Legislative Council with a modest property franchise to protect landed interests. To the Colonial Office this was "the only thorough Benthamite constitution in the empire," but royal assent was given, and South Australia's first Parliament with responsible government met in 1857.

Political independence seemed to exhaust the colonists' ideas but not their powers of resistance. During the first 30 years of self-government there were 34 changes of ministry and very little experimental legislation.

One important early exception was the Real Property Act (1858), subsequently copied by more than 50 countries and sometimes called the Torrens Title System after its sponsor, Sir Robert Torrens (*q.v.*). To make conveyancing simple, quick, and cheap, the act replaced the unwieldy system of title deeds by one of certificates guaranteed by the state. In the courts the act was persistently challenged by Judge Benjamin Boothby, who claimed that it and other colonial enactments, including the constitution, were repugnant to English law. Boothby's stubborn stand soon led to his removal from the bench. To meet his objections the Imperial Parliament passed the Colonial Laws Validity Act (1865), which, by defining "repugnancy," became the legal keystone of the empire until the Statute of Westminster of 1931.

Economic Expansion.—By making titles secure the Real Property Act also simplified mortgaging, especially for wheat growers, whose small freeholds were beginning to force livestock owners farther afield in search of leases. Because pastoral expansion seemed to be checked by a barrier of salt lakes, exploration was stimulated. Slow-moving government parties found gaps between the lakes, and more mobile private explorers passed through them to find vast new pastures. The most notable of these pioneers was John McDouall Stuart, who finally crossed the continent in 1862, paving the way for South Australia's acquisition of the Northern Territory (1863) and for the overland telegraph,

which in 1872 linked Adelaide with the ocean cable terminus at Darwin. Other explorers found traces of minerals. In 1860 the discovery of rich copper lodes near Moonta aroused hopes of general prosperity and encouraged many pastoralists to move their stock beyond the salt lakes. Disaster followed, for the country was poorly watered; in the drought of 1864–65, thousands of sheep and cattle perished. As the drought did not extend very far to the south, Parliament was easily persuaded that all the southern areas were safe for agriculture.

During the gold rushes South Australian grain had fed the diggers in Victoria and followed them into New South Wales and Queensland. From 1853 steamboats carried flour far up the Murray and Darling rivers and returned with cargoes valuable enough to justify a railway and new port at Encounter Bay. In the mid-1860s an even greater market for grain was found in England. To the Parliament in Adelaide the prospect seemed brighter for wheat than for wool. In 1869 the Strangways Act introduced "credit land sales," by which the land hungry might acquire freeholds on expired pastoral leases on easy terms. A decade of prosperity followed. Immigration, heavily subsidized by the state, provided agricultural recruits and labour for the government's new railways. Each year brought new records in cultivation; yet South Australia, with one family in three on its farms, depended too much on agriculture for its economic well-being.

Hope and inexperience carried the farms into unsafe districts, and old overworked districts became "wheatsick." Though prosperity lasted until 1884, two droughts heralded collapse. Bankruptcies mounted as prices fell, banks failed, and copper mines showed signs of exhaustion. Average yields of 15 bu. per ac. dropped to less than 2 bu. Long before the economic depression ended, South Australia lost its lead in wheat growing to Victoria and New South Wales and fell behind Queensland in population. Numerous families migrated to eastern states, and many more were attracted to the towns even though urban industry had been neglected and employment was insecure. Demand for liberal reform stirred Parliament into action. Land in safe districts was offered on easy terms to homesteaders and workingmen. Promises of compensation encouraged pastoralists to improve their ranges and multiply their flocks. Water conservation projects were started; southeastern swamplands were drained; and irrigation and village settlements were begun on the Murray River. The opening of Roseworthy Agricultural College introduced farmers to the use of superphosphates, while scientific techniques began to produce better strains of wheat, freedom from disease, systematic fallowing, and more varied crops. Assisted immigration stopped, but railway building continued. To avoid further borrowing in London—the public debt was already close to £50 per head—in 1884 the government introduced for the first time in Australia direct taxation on incomes and land. When this proved insufficient, tariffs were sharply increased to curtail imports and foster urban industry. The opening of Broken Hill in New South Wales helped South Australia by providing employment for its miners, freight for its railways, and smelting works at Port Pirie, but these gains were at first offset by heavy migration to Western Australian goldfields. The high levels of production in 1884 were not regained again until 1906. Although the population had passed 300,000 by the mid-1880s, it did not reach 400,000 until 1910.

Political Parties.—The most marked advances of the depression years were in politics. Growing discontent brought the eight-hour working day (1884), payment of members of Parliament (1887), workmen's compensation (1889), industrial arbitration and votes for women (1894), and the first Australian use of a referendum (1895). Primary education, which had become compulsory and secular in 1875, was made free in 1891. The depression years also hastened South Australia's progress toward federation.

The tentative beginnings of a Labour Party during the 1890s forced its opponents into closer alliances and more stable ministries. Leading politicians, perplexed by inland tariff barriers, problems of the Murray River irrigation, and annual deficits in the Northern Territory, were eager for intercolonial cooperation. In the federal conventions South Australians played a leading part. Later, when the ablest leaders were drawn into the federal Parliament, local politics became more subdued. The Labour Party grew in strength and in 1905 took office in a coalition government. Its main concern was the reform of unequal electoral constituencies: Adelaide, with half the population, had only one-third of the seats in Parliament. Reform was resisted by the Legislative Council, whose abolition, though advocated by Labour, was not popular with electors. Nevertheless, the Labour Party did form occasional governments, most of them ending in dissension.

The postwar years brought the organization of the Farmers and Settlers' League, the Progressive Country Party, the Nationalists, and the Liberal Union. From these emerged the Liberal Federation and Country Party Alliance, which by 1933 forced Labour into apparently permanent opposition.

Economic Depression.—During the first decade of federation South Australia seemed to prosper. Its share of federal customs revenue paid the annual interest on the public debt, while bumper harvests and rising prices renewed enthusiasm for agriculture. Aided by new railways and superphosphates, farmers opened fresh wheat belts on Eyre Peninsula and in the southeast. Assisted immigration was resumed. Valuable iron ore deposits at Iron Knob were linked by rail with the new port of Whyalla. Port Adelaide was enlarged and an outer harbour added for mail steamers. Electric trams and suburban railways revived land speculation, while government high schools and free secondary education attracted families to the city.

The tide turned in 1910. Although the costly Northern Territory passed under commonwealth control, a new agreement for distributing federal revenue lowered South Australia's income. This was obscured at first by high prices for grain and wool during World War I and the early 1920s, but from 1914 to 1926 the government had to budget for eight deficits. State taxation quadrupled and government borrowing increased the public debt to £A136 per head (compared with Victoria's £A88). Large numbers of workers and some entire industries migrated each year to the eastern states where bigger markets, cheaper power, and lower taxes offered better opportunities. After 1928, four dry years coincided with worldwide depression to reduce the state to a worse plight than any of its neighbours. In the winter of 1931 more than 70,000 unemployed out of a population of 575,000 were dependent on government relief. Thousands of the unemployed sought work in the country, enabling a record crop of 8,000,000 ac. to be sown, although depressed prices gave farmers an average of less than *20s.* per ac. for their harvest. Appeals for federal aid led to the establishment of a Commonwealth Grants Commission and brought substantial assistance earmarked for developmental works. However, most politicians still looked to the land for salvation. To the auditor-general, W. Wainwright, this dependence on rural output seemed disastrous. He advocated more secondary industry: by judicious government guarantees to suitable enterprises the state would attract overseas investment and unlock its own private capital tied up in land. The threatened departure of a large motor-body-building industry forced the government to support Wainwright's plan. Certain wharf dues were abolished, and company tax was halved. The first large project was the planting of pine forests in the southeast, for the state had always depended on imported building timber. Guarantees were made to the cement and chemical industries at Port Adelaide. Elsewhere, industries were encouraged to make equipment for the expansion of roads, water supply, sewerage, and building. To supplement private efforts the government built large reservoirs for Adelaide, expanded technical education, and created a trust for housing industrial workers.

World War II and After.—In 1938 Thomas Playford began his long term as premier, and his accession infused new life into industrial policy. Some of his plans had strategic importance and were implemented during World War II. The Broken Hill Proprietary's new blast furnace and shipbuilding yard led to great expansion at Whyalla which had to be supplied with water by pipeline from the distant Murray River. Near Adelaide large munition works built by the commonwealth government were taken after World War II by overseas firms, some for new industries and others for the development of long-range weapons. The need for an iso-

lated testing site for the weapons brought the town of Woomera into existence, its water piped from the Murray. In 1946 Playford's government took over the private Adelaide Electric Supply Company, turning it into an electricity trust that cut prices and greatly expanded consumption. To make the state less dependent on imported fuels, the government opened a brown coal mine at Leigh Creek, and a large supplementary power station soon followed at Port Augusta. The postwar years also brought great rural expansion. Scientific testing of inferior soils revealed the absence of trace elements, which, when restored, gave fertility to much unused scrubland. Other marginal areas were opened by mechanization and improved techniques. High returns for wool and barley freed farmers from the vagaries of wheat markets, while road transport, diesel rail engines, and bulk handling of grain helped to produce new records in cultivation, pasture improvement, and sheep numbers. The most spectacular expansion, however, was in Adelaide. Its new and growing factories attracted a great influx of labour, which in turn demanded new homes and services. The city itself became dependent on water from the Murray. As the Adelaide Plain rapidly filled with buildings, a new satellite town was created nearby at Elizabeth. (D. H. PE.)

POPULATION

At the 1961 census the population of South Australia was 969,340 (excluding full-blooded aborigines who numbered 2,147), or about 9.2% of the total for Australia, compared with 797,094 at the 1954 census. During 1954–61 the population grew at the high rate of 2½ to 3% per annum, slightly more than half the increase being due to immigration. The fourth largest in population of the states of Australia, South Australia had in 1961 a density of about 2.6 persons per square mile. The population is heavily concentrated in the capital, Adelaide, which ranks fourth in size among Australian cities. The population of the Adelaide metropolitan area in 1961 was 587,957. The only country towns with populations exceeding 5,000 were: Elizabeth 23,326; Mount Gambier 15,388; Port Pirie 14,003; Whyalla 13,711; Port Augusta 9,711; Salisbury 9,349; Port Lincoln 7,508; and Gawler 5,639.

ADMINISTRATION AND SOCIAL CONDITIONS

Administration.—The constitution of South Australia vests legislative power in an elected Parliament comprising a Legislative Council and a House of Assembly. The Legislative Council consists of 20 members, half of whom retire every three years, being replaced by members elected for a term of six years from each of five electoral districts. The House of Assembly, elected for a period of three years, is composed of 39 members representing single electorates. Voting for elections to the House of Assembly is compulsory. Executive power is vested in a governor, appointed by the crown, and an executive council, which comprises the governor and the ministers of the crown. The local government authorities consist of district councils, municipal corporations, and cities.

Education.—South Australia has a comprehensive system of state education, with compulsory schooling until the age of 14 years. In addition to the state school system, a number of private schools have been established under church auspices. The principal institutions of higher education are the University of Adelaide and the South Australian Institute of Technology. There are also an agricultural college and teacher-training colleges.

Health and Welfare.—South Australia enjoys a healthful climate. The infant mortality rate of about 20 per 1,000 births is among the lowest in the world. The state is also fortunate in enjoying high standards of housing, and Adelaide has virtually no slums. At the 1961 census the number of persons per dwelling averaged only 3.5. In the early 1960s there were nearly 70 public and more than 150 private hospitals containing about 4,500 and 3,400 beds, respectively. Of these, the four major hospitals, which were in the Adelaide metropolitan area, contained a full range of facilities for diagnosis and treatment. Incomes in South Australia are comparatively high, and there are no great inequalities in the distribution of incomes. Average wages are about £A23 a week.

THE ECONOMY

Agriculture.—Although it is the third in size of the Australian states, South Australia has a relatively small area that can be utilized for agriculture. More than four-fifths of the total area has an average rainfall of less than ten inches a year and supports fewer than 2,000,000 sheep and 100,000 cattle.

In the higher rainfall areas sheep rearing and cereal production predominate. By the 1960s the sheep population of about 15,000,000 was nearly double the prewar figure and constituted 10% of the Australian total. More land was being put under pasture for livestock as a result of a wider use of subterranean clover, superphosphate, and modern mechanical methods of clearing. By these means, land previously considered useless, or able to carry as little as one sheep to four acres, was able to carry two sheep to the acre. The area sown to pasture increased more than eightfold after 1940.

For many years wheat has been the most important cereal in South Australia. A peak of 4,180,000 ac. was sown to wheat in 1930–31. Livestock then began to assume greater significance in areas formerly devoted exclusively to cereals, with the result that the wheat acreage has greatly declined. This decline has been partly offset by the expansion of the acreage under barley and oats, and in the early 1960s the value of barley output was almost as great as that of wheat. The total cereal acreage in the same period, however, remained only about half that of 1930–31. After World War II, wheat yields rose sharply, and in the early 1960s, at about 20 bu. per ac., they were almost double the yields of the 1930s.

Dairying is carried on mainly on the plains and in the hills within reach of Adelaide. The Angaston, Tanunda, Reynella, and Clare districts are noted for their excellent wines; South Australia produces three-quarters of the Australian output. Oranges, lemons, peaches, apricots, and almonds abound on the plains. By the Murray River citrus fruits, vine crops, and vegetables are grown under irrigation.

Mining.—Although South Australia is not rich in minerals, there are a number of important deposits. Most valuable are the large deposits of iron ore situated in the Middleback ranges, which are thought to amount to about 170,000,000 tons on the surface and so far have provided the bulk of ore for Australia's steel industry. This iron ore is of exceptionally high grade; it averages 62% iron and enables Australia to produce relatively cheap steel. Extensive exploratory campaigns were being undertaken in the 1960s to discover further deposits, and good-grade iron ore has been located underground. Attention was also being directed to the possibility of using the hematite-quartzites, of which the Middleback ranges are largely composed, for iron production. A proportion of the iron ore quarried in South Australia is converted into pig iron at Whyalla, where there is a modern blast furnace. By the late 1960s a fully integrated steelworks was to be established at Whyalla.

South Australia possesses abundant deposits of limestone and dolomite and a wide variety of clays. The state has no large coal deposits; the most important is at Leigh Creek, where about 55,000,000 tons of subbituminous coal is available for extraction by open-cut methods. This coal is used mainly for the production of electricity. South Australia produces four-fifths of Australia's salt requirements and contains the great bulk of the country's gypsum. The state also provides about nine-tenths of Australia's talc requirements and is the only source of opal.

Industry.—The industrialization of South Australia has taken place since World War II, and the manufacturing industry is now the state's main source of employment. At the 1961 census 27% of the work force was engaged in manufacturing, while only 12% was in primary production.

The chief features of South Australian industry are to be found in mineral processing and in automobile and consumer-durable production. Lead-silver concentrates are processed at Port Pirie in one of the world's largest lead smelteries, and sulfuric acid is produced at Port Adelaide. There is large-scale production of fertilizers and cement, and an oil refinery has been established near Adelaide. General Motors Holden and Chrysler Australia are the

main automobile producers, the former having its main Australian body-building works and the latter its principal plant in Adelaide. Around this key industry many subsidiary ones have developed. South Australia is the leading Australian state in the production of the larger household appliances, such as refrigerators, washing machines, and electric stoves, and in the production of tubes and pipes. A number of firms manufacture radio and television receivers and electronic equipment. The shipbuilding yards at Whyalla are the largest and most modern in Australia.

External Trade.—In the early 1960s the total value of South Australia's exports averaged about £A100,000,000 annually. Among the chief exports were wool, lead, wheat, barley, and ores and concentrates. During the same period the total value of imports was about £A60,000,000. Imports included machinery, motor vehicles, electrical equipment, and textile piece goods. (Though these figures are given in £A, Australia adopted a decimal system based on the dollar in 1966.)

Transport and Communications.—There are nearly 4,000 mi. of railway in South Australia, carrying about 15,000,000 passengers and 4,500,000 tons of freight annually. The state has about 87,000 mi. of road, of which about 6,400 mi. is hard surfaced. The cost of building and maintenance of roads is shared between the state and federal governments and the local councils. South Australia possesses several large harbours, the main one being Port Adelaide. Many ports serve small vessels, and there is a considerable coastal trade. In the early 1960s there were nearly 30 airfields in the state and 16 scheduled airline services carrying about 450,000 passengers annually. There were also about 200,000 telephones and 16 radio and 3 television stations.

See also references under "South Australia" in the Index.

(P. H. K.)

Bibliography.—The most complete work on South Australia is *The Centenary History of South Australia*, prepared by the Royal Geographical Society of Australasia, South Australian branch (1936). *The Proceedings of the South Australian Royal Geographical Society* (1885–) contain many valuable studies. *See* also E. Hodder, *The History of South Australia from Its Foundation to the Year of Its Jubilee*, 2 vol. (1894); A. Grenfell Price, *The Foundation and Settlement of South Australia, 1829–1845* (1924); C. E. Fenner, *South Australia, a Geographical Study* (1931); D. H. Pike, *Paradise of Dissent: South Australia, 1829–1857* (1957); G. D. Coombe, *Responsible Government in South Australia* (1957); R. J. Best (ed.), *Introducing South Australia* (1958); Commonwealth Bureau of Census and Statistics, *Official Year Book of the Commonwealth of Australia* (annually).

SOUTH BEND, a city of northern Indiana, U.S., the seat of St. Joseph County, is 90 mi. (145 km.) E of Chicago on the south bend of the St. Joseph River, where it turns northward to Lake Michigan. The locale was visited by Robert de la Salle in 1679; the American Fur Company in 1820 and Alexis Coquillard in 1823 established fur-trading posts at the site. South Bend was platted in 1831, incorporated in 1835, and chartered in 1865.

The population of the city in 1960 was 132,445, that of the standard metropolitan statistical area (St. Joseph and Marshall counties) 271,057. (For comparative population figures *see* table in Indiana: *Population*.)

South Bend's industry is diversified, including manufacture of automobile and truck parts, aviation equipment, guided missiles, farm machinery, tools and dies, clothing, toys, baits and fishing rods, bearings, wallpaper, elastic and surgical goods.

The pioneer industries of South Bend were the Studebaker Brothers Manufacturing Company, the Oliver Chilled Plow Works, and the Singer Sewing Machine Company cabinet works. Studebaker began as a family-operated wagon-building business in 1852, and for 111 years it produced Studebaker wagons and then automobiles in South Bend. Studebaker made its first electric automobile in 1902 and its first gasoline model in 1904; wagon manufacture continued until 1920. It consolidated with the Packard Motor Car Company in 1954 and after 1960 diversified into various nonrelated manufacturing fields. By 1963 only the acquired industries consistently showed profits, and the company shifted all automobile production to its facility in Hamilton, Ont. Corporate offices remained in South Bend, but thousands of hourly rated workers left the city; others were absorbed by the small industries which purchased parts of the Studebaker plant. Readjustment for Studebaker workers was aided by a federal program of retraining the unemployed.

The Singer Company ended its cabinet production in South Bend in 1955. The plant location in 1868 had been determined by the availability of the northern Indiana hardwood forests. By 1901 it was the largest and finest cabinet factory in the world, with an employment peak of 3,000 in 1914. When it closed to move closer to lumber supplies, it was employing 1,200 men.

In 1960 the third of the early South Bend industrial giants underwent change when the Oliver Corporation became a wholly owned subsidiary of the White Motor Company. Oliver began manufacturing plows in 1855. In 1929 it incorporated with other plow manufacturers to broaden its line of farm implements. During the late 1950s much of its manufacturing was shifted from South Bend to other plants. By 1965 only 700 men were employed in the South Bend plant, most of them skilled workers.

About 45% of the South Bend working force is employed in manufacturing industries and 25% in transportation equipment fields. The city is surrounded by rich peppermint-growing muck which also produces potatoes, onions, and truck produce.

To the immediate north of South Bend lie two nationally known Roman Catholic educational institutions, St. Mary's College and the University of Notre Dame. Notre Dame, established in 1842 by Edward Sorin (*q.v.*), has a substantial academic reputation in the natural sciences. Its world-renowned Lobund Germfree Life Laboratory (*see* Germ-Free Life), its unsurpassed football teams, and its famous "gold dome" on the administration building are hallmarks of the university. St. Mary's, a small liberal arts college for women, chartered in 1851, is the mother house of the Sisters of the Holy Cross. South Bend also has an extension centre of Indiana University. Other community cultural programs are provided by the South Bend Symphony Orchestra, the Presbyterian Players, the South Bend Art Association, the Northern Indiana Historical Society, and the public library. South Bend lies near many small lakes and the Lake Michigan resort area. (J. J. D.)

SOUTH CAROLINA, popularly called the "Palmetto state," is an Atlantic coast state of the United States, one of the original 13. The eighth state to ratify the constitution, it entered the union on May 23, 1788. It is bordered on the north by North Carolina, on the south and west by Georgia and on the east by the Atlantic ocean. Its area is 31,055 sq.mi. (80,433 sq.km.), of which 783 sq.mi. (2,028 sq.km.) are water; in size it ranks 40th among the states. The South Carolina state flag has a blue field in the centre of which is represented a palmetto tree and in the upper corner of which, next to the staff, is a crescent. The state flower is the yellow or Carolina jessamine, the state tree the cabbage palmetto, the state bird the Carolina wren, the state song "Carolina." The capital is at Columbia (*q.v.*).

PHYSICAL GEOGRAPHY

Physical Features.—South Carolina (lying between the extremes of approximately 32° 4′ and 35° 12′ N. latitude and 78° 31′ and 83° 23′ W. longitude) is mainly in the coastal plain and Piedmont plateau regions, but in the northwest it extends slightly into the Appalachian mountain region. Locally the coastal plain region is known as the low country, and the Piedmont plateau and Appalachian mountain regions are called the upcountry. The coast is low, and islands increase in size and number toward the Georgia border. For about 10 mi. inland the coastal plain is occupied largely by salt marshes. Although continuing flat, the surface rises at the rate of approximately 2 ft. per mile for 50 mi. or more. Parts of the Blue ridge (*q.v.*) rise abruptly from the foothills to 3,425 ft. (1,044 m.) in Mt. Pinnacle, 3,208 ft. (978 m.) in Caesar's Head and 3,124 ft. (952 m.) in Table Rock. The highest point in the state is Sassafras mountain (3,560 ft. [1,085 m.]) in the Blue ridge and on the North Carolina state line. The mean elevation of the state is about 350 ft. (107 m.).

The principal rivers rise in the Appalachian mountains and flow southeast into the Atlantic ocean. Santee river is formed by the confluence of the Wateree (known above Wateree creek north of Camden as the Catawba; *see* Santee-Wateree-Catawba River System) and the Congaree, which is in turn formed by the Broad

and the Saluda. The basin of this system embraces about one-half the area of the state. In the northeast the Waccamaw and the Pee Dee and its tributaries—the Little Pee Dee and Lynches—are wholly within the coastal plain, but the main stream of the Pee Dee is a continuation below the fall line of the Yadkin river which rises in the mountains of North Carolina (*q.v.*). The Edisto is the principal stream in the southeast.

In the Piedmont plateau there are rapids in the rivers, but in the coastal plain the current becomes sluggish, and in time of high water the rivers spread over wide areas.

Climate.—Along the coast the climate is mild and equable. At Charleston, the mean winter temperature is 51° F. (10.6° C.), the mean summer temperature 80° (26.7°), the mean annual temperature 66° (19°) and the range of extremes from 104° to 7° (40° to −14°). For the state the mean temperature is about 63° (18°). In nearly all sections January is the coldest month and July the warmest. The mean annual rainfall for the state is about 47.75 in. (1,210 mm.) and its distribution is excellent, 17 in. (430 mm.) falling during the summer. Snow is uncommon in the southeast, but elsewhere there may be several inches, occasionally more than a foot.

The frost-free season ranges from 245 days on the coast to 204 days on the uplands. Tornadoes sometimes occur in the west, and the coast suffers from hurricanes.

Soil.—In general the soils of the Piedmont plateau are such as have been formed by the disintegration of the underlying rocks. These consist mostly of granite and gneiss, but in the north-central section there is trap and in the southeast section some slate. On the Piedmont plateau the subsoil is a reddish or yellowish clay. In the upper section of the coastal plain the soil is for the most part a loose sand, but near the coast it becomes more fertile, much being underlain by marl.

Vegetation.—Palmettos grow on the coast but only artificially inland. For some distance from the coast there are magnolias, live oaks draped with long gray moss and reed-covered marshes. In the swamps are cypress, gum and bay trees.

In most of the uplands of the coastal plain the longleaf pine is predominant, but large water oaks and undergrowths of several other oaks and of hickories are not uncommon. On the Piedmont plateau and in some of the more hilly sections below the fall line there is some shortleaf pine, but most of the trees in these areas are hardwoods. Deciduous oaks are commonest, but beech, birch, ash, maple, black walnut, sycamore and yellow poplar also abound. On the mountains are oak, chestnut, laurel, white pine and hemlock. Among indigenous trees, shrubs and vines are the blackberry, grape, persimmon, plum, crab apple, hickory, chestnut and hazelnut.

Animal Life.—Because of temperate climate and good rainfall, South Carolina has a wide variety of wildlife. Hunting has long been a favourite sport of the residents of the state, and many out-of-state persons are attracted by it. A number of large hunting preserves and lodges are maintained, especially in the low country. A state wildlife resources commission maintains a conservation program and supervises hunting and fishing laws. The laws specify the following as game animals, and, except during the open season for hunting, generally protected: deer, fox, mink, muskrat, opossum, otter, rabbit, raccoon, skunk and squirrel. Hunters also find duck, geese, quail and wild turkey in the state.

Though a relatively small state, South Carolina has a variety of fish, ranging from game fish in the offshore waters to trout in the northwestern mountains. The commonest fresh-water game fish are largemouth bass, white bass, bluegill, crappie, jackfish or pickerel, yellow perch and redbreast sunfish. Salt-water species found close to shore and in the inlets include channel bass, croaker, flounder, spot, weakfish and whiting. Farther offshore are amberjack, barracuda, blackfish, bluefish, grouper, mackerel, snapper and tarpon.

Parks, Historic Sites and Recreation.—The state maintains 27 state parks, many of which include facilities for boating and swimming as well as camping. Resorts and vacation places attract both South Carolinians and others in all seasons, especially to the mountains and seashore. Much of the coast of the state is marked by a strand and beaches that draw thousands every summer. An abundance of woods and waters is accompanied by a climate that permits year-round activities in most recreational pursuits.

Historic sites dating from days of Spanish and French exploration are to be found throughout the state. Many battles of the American Revolution and the Civil War took place in the state. Of special historical interest is Charleston (*q.v.*), site of the first permanent settlement and a city that still maintains much of its ante-bellum charm and character along with scores of historic landmarks and structures.

HISTORY

The history of South Carolina may be divided into the periods of discovery and exploration (1521–1670), proprietary rule (1670–1719), royal rule (1720–75) and statehood (from 1776).

The first Europeans to visit the coast were a party of Spaniards from Santo Domingo (Hispaniola) in 1521. The first settlement, most probably at Winyah Bay, was made by Spaniards under Lucas Vazquez de Ayllón in 1526, but after a few months it was abandoned. The Spaniards, again settling in 1566, maintained a fort on Parris Island for about 20 years. In the meantime (1562) French Protestants under Jean Ribault made an unsuccessful attempt to establish a colony on Parris Island (*see* PORT ROYAL). In 1629 Charles I granted to his attorney general, Sir Robert Heath, all the territory lying between the 31st and 36th parallels and extending through from sea to sea, but no settlement was made, and in 1663 the same territory was granted to the earl of Clarendon and seven other favourites of Charles II. A second charter in 1665 extended the limits to 29° and 36° 30′. The proprietors were to legislate for the province "by and with the advice, assent and approbation of the freemen." They were empowered, though not required, to grant religious freedom to Dissenters. Circulars were issued in 1663 and 1665 offering most liberal terms to prospective colonists. In the fundamental constitutions adopted by the proprietary board in 1669, John Locke and Lord Ashley prepared for the province an elaborate feudal system of government that would have been obsolete even in Europe. The colonial assembly refused (as the charter gave it the right to do) to adopt them. They were, nevertheless, an element in arousing the feeling of discontent among the colonists which culminated in the overthrow of proprietary rule, and they encouraged the large plantation system which constituted the foundation of the slaveholding aristocracy.

Proprietary Period.—The first permanent English settlement was made in April 1670 at Albemarle point on the west bank of the Ashley river, but as the situation proved unfavourable the government and most of the people moved over in 1680 to the point between the Ashley and Cooper rivers, the site of the present city of Charleston. The area of settlement was gradually extended along the coast in both directions, but did not penetrate far into the interior until after 1730. There were many English from Barbados and French Protestants, both of whom strongly influenced the history of the province.

South Carolina's political history during the colonial era is the story of a struggle between popular and prerogative interests, between the people and the lords proprietor, then, later, between the people and the crown. From 1670 to 1700 the principal questions at issue were the refusal of the settlers to subscribe to the numerous editions of the fundamental constitutions and disputes over the collection of quitrents. Concessions were finally made which brought the government more directly under popular control. In 1693 the commons house elected by the people secured the privilege of initiating legislation. The truce was followed by a controversy between Anglicans and Dissenters. A test act requiring members of the assembly to conform to the Church of England and to take the Sacrament of the Eucharist according to the rites and usages of that church (1704) was defeated only through the intervention of the Whig house of lords in England. By an act of Nov. 30, 1706, which remained in force until 1778, the Church of England was made the established religion. After a few years of peace and prosperity, the proprietors, acting on the advice of Chief Justice Nicholas Trott, adopted a reaction-

ary policy, vetoed several popular laws and could not give aid in the desperate 1715 Indian war. In 1719 the people rebelled, overthrew the existing government and elected their leader, James Moore, governor.

Royal Rule.—The result of the revolution was accepted in England, and the province at once came under royal control, although the rights of the proprietors were not extinguished by purchase until 1729.

Theoretically South Carolina and North Carolina constituted a single province, but, as the settlements were far apart, their governments were always separate. Until 1691 each had its own governor. From 1691 to 1712 there was usually a governor at Charleston and a deputy for the northern settlements, and after 1712 there were again separate governors. An order of 1730 for a survey of part of the boundary was carried out in 1735, but the definition of the whole boundary was not completed until 1815.

The period 1725–75 was a period of great prosperity, based upon the trade in pelts (mainly deerskins) and more permanently on the rapidly expanding production of rice and indigo. The southern colonies' Indian trade centred mainly in Charleston. This, and a rich agriculture and the large commercial business to serve them, supplied the economic basis of a coast country society of notable intelligence and culture. One of the earliest theatres in America, musical culture, libraries and the education of a large number of youths in England gave the society around Charleston a tone and finish at that time rare in the new world.

Passing under the royal government did not check the encroachments upon the governor and council by the commons house of assembly, which defiantly announced the powers of the British commons as its model. By 1760 the council had almost ceased to exercise any real control over legislation. It rarely initiated or amended a bill and never attempted to change a money bill without incurring violent denunciation from the popular house and paralyzing legislative action, once for years.

The province was unconsciously preparing for independence. Though measures of the British government after 1763 were not especially oppressive to the province, the people were too long accustomed to having their own way. They had, especially in the bitter dispute over the Wilkes fund (1769–75), involving the right of the commons alone to control finance, developed too fully a sense of general American and even British freedom to submit to measures which they regarded as subversive of the principles of liberty.

Delegates were sent to the Stamp Act congress (1765) and to the Continental congresses (1774 and 1775). A council of safety appointed by an extralegal provincial congress virtually took over the government in June 1775. Royal administration ended when, on Sept. 15, the governor dissolved the assembly and fled.

Revolution.—Wars with the Spanish in 1686, 1702–04 and 1740, with the Spanish and French in 1706, with pirates in 1718, with the Yamasee Indians in 1715 and the Cherokees in 1760–61, with aid from British troops only in 1760–61, and a slave rising in 1739 had accustomed the people to arms. The state suffered severely during the American Revolution both from British troops and from the presence of numerous loyalists. A British fleet attempting to capture Charleston was repulsed by Ft. Moultrie, June 28, 1776. Calm prevailed until Gen. Sir Henry Clinton returned in 1780 with an overwhelming force and obtained the surrender of the city from Gen. Benjamin Lincoln on May 12.

Completely overrun, the state experienced two years of fighting involving more battles, though most of them small, than occurred in any other state. A continental army under Gen. Nathanael Greene, assisted by state troops under Thomas Sumter, Francis Marion, Andrew Pickens and others, slowly drove the British back into Charleston and wrecked the plan of British troops from the south to join those from the north to crush Washington. The chief Revolutionary War battles fought in the state were at Ft. Moultrie, the siege of Charleston (March 12–May 12, 1780), Camden (Aug. 16, 1780), King's mountain (Oct. 7, 1780), Cowpens (Jan. 17, 1781), Hobkirk's hill (April 25, 1781) and Eutaw Springs (Sept. 8, 1781).

Early 19th Century.—The period following the Revolutionary War witnessed an ambitious program of transportation development. The Santee canal, connecting Charleston with the whole Santee-Wateree-Broad-Saluda River system, was opened in 1800. Highway building followed. The latter was dropped, however, and the canal was ruined by railroad development. The South Carolina railroad, 136 mi. from Charleston to Hamburg opposite Augusta, Ga., when completed in 1833, was the longest railroad in the world.

The early state period was characterized by a bitter struggle between the older low country and the newer upcountry, the latter settled largely by Scotch-Irish coming down the Piedmont belt from Pennsylvania, Virginia and North Carolina. In 1786 it was necessary, to allay discontent, to consent to the removal of the capital to a newly located site to be called Columbia. Although removal took place in 1790, some state officers until 1865 kept offices both in Columbia and Charleston and the supreme court met in each city to hear appeals from the two sections. The upcountry (then comprising a larger area than now thus designated), containing four-fifths of the white population and one-fifth the wealth, in 1808, with the help of a liberal low country minority led by Joseph Alston, secured a constitutional amendment apportioning one senator to each county and representatives to each county in proportion equally to white population and wealth. This left the control of the upper house to the low country and that of the lower house to the upcountry. Manhood suffrage followed in 1810. The low country's fear for its slave interests was allayed as slavery, fostered by the increasing

THE GRANGER COLLECTION

LINE ENGRAVING OF CHARLESTON HARBOUR, 1739

Charleston, the earliest permanent English settlement in South Carolina, is still the state's chief port

importance of cotton, spread up the state.

The South Carolina college was chartered in 1801 largely to allay sectional antagonisms. The necessity of standing unitedly against antislavery and later radical Reconstruction agitation from the North prevented the masses from forcing democratic changes until late in the 19th century.

Northern antislavery agitation and the adoption of a tariff harmful to southern agriculture united both sections in the passage of the ordinance of nullification, Nov. 24, 1832, forbidding the execution of the tariff in South Carolina (*see* NULLIFICATION). The readiness of a powerful minority to assist Pres. Andrew Jackson, combined with a reduction of the tariff, operated to deter the majority from forcing the issue to armed conflict, but left a bitterness within the state that disappeared only after 1840 in the face of a graver danger from outside.

After a serious slave conspiracy in 1822 and northern abolitionism in the 1830s, the slave code was severely revised (1840–44), even to forbidding the sending of a slave to freedom anywhere in the world. The years 1847–52 saw most of the people of South Carolina deterred from secession only because of lack of co-operation from other southern states. After 1828 Virginia's southern leadership was superseded by the more uncompromising leadership of South Carolina under John C. Calhoun. South Carolina was the first state to secede, its ordinance of Dec. 20, 1860, being almost unanimously approved by its citizens.

Civil War and Reconstruction.—With a white population of 291,300, the state put into the field an estimated 62,838 effectives,

BRUCE ROBERTS FROM RAPHO GUILLUMETTE

PARROTT GUNS AT FT. SUMTER IN CHARLESTON HARBOUR, THE SCENE OF THE FIRST ENGAGEMENT OF THE AMERICAN CIVIL WAR, APRIL 12–14, 1861

of whom 22% were killed or died in prison. Gen. William T. Sherman's march across the state in 1865 was accompanied by enormous destruction, including the burning of Columbia.

The misfortunes of war were more easily borne than the humiliation of Reconstruction. Under Pres. Andrew Johnson's guidance, the white population elected officers who were soon ejected under the congressional plan of Reconstruction, disfranchising many whites and transferring power to unprepared Negroes, northern adventurers (carpetbaggers) and white loyalists of the South (scalawags) joining them. In the spring of 1868 the state was re-admitted to the union, and entered upon a period of eight years of crime and corruption. Much legislation was by bribery. Stealing extended from large blocks of state property to the price of a politician's whisky. Two hundred trial justices were said to be unable to read. Daniel H. Chamberlain, able carpetbagger and would-be reformer of his own party (governor, 1874–76), declared in 1901 that if he had been re-elected in 1876 his party, even with white assistance, could not have given government "fit to be endured." The most urgent pleas on a nonpartisan basis failing to draw the least sympathy in Washington, the whites in 1876, by combined fraud, intimidation and persuasion of Negroes, elected Gen. Wade Hampton (1818–1902) governor by a narrow majority. Pres. Rutherford B. Hayes's withdrawal of troops in March 1877 marked the collapse of "radical" rule.

Political Developments.—In 1878–80, under the leadership of M. W. Gary (1831–81), the old conflict between upcountry and low country became a fight mainly between the poorer masses and the propertied classes.

The triumph of the former under Benjamin R. Tillman (governor, 1890–94; U.S. senator, 1895–1918), prominent farmers' movement leader, was facilitated by acute agricultural distress. Tillman's victory had come in the state Democratic convention, which had been captured by his agrarian rebels. In the subsequent general election, Col. A. C. Haskell, a conservative representing the Bourbon class who had ruled the state since 1876, bolted the party and campaigned against Tillman, making his appeals to both conservatives and Negroes. His efforts failed and served only to embitter feelings among the whites. The farmers', or more broadly "reform," movement was marked by the establishment of Clemson Agricultural college (1889), the chartering of Winthrop college for women (1886), the dispensary system of state liquor monopoly (later abandoned because of corruption and failure to restrain intemperance) and the work of the constitutional convention of 1895 disfranchising so far as possible the Negro. Direct primary elections, long in use in most of the counties, beginning in some in 1878, were adopted in 1896 for naming the U.S. senators, governor and state officers.

State politics after 1890 presented a strange combination of rancorous personal politics, mass conservatism and dislike for "aristocratic" influence. Class feeling was strongly manifested in the campaigns of Coleman L. Blease (governor, 1911–15). R. I. Manning's administrations (1915–19) were notable for progressive legislation. Factionalism subsiding, Blease was in 1924 elected senator, after having been badly defeated in 1918 on a platform condemning U.S. participation in World War I.

Agriculture and banking suffered severely in the deflation of 1920–21, and all business still more from 1929 to 1937. North and South Carolina were the only states voting against the repeal of national prohibition in 1933. In 1935 Gov. O. D. Johnston climaxed his long hostility to the highway department and its chief commissioner, Ben M. Sawyer, by military seizure of the department until condemned by both legislature and supreme court. Public esteem for the construction by the department seemed justified by the failure to discover any wrong. Extensive violent strikes in 1935–36 caused a conservative reaction. Pres. Franklin D. Roosevelt's attempt to have Sen. E. D. Smith defeated resulted in Smith's overwhelming re-election in 1938, although the state supported Roosevelt's general policies.

From 1876 almost the entire white population was Democratic, partly from historical reasons and partly because of the conviction that union under the only national party that had defended them against Negro rule in the past was necessary for the maintenance of white supremacy. In 1948 the state bolted the Democratic party, casting its eight electoral votes for Gov. J. Strom Thurmond, of the dissident States' Rights Democratic party. In 1952, 1956 and 1960 the state gave its strongest support since Reconstruction to the Republican presidential candidates, and in 1964 cast its electoral votes for Republican Barry Goldwater.

In 1950 new election laws were passed controlling general elections and providing the secret (Australian) ballot in general elections for the first time in the history of the state. Conduct of party primaries was again put under state control (after a 1944 ruling by the U.S. supreme court that had opened primaries to Negroes if conducted under state authority, the legislature had repealed all primary election laws to make primaries voluntary and not subject to federal interference). In 1951 the state abolished the poll tax as a voting requirement and also undertook to curb activity of the Ku Klux Klan.

Economic Changes.—As other parts of the south, the state began an industrial revolution in the 1940s, to a large extent attributable to expansion and building related to World War II. This trend in South Carolina was accelerated following the war and was officially encouraged by the State Development board (established 1945). Changing the character of the state has been the growing diversification of its industry. Considerable growth occurred after 1939 in clothing, furniture, meat packing and the

chemical and paper and pulp industries. The U.S. Atomic Energy commission also established its huge Savannah River plant in the state. (*See* also *The Economy*, below.) Accompanying the 20th-century industrial revolution was a relative decline of the agriculture which once dominated the state's economy (*see* also *Agriculture*, below). Despite the rapid expansion of industry the large number of marginal agricultural workers kept the per capita income of the state low (about $1,600 in the 1960s as compared with the U.S. average of $2,400).

GOVERNMENT

South Carolina was governed from 1670 to 1719 under the provincial charter of 1665, from 1719 to 1775 under commissions and instructions from the crown and from 1776 under constitutions of 1776, 1778, 1790, 1865, 1868 and 1895. An amendment to the constitution may be proposed in either house of the legislature. If approved by two-thirds of the members elected to each it must be submitted to the people at the next election for members of the house of representatives. If it is approved by a majority of those voting upon it and subsequently by a majority vote in each house of the general assembly chosen at that election, it becomes a part of the constitution. A convention to revise the constitution may be called by a two-thirds vote in each house ratified by a majority of those voting for representatives at the next election.

A voter must be 21 years old, a resident of the state two years, of the county one year and of the voting precinct four months; and must obtain a certificate of registration conditioned on showing that he or she is able to read and write the constitution, or the past year paid taxes on property assessed at $300 or more. Paupers, idiots, the insane, prisoners and persons ever convicted of certain crimes, unless pardoned, are disfranchised. The application of the literacy test or other restrictions in such a manner as to prevent Negroes from registering or voting was prohibited by the Voting Rights act of 1965.

Until the slow emergence of a two-party system in the 1960s, the Democratic nomination was equivalent to election. Voting in the Democratic primary was confined to whites without other qualification than age and residence until federal courts in 1948 ordered Negroes admitted to party membership. An increasing number of Negroes voted after that date, with voting qualifications the same after 1950 in both primaries and general election.

Executive.—Constitutional amendments changed the terms of the governor and other state officers from two to four years, beginning with those elected in 1926. The governor may not succeed himself. He may veto "any one or more of the items or sections contained in any bill," but may be overruled by a two-thirds majority of those voting in each house. In practice the item veto has been used only for appropriation bills. The governor cannot grant pardons except on the recommendation of the pardon board, but he may commute a death sentence to life imprisonment.

Legislative.—The general assembly, meeting annually, is composed of the senate (one member from each of the 46 counties elected for four years) and the house of representatives (124 members elected for two years from the counties in proportion to population). The state elects six congressmen.

Judiciary.—The 5 supreme court judges and the 14 circuit court judges are elected by the legislature for ten and four years respectively. Magistrates (justices of the peace) are appointed by the governor with senate confirmation, although the recommendation of the senator from the county or of the primary election practically determines the selection.

In a number of counties, a county court (with a judge usually elected) disposes of a large part of the lesser civil and criminal cases elsewhere burdening the state circuit courts. Special juvenile courts have been established in some counties; elsewhere the probate judge has this duty. The state is divided into two federal court districts.

Local Government.—The unit of local government is the county. Counties are divided into townships, which have no governments and exist largely for the assessment of taxes. Although the forms of municipal government are provided under general statutes and enjoy certain constitutional rights, they are considered subject to state legislative authority. New industries are encouraged to locate in the state by exemptions from certain municipal and county taxes under conditions set by the state constitution and statutes.

The county is administered by a commission, almost always popularly elected. Road or hospital commissions, etc., are frequently appointed by the governor, usually upon recommendation of the county legislative delegation. Most of the county officers are elected, with the exception of the auditor and treasurer, who are appointed by the governor with senate confirmation; however, the governor almost always names the successful primary election candidates without question.

Since there is no legislative body on the county level, the state legislature is oppressed with a mass of local legislation. In actual practice it enacts without question whatever the senator and representatives of the county concerned agree upon. The delegation thus forms in effect a county legislature. The senator of each county, by controlling action in the senate, exercises an enormous power and in effect has a veto over any legislation concerning his county.

Finances.—The state tax commission, created in 1915, supervises all tax administration and directly administers those taxes from which most of the state revenue is derived, such as taxes on incomes, gasoline, retail sales, soft drinks, beer and liquor, electric power, corporations and business. There exists in practice a considerable separation of state and local tax sources, the local governments depending mainly on the general property tax; however, the counties receive a share of state gasoline and income taxes, and both county and municipality share the state tax on alcoholic beverages.

POPULATION

The population of South Carolina in 1790 was 249,073; in 1840 it was 594,398; in 1880, 995,577; in 1910, 1,515,400; in 1940, 1,899,804; in 1950, 2,117,027; and in 1960, 2,382,594. The 1960 figure represents an increase of 12.5% over the population in 1950. The population per square mile in 1960 was 78.7, as compared with

*South Carolina: Places of 5,000 or More Population (1960 census)**

Place	Population				
	1960	1950	1940	1920	1900
Total state	2,382,594	2,117,027	1,899,804	1,683,724	1,340,316
Abbeville	5,436	5,395	4,930	4,570	3,766
Aiken	11,243	7,083	6,168	4,103	3,414
Anderson	41,316	19,770	19,424	10,570	5,498
Beaufort	6,298	5,081	3,185	2,831	4,110
Belton	5,106	3,371	2,119	1,780	826
Bennettsville	6,963	5,140	4,895	3,197	1,929
Camden	6,842	6,986	5,747	3,930	2,441
Cayce	8,517	3,294	1,476	746	—
Charleston	65,925	70,174	71,275	67,957	55,807
Cheraw	5,171	4,836	4,497	3,150	1,151
Chester	6,906	6,893	6,392	5,557	4,075
Clinton	7,937	7,168	5,704	3,767	1,869
Columbia	97,433	86,914	62,396	37,524	21,108
Conway	8,563	6,073	5,066	1,969	705
Darlington	6,710	6,619	6,236	4,669	3,028
Dillon	6,173	5,171	3,867	2,205	1,015
Easley	8,283	6,316	5,183	3,568	903
Florence	24,722	22,513	16,054	10,968	4,647
Gaffney	10,435	8,123	7,636	5,065	3,937
Georgetown	12,261	6,004	5,559	4,579	4,138
Greenville	66,188	58,161	34,734	23,127	11,860
Greenwood	16,644	13,806	13,020	8,703	4,824
Greer	8,967	5,050	2,940	2,292	648
Hartsville	6,392	5,658	5,399	3,624	704
Lake City	6,059	5,112	2,522	1,606	375
Lancaster	7,999	7,159	4,430	3,032	1,477
Laurens	9,598	8,658	6,894	4,629	4,029
Marion	7,174	6,834	5,746	3,892	1,831
Mount Pleasant	5,116	1,857	1,698	1,575	2,252
Mullins	6,229	4,916	4,392	2,379	828
Myrtle Beach	7,834	3,345	1,597	—	—
Newberry	8,208	7,546	7,510	5,894	4,607
North Augusta	10,348	3,659	2,629	1,742	—
Orangeburg	13,852	15,322	10,521	7,290	4,455
Rock Hill	29,404	24,502	15,009	8,809	5,485
Seneca	5,227	3,649	2,155	1,460	920
Shannontown	7,064	5,828	—	—	—
Spartanburg	44,352	36,795	32,249	22,638	11,395
Sumter	23,062	20,185	15,874	9,508	5,673
Union	10,191	9,730	8,478	6,141	5,400
Walterboro	5,417	4,616	3,373	1,853	1,491
West Columbia†	6,410	1,543	1,744	1,793	1,089

*Populations are reported as constituted at date of each census. †Name changed from Brookland in 1938.
Note: Dash indicates place did not exist during reported census, or data not available.

68.2 in 1950 and with 49.6 for the United States in 1960.

Of the 1960 population, 739,300 or 31% lived in incorporated places of 2,500 or more, as compared with 28.8% in 1950. The entire urban population amounted to 981,386, or 41.2% of the state total. The state has three standard metropolitan statistical areas, which are Charleston, Columbia and Greenville and part of a fourth (Augusta, Ga.). These areas had a total population of 768,024, or 32.2% of the total population of the state in 1960.

The number of households in 1960 was about 604,000, as compared with 515,000 in 1950. The average population per household had declined from 4.0 in 1950 to 3.8 in 1960.

The population of the state was distributed by colour and nativity in 1960 as follows: 64.7% native white; 0.4% foreign-born white; and 34.9% nonwhite, practically all Negro. There were 100.1 males per 100 females in the native white population and 92.9 in the Negro population; 6.2% of the population was 65 years old or over; and 66.3% of the population 14 years old and over was in the labour force. Of the total number of employed, 11.7% were engaged in agriculture, 6.4% in constuction, 32.0% in manufacturing and 2.3% in transportation.

EDUCATION

Public Schools.—As early as 1710 the state made meagre provision for the education of indigent children, and in 1811 an effort was made to inaugurate a genuine public school system. Because of small appropriations and a law that paupers be given first preference, this system proved very inadequate and most education continued to be by private schools and academies in the antebellum period.

The modern free school system was established in 1868. The educational system is under the supervision of a state superintendent of education, with the assistance of a board composed of the governor, the superintendent of education and seven other persons appointed by the governor. State support and regulation of high schools were greatly increased in 1907.

In 1951 Gov. James F. Byrnes successfully sponsored a long-range expansion program designed to improve and equalize physical facilities for white and Negro schools. State bond issues up to $75,000,000 (later increased) were authorized and a 3% retail sales tax earmarked for education was adopted. Prior to this, white children, who in 1950 constituted 55% of the school population, had occupied school property constituting 71.3% of the total value; in 1950 about 67.5% of new construction had been for white schools. Both this ratio and the total expenditure were altered by the 1951 reform efforts. In the first year of the program 78% of the funds for school construction was spent on Negro schools. The value of school property increased from $41,600,760 in 1930 to $90,720,547 in 1950, and to more than $400,000,000 in the 1960s. Annual per pupil cost to the state increased from $35.82 in 1930 to more than $200 in the mid 1960s. In the 1960s per capita average spending for education was about $165 a year in state and local money, in comparison with the national average of a little over $150. While it ranked 49th among the states in per capita income of its inhabitants, South Carolina ranked 29th among all states in the amount contributed toward public education, and was near the top in the percentage of total expenditures devoted to education.

The 1954 decision of the U.S. supreme court holding compulsory racial segregation in public schools unconstitutional centred political activity and public interest on the racial issue. The state constitution was immediately amended to remove the requirement of state support for schools. In 1955 the legislature repealed the compulsory attendance law and the law regulating school terms. Local school boards of trustees were given increased authority and freedom, including the exclusive authority to operate or not operate schools. Beginning in 1963, policy showed signs of change as Negroes were admitted peacefully to the public schools in Charleston and to two state-supported colleges, Clemson and the University of South Carolina. Accompanying these changes were numerous instances of quiet desegregation of various public accommodations and businesses.

Enrollment in public elementary schools at that time was about 406,000, of whom more than 225,500 were white and 180,000 Negro; in high schools there were about 234,000, including 146,000 white and 87,000 Negro. (Negro high school enrollment in 1930 had been only 35,000.)

Higher Education: *State Institutions.*—The state government supports several institutions of higher learning, chief among which is the University of South Carolina at Columbia. The university, chartered as South Carolina college in 1801 and opened for instruction in 1805, comprises schools of arts and sciences, business administration, education, engineering, journalism, law, nursing and pharmacy. It also has regional campuses at Aiken, Beaufort, Conway, Florence, Lancaster, and Union. By the 1960s the university enrolled about 7,000 students annually.

Clemson university, formerly agricultural college, at Clemson (incorporated 1889) is a university and land-grant college, with schools of agriculture, architecture, engineering and textiles, arts and sciences, and graduate school. Winthrop college at Rock Hill (established 1886) is a state college of liberal arts for women, specializing in teacher training. At Charleston is located the Citadel, the Military College of South Carolina (1842), one of the few state-supported military colleges. The South Carolina State college at Orangeburg (1896), for Negroes, has schools of arts and sciences, agriculture and home economics, education, industrial education, graduate studies, and law, and graduate extension and summer schools.

Private Institutions.—Among the private institutions are Columbia college, for women, at Columbia (Methodist, 1854); Erskine college, at Due West (Presbyterian, 1839); Furman university, at Greenville (Baptist, 1826); Newberry college, at Newberry (Lutheran, 1856); Presbyterian college, at Clinton (Presbyterian, 1880); Central Wesleyan college, at Central (Wesleyan Methodist, 1906); Wofford College, for men, at Spartanburg (Methodist, 1854). There are also several nondenominational colleges: Bob Jones university, at Greenville (1927); Coker college, for women, at Hartsville (1908); College of Charleston (1770); Converse college, for women, at Spartanburg (1889); Lander college, at Greenwood (1872); and Limestone college, for women, at Gaffney (1845).

In addition to the state college, private colleges established for Negroes, included Allen university, at Columbia (African Methodist, 1870); Benedict college, at Columbia (Baptist, 1870); Claflin university, at Orangeburg (Methodist, 1869); and Morris college, at Sumter (Baptist, 1908).

HEALTH, WELFARE AND CORRECTIONS

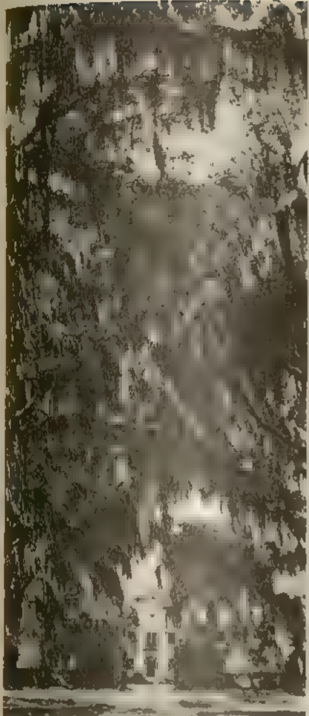
The state has maintained a state board of health since 1879, and beginning in the 1880s similar agencies were established on the municipal level and later by counties. The state board conducts programs for the control of preventable diseases and for maternal and child care. Its program was expanded considerably following passage of the Social Security act in 1935. The state also maintains a mental health commission. Most mental patients are treated at the state hospital, Columbia, although there are four other smaller institutions for such care as well as five mental health clinics.

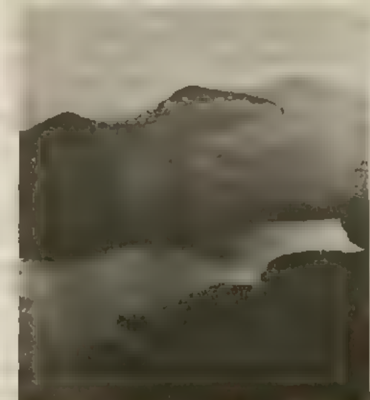
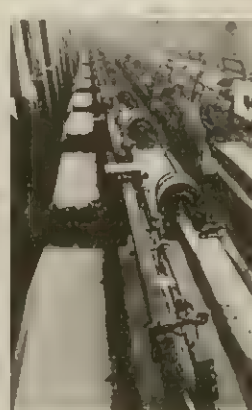
The state department of public welfare (established 1937) is concerned mainly with dependent children, the blind and the aged. South Carolina in 1964 inaugurated a special antipoverty program to reduce illiteracy and unemployment, becoming the first state to use federal funds made available for this purpose under the Manpower Development and Training act (1962).

The state and counties share in maintaining penal institutions for convicted criminals. The state maintains a penitentiary, reformatories and industrial training schools for delinquent minors.

THE ECONOMY

Agriculture.—South Carolina's farm products are normally valued at much less than its factory output. Thus, in the second half of the 20th century, the value of manufactures was over six times greater than the gross income from agriculture. The average size of farms increased, mechanization was widely adopted and farmers generally shared in prosperity. From 1920 through 1950,



(LEFT, CENTRE) TOM HOLLYMAN—PHOTO RESEARCHERS, INC.; (TOP) FREDERICK C. BALDWIN—PHOTO RESEARCHERS, INC.; (RIGHT) BRUCE ROBERTS FROM RAPHO GUILLUMETTE

(Left) Live oaks, festooned with Spanish moss, at Boone Hall plantation near Charleston; (top) small boy fishing on the Edisto river, principal stream of southeastern South Carolina; (centre) Table Rock and lake in the Blue Ridge Mountains, northern South Carolina; (right) spinning room in a Piedmont cotton mill: South Carolina is the largest producer of textiles in the U.S.

the number of farms had declined by more than 25% and the farm population by more than one-third. South Carolina agriculture, like industry, became more diversified in the 20th century. This was particularly noticeable in the case of cotton, which declined 68.6% in value during the period, and in that of peaches, production of which increased by 737.8%.

Total farm acreage harvested in the 1960s was about 2,500,000, in comparison with 4,000,000 in 1950. In the 1960s less than 30% of farm operators were tenants, as compared with 39% in 1954 and 65% in 1930.

Cotton, long dominant in agriculture in South Carolina, took second place to tobacco in crop value for the first time in 1955. In value of production in the 1960s, tobacco led, followed by cotton, soybeans, corn, peaches and hay. These six principal crops accounted for 90% of all crop income.

The coastal region, with a mild climate and a soil responding quickly to intensive cultivation, is especially suited to commercial truck farming. The chief products, exclusive of potatoes and sugarcane grown for sirup, are watermelons, cabbages, cucumbers, string beans, beets, asparagus, tomatoes and lettuce. Strawberries, dewberries, blackberries, figs and grapes add to the fruit wealth of the state.

Dairying and livestock industries were receiving much greater attention by the 1960s. Between 1920 and 1950 the value of cattle increased 62% while the value of hogs increased 27%. Leading livestock and products include poultry and eggs, meat animals and dairy products.

Forests.—Forests make up about 60% of the area of South Carolina, and timber resources are estimated at about 15,000,000,000 bd.ft. in softwoods and 10,700,000,000 bd.ft. in hardwoods. A great deal of emphasis has been placed on pulpwood production for use in the paper industry.

Minerals.—South Carolina's mining output is very small, amounting to 0.19% of the total mineral value of the United States. Chief minerals are kaolin (china clay), vermiculite (produced in substantial quantities in only one other state), stone, sand and gravel. The stone quarried is chiefly granite and granite-gneiss. High-grade granite, the Winnsboro Blue, is quarried in Fairfield county.

Manufactures.—An industrial revolution in South Carolina that began in the 1930s continued into the 1960s, considerably overshadowing agriculture. By the 1960s there were 2,500 manufacturing establishments (up 150% over 1936) employing 250,000 workers (up 100%), and annually producing goods worth $4,000,000,000 (up more than 1,000%). Besides new local industries, many manufacturing operations were moved to South Carolina, and many industries added new plants in the state. Older industries also undertook expansion. With 36% of the spindle hours in the nation, textiles dominate industry in the state (about 50% of total value added by manufacture) and make South Carolina the largest textile producer in the United States. In value added by manufacture, the leading industries of the state are textiles, chemicals, clothing and apparel, paper products, lumber and wood products, and tobacco products.

Textiles consume about 2,500,000 bales of cotton annually. Almost every grade of cotton cloth is produced, and there is also widespread use of synthetic materials, many of which are made in the state. Factors contributing to the rapid development of the cotton textile industry in South Carolina were a suitable climate, dependable labour and an abundance of cheap hydroelectric power. The Piedmont section has led in the textile industry, chief producing counties being Spartanburg, Greenville, Anderson, York, Kershaw and Greenwood.

Charleston and Richland counties are important for diversified industry. Industries showing the greatest growth between 1936 and the 1960s were clothing, furniture, meat packing and paper and pulp.

Hydroelectric development and new steam generating plants greatly aided industry and improved living conditions. Of the numerous hydroelectric installations the greatest is the publicly owned Santee-Cooper project, producing approximately 133,000 kw. Its two reservoirs (Lakes Marion and Moultrie) have a combined area of 160,000 ac. (64,800 ha.). The Clark Hill dam supplies the U.S. Atomic Energy commission's Savannah River plant in Aiken and Barnwell counties.

Lake Murray dam, about 10 mi. (16 km.) above Columbia on the Saluda river, 208 ft. (63 m.) high and 7,000 ft. (2,100 m.) long (one of the largest earth dams), forms a lake about 50,000 ac. (20,200 ha.) in area and develops 130,000 kw. The state's total installed generating capacity (both privately and publicly owned) is about 3,000,000 kw.

Transportation, Commerce and Communication.—Main-line railway mileage in South Carolina was reduced from 3,780 mi. (6,083 km.) at the end of 1930 to 3,250 mi. (5,230 km.) in the 1960s. During the same period there was a steady increase in the mileage of paved roads built by the state government. By the 1960s there were more than 30,000 mi. (48,000 km.) of state highways; the post-World War II period brought 15,000 mi. (24,000 km.) of paved secondary roads also.

Ocean traffic through Charleston (*q.v.*) and Georgetown expanded greatly after World War II, promoted by the State Ports authority, which began a large expansion of port facilities in 1958. Charleston ranked 15th among the nation's ports in the 1960s, and leads all South Atlantic ports in the dollar value of imports and exports.

See also references under "South Carolina" in the Index.

BIBLIOGRAPHY.—J. H. Easterby, *South Carolina Bibliographies* (No. 1, 1949, Topical Lists; No. 2, 1950, General Classified Bibliography); D. D. Wallace, *History of South Carolina*, 4 vol. (1934) and *South Carolina* (1951); Edward McCrady, *History of South Carolina, 1670–1783*, 4 vol. (1897, 1899, 1901–02); W. R. Smith, *South Carolina as a Royal Province, 1719–1776* (1903); R. L. Meriwether, *The Expansion of South Carolina, 1729–1765* (1940); C. G. Singer, *South Carolina in the Confederation* (1941); J. H. Wolfe, *Jeffersonian Democracy in South Carolina* (1940); C. S. Boucher, *Nullification Controversy in South Carolina* (1916); A. G. Smith, *Economic Readjustment of an Old Cotton State: South Carolina, 1820–1860* (1958); P. M. Hamer, *Secession Movement in South Carolina, 1847–1852* (1918); H. S. Schultz, *Nationalism and Sectionalism in South Carolina, 1852–1860* (1950); C. E. Cauthen, *South Carolina Goes to War, 1860–1865* (1950); F. B. Simkins and R. H. Woody, *South Carolina During Reconstruction* (1931); F. B. Simkins, *Tillman Movement in South Carolina* (1926) and *Pitchfork Ben Tillman, South Carolinian* (1944); S. M. Derrick, *Centennial History of the South Carolina Railroad* (1930); C. J. Milling, *Red Carolinians* (1940); W. A. Schaper, "Sectionalism and Representation in South Carolina," in Amer. Historical Assocn. *Annual Report*, I, 1900; R. H. Taylor, *Ante-Bellum South Carolina* (1942); G. B. Tindall, *South Carolina Negroes, 1877–1900* (1952); J. G. Van Deusen, *Economic Bases of Disunion in South Carolina* (1928); D. D. Wallace, *South Carolina Constitution of 1895* (1927); J. K. Williams, *Vogues in Villainy* (1959). Selected biographies: A. K. Gregorie, *Thomas Sumter* (1931); T. D. Jervey, *Robert Y. Hayne and His Times* (1909); L. A. Kibler, *Benjamin F. Perry, South Carolina Unionist* (1946); Broadus Mitchell, *William Gregg, Factory Master of the Old South* (1928); D. D. Wallace, *Life of Henry Laurens* (1915); Laura A. White, *Robert Barnwell Rhett: Father of Secession* (1931); C. M. Wiltse, *John C. Calhoun*, 3 vol. (1944–51); Margaret L. Coit, *John C. Calhoun, American Portrait* (1950); H. T. Cook, *Life and Legacy of David Rogerson Williams* (1916). *Elementary books:* H. K. Hennig, *Great South Carolinians*, 2 vol. (1940–1949); Mary C. S. Oliphant, *History of South Carolina* (numerous late editions).

Current statistics on production, employment, industry, etc., may be obtained from the pertinent state departments; the principal figures are summarized annually in the *Britannica Book of the Year*.

(D. D. W.; C. E. CN.; L. P. J.)

SOUTH CHARLESTON, a city of West Virginia, U.S., across the Kanawha River from Charleston (*q.v.*), is an important chemical manufacturing centre. It was founded in 1916 and incorporated in 1917. Two chemical companies were established there in 1916, and a plant built in 1920 to manufacture antifreeze has become the city's largest single industry, employing over 10,000 workers and participating in the manufacture of about 450 articles. A U.S. naval ordnance plant in South Charleston from 1917 also stimulated the economy. In the heart of the city is an Indian mound 33 ft. (10 m.) in height and 550 ft. (168 m.) in circumference. In 1883 it was explored to a depth of 31 ft. by archaeologists under the direction of the Smithsonian Institution at Washington, D.C. For comparative population figures *see* table in WEST VIRGINIA: *Population*. (K. K. McC.)

SOUTHCOTT, JOANNA (1750–1814), English prophet, whose utterances, dated and sealed on the day she received them, were preserved in a box that, by her wish, was to be opened by 24 bishops a century after her death. The box has never been opened, though as late as 1935 a petition to do so, bearing nearly 29,000 signatures, was presented to the archbishop of Canterbury. Born in April 1750 at the East Devonshire village of Gittisham, Joanna worked as a maidservant and upholsterer and lived a conventionally normal life until, in 1792, her hand began to write automatically. The words were largely biblical, for she was a devout Anglican who also attended Methodist meetings. Assuming that the thoughts came from God, she had them printed. Seven disciples of Richard Brothers (*q.v.*) were impressed and brought her to London in 1802. Her main message was "Christ's Second Coming is nigh." She enrolled 40,000 disciples, including clergy. After puzzling much over the prophecy about the woman whose son was "to rule all the nations" (Rev. 12:5), she developed signs of pregnancy although elderly and chaste (1814). Prominent London doctors assured her that she had conceived a child, and she expected a virgin birth. The child was to be Shiloh, a name connected with an obscure biblical prophecy (Gen. 49:10). But on Dec. 16, 1814, her doctors found that every sign of her phantom pregnancy had vanished, and she died of unknown cause on Dec. 27.

See G. R. Balleine, *Past Finding Out* (1956). (G. R. B.)

SOUTH DAKOTA, popularly called the "Coyote" or "Sunshine state," is one of the north-central plains states of the United States, bordered on the north by North Dakota, east by Minnesota and Iowa, south by Nebraska and west by Wyoming and Montana. It is 372 mi. (599 km.) (east-west) by 204 mi. (328 km.), with an area of 77,047 sq.mi. (199,552 sq.km.), of which 669 sq.mi. (1,733 sq.km.) are covered by rivers and lakes; in size it ranks 16th among the states. Its name is taken from the Dakota or Sioux Indians. The geographical centre of the United States is in the northwestern part of the state. South Dakota was admitted to the union in 1889 as the 40th state. Its capital is at Pierre (*q.v.*). The state motto is "Under God the People Rule," the flower is the American pasqueflower, the bird, the ringnecked pheasant, the song (unofficial) "South Dakota." The state flag has a field of blue on which is represented a blazing sun in gold surrounded by the words "South Dakota the Sunshine State."

PHYSICAL GEOGRAPHY

Physical Features.—South Dakota (extreme points 42° 29′ and 45° 57′ N. lat. and 96° 26′ and 104° 3′ W. long.) has some of the features of regions that bound it: the mountains of Wyoming to the west, the lakes of Minnesota to the east and the rich corn lands of Iowa to the southeast. With its other border states, Nebraska and North Dakota, it is part of the continental Great Plains (*q.v.*), a land that rises gradually to the foothills of the Rockies, accompanied by a climatic change from humid to semiarid. The half of the state east of the Missouri, which originally flowed in the valley of the modern James river, was subjected to glacial action and is largely flat prairie. In the northeastern part of the state there are numerous glacial lakes. Oddly, the lowest point in the state, at the northeastern edge, is a continental divide. From this point waters flow northward to Hudson bay and southward into the Gulf of Mexico. Apart from this extreme corner, the entire state is drained by the Missouri river and its tributaries.

The Missouri, which bisects South Dakota north-south, flows through a veritable gorge except for a short stretch above Sioux City, Ia., at the southeastern corner of the state. The channel is 300 to 500 ft. below the bordering uplands, with the river plain seldom more than two miles and often less than one mile in width. This feature has made possible the location of four federal dams, primarily for hydroelectric power and flood control, along the South Dakota stretches of the Missouri. Total installed capacity is set at 1,232,000 kw. The dams form a series of lakes roughly equal in area to the state's natural water area.

The valley of the James river, on the other hand, averages 60 mi. in width, and the river drops only 100 ft. in its course from north to south across the state (east of the Missouri). The *coteaux* on either side of the James are largely undrained, reflecting their glacial history. The Big Sioux river is at the eastern edge of the state, where, for the last part of its course, it forms the state boundary.

The land west of the Missouri is largely rolling grassland, cut by a number of streams of considerable size. On either side of the White river are the Bad Lands (*q.v.*), the eroded sections of which, banded in soft colours and carved into unusual shapes, are interspersed with numerous well-grassed flat-topped tablelands.

From the James river to the Black hills at the western edge of the state the land rises gently from 2,000 ft. above sea level to 3,700 ft. This area is subject to recurrent periods of drought, the worst occurring in the early 1890s and in the 1930s.

The Black hills (*q.v.*) are actually well-eroded mountains, Harney peak (7,242 ft. [2,207 m.]) being the highest point in the United States east of the Rocky mountains. The region is virtually the only wooded section of the state and the source of a wide variety of minerals. There are three national forests in the western part of the state, covering nearly 2,000,000 ac.

Climate.—South Dakota straddles the 100th meridian, the usual dividing line between the arid and more humid sections of the United States. Rainfall is heaviest in the southeast, averaging 24 in., but tapers off to the west and north to a low of 14 in. Variability in the weather and a wide range of temperature help

to make the climate invigorating. Highest temperature recorded is 120° F.; the lowest −58°. Droughts of varying intensity marked the periods 1886–94, 1910–11, 1930–36 and the mid-1950s. It is estimated that much of the state receives less than 75% of its normal rainfall 20% of the time. The average snowfall is 36 in., ranging from 22 in. to about 59 in. Three-fourths of all precipitation occurs during the growing season. Contrary to popular belief, blizzards are not common. The severest storms on record are those of 1881 and 1949.

Soil.—Most of the eastern half of the state lies in the chernozem (black earth) belt of the Great Plains. The soil varies in depth from 18 to 28 in., is rich in lime and highly fertile. The chestnut soils of western and northwestern South Dakota range from 6 to 14 in. in depth and are generally fertile, but climatic factors limit their use. Soils in the Black hills are thin and vary widely in type. The parent materials of most soils in the state are soft rocks, loess, glacial till and alluvium. Clay and clay shale are important parent materials in the west-central section where gumbo soil or Pierre clay is found. Fringe areas on the northern and southern borders of the state have soils derived from sand.

Vegetation and Animal Life.—A wide diversity of trees is native to the eastern section, including elm, oak, box elder, cottonwood and hackberry. Small cedars are found in much of the rough country west of the Missouri. In the Black hills are ponderosa pine, spruce, aspen, white birch and cedar. Animals widely found include deer and antelope. Buffalo and prairie dogs can still be seen in Black hills preserves. Pheasants can also be found.

Parks, Historic Sites and Recreation.—The state has 170 state and federal parks, recreational areas, camps and picnic grounds. Included in this number are the Mt. Rushmore National memorial, Wind Cave National park, Custer State park in the Black hills area, the Badlands National monument and the Gavins Point dam and Fort Randall dam recreation areas.

NED HAINES FROM RAPHO-GUILLUMETTE

SCENE IN THE BADLANDS NATIONAL MONUMENT SHOWING EFFECTS OF EROSION

The gigantic sculptured faces of Washington, Jefferson, Lincoln and Theodore Roosevelt at Mt. Rushmore are the greatest single tourist attraction in the state, followed by the weird landscape of the Bad Lands of the White river. Camping is especially popular in the Black hills and at federal dams along the Missouri. There are extensive facilities for boating and fishing in the Black hills, in the northeastern region of the glacial lakes, at Lewis and Clark lake above Gavins Point dam and at the Fort Randall dam. The state boasts some of the best pheasant hunting in the nation and is also on the flyway of ducks and geese. Deer are hunted in the Black hills and in some eastern counties. Antelope, once the most popular game animal at frontier army posts, are hunted in western sections of the state.

Sites of great historic interest include a number connected with the periods of exploration and Indian warfare. At Ft. Pierre a small monument marks the place where Louis and François Verendrye buried a lead plate in 1743, claiming the region for the king of France. Ft. Wadsworth, near the Sisseton Indian reservation, has been restored and is maintained by the state. Points visited by Lewis and Clark include Spirit mound, near the town of Vermillion. The site of the Wounded Knee massacre of 1890 is on the Pine Ridge reservation. Yankton, the capital of Dakota Territory from 1861 to 1883 and onetime river port, has many fine old homes as well as a Territorial museum.

HISTORY

Mound Builders resided in the area that is now South Dakota prior to A.D. 1200. Evidence of their occupation has been found in 22 counties east of the Missouri. The Arikara, an agricultural and village dwelling Indian people, inhabited the region after 1600. The first white men to visit the area were the Verendrye brothers, who came into northwestern South Dakota via the Mandan villages of North Dakota in 1742 and may have visited the Black hills. On March 30, 1743, they planted a lead plate on a hilltop on the site of what was later Ft. Pierre, claiming the land for France. The plate was recovered by three school children in 1913.

In 1762 the area was ceded to Spain, and it was under Spanish rule that fur traders began visiting this part of the St. Louis trade empire. The arrival of fur traders coincided with the arrival of the Sioux or Dakota Indians, pushing westward from the woodlands of Minnesota and driving the Arikara upriver before them. Under Napoleon France reclaimed Louisiana, but held it only three years before selling it to the United States in 1803. Lewis and Clark crossed the state in the course of their explorations.

The fur trade on the upper Missouri reached its peak in the 1830s and 1840s. Ft. Pierre, started as a fur trading post in 1817, became the capital of a fur territory of 250,000 sq.mi. Among the principal traders were Manuel Lisa, Pierre Chouteau and Registre Loisel. In support of fur traders Col. Henry Leavenworth in 1823 led a small party of U.S. soldiers and a motley array of trappers, traders and Indian allies against the Arikara at the mouth of the Grand river. In 1831 the "Yellowstone," the first steamboat on the upper Missouri, traversed the state. For more than 50 years thereafter steamboats provided the readiest means of access to the Dakota frontier.

Settlement.—No attempt at settlement by persons other than traders was made until 1856, when land speculators from Iowa and Minnesota, drawn by the prospects of water power at the falls of the Big Sioux river, located there. Indian title had been surrendered by the treaty of Traverse des Sioux of 1851. These settlements were abandoned at the time of the War of the Indian Outbreak in 1862. A permanent settlement was effected at Yankton in 1859, following a treaty with the Yankton Sioux. Three years later, with the creation of Dakota Territory, Yankton became its capital. In the next decade fewer than 12,000 people migrated to the territory, establishing settlements largely along the Missouri and its lower tributaries, the Big Sioux, Vermillion and James rivers.

Rate of settlement increased with the completion of a railroad from Sioux City to Yankton in 1872 and to Sioux Falls and Watertown in 1878. The Black hills expedition of Gen. George A. Custer in 1874, which confirmed rumours that gold was to be found there, also hastened the development of the territory; this invasion of unceded lands contributed directly to the warfare in which Custer lost his life in 1876. The richest gold deposits were found in the area of Deadwood (*q.v.*) and Lead. With favouring rains, settlement reached boom proportions in the 1880s, tripling the territorial population and filling in the lands east of the Missouri, largely with north Europeans.

Statehood.—Efforts to organize as a state were hampered by Democratic opposition in congress, it being known that South Dakota as a state probably would vote Republican. The Republican party had emerged victorious in 1876 by such a narrow margin that its opponents were reluctant to add to Republican strength in congress or in the electoral college. In the election of 1888, however, Republicans captured control of both houses and the presidency, making statehood a certainty and influencing Pres. Grover Cleveland to sign the enabling act under which North and South Dakota were subsequently admitted to the union, simultaneously, on Nov. 2, 1889.

(Left) Lieut. Col. George A. Custer and Sioux scouts on 1874 survey of Black Hills on which gold was discovered; (below) Deadwood, largest of the mining camps, in 1876; (right) 1890 poster designed to attract settlers

(TOP LEFT; RIGHT) THE BETTMANN ARCHIVE; (BOTTOM LEFT) CULVER PICTURES

its support to higher education. Also during the post-World War II period, the Black hills became a major tourist attraction. The gigantic sculptured monument of Washington, Jefferson, Lincoln and Roosevelt, carved on the face of Mt. Rushmore by Gutzon Borglum, was opened to the public in 1942.

GOVERNMENT

The 1889 constitution provides for a bicameral legislature (35 senators and 75 representatives). It met annually from 1964, when it held its first annual session. The governor is nominated by direct primary while most other state officers are nominated by party convention. The constitution is notable for its detailed provisions, a reflection of a populistic distrust of legislatures. In addition, the constitution can be amended by a simple majority vote of legislature and electors. Amendments have been added at the rate of approximately one per year. Most notable is the 1898 amendment which provides for initiative and referendum, South Dakota having been the first state to adopt such measures. The state's 5 supreme court judges, serving for 6 years, and its 20 circuit judges, serving for 4 years, are elected on a nonpolitical ballot.

The state has no personal income tax and ordinarily levies no general property tax. Among the states, South Dakota ranks fourth to seventh from the bottom in per capita payment of state taxes. Sales and use taxes and motor fuel taxes account for more than 60% of all revenue from taxes. After World War II expenditures for highways amounted to roughly one-third of all state expenditures. The other leading items of expenditure were education and social welfare. With the retirement of rural credit bonds in 1954, the state became free of debt.

The next decade was marked by drought and a national economic depression. The distress of Dakota farmers is well described in the works of Hamlin Garland, whose parents had settled near Aberdeen. Farm discontent found expression in wide support of the Populist movement, including the election of a Populist governor and a sweep of the state for William Jennings Bryan in the "free silver" presidential campaign of 1896 (*see* POPULIST PARTY). With the return of prosperity, however, the state resumed its normal Republican course in politics, to be disturbed again only by a flurry for Theodore Roosevelt when he ran as a Progressive in 1912, and by Democrats Franklin D. Roosevelt in 1932 and 1936 and Lyndon B. Johnson in 1964. Democratic governors were elected in periods of farm protest, W. I. Bulow in 1926, re-elected in 1928, Tom Berry in 1932, re-elected in 1934 and 1936, and Ralph Herseth in 1958.

The Republican party frequently has been divided between standpatters and progressives. The latter faction was responsible for constitutional amendments that permitted the state (1915–25) to enter into a variety of business affairs, including rural credits, state bonding, state coal mining, state hail insurance, state cement plant and state guarantee of bank deposits. The state cement plant proved to be highly remunerative, but the other ventures, particularly rural credits, left the state in debt for the next 30 years.

Settlement of the lands west of the Missouri awaited the extension of rail lines and the concentration of Indians upon smaller reservations in the decade before World War I. Further development of the state also was advanced by the discovery of dry farming techniques and the development of plant types adapted to the region. Highway bridges across the Missouri were not constructed until the 1920s. South Dakota suffered greatly during the period of drought, grasshoppers and low farm prices of the 1930s. The dust storms, whipped up from areas where several years of above-normal rainfall had encouraged farmers to plow under the native buffalo grass, were worst in 1934 and 1935. Cash farm income fell to one-third of its former level. In a single year (1933) one farm out of every ten in the state was taken from its owner through foreclosure or tax sale. After World War II, on the other hand, only 1 farm in 1,000 was ordinarily foreclosed or sold for taxes in any one year. At one time nearly two-fifths of the people of the state were on government relief.

By mid-century life on South Dakota farms had improved in many ways. Tenancy and mortgage indebtedness were at a record low. In 1930 seven-eighths of all farms were located on dirt or unimproved roads. With more adequate revenues, the state legislature after 1945 appropriated large sums for interstate and farm-to-market roads. Under Gov. Joe Foss the state greatly increased

POPULATION

The area which now forms South Dakota had 11,776 inhabitants in 1870 and 348,600 in 1890. The population of the state in 1910 was 583,888; in 1940, 642,961; in 1950, 652,740; and in 1960 680,514. This last figure represented an increase of 4.3% over the population in 1950. The population per square mile in 1960 was 8.8, as compared with 8.5 in 1950 and with 49.6 for the United States in 1960.

Of the 1960 population 262,180, or 39.3%, lived in incorporated places of 2,500 or more, as compared with 33.1% in 1950 and 24.6% in 1940. The state has one standard metropolitan statistical area, which is Sioux Falls. This area had a total population of 86,575 or 12.7% of the total population of the state in 1960.

*South Dakota: Places of 5,000 or More Population (1960 census)**

Place	Population				
	1960	1950	1940	1920	1900
Total state	680,514	652,740	642,961	636,547	401,570
Aberdeen	23,073	21,051	17,015	14,537	4,087
Brookings	10,558	7,764	5,346	3,924	2,346
Huron	14,180	12,788	10,843	8,302	2,793
Lead	6,211	6,422	7,520	5,013	6,210
Madison	5,420	5,153	5,018	4,144	2,550
Mitchell	12,555	12,123	10,633	8,478	4,055
Pierre	10,088	5,715	4,322	3,209	2,306
Rapid City	42,399	25,310	13,844	5,777	1,342
Sioux Falls	65,466	52,696	40,832	25,202	10,266
Vermillion	6,102	5,337	3,324	2,590	2,183
Watertown	14,077	12,699	10,617	9,400	3,352
Yankton	9,279	7,709	6,798	5,024	4,125

*Populations are reported as constituted at date of each census.

The number of occupied dwelling units (or households) in 1960

was approximately 217,000, as compared with 190,000 in 1950 and 165,000 in 1940. The average population per household had declined from 3.9 in 1940 to 3.6 in 1950 and to 3.1 in 1960.

The population of the state was distributed by colour and nativity in 1960 as follows: 93.3% native white; 2.7% foreign-born white; and 4.0% nonwhite, almost all Indians. Most of the Indians (nearly all Sioux) live on the state's eight reservations. Of the foreign-born white population a large proportion were born in Scandinavian countries. Germany and Russia also provided significant numbers of immigrants.

In 1960 there were 102.4 males per 100 females in the total white population, and 101.9 in the total nonwhite. Of the total population, 10.5% were 65 years old or over, and 54.8% of the population 14 years old and over were in the labour force. Of the total number of employed persons, 30.5% were engaged in agriculture, 6.4% in construction, 6.6% in manufacturing and 18.9% in wholesale and retail trade.

EDUCATION

Public Schools.—Approximately nine-tenths of the state's primary and secondary school pupils attend public schools. State aid amounts to about 10% of the cost of public education. A high proportion of local property taxes goes to its support, and in many localities school authorities have found it difficult to operate within prescribed mill levy limitations. School reorganization in rural districts is not compulsory, and many common school districts do not operate schools, preferring to pay tuition at other schools for the limited number of children in their own districts. This is particularly true for high school students.

Higher Education.—The University of South Dakota, at Vermillion, is the chief state institution of higher education. Established in 1882, it is governed, along with the other state institutions, by the Regents of Education, State of South Dakota, a seven-member board appointed by the governor with the consent of the state senate. The university comprises schools and colleges of arts and sciences, business, education, fine arts, law, medicine and nursing. It enrolls about 2,500 students annually; the faculty numbers almost 300.

The state also supports the South Dakota State College of Agriculture and Mechanic Arts at Brookings (1883); the South Dakota School of Mines and Technology at Rapid City (1885); and state teachers' colleges at Aberdeen (1901), Spearfish (1883), Springfield (1881) and Madison (1881).

Private institutions of higher education include Augustana college at Sioux Falls (Lutheran, 1860), Dakota Wesleyan university at Mitchell (Methodist, 1885), Huron college at Huron (Presbyterian, 1883), Sioux Falls college (Baptist, 1883) and Yankton college at Yankton (Congregational, 1881).

HEALTH, WELFARE AND CORRECTIONS

The South Dakota department of public welfare provides old-age assistance, aid to dependent children, the disabled and the blind, and general assistance. The state maintains one prison and one training school.

THE ECONOMY

Agriculture.—South Dakota leads all the states in the proportion of personal income derived from agriculture (more than 28%). The economy of the state is thus subject to wide fluctuations, varying with both the market and the weather. Approximately two-thirds of cash farm income is derived from livestock and products; alfalfa, corn and oats are commonly sold "on the hoof" rather than as cash crops. Only in the northeastern and north-central sections do cash crops amount to as much as one-half of cash farm income. The average size of farms in the southeastern section is more than 225 ac.; in the northern wheat areas it is 800 ac.; and in the western range sections, 3,000 ac. In all sections the trend is toward larger farm units and fewer operators.

The state's leading crop is corn, nearly all hybrid strains that have helped to increase yields and to extend the corn area north and west within the state. Sorghum grain is planted as a substitute for corn during extended periods of drought, and in small amounts at other times. Production of corn averages a little over 100,000,000 bu. on approximately 4,000,000 ac. of land. The state usually ranks ninth in the production of corn.

Wheat production, both in acres harvested and in total yield, varies more widely than does corn production. During years of good rain large areas of range land have been plowed up and planted to wheat. This was particularly true during World Wars I and II, which were also periods of high wheat prices. As much as 4,000,000 ac. of land have been planted to wheat, but government quotas reduced this to between 2,000,000 and 3,000,000 ac. In spite of these limitations, total production increased. The state usually ranks tenth in the production of all wheat and second in the production of durum wheat. South Dakota is usually among the first five states in numbers of beef cattle and calves and among the first ten in numbers of hogs and sheep. Milk production is concentrated in the eastern quarter of the state, near to urban centres, while sheep are raised primarily in the northwestern counties. The state ranks tenth in wool production.

Irrigation is carried on extensively in only a few localities. The largest federal reclamation project is Oahe dam on the Missouri, designed to provide water for lands in the James river valley. The Belle Fourche project dates from 1908, the Shadehill dam from 1950 and the Angostura dam from 1952. Other irrigation is carried on along the arid fringes of the Black hills and, by means of wells, in scattered localities in the humid eastern parts of the state.

Manufacturing.—Meat packing and processing of dairy and poultry products are the leading industries, accounting for nearly two-thirds of all workers in manufacturing. Printing and publishing rank next, followed by lumber and wood products.

Mining.—South Dakota ranks first among the states in the production of gold and beryl, second in feldspar, columbium-tantalum and lithium minerals and third in lignite. Between $16,000,000 and $20,000,000 worth of gold is mined annually. The Homestake mine at Lead is the largest gold mine in the western hemisphere. Sioux quartzite and a mahogany-coloured granite are quarried in the eastern part of the state. Uranium ores were discovered in the Black hills in 1951. There are a small number of oil wells, the first dating from 1953. There are undeveloped deposits of low-grade manganese ore near Chamberlain and low-grade uraniferous lignite in the northwestern section of the state.

Transportation and Communication.—The first railroad reached the southeastern corner of South Dakota in 1872, but it was 1881 before the Chicago and North Western and Milwaukee lines reached the central part of the state; it was 1907 before the Missouri river was bridged and east-west rail connections established to the Black hills. There are nearly 4,000 mi. of rail lines in the state. Yankton, Dakota Territory, was a thriving port for steamboats prior to 1881, but the new railroad facilities of that year reduced river traffic to short hauls. The construction of a network of highways, begun in the 1920s, was interrupted by depression and war, but after World War II highway construction claimed the largest share of the state budget. The completion of five Missouri river bridges (1924–26) was a major milestone in establishing the state trunk system which, by the 1960s, consisted of more than 7,000 mi. of highway. Eleven cities are served by commercial airlines; the state has more than 30 public airports.

See also references under "South Dakota" in the Index.

BIBLIOGRAPHY.—Robert J. Casey, *The Black Hills and Their Incredible Characters* (1949); George E. Hyde, *Red Cloud's Folk*, rev. ed. (1957) and *A Sioux Chronicle* (1956); G. W. Kingsbury, *History of Dakota Territory* (1915); Bruce Nelson, *Land of the Dacotahs* (1946); M. L. Reese, *South Dakota, a Guide to the State*, American Guide Series, rev. ed. (1952); Herbert S. Schell, *History of South Dakota* (1961); S. S. Visher, *Geography of South Dakota* (1918).

Current statistics on production, employment, industry, etc., may be obtained from the pertinent state departments; the principal figures are summarized annually in the *Britannica Book of the Year*. (E. W. St.)

SOUTHEAST ASIA usually denotes the area comprising the East Indies (Greater and Lesser Sunda islands, Moluccas, and Philippines) and the Indochinese peninsula of continental Asia. Mainland Southeast Asia consists principally of a series of roughly parallel, north–south-trending ranges and valleys, associated with some of Asia's larger rivers; *e.g.*, the Mekong and the Irrawaddy

(*q.v.*; the principal river in Burma). Insular Southeast Asia includes several geologically unstable island-arcs, possessing an abundance of active volcanoes, as well as a stable zone (Borneo, eastern Sumatra, and nearby areas) resting on the Asian continental platform. The climate is predominantly tropical, with annual rainfall generally above 40 in. (1,000 mm.). The natural vegetation is forest.

The area is highly diversified culturally, although Malaya and the East Indies, with strong affinities among their languages, are a partial exception. Rice and maize are the principal dietary staples. Southeast Asia is an important source of rubber, rice, coconut oil, sugar, teak, Manila hemp, tea, coffee, tobacco, and pepper, and of tin, petroleum, bauxite, and other minerals. Industry is poorly developed. Prior to the 16th century, Southeast Asia had commercial and other contacts with India and China, and in several places, *e.g.*, Cambodia, Sumatra, Java, had evolved high civilizations. From the 16th to the 20th centuries the area was influenced by the West, all portions except Thailand becoming colonies of western powers. By 1963, however, all parts of Southeast Asia had achieved independence except Portuguese Timor; Brunei was independent under British protection.

Turbulence, crises, and change marked the 1960s in much of Southeast Asia. Tensions over positions there continued between the West and Communist China and between the West and the U.S.S.R. For the war and political events in Vietnam, *see* VIETNAM: *History;* the role of the U.S. in the war is also covered in UNITED STATES: *History. See* further LAOS and CAMBODIA. *See* also BANDUNG and SOUTHEAST ASIA TREATY ORGANIZATION. For history of the other political units in Southeast Asia, *see* BRUNEI; BURMA; INDONESIA, REPUBLIC OF; MALAYSIA; PHILIPPINES, REPUBLIC OF THE; SINGAPORE; and THAILAND.

BIBLIOGRAPHY.—E. H. G. Dobby, *Southeast Asia* (1950); Karl J. Pelzer, *Pioneer Settlement in the Asiatic Tropics* (1945); Charles Robequain, *Malaya, Indonesia, Borneo, and the Philippines* (1954); Rupert Emerson, *Representative Government in Southeast Asia* (1955); Russell H. Fifield, *The Diplomacy of Southeast Asia: 1945–1958* (1958); D. G. E. Hall, *A History of South-east Asia* (1955); W. Henderson, *Southeast Asia* (1964). *See* also "Southeast Asia" and articles on the various political units in *Britannica Book of the Year*.

(J. M. BT.; X.)

SOUTHEAST ASIAN LANGUAGES. Southeast Asia is one of the great linguistic treasure houses of the world, both in number of languages and in the complexity of their interrelationships. Burma alone is reported to have between 125 and 140 languages spoken within its borders. Although national languages like Vietnamese, Cambodian, Thai, Malay, and Burmese are literary languages with a centuries-old tradition, others, like the Karen languages and Shan, have acquired a writing system only within the past hundred years. Until the mid-20th century little was known about most of the remaining languages. Since World War II, however, both Western and Japanese scholars have concerned themselves increasingly with many hitherto neglected nonliterary languages.

Linguistic diversity of this degree is the result of the fact that Southeast Asia has been, for untold millennia, a crossroads between India and China and between these and the great islands of the Pacific. At the same time long contact among diverse families has given rise to certain broad similarities characteristic of a linguistic area. Prominent cross-genetic features include the widespread use of status and kin-term pronominals, of numeral classifiers, and of complex verbal strings (*e.g.*, take-come-give, meaning "bring it to me"). As a consequence, even quite unrelated languages like Thai and Cambodian offer greater ease of intertranslatability than either does with English.

Three important linguistic families are represented: (1) Austronesian or Malayo-Polynesian; (2) Mon-Khmer; and (3) Sino-Tibetan. Actual linguistic reconstructions based on well-attested phonetic correspondences have been worked out only for the Austronesian family. Malay (Malaya) is the principal representative of this family on the Southeast Asian mainland. Mon-Khmer, in its narrowest sense, includes Mon (Burma) and Cambodian, or Khmer (Cambodia). Sino-Tibetan may be conservatively described as having three branches: (1) Tibeto-Burman; (2) Chinese; and (3) Tai. Tibeto-Burman, named after the two great literary languages, Tibetan (Tibet) and Burmese (Burma), also includes many nonliterary languages, *e.g.*, Lushai, Garo, and Bodo (Assam), Jingphaw or Kachin (Burma), and possibly the languages of the Chins and Nagas (Burma) and the Lolo (southwestern China). The Karen languages (Burma and Thailand) may also be related. The best-known Tai languages are Thai, or Siamese (Thailand), Lao (Laos), and Shan (Burma). Many lesser-known languages also unquestionably belong here, including the extinct Ahom (Assam), and Dioi, Tho, Nung, Lung-chow, and Wu-ming (northern Vietnam and southern China), as well as others.

The above description of family affiliations leaves many questions unanswered. One theory places Chinese and Tai together in a single subfamily on a par with Tibeto-Burman in Sino-Tibetan. According to another theory, however, Tai is not related to Chinese at all (or to Tibeto-Burman) but is affiliated with Austronesian. Much the same kind of confusion is found with other Southeast Asian languages. Vietnamese is classified now as a Mon-Khmer language, now as a Tai affiliate, but has never been proved related to any other language. The much-debated Austroasiatic hypothesis attempts to encompass the Munda languages (India), Khasi (Assam), Nicobarese (Nicobar Islands), Semang, and Sakai (Malaya and southern Thailand), as well as Mon-Khmer, Vietnamese, and Muong. Miao and Yao (China) are sometimes added, but other students favor a Tai or Tibeto-Burman affiliation.

The Austroasiatic hypothesis, tenuous as it is, has appealed to many as the starting point for other connections; *e.g.*, Austroasiatic plus Austronesian (the "Austric" theory), Austroasiatic plus Japanese, and Austric plus Sino-Tibetan. On a different axis, transpacific relationships with American Indian languages have also been suggested. Possibly deserving of further investigation is Edward Sapir's proposal, in 1925, that Sino-Tibetan (including Tai) may be related to Nadene (the Athapascan languages plus Tlingit and Haida).

Only by continuing to intensify linguistic work in Southeast Asia can we hope to reach a stage at which the interrelationships of these languages can be cogently described. Urgent needs include the linguistic reconstruction of such families as Tibeto-Burman and Mon-Khmer according to the same rigorous methods used to arrive at Proto-Austronesian and Proto-Indo-European. Progress has been made on Proto-Karen and Proto-Bodo, subbranches of Tibeto-Burman. Details regarding the consonants, consonant clusters, and tone classes of Proto-Tai are also fairly well known.

See also the articles on individual languages and language groups.

BIBLIOGRAPHY.—Robbins Burling, "Proto-Bodo," *Language*, 35:433–453 (1959); Robert B. Jones, *Karen Linguistic Studies*, University of California Publications in Linguistics, vol. 25 (1961); Fang-Kuei Li, "Consonant Clusters in Tai," *Language*, 30:368–379 (1954); Thomas A. Sebeok, "An Examination of the Austroasiatic Language Family," *Language*, 18:206–217 (1942).

(M. R. H.)

SOUTHEAST ASIA TREATY ORGANIZATION, usually known as SEATO, is the regional defense organization created by the Southeast Asia Collective Defense Treaty signed at Manila, Sept. 8, 1954, by the representatives of Australia, France, New Zealand, Pakistan, the Philippines, Thailand, the United Kingdom, and the United States. The treaty entered into force Feb. 19, 1955.

The formation of SEATO was a response to the demand that the area be protected against Communist expansionism, especially as manifested through military aggression in Korea and Indochina and through subversion backed by organized armed forces in Malaya and the Philippines. The Republic of Vietnam and the kingdoms of Cambodia and Laos (the successor states of Indochina) were not considered for membership in SEATO for reasons largely relating to the 1954 Geneva agreements ending the fighting in Vietnam. These states were, however, accorded military protection by a protocol. Other nations of south and Southeast Asia preferred to retain their foreign policies of nonalignment.

The treaty area is the general area of Southeast Asia, including the entire territories of the Asian members and the general area of the Southwest Pacific south of latitude 21° 30′ N. The treaty defines its purposes as defensive only and includes provision for self-help and mutual aid in preventing and countering subversive

activities directed from without, and cooperation in promoting economic progress and social well-being.

SEATO has no standing forces but is based upon the mobile striking forces of its member states, which regularly engage in combined military exercises. A secretary-general, acting under the direction of the council or its representatives, heads an international staff in Bangkok which supplements the work of the organization's expert committees on public relations, area security, cultural affairs, and certain economic matters. (R. W. Z.)

SOUTHEND-ON-SEA, a county and parliamentary borough and seaside town of Essex, Eng., 39 mi. (63 km.) E of London Bridge by road, on the Thames estuary facing south. Pop. (1961) 165,093. Area 16.1 sq.mi. (42 sq.km.). The borough includes Leigh-on-Sea, Westcliff-on-Sea, Thorpe Bay and Shoeburyness. With its 7-mi. seafront it is one of England's largest seaside towns and the nearest to London. It has gardens, beaches, an open-air theatre, yachting, a carnival week, magnificent illuminations from August to late October, and a pier (1⅓ mi.) which accommodates pleasure steamers. The oldest part of the borough is Prittlewell where the priory (1110) has housed the museum since 1917. Southchurch Hall (13th century) is a public library and Porters (Tudor) is the civic house. Southend became fashionable as a seaside resort when visited by Princess Charlotte of Wales in 1801 and her mother Princess Caroline in 1803.

It was incorporated in 1892, became a county borough in 1914, and a parliamentary borough in 1918. It returns two members. Besides providing for more than 5,000,000 annual holiday makers, Southend's many industries, mostly on the new industrial estate, include radios and televisions, electrical apparatus, glass blowing, and engineering.

SOUTHERN ALPS, New Zealand: *see* New Zealand: *Physical Geography.*

SOUTHERN AND ANTARCTIC LANDS, FRENCH (Terres Australes et Antarctiques Françaises). These barren lands, which were linked for administrative purposes with Madagascar (now the Malagasy Republic) from 1924, were not effectively occupied till 1949 (mainly by scientific personnel). In August 1955 they became a territory of the French Union, later French Community, governed under special statute by a senior administrator who is partially resident in Paris. The lands comprise: (1) in the south Indian Ocean, and in a temperate cold climate, the Islands of Saint Paul and Nouvelle Amsterdam; (2) in a subantarctic climate, the Kerguelen and Crozet groups of islands; and (3) Adélie Land on the Antarctic continent.

Nouvelle Amsterdam and Saint Paul.—These are two volcanic islands (Nouvelle Amsterdam, area 19 sq.mi. [50 sq.km.] in latitude 37° 52′ S, longitude 77° 32′ E; Saint Paul, area 2.7 sq.mi. [7 sq.km.] in 38° 43′ S, 77° 31′ E). Nouvelle Amsterdam is an extinct volcano (2,953 ft. [900 m.]); the crater of Saint Paul has been filled by the sea. The temperature varies from 5° to 20° C (41° to 68° F) with an average of 12° (53.6°). Fog is frequent and the westerly winds are strong. Mosses, bracken, grasses, and a scattering of trees (*Phylica arborea*) grow there. Animal life consists mainly of seabirds, fish, and crayfish. Access to the islands is difficult.

Magellan's companions discovered Nouvelle Amsterdam in 1522. Fishermen from Réunion used to visit the islands, and France annexed them in 1843. In 1949 a permanent radio-meteorological station, Camp Heurtin, was established at Nouvelle Amsterdam. Research stations, an administrative office, and a hospital were also established there.

Kerguelen.—Kerguelen, lying in latitude 49°–50° S and longitude 69°–70° E, covers about 2,700 sq.mi. (7,000 sq.km.). It is volcanic and has a certain similarity to Iceland. It comprises one main island and about 300 small islets. The main island is broken into ravines by glacial erosion, and has many peninsulas and fjords, chiefly in the east. The Cook Glacier occupies the west; Mont Ross in the south reaches 6,430 ft. (1,960 m.). Temperatures are not excessively low, averaging 0° C (32° F) in the winter and 6° (42.8°) in the summer. Rain and snow are frequent, the west wind is strong, and the sea is often rough. The vegetation consists solely of two flowering bushes (*Acaena* and *Azorella*), some enormous seaweeds, and the edible "Kerguelen's Land cabbage" (*Pringlea antiscorbutica*). There are numerous seabirds, including penguins, petrels, cormorants, sheathbills, skuas, and albatrosses. The chief of the several species of seal that go to Kerguelen to breed is the huge elephant seal (*Mirounga leonina*). The rabbit, introduced by man, has become a pest.

Kerguelen was discovered in 1772 by the French navigator Yves Joseph de Kerguelen-Trémarec who took possession of it. The island was explored in 1776 by Capt. James Cook, who called it the Island of Desolation, and since then it has been frequented by seal hunters. In 1950 a permanent base and scientific centre, Port-aux-Français, was established on the east coast. It has a hospital, military camp, and seal-oil plant. Reindeer, sheep, poultry, and vegetables have been introduced. There are lignite beds and peat bogs, and some semiprecious stones have also been found. About 100 persons (relieved at regular intervals) live on the island, which is served by chartered vessels and also has an airstrip.

Crozet Islands.—Discovered by Capt. Marion Dufresne and his mate, Crozet, in 1722, these comprise two groups of volcanic islands between latitude 45°–47° S and longitude 50°–52° E: (1) the Islands of Cochons, Pingouins, and Apôtres, together with several small islands; (2) Île de la Possession, which rises to 6,560 ft. (2,000 m.), and Île de l'Est, which is glacier covered. The total area is 193 sq.mi. (500 sq.km.). The coasts are steep. Climate, vegetation, and animal life are similar to those of Kerguelen. The islands, uninhabited, have been made a national park.

Adélie Land.—Adélie Land (Terre Adélie), discovered by J.S.C. Dumont d'Urville in 1840, is situated south of Australia on the edge of Antarctica between longitude 136° and 142° E. It is a plateau of ancient rocks covered by ice and has an area of about 150,600 sq.mi. (390,000 sq.km.). It can be reached only in summer, during which the average temperature is 0° C (32° F). In winter the temperature averages −30° C (−22° F), and blizzards reach 125 mph. Whales, seals, petrels, and penguins were the sole inhabitants until 1950 when A. F. Liotard set up the first station at Port Martin. D'Urville Base was established at Pointe Géologie in 1952. Exploratory teams have penetrated more than 300 mi. into the interior.

See also Antarctica.

See Edgar Aubert de la Rue, *Les Terres australes* (1953); *Terres australes et antarctiques françaises* (quarterly). (Hu. De.)

SOUTHERN DESERT TERRITORY (Al Badiyah al Janubiyah), a region in the extreme southwest of Iraq, bordering Saudi Arabia. The region was in 1947 formed into one of three *liwa*'s or provinces (along with the Northern Desert Territory and Al Jazirah) from what had previously been termed tribal areas and left undemarcated. Area 29,399 sq. mi. (76,143 sq.km.). The population is of mixed tribal allegiance, mainly Anaiza ('Unayzah) and Amarat ('Imarat), with many Dhafir and Muntafiq and some others. The region lies astride a main pilgrimage route southwest from Karbala but otherwise has few international connections and trade links. The population is almost entirely nomadic, and there are no important settled centres. By 1960 the territory had ceased to be a *liwa'*, and its area and population had been administratively attached to the settled *liwa*'s. (W. B. Fr.)

SOUTHERNE, THOMAS (1660–1746), Irish-born dramatist, was long famous for two sentimental tragedies successfully acted until well into the 19th century—*The Fatal Marriage, or The Innocent Adultery* (first produced in 1694 and known after Garrick's adaptation in 1757 as *Isabella, or The Fatal Marriage*), and *Oroonoko* (1695). Both were based on novels by Aphra Behn (*q.v.*); both owed much to Thomas Otway in their mingling of pathos with a sometimes flaccid rhetoric. *The Fatal Marriage* anticipated 18th-century domestic tragedy, while *Oroonoko* showed affiliations with the earlier heroic plays of Dryden. Isabella, first acted by Mrs. Barry, gave Mrs. Siddons one of her major successes a century later. The character of Oroonoko, an African prince enslaved in the English colony of Surinam, marks one of the first literary appearances of the "noble savage."

Southerne was born at Oxmantown, Dublin, and entered Trinity College, Dublin, in 1676. Probably in 1680 he went to London and in July of that year was admitted to the Middle Temple, where

he presumably remained until 1682, when his first play, *The Loyal Brother, or The Persian Prince*, was produced at Drury Lane. In 1685 he received a commission in Princess Anne's Regiment of Foot, but his military career ended with the Revolution of 1688. As well as writing seven other plays—lively comedies of manners and frigid tragedies in Roman settings—he made some contribution to Dryden's tragedy *Cleomenes* (1692), and was well known for recommending plays to the theatre, including Congreve's *The Old Bachelor* (1693) and Colley Cibber's *Love's Last Shift* (1696). He died in London on May 26, 1746.

BIBLIOGRAPHY.—*Plays . . .* , 3 vol. (1774); modern editions of *Oroonoko* in *Five Restoration Tragedies*, ed. by B. Dobrée (1928); and of *The Loyal Brother*, ed. by P. Hamelius (1911). *See* also J. W. Dodds, *Thomas Southerne, Dramatist* (1933). (C. LE.)

SOUTHERN OVERLAND MAIL, the service established in the United States in 1858 for carrying mail and passengers by stagecoach from St. Louis to San Francisco. Earlier U.S. mail service to the Pacific coast had been slow and unreliable. Beginning in 1849, the year of the gold rush, mail was carried by steamboat to the Isthmus of Panama, over the isthmus by mule or canoe, and then up to California by steamboat. Inadequate overland mail service was provided along a central route through Salt Lake City, Utah, in 1851. Thereafter, demands mounted for greater government support for overland service. The problem was complicated by the bitter quarrels of the period between Southerners and Northerners in Congress. In March 1857 Congress passed an overland mail bill that left the choice of route to the postmaster general, Aaron V. Brown of Tennessee. He chose a southern route, defending it as the best all-weather route. In September 1857 John Butterfield of Utica, N.Y., and several associates signed a contract with the U.S. government to transport letter mail over the southern route from St. Louis, Mo., and Memphis, Tenn., by way of Fort Smith, Ark., El Paso, Tex., Tucson and Yuma, Ariz., and Los Angeles to San Francisco, Calif., and back "twice a week in good four-horse post-coaches or spring wagons, suitable for the conveyance of passengers as well as the safety and security of the mails." Butterfield and his associates were to receive a government subsidy of $600,000 a year for six years, 1858–64. The first eastbound mail on the new route left San Francisco on Sept. 15, 1858, and the first westbound mail left St. Louis on the following day.

Although northern critics objected to the circuitous route—2,795 mi. from St. Louis to San Francisco—the service proved satisfactory. The contractors were able to provide the 25-day service they promised. Stations were built at intervals of 15 mi. or less, and the coaches carried passengers as well as mail. Unfortunately for the contractors the American Civil War interrupted service and made it necessary to move the line to a central route on July 1, 1861. The contract was transferred to Ben Holladay in 1862, and he in turn transferred his interests to Wells, Fargo and Company in 1866. The completion of a transcontinental railroad in 1869 brought cancellation of the last contract for carrying overland mail by stagecoach. *See* also PONY EXPRESS.

BIBLIOGRAPHY.—R. P. and M. B. Conkling, *The Butterfield Overland Mail, 1857–1869*, 3 vol. (1947); L. R. Hafen, *The Overland Mail 1849–1869* (1926); L. H. Wright and Josephine M. Bynum (eds.), *The Butterfield Overland Mail* (1942). (T. A. LN.)

SOUTHERN PROVINCE, CEYLON, covers an area of 2,146 sq.mi. (5,558 sq.km.) of southernmost Ceylon and had a population (1953) of 1,129,308. It consists for the most part of coastal lowland crossed by rivers, the more important of which are the Nilwala Ganga and Walawe Ganga. In the western part of the province the relief is more prominent and more confused; it includes the southern part of the Rakwana Hill mass, which rises to more than 4,000 ft. Apart from coastal sands and alluvium, Southern Province is made up entirely of ancient crystalline rocks.

The western part of the province (Galle and Matara districts and Giruwa Pattu west of Hambantota District) lies in the wet zone of Ceylon and has a rainfall of 70 in. (1,800 mm.) or more, well distributed throughout the year; the hillier areas of the interior are particularly wet. The rest of the province is in the dry zone and is subject to a searing seasonal drought from May to September. The extreme east is particularly dry. Natural vegetation has largely disappeared from the wet zone, though some forests and fernlands remain; much of the dry zone is still under tropophilous forest, which passes into poor scrub as the driest areas are approached. In this region is Ruhunu (Yala) National Park (50 sq.mi. [130 sq.km.]), where elephant, bear, leopard, deer, and other wild animals may be studied. (*See* also CEYLON: *Physical Geography.*)

Ruhunu is the traditional name for an area wider than the present Southern Province, and it functioned as one of the two main core areas of ancient Sinhalese civilization. The part of Southern Province lying in the dry zone, together with the belt transitional to the wet zone, was of great importance, as is witnessed by the many surviving ruins, irrigation works, and inscriptions. The ruins of the ancient city of Magama are at Tissamaharama. Wet zone places, such as Matara, later became more prominent. It was the area in the wet zone that was particularly valuable to the Portuguese and Dutch, who used Galle and Matara, in particular, as fortified trading posts.

Low-country Sinhalese form the majority of the population. The wet zone is densely peopled, but apart from occasional areas of denser settlement the dry zone has a sparse population, the extreme east being virtually empty. The main towns are Galle (*q.v.*), capital of the province (55,848), Matara (*q.v.*; 27,641), and Hambantota (4,299), all district capitals; others are Weligama (12,834), Ambalangoda (10,554), and Tangalla (6,823).

The province has a varied economy. The wet zone produces rice, coconut, rubber, tea, cinnamon, and citronella. In the dry zone rice growing under new or restored irrigation works is of increasing importance. Fishing is also carried on. A coastal railway runs from Colombo to Galle and Matara, the latter being the railhead. The road network is adequate in the wet zone and rather rudimentary in the dry zone. (B. H. F.)

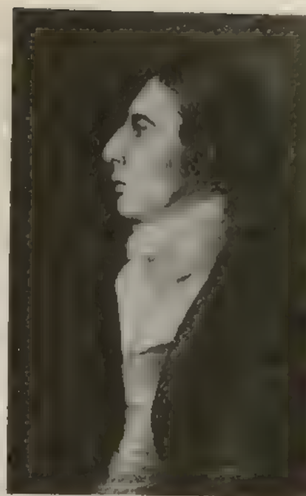
SOUTHEY, CAROLINE ANNE (née BOWLES) (1786–1854), English writer, the second wife of Robert Southey (*q.v.*), whose correspondence with her reveals both at their best. The daughter of an East India Company captain, she was born at Buckland Cottage, near Lymington, Hampshire, on Oct. 7, 1786. After her mother's death (1816), a guardian's fraudulence left her almost penniless, and, knowing of Southey's kindness to poor poets, in May 1818 she sent him a metrical tale, *Ellen Fitzarthur* (anon. 1820). Thus began a correspondence lasting until their marriage, on June 4, 1839. Southey encouraged her writing and at their first meeting (London, 1820) suggested subjects and treatment; she, on her side, showed unfailing sympathy and understanding of his work, criticizing it with lively wit and fancy. She stayed at Greta Hall (1823), and Southey visited her at Buckland. During his wife's illness, and after her death (1837), her friendship comforted him, but he lived only four years after their marriage. After his death she returned to her old home: her step-daughter Kate Southey's jealousy had provoked a family quarrel. She died at Buckland on July 20, 1854.

Of her writings, most notable are the poems in *The Widow's Tale* (1822) and *The Birthday* (1836); the verse *Tales of the Factories* (1833); the prose stories in *Chapters on Churchyards* (2 vol., 1829); and *Robin Hood*, on which Southey collaborated, published as a fragment (1847). Southey's comment, in his first letter, "You have the eye, the ear and the heart of a poetess," had truth; but her talent was slender, and her charming personality is only fully seen in her letters.

See J. F. Burnet in the *Fortnightly Review*, CLXXIII (1953); and bibliography to SOUTHEY, ROBERT.

SOUTHEY, ROBERT (1774–1843), English writer who is remembered chiefly for his association with Wordsworth and Coleridge (*qq.v.*) as one of the Lake poets, though his poetry has little in common with theirs and is now esteemed less highly than his prose. The son of a linendraper, he was born in Bristol on Aug. 12, 1774, but spent much of his childhood at Bath in the care of Elizabeth Tyler, his mother's eccentric half sister. In 1788 he entered Westminster School, London, but was expelled in 1792 for an attack on corporal punishment in *The Flagellant*, a school magazine that he and his friends had founded. This in-

justice roused his rebellious nature, and, already enthusiastic for the ideas of the French Revolution, when he entered Balliol College, Oxford, in December 1792, soon after his father's bankruptcy and death, he expressed ardent sympathy for it in a long poem, *Joan of Arc* (published 1796).



BY COURTESY OF THE NATIONAL PORTRAIT GALLERY, LONDON

ROBERT SOUTHEY, PENCIL AND WATER-COLOUR PORTRAIT BY R. HANCOCK, 1796

In June 1794, Coleridge, on a visit to Oxford, met Southey and outlined his plans for a "pantisocracy," or utopian agricultural community, on the banks of the Susquehanna in the United States. They met again at Bristol in August, and Southey (who had left Oxford without taking a degree) introduced Coleridge to Sara Fricker, whom he was later to marry. Robert Lovell, another friend (who collaborated with Southey in a volume of poems, published in the autumn, 1794; titlepage, 1795), had already married Mary Fricker, and Southey himself soon became engaged to a third sister, Edith. Miss Tyler objected both to the engagement and to "pantisocracy"; she barred Southey from her house and never saw him again. To raise funds for their plan, Southey and Coleridge turned to lecturing in Bristol and to journalism: they had already produced a verse drama, *The Fall of Robespierre*, published in 1794 as by Coleridge. The dream of "pantisocracy" faded, and, soon after secretly marrying Edith Fricker at Bristol (Nov. 14, 1795), Southey left for Portugal with his uncle, Herbert Hill, who was chaplain to the British colony at Lisbon. He had already decided, against his uncle's wishes, not to take holy orders. While in Portugal he wrote the letters published as *Letters Written During a Short Residence in Spain and Portugal* (1797) and learned to thank God that he was an Englishman, thus beginning the change in his political opinions that turned him from revolutionary to Tory. On his return (May 1796) he made public his marriage and in February 1797 began to study law in London. In 1797 he also began to receive an annuity of £160, paid for nine years by a Westminster schoolfriend, Charles Watkyn Williams Wynn, who later became a cabinet minister. In 1797–99 he published a second collection of poems, in two volumes, including some from the 1794 volume.

The Southeys soon left London, and Southey gradually abandoned the law. They lived partly at Westbury-on-Trym, near Bristol, where (1798–99) he wrote many of his best short poems and ballads, among them "The Ebb-Tide," "Bishop Hatto," and "The Battle of Blenheim"; and partly at Burton, Hampshire. During this time Southey was contributing regularly to the *Morning Post* and other newspapers and reviews, translating, working on the epic *Madoc* (published, fully revised, 1805), and editing Chatterton's works (published 1803). On a second visit to Portugal (1800–01), made partly to recover from serious illness, he gathered material for a history of that country and completed his epic poem *Thalaba the Destroyer* (1801). Life abroad deepened his conservatism, finally confirmed by the Peace of Amiens (1802) and replacement of Pitt (whom he hated) by Addington.

Having returned to England (May 1801), in August the Southeys visited Greta Hall, Keswick, where Coleridge (to whom Southey, after an estrangement, was now reconciled) was living. This, his first visit to the Lake District, made little impression, and after a fortnight he set off for a tour in north Wales. Early in September he accepted the post of secretary to Isaac Corry, the Irish chancellor of the exchequer. Still undecided about his future, he gave up the job in May 1802, and, depressed by the deaths of his mother and (August 1803) of his first child, a daughter (b. 1802), on Sept. 7, 1803, he and his wife joined the Coleridges and Mrs. Lovell (whose husband had died in 1801) at Greta Hall. They were to remain there for the rest of their lives.

Deciding temporarily to abandon poetry, which he found too exciting to his nature, he took up his *History of Portugal*, of which only the *History of Brazil* (3 vol., 1810–19) was published. The birth of five daughters (1804–12), of a second son, Herbert (1806), and the dependence on him of the household, especially after Coleridge left his family, forced him to apply himself conscientiously and unremittingly to authorship—producing poetry, history, biography, criticism, and journalism of various kinds. In 1807 he published, anonymously, the *Letters from England: By Don Manuel Alvarez Espriella*, a fictitious Spaniard; and, in the same year, an edition of the works of Henry Kirke White (2 vol.; vol. III, 1822), a young protégé who had died in 1806. In 1809 he became a regular contributor to the new Tory *Quarterly Review*, for which he wrote, anonymously, a series of 95 political articles for the next 29 years: at first concerned mainly with foreign policy, these later expounded his plans for social reform, authoritarian but farseeing in demanding state provision of "social services." In 1813, through the influence of Sir Walter Scott, he was appointed poet laureate, and in 1835 the government pension of £160, secured for him by Wynn in 1807, was increased to £300. His fame as a poet and critic grew steadily. An interval of domestic happiness was interrupted by the death of the brilliant, much-loved Herbert (1816), although the birth of Charles Cuthbert (1819) somewhat restored his naturally buoyant spirits.

In 1817, an unauthorized publication of an early verse drama, *Wat Tyler*, was used by the Radicals to remind the public of his youthful republicanism. His change of principle had irritated Byron (*q.v.*), who had already attacked Southey in *English Bards and Scotch Reviewers* (1809), and who, in 1818, dedicated *Don Juan* to the laureate in a set of ironic stanzas. Byron believed, mistakenly, that Southey had spread scandal about himself and Shelley, so that when, in the introduction to *A Vision of Judgement* (1821), Southey continued the quarrel by denouncing Byron as belonging to a "Satanic school" of poetry, he replied in the brilliant satire of *The Vision of Judgement* (1822).

Southey's later years were saddened by his wife's illness: she became insane in September 1834 and died on Nov. 16, 1837. In 1839 he married Caroline Bowles (*see* SOUTHEY, CAROLINE ANNE), but his mental and physical powers were already failing. He died at Greta Hall on March 21, 1843, and was buried at Crosthwaite, Keswick. Family quarrels about his second marriage resulted in the erection of three memorials: the monument in Crosthwaite Church bears an inscription by Wordsworth, referring to his "wide aims" and "life to heaven vowed."

Critical Assessment.—Southey's poetry is little read today, though the two longer poems, *Thalaba* and *The Curse of Kehama* (1810), both in irregular verse, the former unrhymed, the latter rhymed, are good and readable examples of the Oriental verse tale. Of his minor poems, such lyrics as "My Days Among the Dead Are Passed" are still read, and his comic-grotesque ballads, *e.g.*, "The Inchcape Rock," are unexcelled in English except by *The Ingoldsby Legends*. His poetic theory is expounded in the prefaces to the collected edition of 1837–38. He regarded poetry as the noblest of all vocations and constructed his poems on high moral, even heroic, principles; but they lack the creative genius of Wordsworth and Coleridge.

His more ambitious prose works are not entirely satisfactory. *The Life of Wesley; and the Rise and Progress of Methodism* (1820), a standard, though controversial, work in its day, and his valuable edition of Cowper (15 vol., 1835–37) have been superseded, as have most of his historical works. Only *The Life of Nelson* (2 vol., 1813; often reprinted) remains a classic. It is in his less formal work that his gifts are most happily displayed: in *Sir Thomas More: or, Colloquies on the Progress and Prospects of Society* (1829); in the anonymously published *The Doctor* (7 vol., 1834–47), a fantastic, rambling novel miscellany, packed with anecdotes (including the famous "Story of the Three Bears"), comments on life and literature, and quotations from his wide reading; and in *Espriella's Letters*, which gives a lively picture of English life. In these, and in his correspondence, Southey's prose is masterly, modeled on that of Defoe and Swift, with an ease and

clarity found in no other writer of his time. Byron called his style "perfect," Hazlitt said that it could "scarcely be too much praised," and it has been particularly admired by such modern critics as Sir Herbert Read and Jack Simmons.

BIBLIOGRAPHY.—*Editions and Correspondence: The Poetical Works,* 10 vol. (1837–38; 1 vol. ed., 1846); selections by E. Dowden, in the "Golden Treasury Series" (1895) and by M. H. Fitzgerald (1909); *Life of Wesley . . .*, ed. by M. H. Fitzgerald, 2 vol. (1925); *Life of Nelson,* ed. by E. R. H. Harvey (1953); *Letters from England,* ed. by J. Simmons (1951); *Life and Correspondence,* ed. by C. C. Southey, 6 vol. (1849–50); *Selections from the Letters . . .*, ed. by J. W. Warter (Southey's son-in-law), 4 vol. (1856); *Correspondence . . . with Catherine Bowles, etc.,* ed. by E. Dowden (1881); *Letters . . . A Selection,* ed. by M. H. Fitzgerald, "World's Classics" (1912).

Biography and Criticism: J. Simmons, *Southey* (1945), the best modern biography, includes a short bibliography of Southey's works. *See* also E. Dowden, *Southey,* "English Men of Letters Series" (1874); W. Haller, *The Early Life of Robert Southey, 1774–1803* (1917). There are reminiscences by contemporaries, references in Wordsworth's correspondence and in that of Byron and Coleridge (*see* bibliographies to the articles on them), and criticism in Coleridge's *Biographia Literaria,* Hazlitt's *Spirit of the Age,* and Landor's *Imaginary Conversations.* For Southey's political opinions, *see* G. D. Carnall, *Robert Southey and His Age* (1960). (N. C. N.)

SOUTH GATE, an industrial and residential city of central Los Angeles County, Calif., U.S., is 12 mi. (19 km.) S of the Los Angeles civic centre and a part of the Los Angeles metropolitan area (*see* LOS ANGELES). Prior to 1915 the area of South Gate was devoted to truck farming. Heavy industry was encouraged to come in, and South Gate's growth has been continuous since. It was incorporated as a city Jan. 20, 1923, and in 1958 established a council-manager form of government. Manufactured products include automobiles, fibreboard products, auto tires, chemical products, and airplane parts. Pop. (1960) 53,831. For comparative population figures *see* table in CALIFORNIA: *Population.* (J. M. Wo.)

SOUTH GEORGIA: *see* FALKLAND ISLANDS AND BRITISH ANTARCTIC TERRITORY.

SOUTH HOLLAND (ZUIDHOLLAND), the most densely populated province of the Netherlands, borders the North Sea and adjoins the provinces of North Holland to the north, Utrecht and Gelderland to the east, and North Brabant and Zeeland to the south. Area 1,234 sq.mi. (3,196 sq.km.). Pop. (1960) 2,706,810. The province consists of a coastal strip of dunes and sandy soil (*geestgronden*), a low-lying inland region of peat and fluviatile clay, and, in the south, an archipelago (in rapid transition under the Delta plan; *see* NETHERLANDS, THE: *Dikes and Polders*) formed by the ramifications of the Lek, Waal, and Maas (Meuse) rivers, the largest islands being Dordrecht, Voorne-Putten, Hoeksche Waard, and Goeree-Overflakkee, all with soils of marine clay. On the inner side of the dunes, which was early populated, are the old towns of Leiden, The Hague, and Delft (*qq.v.*). The Hague is the capital of the province and the seat of the Netherlands government.

Whereas the dunes are mainly used for recreational and residential purposes (Scheveningen, Katwijk aan Zee, and Noordwijk aan Zee are important seaside resorts), the geest is mostly devoted to horticulture. Northward from Leiden, bulbs are grown exclusively, and at Keukenhof, near Lisse, there is a famous annual exhibition of flowering bulbs. Northwest of Leiden, in the Rijnsburg area, vegetables and flowers are grown. South of The Hague, as far as the New Waterway, is the Westland, a district of intensive horticulture where fruit and vegetables (especially tomatoes, cucumbers, and grapes) are grown under glass.

By far the most important part of South Holland is the port and industrial district of Rotterdam (*q.v.*), which extends along the New Waterway and the Lek, Noord, and Merwede rivers, and of which The Hague, Delft, Gouda, Dordrecht, and Gorinchem are the outposts. Industries include shipbuilding, engineering, oil refining (Pernis and Europoort), and chemical manufacture (Vlaardingen, Rozenburg, and Botlek).

The inland clay and peat region is given to agriculture in the reclaimed lakes and to cattle farming on the older soils, cheese making being of particular importance in the eastern districts. Pot plants and shrubs are grown at Boskoop. A number of interesting old towns, *e.g.,* Vianen, Schoonhoven, Oudewater, and Woerden, are situated on the rivers; Gouda (*q.v.*) is the service centre for this region.

Arable farming predominates in the archipelago; local specialities include onions (Overflakkee). An old town in this region, Brielle, an ancient port (the capture of which by the *Watergeuzen* [*see* GUEUX] in 1572 was the signal for the revolt against the Spaniards), has become a recreation centre for Rotterdam since the closing of the Brielse Maas in 1950. *See* also HOLLAND, COUNTY AND PROVINCE OF. (H. J. KE.)

SOUTH INDIA, CHURCH OF, came into existence in September 1947, as the result of the merger, after 28 years of negotiation, of the Anglican, Methodist, Presbyterian, and Congregationalist groups with the Basel Mission (Lutheran and Reformed), in the area of India south of the Vindhya Mountains. The union was based on the acceptance of the Holy Scriptures as the supreme authority in faith and life, of the so-called Nicene Creed as the authorized summary of the faith, of the sacraments of baptism and Holy Communion, and of the historic episcopate in constitutional form. Provision was made for a 30-year period of growing together, in the course of which it was hoped that the union would become complete. On the day of the inauguration, nine new bishops, drawn from all the contributing traditions, were consecrated to serve with the five Anglican bishops already in office, in the 14 (later 16) dioceses of the united church. None of the ministers in office were reordained, and all were regarded as equally eligible for election to any office in the church.

No attempt was made to impose uniformity at once on all the local churches, which were to continue to use their accustomed liturgical forms until genuinely Indian forms of worship could be worked out by agreement. The church later issued orders for Holy Communion, baptism, ordination, and other services; these are not mandatory, but use of them expanded steadily. The order for Holy Communion has been highly praised by liturgists but has nothing of specially Indian character. Yet the church is unmistakably Indian, independent of control from other church bodies outside India.

The church has a membership of more than 1,000,000 baptized adherents; their number has grown steadily, though slowly, and without any notable new mass movement toward Christianity. The church has full communion with the nonepiscopal bodies from which in part it sprang, but not with all Anglican churches, though Anglican clergy or laity serving in the Church of South India do not enter into schism.

The Church of South India marked a new beginning, as it was the first instance since the Reformation of union between episcopal and nonepiscopal churches. As such it aroused passionate controversy, which still continues. In its intention to extend the union, it engaged in conversations with the Baptists and Lutherans of the area. Conversations with the Baptists lapsed, but agreement was reached with the Lutherans on doctrinal points, though not on all questions of organization.

See also CANON LAW: *Church of South India.*

BIBLIOGRAPHY.—B. G. M. Sundkler, *The Church of South India* (1954); A. M. Ward, *The Pilgrim Church* (1953); R. D. Paul, *The First Decade* (1958); *Book of Common Worship of the Church of India* (1963). (S. C. N.)

SOUTH MOLTON, a municipal borough of Devon, Eng., lies 12 mi. (19 km.) SE of Barnstaple by road, on the River Mole. Pop. (1961) 2,993. The parish church of St. Mary Magdalene is mainly Perpendicular, and the market square has many Georgian houses. In 1357 the town was granted a charter to hold a market and annual fair, which still take place, and there is also a large annual sheep fair. Incorporated in 1590, the town was granted another charter in 1684. South Molton is an ideal tourist centre for Exmoor and the Doone Valley. Its industries (besides textile factories) are mainly agricultural. (L. J. WE.)

SOUTHPORT, a municipal (1867), county (1905), and parliamentary borough, Lancashire, Eng., is a popular seaside resort. It also has a large residential population, being 20 mi. (32 km.) N of Liverpool and 39 mi. NW of Manchester and connected by rail to both cities. Pop. (1961) 82,004. The town developed in the 19th century on the coastal sand dunes, the ground inland being

marshy. There are extensive beaches, pleasure gardens, a long pier, and several parks and handsome thoroughfares, notably Lord Street. An important annual flower show is held. Southport is a golfing centre, the courses including Royal Birkdale. Light engineering is the most important industry.

SOUTH PORTLAND, a city of Cumberland County in southwestern Maine, U.S., separated from Portland (*q.v.*) by Portland Harbour and tidal Fore River. South Portland was originally part of Cape Elizabeth, which divided over the offer of the Portland Water District to supply the town with Sebago Lake water; some of the Cape Elizabeth people were opposed, but residents of the salt-water-lapped southern section, who were unable to dig enough freshwater wells, welcomed the offer. The region was set off and incorporated as the town of South Portland on March 15, 1895, and was granted a city charter a week later. A city-manager system was instituted Jan. 1, 1934. During World War II, Liberty ships were built in South Portland; the shipyards were later taken over by three steel-fabrication firms and by a Canadian company for an oil pipeline terminal. The city also became the site of one of the largest Atlantic coast oil-tank farms. For comparative population figures *see* table in MAINE: *Population*. (A. K. GL.)

SOUTH SAN FRANCISCO, an industrial city of California, U.S., 9 mi. (14 km.) S of San Francisco (*q.v.*) and just north of San Francisco International Airport, on San Francisco Bay. The city is built upon Spanish land bought in 1869 for the "cattle empire" of Henry Miller and Charles Lux; in the 1890s financiers and industrialists led by G. F. Swift of the meat-packing firm planned a workers' community there. The tract was subsequently subdivided and incorporated in 1908. The location offered excellent rail and water transportation and inexpensive factory sites, and industrialization continued. By the city's 50th anniversary important industries included steel, paint and chemicals, and meat-packing. Housing expanded rapidly during and following World War II. For comparative population figures *see* table in CALIFORNIA: *Population*. (J. H. ST.)

SOUTHSEA: *see* PORTSMOUTH.

SOUTH SEA BUBBLE, the name given to a speculation mania that ruined many English investors in 1720, closely following the French financial crisis of the previous year. This so-called bubble (or hoax) centred on the fortunes of the South Sea Company founded by Robert Harley in 1711 to trade with Spanish America, on the assumption that the War of the Spanish Succession would end with a treaty under which such a trade (mainly in slaves) might be opened. The company was given a monopoly of trade to South America and the Pacific. Holders of navy and ordnance bonds were offered stock, with a guaranteed interest of 6%, and the stock was accepted willingly. But the peace treaty, when made, conveyed only a very limited trading concession, restricted to a single annual ship that could not enter the Pacific. The success of the first voyage in 1717 was only moderate, but in the following year the king became governor of the company, creating fresh confidence in an enterprise that soon yielded up to 100%. There had been a first wave of speculation when the Bank of England was founded in 1694, but the postwar boom of 1713–19, paralleled by the Mississippi Company mania in France, now gave rise to extravagant hopes. Those who were unable to buy South Sea stock (which was genuine, if overcapitalized) were consoled by optimistic investors and swindlers with the offer of shares in a hundred dubious ventures.

South Sea stock itself stood at 128½ in January 1720, rose to 330 in March, 550 in May, 890 in June, and stood at 1,000 in July and August. In September it dropped to 175 and in December to 124. Other shares followed a similar trend, Bank of England stock falling from 263 to 145.

The occasion of the boom in South Sea shares was the scheme by which in 1720 the company took over the whole of the national debt except that owed to the Bank of England and the East India Company and persuaded the investors to exchange their state annuities for South Sea stock, sold at a tremendous premium. Though a few investors profited greatly, far more were ruined. Some fled from their creditors and not a few committed suicide.

This financial disaster was followed by an inquiry undertaken by a committee of the House of Commons. Its report, rendered in February 1721, showed that at least three ministers had both accepted bribes and speculated in the stock. John Aislabie (1670–1742), chancellor of the exchequer, was expelled from the House and imprisoned. The Earl of Sunderland and Charles Stanhope were acquitted by small majorities, and the estates of the company's directors were mostly confiscated. The company itself continued to exist, making fair profit until 1732, selling most of its rights to the Spanish government in 1750 but retaining some privileges until 1807 and its existence until 1853. In its later years it was concerned mainly with the Greenland whale fishery and for a short time numbered Charles Lamb among its clerks.

The scandal surrounding this company in 1720 weakened the whole system of chartered companies (*see* CHARTERED COMPANY) and provided arguments for free trade. Another result of the affair was the destruction, indirectly, of the London slave trade, to the immediate profit of Bristol and to the eventual profit (on a far greater scale) of Liverpool.

BIBLIOGRAPHY.—Jean O. McLachlan, *Trade and Peace with Old Spain, 1667–1750* (1940); Lord Erleigh, *The South Sea Bubble* (1933); Lewis S. Benjamin (L. Melville, pseud.), *The South Sea Bubble* (1921); D. Hannay, *The Great Chartered Companies* (1926). (C. N. P.)

SOUTH SHIELDS, seaport, municipal, parliamentary, and county borough, Durham, Eng. Pop. (1961) 109,521. Area 7.6 sq.mi. (20 sq.km.). It is on the south side of the mouth of the River Tyne, there protected by the massive South Pier, nearly 1 mi. long. Ferries connect the town with North Shields. Shipbuilding and repairing and marine engineering are important, and there is considerable export of coal and iron. The Tyne dock has a water area of 50 ac., and the tidal basin is 5¾ ac. Roman Remains Park, near the northern end of the town, may have been an island in the Tyne; the remains comprise storehouses, etc. The town was founded by the convent of Durham in the 13th century. In the 17th and 18th centuries it had large glassworks and salt pans. The first self-righting lifeboat was launched there in 1790. The town has returned one member to Parliament since 1832, and in 1850 it was incorporated, being raised to county borough status in 1889.

SOUTHWARK, one of the 32 boroughs constituting Greater London, Eng., established in 1965 under the London Government Act 1963 (*see* LONDON) and formed by the amalgamation of three former metropolitan boroughs, Bermondsey, Camberwell, and Southwark. Area 11.5 sq.mi. (30 sq.km.). Pop. (1961) 312,761. The borough stretches from Sydenham Hill to the south bank of the Thames, which is crossed by Blackfriars, Southwark, London, and Tower bridges, and Cannon Street Railway Bridge. It consists of three parliamentary constituencies.

Historically, Southwark's northern area has always been important as a meeting place of roads from the south and as the commanding point to the approach to London Bridge. There was a Roman settlement there, and the main roads (Old Kent Road, Borough High Street, Newington Causeway, and Newington Butts) follow Roman routes. After the building of the other Thames bridges that connect Southwark with the City, the Elephant and Castle area became a major traffic junction. Southwark's river frontage is lined with wharves and warehouses. The Surrey Commercial Docks (381½ ac. [154 ha.]) are at Rotherhithe. Because of its position Southwark has long been a busy commercial area. The docks and warehouses handle much of London's food. The London Provision Exchange is at No. 1 London Bridge, and several famous firms are engaged in food processing and brewing. The Borough Market (fruit and vegetables) dates from the 18th century. A traditional Bermondsey industry is leather, and the National Leathersellers' College is in Southwark. There are also various printing and engineering firms. Dulwich and parts of Camberwell are largely residential.

The borough is served by Southern Region and Underground railways. The Metropolitan Line which links Rotherhithe with Wapping (in Tower Hamlets) uses a tunnel constructed (1825–43) beneath the Thames by Sir Marc Isambard Brunel.

After the Reformation the medieval Augustinian priory of St. Overie became the parish church of St. Saviour; in 1905 it became

the cathedral for the Southwark diocese. The church was rebuilt in the 13th and 15th centuries and the nave restored in the 19th. The tower is 15–16th century. The Harvard Memorial Chapel commemorates John Harvard, the founder of Harvard University, who emigrated to New England from Southwark in 1637. St. George's Roman Catholic Cathedral, designed by A. W. N. Pugin in 1848, was largely rebuilt after severe damage sustained during World War II air raids. The Camberwell parish church of St. Giles, designed by Sir George Gilbert Scott, occupies the site of an earlier church that was destroyed by fire in 1841. The Bermondsey parish church of St. Mary Magdalene was rebuilt in 1680. St. Mary's Church at Rotherhithe, a riverside landmark, has many memorials to seamen, including Christopher Jones, captain of the "Mayflower." The Pilgrim Fathers' Memorial Church has been rebuilt in Great Dover Street. The Metropolitan Tabernacle at the Elephant and Castle, built 1861 for Charles Haddon Spurgeon, the Nonconformist preacher, has been reconstructed but retains its impressive facade.

Guy's Hospital was opened in 1725. The chapter house of Southwark Cathedral was formerly the chapel of St. Thomas' Hospital, founded 1213 and removed to Lambeth in 1868. Maudsley Hospital is on Denmark Hill.

Educational institutions include Dulwich College (founded 1619, rebuilt 1870), Wilson's Grammar School (1615), St. Olave's and St. Saviour's Grammar School (originally two separate 16th-century foundations), the Camberwell School of Arts and Crafts (1898), the London College of Printing (at the Elephant and Castle), and the William Booth Memorial Training College (1929) on Denmark Hill. The South London Art Gallery and Cuming Museum are both administered by the borough. The Cuming Museum deals mainly with local history. The Dulwich College Picture Gallery was the bequest mainly of Sir Francis Bourgeois (*see* DULWICH). The Imperial War Museum occupies buildings originally erected in the early 19th century to house the Bethlehem Royal Hospital (*see* BEDLAM). Open spaces include Peckham Rye Common and Park (103 ac. [42 ha.]), Dulwich Park (75 ac.), and Southwark Park (66 ac.).

Chaucer's Canterbury pilgrims started from the Tabard Inn in Borough High Street. Many of Shakespeare's plays were first produced at the Globe Theatre on Bankside, the site also of the Rose (noted for plays by Marlowe), the Swan, and the Hope theatres. An annual Shakespeare Festival is held in the borough. Many scenes in Dickens' novels are set in Southwark, and St. George's in Borough High Street is familiarly known as the Little Dorrit Church. Robert Browning was born in Camberwell, John Ruskin lived at Herne Hill and Denmark Hill, and Michael Faraday was born at Newington Butts.

Bermondsey, Camberwell, Southwark, Rotherhithe, Peckham, and Walworth all appear in Domesday Book. Old Southwark, traditionally known as "The Borough," was from 1295 to 1547 the only town besides the City in the present Greater London area to be represented in Parliament. In 1327 Edward III granted Southwark to the citizens of London, and it became known as "The Bridge Ward Without." Several great ecclesiastics had houses on Bankside, notably the bishop of Winchester. The many famous inns in Borough High Street have been rebuilt except for the south wing of the George (rebuilt 1676), the last surviving galleried inn in London (now owned by National Trust). The medieval history of Bermondsey was dominated by its great Cluniac abbey, founded in 1082; Bermondsey Square approximately marks its site. Until the 19th century much of Southwark was rural, with farms and market gardens. A spa and various country inns were places of resort for Londoners. The Thames bridges, the railways, and the general increase in population led to rapid urbanization, often accompanied by poverty and overcrowding. The architecture where not post-World War II is predominantly Victorian; wartime destruction and large-scale redevelopment schemes completely changed parts of the old borough, notably the Elephant and Castle area. (M. J. Bo.)

SOUTHWELL, ROBERT (1561–1595), English poet and martyr, remembered for his saintly life as a Jesuit priest and missionary in England during a time of persecution, his heroic endurance of torture and death, and his religious poetry, which anticipates that of George Herbert and Richard Crashaw (*qq.v.*). Born at Horsham St. Faith, near Norwich, in 1561, he was brought up as a Catholic and in 1576 was sent to the Jesuit English College at Douai. From there he went to Clermont, the Jesuit college in Paris, and in 1577 he was sent to Rome, where in October 1578 he became a novice. In 1585 he was ordained priest and made prefect of studies at the English College at Rome.

He had determined, however, to return to England as a missionary and in 1586 persuaded his superiors to allow him to accompany Henry Garnett (*q.v.*) to London. They arrived in July, and, after spending some time in hiding at the house of Lord Vaux at Hackney, Southwell became chaplain to Anne Howard and spiritual adviser to her husband, the first earl of Arundel, who, as a recusant, was imprisoned in the Tower. He lived in concealment at Arundel House in the Strand, London, writing letters of consolation to the earl and other persecuted Catholics and making pastoral journeys to Catholic families outside London. One of his letters, *An Epistle of Comfort . . .* , was printed on an illicit press in 1587; others circulated in manuscript.

After many narrow escapes, Southwell was arrested when celebrating mass in a Catholic household at Uxendon, near Harrow, Middlesex, on July 5 (new style; June 26, old style), 1592. He was taken first to the house of Richard Topcliffe, the head of the pursuivants, as those employed in seeking out Catholics were called, where he was tortured in an attempt to make him reveal the whereabouts of his fellow priests. He was then transferred to the Gatehouse, a prison in Westminster, and, after his father had petitioned the queen that he might be put to death rather than stay in such a "filthy hole," to the Tower. There he was kept for three years in solitary confinement. At last, in 1595, he was brought to trial on a charge of treason under the anti-Catholic penal laws of 1585, and on March 4 (N.S.; Feb. 22, O.S.) he was executed at Tyburn.

Southwell's theme is always religion, and his devotional lyrics and prose treatises and epistles—mainly written at Arundel House and published only after his death—reflect the ardent piety of his life. His best works achieve an unusual directness and simplicity, and in his use of paradox and parallelism and of striking imagery he is akin to the 17th-century metaphysical poets. He has a secure place as the foremost representative of Catholic letters in Elizabethan England.

BIBLIOGRAPHY.—Southwell's poems were edited, with introduction, by A. B. Grosart (1872). Selections in L. I. Guiney (ed.), *Recusant Poets* (1938); shorter poems ed. by C. M. Hood, as *The Book of Robert Southwell* (1926); *An Humble Supplication* (1600, dated 1595; ed. by R. C. Bald, 1953). *See* also P. Janelle, *Robert Southwell the Writer* (1935), the first authoritative biographical study; C. Devlin, *The Life of Robert Southwell* (1956).

SOUTHWELL, a town and rural district of Nottinghamshire, Eng., 14 mi. (22 km.) NE of Nottingham by road. The rural district of 55 parishes had a population (1961) of 45,850; area 185.2 sq.mi. (480 sq.km.). It has hosiery works, a flour mill, and an interest in the neighbouring market-garden area (Bramley seedling apples are said to come from Southwell), but is chiefly known for its minster, a successor of a 7th-century church generally attributed to Paulinus; certainly there was a church there in the 8th century.

The present building was begun in the 12th century and has a Norman nave, transepts and towers, an Early English choir and a Decorated octagonal chapter house. It is cruciform and is 306 ft. long. There are no cloisters. Particularly noteworthy are the magnificent 14th-century stone rood screen, intricately carved, and the carved animals and foliage on the doors of (and within) the chapter house.

In 958 land at Southwell was granted to the archbishop of York by Edwy and a detailed description of the manor appears in Domesday Book. Southwell was a collegiate establishment and remained under the lordship of the see of York until it was taken over by the ecclesiastical commissioners. Remains of the archbishop's palace (15th-century; "great chamber" restored 1880) are incorporated in the bishop's palace.

The episcopal see founded in 1884 included the counties of Not-

tingham and Derby, but the latter was detached in 1926 when the see of Derby was established.

SOUTHWEST, THE. This term has been in use in the United States for more than 150 years to denote a geographic area, and its meaning has changed over the years as the nation has expanded. After the War of 1812, the Southwest generally meant Missouri, Arkansas, and Louisiana. After Texas was annexed, it too, was included. In the wake of the war with Mexico, the Southwest embraced most, but not all, of the territory acquired under the Treaty of Guadalupe Hidalgo (1848; *see* MEXICAN WAR). California's littoral and great interior valleys were so different in physiography and development that the term never settled easily upon them. Rather, the Southwest came to mean, by and large, the 900-mi. stretch of arid land between the 100th meridian and the deserts athwart the lower Colorado River; *i.e.*, New Mexico, Arizona, and such contiguous parts of Texas, Oklahoma, Colorado, Utah, and Nevada as suited the convenience of the person using the term.

Early explorers found a diverse and difficult topography. The high, dry plains of Texas, sliced in the north by such unexpected abysses as Palo Duro Canyon, rose imperceptibly to the Pecos Valley of New Mexico, a dividing line that the myriads of buffalo farther east seldom crossed. Beyond the Pecos and embracing its headwaters were the southern spurs of the Rocky Mountains, cool with evergreens. West of the Rockies came vast horizontal strata of highly coloured sandstone; in places these beds formed broad peneplains whose dry monotony was broken by occasional mesas or buttes. Where the beds were higher they were cut by spectacular gorges, three of which—Zion and Bryce canyons in Utah and the Grand Canyon of the Colorado River in Arizona—became national parks. Still westward, beyond a forested uplift that cut across Arizona from northwest to southeast, lay the true deserts marked by prickly plant growths and parallel chains of north-south mountains, gaunt and almost devoid of vegetation.

The common denominator of the land was aridity. Although the highest peaks of New Mexico and Arizona caught 30 in. (750 mm.) of rain or more a year, the annual average was less than half that amount; indeed, the western deserts might receive only 2 in. (50 mm.) or less. Except for the snow on the mountains, the bulk of this scanty precipitation came during violent thunderstorms; the flash floods they produced did much to give the region its scarred and angular topography.

High temperatures and rapid evaporation intensified the dryness. Precipitation that would support dry farming in the Dakotas or Nebraska proved, generally speaking, inadequate in the Southwest: crops could be grown there only with irrigation. Yet the region's two principal rivers, the Colorado and the Rio Grande (both of which rose outside the area), as well as the smaller streams of the local mountains could be tapped at only a handful of widely scattered spots. This overwhelming geographic fact turned some of the prehistoric Indians into sedentary agriculturalists (for these aboriginal peoples, *see* INDIAN, NORTH AMERICAN: *Culture Provinces*), determined the shape of the Spanish occupation, and ultimately led the American inhabitants to an alliterative folk description of their economy as one of "copper, cattle, cotton, and climate."

Despite oversimplification, there is much to justify this summation. Geologic evidence even suggested that aridity might have had some bearing on the Southwest's huge deposits of low-grade copper ore, the so-called porphyry coppers. Mexicans were working the richer streaks of the long-productive ore body at Santa Rita, N.M., before 1800; other veins were opened in southeastern Arizona following the Civil War. However, the great development came after 1910 when the handling of enormous tonnages by power shovels in open-pit mines lifted Arizona to first place among the copper-producing states. Compared with this outpouring of copper, the Southwest's brief stampedes for gold and silver and later for oil and uranium achieved little more than regional significance.

The dryness of the land enforced a pastoral agricultural economy that began in 1598 when Juan de Oñate established in the Rio Grande Valley of New Mexico the first permanent colonies of the western United States. (San Antonio, Tex., and San Xavier del Bac near Tucson, Ariz., were not founded for another century; San Diego, Calif., not until 1769). Because of the difficulty of supervision cattle were not popular with the first settlers. Horses, however, multiplied rapidly and were soon spread among the Indians by barter and theft. Sheep ranches grew to enormous size—in the early 1800s New Mexico's governor, Bartolomé Baca, owned 2,000,000 head—and social cleavage was largely between those who owned sheep and those who herded them. The Pueblo Indians (*q.v.*) took readily to the animals and learned to weave wool just as they were already weaving a native cotton. The Navaho learned the craft from the Pueblo, developed their famous blankets, and as late as World War II accounted for half the sheep raised in Arizona (*see* NAVAHO).

During the period of Spanish ascendancy, the sheep were driven to market among the silver mines of Mexico. In the 1850s the California goldfields offered an outlet for those owners willing to brave the long trek across mountains and desert to the coast. After the Civil War the demand of the new Southwestern Indian reservations and army posts for beef lured herds of longhorns west from Texas. When the Apaches of the mountains were brought under control, permanent establishments took root in New Mexico and Arizona, sometimes in violent conflict with rival sheep ranches. High profits made overgrazing of the public domain a more serious threat to the economy of the Southwest than the Indians ever were. Only federal control, exerted through an increasing number of agencies, was able to check the trend and start restoring the ranges.

Other federal intervention came in the management of water resources. Irrigation had been practised before A.D. 1000 by the Pueblo Indians in New Mexico and central Arizona. The Spanish added a few *acequias* or "ditches" of their own, often as part of the huge land grants belatedly made by the Mexican government in the hope of establishing agricultural communities as buffers against the encroaching Americans. In their turn these Americans, individually and through cooperatives, struggled to increase the paltry stream diversions; but measures commensurate with the problems were not possible until the Reclamation Act of 1902 made available the resources of the national government.

The Reclamation Service's first successful large-scale project in the nation was Theodore Roosevelt Dam near Phoenix completed in 1911. The largest project in the Southwest was Hoover Dam on the Colorado River, completed in 1936 (though the Glen Canyon Dam, upriver from Hoover, completed in 1964, closely approached it in size). But the water proved expensive and led to intensive cultivation of high-income crops: alfalfa, citrus, and long-staple cotton. It also led, in the case of streams crossing state boundaries, to complex legal disputes, the bitterest of which concerned the diversion of the Colorado River water far from its natural channels. Various state compacts were entered into by the claimants, but sectional jealousies militated against a satisfactory solution. After World War II the water shortage was made even more critical when developments associated with aviation and nuclear energy brought an unprecedented influx of industry and population into all of the Southwest's urban areas.

Though the crisp climate and scenic landscapes were a curse to agriculture, they were a boon to businesses catering to tourists and health-seekers. These visitors, it was shortly discovered, had a lively interest in the once-despised cultures of the Indians and Spanish-Americans. The heavily beamed, flat-roofed Indian pueblo, as modified by the Spanish, became the inspiration for the region's new architecture. Indian dances, Spanish fiestas, and cowtown rodeos, catching hold first as tourist attractions, became the year's social and economic high points for scores of communities. Although the showier forms of this evolving culture were not truly typical of any of its founding strains, they fixed themselves, in the popular mind at least, as an even more distinctive feature of the American Southwest than the ancient angular topography and dazzling sunlight.

For bibliography, *see* the articles on the several states and on the main rivers, COLORADO RIVER and RIO GRANDE. (D. S. L.)

SOUTH WEST AFRICA (SUIDWES-AFRIKA), a territory of southern Africa that came into being as a German protectorate in 1890 and developed under German control until it was invaded by forces from the Union of South Africa in 1915. After the

Treaty of Versailles the territory was mandated to the Union. It is still administered as a mandated territory of the Republic of South Africa. Extending from the Orange River in the south to the Kunene and Okovango (Okavango) rivers and the parallel of 17° 23′ S in the north, the territory extends to the Chobe and Zambezi rivers in the Caprivi Zipfel (Caprivi Strip) between northern Botswana (Bechuanaland) and southeastern Angola. Its eastern boundary south of this strip lies on the meridian of 21° E as far south as the 22nd parallel and then on the meridian of 20° E to the Orange River. The area of the territory is 317,887 sq.mi. (823,328 sq.km.), not including the Walvis Bay enclave of 374 sq.mi. (969 sq.km.), which is an integral part of Cape Province in the Republic of South Africa, although administered as a part of South West Africa. The capital is Windhoek (*q.v.*).

Physical Geography.—*Geology and Structure.*—The territory includes four of the major geological formations of southern Africa. The Basement Complex, of old granite and gneiss, with metamorphosed sediments, forms much of the plateau surface between Windhoek and Outjo, extending from this main area northwestward into the Kaokoveld Mountains and southward along the plateau edge to the Orange River Valley. The Precambrian Damara System, of dolomite, limestone, sandstone, and metamorphosed rocks, occupies most of the plateau surface from the Khomas Highland in the Windhoek area to the Kunene on the northern border. Much of this system is strongly folded, the axes trending generally north by east. In the south of the territory the Nama System, of quartzite, shale, and black limestone, dips gently eastward. The Karroo System, comprising almost horizontal beds of tillite sandstone, shale, and overlying lava, forms the surface in the area extending from the Orange River in the south to near Gobabis in the middle of the territory. In the north it forms local plateaus such as the Waterberg and great areas of lava in the Kaokoveld. Young granites have been intruded into these beds, forming such prominent features as the Brandberg, in the Namib, and the Erongo Mountains.

The youngest formations are in the east and west. In the east the Tertiary Kalahari System, of consolidated gravels and sands at the base and a varying thickness of overlying loose sand, covers about half the territory to the north of 22° S and a considerable area to the west of the southeastern boundary. In the west of the territory the sands and clays of the Namib, together with the diamond-bearing shingle of the southern coastal strip, are probably the youngest formations.

Physiography.—Two major divisions can be recognized: the plateau surface, and the marginal zone between the plateau and the coast. Separating these divisions is the plateau edge or Great Escarpment, the highest parts of which are known in the south as the Huib (highest point 5,860 ft. [1,786 m.]), the Tsaris (6,486 ft. [1,977 m.]), and the Naukluft-Hakos (6,200 ft. [1,973 m.]) escarpments and, in the north, as the Joubertsberge (5,856 ft. [1,785 m.]) and the Baynes Mountains (7,200 ft. [2,195 m.]).

On the plateau the highest area is the Khomas Highland, whose ridges trend east-northeast through the Windhoek area. The main watershed of the territory is formed in this highland by the Auas Mountains (Moltkeblick 8,153 ft. [2,485 m.]). In the north the Otaviberge (Nageib 7,001 ft. [2,134 m.]) trend east-west in the Grootfontein area, bending northward in the Joubertsberge. In the south of the plateau area the main orographic features are the Klein and Gross Karas Mountains (Schroffenstein Berg 7,224 ft. [2,202 m.]) and Mount Brukkaros (5,203 ft. [1,588 m.]), an ancient volcanic crater, about 9 mi. (14 km.) N of Berseba, which rises 1,500 ft. (457 m.) above the Namaqualand Plain.

Most of the marginal zone is known as the Namib, generally flat and true desert near the coast, passing, as it ascends to the plateau edge, into a semidesert zone generally known as the Inner Namib. (*See* South Africa, Republic of: *Geographical Regions.*) The highest point in the territory is the Brandberg in the Namib, a young granite mass that rises precipitously to an altitude of 8,458 ft. (2,578 m.).

Five geographical regions can be recognized. (1) The Etosha Plain in Ovamboland, to the north of the Otaviberge, extending eastward to the south of the Okovango River. (2) Damaraland (or Hereroland) between the Otaviberge and the Auas Mountains, to the south of Windhoek. This is a plain of fairly even surface, which, however, is broken by many island-mountains, the highest being the Great Omatako (7,510 ft. [2,289 m.]). (3) Namaqualand (or Namaland), lying between the Auas Mountains and the Orange River, rather flat in the north but deeply dissected in the south by the Fish River and other smaller tributaries of the Orange. (4) The Kalahari surface along the eastern border, which there forms the western part of the southern Kalahari sandveld. (5) The Namib between the plateau edge and the coast, predominantly sandy in the south but behind the coastal belt becoming rocky in the north where it forms the western part of the Kaokoveld.

Drainage.—Except for the Orange in the south and the Kunene and Okovango at the northern boundary, all the rivers are of intermittent flow, "coming down" generally during two or three months of the year after the seasonal rains. The most spectacular scenery in the territory is the canyon of the Fish River, the southward-flowing tributary of the Orange. Cut in Nama beds and Archean rocks, the canyon is between 1,500 and 2,000 ft. (460 and 600 m.) deep and lies between about 27° and 27° 75′ S. In the east the Auob River has artesian conditions in its upper course. Pans (shallow temporary ponds or lakes) occur in many areas, especially on the Kalahari sand. The largest is the Etosha Pan, about 70 by 30 mi. (115 by 50 km.), which receives the summer drainage of Ovamboland.

Climate.—Because of the northward sweep of the cold Benguela Current, from the south Atlantic, along the coast, temperatures are consistently low in the Namib. Walvis Bay has a mean temperature for the coolest month (August) of 14° C (57° F) and for the warmest month (February) of 19° C (67° F). On the plateau the temperature range is greater. Windhoek, at 5,700 ft. (1,740 m.), has a mean for the coldest month (July) of 13° C (55° F) and for the warmest month (January) of 23° C (74° F). Rainfall is least in the western Namib and in the lower Orange River Valley, where the annual precipitation is less than 2 in. (51 mm.). On the plateau there is an increase of rainfall from southwest to northeast. Warmbad, in southern Namaqualand, has a mean annual rainfall of 3.4 in. (86 mm.), Windhoek 15.1 in. (384 mm.), Tsumeb 20.7 in. (526 mm.), and Katima Mulilo in the Caprivi Strip 26 6 in. (676 mm.). Except in the extreme southwest the rainfall of the territory occurs mainly in the summer months, the percentage of summer precipitation increasing steadily from southwest to northeast.

Vegetation.—In the Namib coastal zone the only vegetal types are those adapted to extreme drought. Among these are the indigenous tumboa (*Welwitschia bainesii*), the tortoise bush (*Zygophyllum stapfii*), and the edible naras (*Acanthosicyos horrida*), which grows between the dunes near Walvis Bay. Behind the coastal zone various succulents occur that give place near the plateau edge to desert grasses and aloes. In the southern part of the plateau the vegetation is predominantly of the Karroo type, with desert shrubs and succulents passing northward into grass steppe. In Damaraland and on the eastern sandveld of the Kalahari, thorn savanna (*Acacia*) with grass is the predominant type, merging to the north of the Etosha Pan into dry forest in which the mopane (*Colophospermum mopane*) is associated with large fig trees (*Ficus hereroensis*), *Hyphaene* palms, and baobabs. In the *oshanas*, or shallow floodwater channels, of Ovamboland the floors are covered with *Aristida* grasses.

Animal Life.—Primates are represented by the gray vervet monkey, the chacma baboon, and, in the north, by the bush baby (*Galago senegalensis*). Carnivores include the lion in the northwest and the northeast. The leopard and the cheetah are widespread, but the most numerous of all the larger carnivores is the brown hyena, which roams even the coastal belt of the Namib, where it lives on marine refuse and shellfish. Jackals (black-backed and side-striped) are widespread. The elephant roams the Kaokoveld, the Caprivi, and, occasionally, the Okovango area. The black rhinoceros (*Diceros bicornis*) is found in small numbers in the Kaokoveld. The giraffe and the hippopotamus also are found only in the north. Of the antelopes some, such as the koodoo (kudu), gemsbok, and springbok, are found over most of

the territory; others are limited to the northern areas, including the sassaby (tsessebe), lechwe, puku, impala, roan antelope, and the sitatunga, or marshbuck (*Tragelaphus spekei*). The hartebeest is found in the northeast and the buffalo (*Syncerus caffer*) only in the northern areas. On many of the islands off the coast to the south of Walvis Bay, and at Cape Cross farther north, the Cape sea lion or fur seal (*Arctocephalus pusillus*) is abundant.

The People.—*Anthropology.*—The nonwhite people of the territory may be divided into two main types, the Southwestern Bantu and the Khoisan types (the Hottentots and Bushmen). The Bastaards of Rehoboth Gebiet are the descendants of mixed European and Hottentot unions. The strongest Bantu groups in pre-European times were the Ambo of the north and the Herero, called Damara by the Hottentots. Occupying Hereroland (or Damaraland) to the north of the Khomas Highland, the Herero were in constant conflict with the Hottentots who occupied areas to the south of the Khomas. In the early period of German occupation the Herero, with their northern offshoot the Ovashimba, numbered probably about 100,000; they were seriously reduced under German occupation and now number about 35,000. A more primitive Bantu group, the Bergdama, or Berg (Mountain) Damaras, formerly vassals of the Herero and Hottentots, had no territory of their own but have been allocated reserves. They number about 44,000 and are probably the one native group that openly rejoices at the European occupation of the territory. They speak the language of the Nama Hottentots, who subjugated them. Hottentots, in reserves mainly in the southern part of the territory, number about 35,000. The total Bushman population in the territory is estimated at 12,000. Under German rule the area within the police zone allocated as native reserves was about 4,030 sq.mi. (10,440 sq.km.). Under the administration of South Africa the reserves were enlarged to about 79,700 sq.mi. (206,420 sq.km.), in addition to Rehoboth Gebiet, 5,060 sq.mi. (13,105 sq.km.), containing about 11,000 Bastaards, who speak Afrikaans. (*See* also AFRICA: *Ethnography* (*Anthropology*); *Southern Africa;* AMBO; BERGDAMA; BUSHMAN; HERERO; HOTTENTOT.)

Population.—At the census of 1921 the total population was 227,988, made up of 19,372 whites and 208,616 Africans and Coloured. In 1960 the total population of 526,004 comprised 73,464 whites and 452,540 Bastaards, Coloureds, and Bantu. In 1960 Windhoek had a population of 36,051, including about 19,000 whites. Populations of most of the other towns had increased considerably since 1951. The Bantu population may be divided into two groups, those in the police zone and those in the northern native territories of Ovamboland and Okovango, with small groups in the Kaokoveld and the Caprivi Strip. These territories are generally referred to as "outside the police zone." In Ovamboland (Amboland) the Ambo number about 231,000; in the Okovango territory the tribal Bantu number about 28,000. Within the police zone the native reserves in 1960 had about 26,000 Bantu, Hottentots, and Coloureds. (J. H. WN.)

History.—The first European to set foot in the territory was the Portuguese navigator Diogo Cam, who erected a pillar at Cape Cross in his voyage of 1485–86. At the end of 1487 his compatriot Bartolomeu Dias de Novais touched the coast at Walvis Bay and at Dias Point, where he erected a cross near Angra Pequena (Lüderitz). The Dutch East India Company's ship "Grundel" sailed up the coast in 1670 and was followed by the same company's "Bode" in 1677. Though the interior still remained unknown, the coastal waters were not neglected; sealing and whaling attracted ships of various nationalities. In 1793 Capt. F. R. Duminy, of the Dutch ship "Meermin," proclaimed certain areas to be the property of the Dutch government. In 1795 Capt. Thomas Alexander of H.M.S. "Star" hoisted the British flag at various points on the coast, but his action was later disowned by his government. Meanwhile, in 1760, Jacobus Coetzee found a way into Great Namaqualand by the overland route from the Cape. He was followed in 1761–62 by Henry Hop and thereafter by a fairly steady stream of explorers, of whom the most famous was C. J. Andersson, a Swede, who in 1853 crossed from Walvis Bay to Lake Ngami in Bechuanaland. The London Missionary Society was represented on the Orange River as early as 1802 and in 1807 established a station at Warmbad. This society withdrew and was succeeded by the Methodists. In the 1840s the German connection with the territory began with the arrival of the Rhenish Missionary Society.

In 1868 tribal wars prompted the missionaries to appeal for British protection. In this they were actually supported by the German government, but the British government turned the proposal down. In 1876 the Cape Colony sent a special commissioner, W. C. Palgrave, who made treaties with chiefs anxious to secure British protection. This placed the whole territory as far as the Portuguese colony of Angola under British control. The British high commissioner in South Africa approved Palgrave's action, but the British government was unwilling to assume these new responsibilities and was only with difficulty persuaded to consent to the annexation of Walvis Bay and some adjacent territory (1878). The guano islands off the coast had been annexed in 1867. In 1880 tribal war again broke out, and the German government itself asked Great Britain to protect the life and property of its subjects; the British reply was negative.

German Possession.—In 1883 Heinrich Vogelsang, agent of F. A. E. Lüderitz, a merchant of Bremen, obtained from the Hottentot chief Joseph Frederick of Bethany a cession of land at Angra Pequena, later known as Lüderitz. Bismarck, still unconverted to a colonial policy, informed the British government beforehand of the project in words that were almost an invitation to Great Britain to assume sovereignty over the territory and to act as a protecting power. This opportunity, like the previous one, was allowed to slip away, and in April 1884 Bismarck took the initiative and assured Lüderitz and his establishments of German protection, which later developed into full-fledged German annexation of the whole territory.

The first German representative, sent to the territory in 1885, was H. E. Göring. He was succeeded by Kurt von François, who transferred the capital from Otjimbingue to Windhoek. It was there that the first German farmers settled in 1892. In 1893 there was trouble with the Hottentot chief Hendrik Witbooi. The Germans attacked Witbooi's village, and 150 of his subjects, including women and children, were killed. François was succeeded in 1896 by Theodor Leutwein, who tried to treat the Africans with consideration. However, the latter found much cause for dissatisfaction, and it required only a spark to set the country ablaze. The Bondelswaartz Hottentots rose in 1903. This rising was suppressed, but early in 1904 the Herero revolted and killed a number of German (but not British or Boer) settlers. The rebellion was speedily quelled, but in the "cleaning up" operations the Herero were ferociously harried by Gen. Lothar von Trotha, Leutwein's successor from 1904, and in the end were reduced from a tribe of 80,000 people to 15,000 starving refugees. Various groups of Hottentots who had joined in the rebellion still held out until early in 1907. In the years that followed these troubles the depopulation resulting from the methods used in suppressing the rebellion caused a labour shortage that hampered the development of the territory. The discovery of diamonds in 1908 near Lüderitz Bay led to a considerable increase in the European population, which rose to nearly 15,000 by 1913.

On Aug. 6, 1914, after the outbreak of World War I, the government of the Union of South Africa undertook to assume all obligations resting upon the British regular garrison in South Africa, and on Aug. 10 to send a military expedition of its own to German South West Africa. The surrender of the Germans was forced on July 9, 1915. At the end of the campaign the German troops were interned and later (in 1919) repatriated with about 600 undesirables. German civilians, on the other hand, were allowed to return to their homes in the territory and to continue their ordinary pursuits.

The South African Mandate.—Under the Treaty of Versailles, Germany ceded its colonies to the principal Allied and associated powers. South West Africa was designated as a C mandate and was conferred on "his Britannic majesty for and on behalf of the government of the Union of South Africa." The newly created League of Nations was charged, *inter alia,* with the supervision of mandates. Subject to the mandate, South Africa was given full

powers of legislation and administration over the territory as an integral part of the Union. The exercise of these powers was vested by the House of Assembly in the governor-general (Union act no. 49 of 1919), who later delegated them to a resident administrator. The S.W.A. Naturalization of Aliens Act of 1924 provided for the automatic naturalization of all German adult males domiciled in the territory on Jan. 1, 1924, unless they specifically objected.

During the first few years of the mandate the territory was administered after the fashion of a crown colony, with an administrator assisted by an advisory council. In 1926 a Legislative Assembly of 12 elected and 6 nominated members, with an executive committee, was set up to legislate for such needs as roads and bridges, agriculture, and taxation, while the rest of the administrative functions remained in the hands of the administrator.

Until the rise of Hitler the Germans in the territory had been reasonably cooperative, but by 1939 the grip of the Nazi Party on all Germans in the territory was practically complete. At the outbreak of World War II the Union government acted with dispatch. By December 1939 about 150 Nazi leaders were arrested and the organizations themselves proscribed. By October 1940 the number of persons interned was 1,200, or one-third of the German adult male population. Under the Naturalization and Status of Aliens Amendment Act of 1942 most of the automatically naturalized Germans reverted to their former German status.

Trusteeship or Incorporation?—With the demise of the League of Nations the Union government held that there was nothing in the covenant that empowered it to transfer or delegate its powers, as regarded mandates, to any organization. In this contention South Africa was supported in 1950 by an opinion of the International Court of Justice, which held that the Union was under no obligation to place the territory under a United Nations trusteeship, but that on the other hand an obligation rested on South Africa to submit reports on the administration of the territory to the UN. In spite of pressure South Africa firmly refused to agree to the inclusion of the territory in the international trusteeship system but was willing to submit reports to Great Britain, the United States, and France as a matter of courtesy.

The proposal that South West Africa should be incorporated in, or should become a fifth province of, the Union arose in 1946 when the subject was referred to a series of tribal meetings and it was reported that the majority of Africans was in favour of the proposal, the Herero being the only substantial body against it. Later in the same year Gen. J. C. Smuts asked the United Nations that the territory be incorporated in accordance with the wish of its people, but the request was refused. The South African government did not press the demand but declared its intention, and this applied both to trusteeship and to incorporation, to continue to administer South West Africa in the spirit of the mandate. However, the association of the two countries, already close, was made much closer by the South West African Affairs (Amendment) Act of 1949 (see *Administration and Social Conditions*, below). The South African Citizenship Act of 1949 made it possible for those who had been "denaturalized" in 1942 to apply once more for naturalization. In 1955 a commission was appointed to consider the question of German language rights. Its report, produced in 1957, stated that the 20,000 German-speaking people, *i.e.*, 35% of the white population, wanted equal status for German with English and Afrikaans. A German-speaking person may address the South African state departments in German but cannot demand a reply in the same language. German is, however, fully recognized in the local Legislative Assembly.

The deadlock regarding trusteeship continued in discussions at the UN in the following years. A UN committee (appointed to consult with the South African government on a possible basis for agreement) suggested in its report in September 1958 that the territory might be partitioned. The northern portion of South West Africa, containing the great majority of the Bantu population, was to be administered by the South African government as an integral part of the Union under a trusteeship agreement with the UN, while the rest of the territory was to be annexed to the Union. The Union government announced that this did not mean that it had agreed to partitioning the territory but that it would be prepared to investigate the possibilities. However, the UN later rejected partition as a possible basis for agreement.

A deputy minister for South West Africa was appointed for the first time in August 1961. Debates on the future of the mandate continued in the UN. In 1964 the report of the Odendaal Commission of South West Africa was published. It recommended proposals for constitutional change, including extension of the principles of apartheid to South West Africa, and economic development with South African financial help. In 1960 Ethiopia and Liberia instituted proceedings in the International Court of Justice in which they asked the court to declare, *inter alia*, that the 1920 mandate was still in force and that South Africa had failed to fulfill its obligations under the mandate. In 1962 the court held that it had jurisdiction to adjudicate in the case, and in July 1966 it rejected the Ethiopian and Liberian complaints on the ground that these two countries had not established any legal right or interest in the subject matter of their claims. In October the UN General Assembly voted to end South Africa's mandate.

(Av. Sv.; X.)

Administration and Social Conditions.—*Government.*—In 1920 the Council of the League of Nations conferred on the British sovereign a mandate over the former German colony of South West Africa to be exercised by the governor-general of the Union of South Africa, who, by proclamation in 1921, delegated his powers to an administrator appointed by the Union government. After May 31, 1961, the state president of the Republic of South Africa exercised the mandate and delegated his powers to an administrator appointed by the government of the republic. An executive committee of 4 members (presided over by the administrator) and a Legislative Assembly of 18 members are elected as in the republic's provinces and with similar powers. (*See* South Africa, Republic of: *Administration and Social Conditions.*) In addition, the South West African Affairs Amendment Act, 1949, provides for six electoral divisions of the territory, each returning one representative to the South African House of Assembly, and for representation in the Senate by four senators, two of whom are nominated (one for the Coloured population).

Urban local government is undertaken by elected town councils and appointed village management boards, while rural administration is a matter for local magistrates. Windhoek is the seat of the administration. Administration of Bantu affairs (subject to the Native Urban Areas Proclamation, 1951, which followed the Union Native [Urban Areas] Act, 1945) was taken over by the South African Department of Bantu Administration and Development in 1955. Extensive Native reserves have been established in the rural areas.

Generally speaking, taxation accords with that of the Republic of South Africa, a state income tax having been introduced in 1942 and the old municipal income tax replaced by rates on property in 1920. The main sources of revenue are income tax, customs and excise duties, and mining taxation.

Living Conditions.—Machinery for regulating wages, registering labour unions, and settling disputes was established in 1953. There was full employment, and no unemployment benefit plans were found necessary. African labour is generally adequate, but in some cases there is a shortage. Housing for the white population is financed by building societies. In the African townships municipalities provide subsidized housing with state aid.

A report of the Special Committee on Colonialism, presented to the UN General Assembly in January 1965, alleged that foreign monopolistic combines operating in South West Africa were exploiting African workers, through restrictive contracts and conditions, in a way that amounted to virtual slavery. These allegations were denied by South Africa. African labour in the territory is regulated by masters' and servants' laws, pass laws, and influx control in the urban areas.

Health.—Health and hospital services parallel those of South Africa. Hospitals for whites are state aided and run by boards responsible to a director of health services. Hospitals for Africans are directly controlled and maintained by the administration.

Justice.—The judiciary consists of a High Court and circuit

courts (from which appeals lie to the appellate division of the South African Court) and magistrates' courts. The police and defense forces are incorporated in the South African forces.

Education.—Primary and secondary education is compulsory for all white children between 7 and 16 years of age, and German is a medium in some schools. The Cape syllabus and system are followed. Education of Coloureds and Africans is undertaken mainly by mission schools subsidized by the administration.

(L. H.)

The Economy.—*Agriculture.*—There is little crop cultivation in the territory except in the northern areas where the rainfall is just sufficient in most years for small crops of maize (corn) and kafir (sorghum), varying from about 30,000 to a few hundred thousand bags (of 200 lb.). Under irrigation a few thousand bags of wheat are produced, especially in the artesian valleys of the Nossob-Auob river system. The territory is mainly pastoral, with about 2,400,000 cattle, 3,000,000 karakul sheep, 700,000 other sheep, and about 1,500,000 goats. The pelts of the newborn karakul lambs are among the most valuable of the territory's exports.

Fisheries.—Sea fisheries are based mainly on Walvis Bay, with crawfish or rock lobster fishing mainly at Lüderitz. The pilchard is the all-important shoal fish. The fishing fleet includes about 100 well-equipped boats with diesel engines, echo-sounding devices, radio telephones, and radio direction-finders. The annual catch is about 600,000 to 700,000 tons. Whitefish and snoek are also taken in small quantities. At Walvis Bay there are seven factories for processing pilchards and three for whitefish and snoek, the fish meal and body oil being of more value than the canning product.

At Lüderitz six factories can rock lobster and produce meal and frozen tails, and one processes pilchards. Products of sealing on the offshore islands and at Cape Cross include annually about 10,000 pups (for furs), 3,000 wigs (skins of bulls), 25,000 pelts, and between 40,000 and 50,000 gal. of oil.

Mining.—The most valuable mineral is the diamond, which is found along the coastal strip and offshore between Lüderitz and the Orange River mouth. The principal fields are now the Oranjemund diggings, along the coast for about 60 mi. (100 km.) N of the Orange River. The annual production is about 1,500,000 k., of which about 98% are gem stones and 2% industrials, the value being about £30,000,000 ($84,000,000). From the Tsumeb mine the annual production of about 30,000 tons of blister copper, 45,000 tons of refined lead, 100,000 tons of lead, lead-vanadium, and lead-zinc ores, with 16,000 lb. of germanium dioxide, is valued at about £15,500,000 ($43,400,000). Tin and wolfram concentrates, from the vicinity of the Erongo Mountains and the Brandberg, vary in quantity from about 450 tons to about 1,500 tons, the value being about £400 ($1,120) a ton. Other minerals have much lower values: lithium about £15,000 ($42,000), beryllium £4,000 ($11,200), sillimanite £6,000 ($16,800). Salt is an important mineral from the coastal pans, the annual production of coarse and refined salt being about 100,000 tons, the value between £150,000 and £240,000 ($420,000 and $672,000).

Trade.—The bulkier items of export include, except for diamonds, the minerals mentioned above, the export being mainly from Walvis Bay, where 200,000 to 300,000 tons of ore and other minerals are shipped annually, mainly for European and American destinations. Karakul pelts are usually sent to the European auctions by air, their value generally exceeding £13,000,000 ($36,400,000) yearly. Fish products (canned, salted, fish meal, and oil) realize about £17,000,000 ($47,600,000).

To the Republic of South Africa the territory sends supplies of dairy products (butter and cheese) and cattle on the hoof. Imports include machinery, metal goods, building materials, manufactured goods, and oil. Of the cargoes landed at Walvis Bay the main commodities are gasoline, oil, timber, cement, and coal.

Finance.—In the 1950s and 1960s the territorial revenue fluctuated between about £15,000,000 ($42,000,000) and £20,000,000 ($56,000,000), the main items of revenue being income tax (most of which was paid by the mining companies and only about 10% by the non-Bantu inhabitants), customs and excise, and diamond export duty and profit tax.

Expenditure fluctuated during the same period between about £14,000,000 and £21,000,000, but these amounts included the annual appropriations to the Territorial Development and Reserve Fund, which varied between £7,000,000 and £20,000,000. Apart from these appropriations the main items of expenditure were: administration, education, public works, health, and posts and telegraphs. The net accumulated surplus in the territory revenue fund fluctuated between £1,000,000 and £7,000,000.

Communications.—The 1,453 mi. (2,338 km.) of railway line in the territory is all of 3 ft. 6 in. gauge. Coal-burning locomotives have been replaced by diesel. Of the 20,000 mi. (32,000 km.) of roads, 2,144 mi. are trunk roads, of which 500 mi. are tarred, 5,000 mi. are gravel-surfaced main roads, and 13,000 are district roads with varied surfaces. Road motor services include a bus route from the railhead at Grootfontein to the Angola border, and from Gobabis to Bechuanaland (Botswana). There are frequent air services between Windhoek and Keetmanshoop in the territory and Johannesburg and Cape Town, linking up with South Africa's international air services. The port of Walvis Bay, with an entrance channel 33 ft. (10 m.) deep, accommodates about 400 oceangoing vessels a year, with a gross tonnage of about 2,300,000. South West Africa is a member of the African Postal and Telecommunications Union. Telex services were introduced in 1956. There are radio-telephone links between point-to-point radio stations where the use of wire communication is uneconomical.

Bibliography.—*Yearbook of the Republic of South Africa;* G. N. G. Hamilton and H. B. S. Cooke, *Geology for South African Students,* 4th ed. (1960); P. A. Wagner, *The Geology and Mineral Resources of South-West Africa,* Geological Survey Memoir no. 7 (1916); F. Jaeger and L. Waibel, *Beiträge zur Landeskunde von Südwestafrika,* Mitt. Deutsch. Schutzgebieten, Ergänzungsheft, nr. 14 (1920); Weather Bureau, Pretoria, *Climate of South Africa,* parts i–iv (1957); Drought Commission Report, *South West Africa* (1924); H. Vedder *et al., Native Tribes of South-West Africa* (1928); H. Walter and O. H. Volk, *Grundlagen der Weidewirtschaft in Südwestafrika* (1954); H. Walter, *Die Farmwirtschaft in Deutsch-Südwestafrika,* Heft 1–4 (1941); South-West Afrika, Long Term Agricultural Policy Commission, *Report* (1949); Sir Charles Dundas, *South-West Africa: the Factual Background* (1946); Odendaal Commission, *Report on South West Africa* (1964); G. Lawrie, *New Light on South West Africa* (1964). Since 1955 all official South West African statistics have been merged in government reports of the Union or Republic of South Africa.

(J. H. Wn.)

SOUTINE, CHAIM (Haim) (1893–1943), Expressionist painter, was born in Smilovitchi, near Minsk in White Russia, the 10th of 11 children of a poor tailor. As a child he caricatured his family and neighbours. At the age of 16 he went to Vilna (Vilnius), where a friendly doctor helped him attend the school of fine arts for three years. In 1911 he emigrated to Paris and there met the sculptor Jacques Lipchitz and the painter Amedeo Modigliani, whose portraits he painted. In Céret and elsewhere in the south of France, he developed his characteristic style of painting with thick pigment and convulsive rhythms, twisting the human face and figure and tilting the forms of nature as in an earthquake. Though a leader of the Expressionist movement, he only admired earlier masters, especially Tintoretto, Rembrandt, and Gustave Courbet, whose compositions he adapted in certain famous canvases. He exhibited little during his lifetime, but his jewel-bright landscapes, studies of valets, cooks, and choirboys, and still-life paintings of poultry and carcasses of beef entered into French and American private collections and U.S. museums. While hiding in Touraine during World War II, he fell ill and was brought to Paris for an operation, where he died on Aug. 9, 1943. The Museum of Modern Art, New York City, exhibited his lifework in 1950, and 100 of his paintings were shown in the Galerie Charpentier in Paris in the summer of 1959.

See Monroe Wheeler, *Soutine* (1950). (M. Wh.)

SOUTSOS, ALEXANDROS (1803–1863), Greek poet, born in Constantinople, Turk., was the founder of the Greek Romantic school of poetry. He studied in Chios and later in Paris, where he was influenced by the French Romantics and by liberal political opinion. His verse satires, modeled on those of Pierre Béranger, are his liveliest writings and inspired the early development of modern political liberalism in Greece. His dramas and one long prose work, the *Exoristos* ("The Exile"), were considered

frigid and artificial, but his numerous lyrics were admired by his contemporaries in spite of their lack of originality and their rhetorical exuberance. Byron's *Childe Harold* was the model for his longest poem, *Periplanomenos* ("The Wanderer").

Soutsos' collected works were published in 1916.

See C. A. Trypanis, *Medieval and Modern Greek Poetry* (1951). (Ce. A. T.)

SOUZA BOTELHO, ADÈLE MARIE ÉMILIE FILLEUL, Marquise de (1761–1836), French writer of sentimental novels of a high moral tone, was born in Paris on May 14, 1761. Her mother was the wife of one of Louis XV's secretaries and may have been one of the king's mistresses. After the marriage of her sister Julie in 1767 to the marquis de Marigny (Abel Poisson, Mme de Pompadour's brother), Adèle in 1779 was married to the old comte de Flahaut (Alexandre de Flahaut de La Billarderie, 1726–1794). One of the many visitors to her salon, however, was Talleyrand, who was generally recognized to be the father of her son Charles, later celebrated as comte de Flahaut (*q.v.*). She went to England in 1792 (her husband stayed in France and died imprisoned by the revolutionaries), and published a novel, *Adèle de Senange* (1794), which had some success. Passing then to Switzerland, she attached herself to the duc d'Orléans (Louis Philippe, the future king of the French) and went with him in 1795 to Hamburg, where she had to work as a milliner. She returned to Paris in 1798. Married in 1802 to the Portuguese envoy José Maria, marques de Souza Botelho (1758–1825), she enjoyed some influence in society under Napoleon I. Later she helped in the education of the future duc de Morny (*q.v.*), the child of her son Flahaut and Queen Hortense. Her collected works, 6 vol. (1821–22), contained *Adèle* and six other stories. Her last work, a play, *La Duchesse de Guise*, was published in 1832. She died in Paris on April 19, 1836.

See A. de Maricourt, *Madame de Souza et sa famille* (1907).

SOVEREIGNTY. The concept of sovereignty, one of the most controversial ideas in political science and international law, is closely related to the difficult concepts of state and government, of independence and democracy. Originally, as derived from the Latin term *superanus* through the French term *souveraineté*, sovereignty was meant to be the equivalent of supreme power. It has departed, however, quite often from this traditional meaning.

Historical Development.—In 16th-century France, Jean Bodin (*q.v.*) used the new concept of sovereignty to bolster the power of the French king over the rebellious feudal lords; the transition from feudalism to nationalism was thus facilitated. The theories of John Locke at the end of the 17th century, and of Jean Jacques Rousseau in the 18th century, that the state is based upon a compact of its citizens, through which they entrust such powers to a government as may be necessary for common protection, led to the development of the doctrine of popular sovereignty that found expression in the American Declaration of Independence in 1776. (*See* Social Contract.) Another twist was given to this concept by the statement in the French constitution of 1791 that "Sovereignty is one, indivisible, unalienable and imprescriptible; it belongs to the Nation; no group can attribute sovereignty to itself nor can an individual arrogate it to himself." The idea of popular sovereignty exercised primarily by the people became thus combined with the idea of national sovereignty exercised not by an unorganized people in the state of nature, but by a nation embodied in an organized state. Going one step further and investigating who in the name of the people or of the state exercises sovereignty, John Austin (*q.v.*) came to the conclusion that sovereignty was vested in a nation's parliament. This was the supreme organ that enacted laws binding upon everybody else, but that was not itself bound by the laws and could change these laws at will. This particular description again fitted only a particular system of government, such as prevailed in Great Britain in the 19th century.

When this idea of legislative sovereignty crossed the Atlantic ocean, it did not really fit the American situation. The Constitution of the United States, being the fundamental law of a federal union, did not endow the national legislature with supreme power but imposed important restrictions upon it. A further complication was added when the Supreme Court of the United States asserted successfully its right to declare laws unconstitutional. While this development did not lead to a concept of judicial sovereignty, it seemed to vest the sovereign power in the fundamental document itself, the Constitution. This system of constitutional sovereignty was made more complex by the fact that the authority to propose changes in the Constitution and to approve them was vested not only in Congress but also in the several states and special conventions called for that purpose. It could be argued, therefore, that sovereignty continued to reside in the states or in the people, who under the terms of the 10th Amendment to the Constitution retained all powers not delegated to the United States by the Constitution or expressly prohibited by it. Consequently, the claims by the states' rights (*q.v.*) advocates that states continued to be sovereign were bolstered by the difficulty of finding a sole repository of sovereignty in a complex federal structure; and the concept of dual sovereignty of both the union and the component units found a theoretical basis. Even if the competing theory of popular sovereignty were accepted, vesting sovereignty in the people of the United States, it still might be argued that this sovereignty need not be exercised on behalf of the people solely by the national government, but could be divided on a functional basis between the federal and state authorities.

Another assault from within on the doctrine of state sovereignty was made in the 20th century by those political scientists (*e.g.*, Léon Duguit, Hugo Krabbe, and Harold J. Laski) who developed the theory of pluralistic sovereignty exercised by various political, economic, social, and religious groups that dominate the government of each state. According to this doctrine, sovereignty in each society does not reside in any particular place but shifts constantly from one group (or alliance of groups) to another. This pluralistic theory went even so far as to contend that the state is but one of many examples of social solidarity and possesses no special authority in comparison with other components of society.

Sovereignty and International Law.—While the doctrine of sovereignty has had an important impact on developments within states, its greatest influence has been in the relationships between states. The difficulties here can be traced back to Bodin's statement in 1576 that the sovereign who makes the laws cannot be bound by the laws he makes (*majestas est summa in cives ac subditos legibusque soluta potestas*). This statement has been often interpreted as meaning that a sovereign is not responsible to anybody and is not bound by any laws whatever. But closer reading of Bodin's writings does not support this interpretation. He emphasizes that even with respect to his own citizens a sovereign is bound to observe certain basic rules derived from the divine law, the law of nature or reason and the law that is common to all nations (*jus gentium*); as well as the fundamental laws of the state that determine who is the sovereign, succession to sovereignty, and limitations on the sovereign power. Thus Bodin's sovereign was restricted by the constitutional law of the state and by the higher law that was considered as binding upon every human being. In fact, Bodin discussed as binding upon states many of those rules that were later woven into the fabric of international law. Nevertheless, his theories have been used as justifying absolutism in the internal political order and irresponsible anarchy in the international sphere.

This interpretation was developed to its logical conclusion by Thomas Hobbes in the *Leviathan* (1651), in which the sovereign was identified with might rather than law. Law is what the sovereign commands, and it cannot limit his power; sovereign power is as absolute as men can make it. In the international sphere this condition led to a perpetual state of war, one sovereign trying to impose his will by force on all other sovereigns. This situation changed but little over the next two centuries; and sovereign states have continued to claim the right to be judges in their own controversies, to enforce by war their own conception of their rights, to treat their own citizens in any way that suited them, and to regulate their economic life with complete disregard for possible repercussions in other states.

During the 20th century important restrictions on the freedom of action of states started to appear. The Hague conferences (*q.v.*) of 1899 and 1907 established detailed rules governing the conduct of wars on land and at sea. The Covenant of the League of Nations (*q.v.*) restricted the right to wage war, and the Briand-Kellogg Pact of 1928 condemned recourse to war for the solution of international controversies and its use as an instrument of national policy. They were followed by the Charter of the United Nations (art. 2), which imposed the duty on member states to "settle their international disputes by peaceful means in such a manner that international peace and security, and justice, are not endangered," and supplemented it with the injunction that all members "shall refrain in their international relations from the threat or use of force. . . ." At the same time, however, the Charter listed as one of the basic principles of the UN "the principle of sovereign equality of all its Members."

In consequence of these developments, sovereignty ceased to be considered as synonymous with unrestricted power. In the first place, all states being equal, no state was allowed to impose its will upon another by the use of force. Sovereignty of each state became entitled to protection by the international community, as organized in the UN, against the designs of greedy neighbours; and the organization was authorized "to take effective collective measures for the prevention and removal of threats to the peace, and for the suppression of acts of aggression . . ." (*United Nations Charter*, art. 1). Though the means given to the UN to protect the sovereignty of its members were not sufficient to prevent all violations of national sovereignty and to settle all disputes by peaceful means, it cannot be doubted that important limitations have been imposed on the right of each state to do what it pleases in relation to other states. These limitations on the sovereignty of each state were balanced, however, by the increase in the freedom of each state from interference by other states. As far as matters essentially within the domestic jurisdiction of a state were concerned, each state regained that measure of internal sovereignty that had suffered so many actual restrictions in the period of constant use of force by one nation against another. Only to the extent that use of force has been effectively restricted was it possible for a state to expect that its personality, territory, and political independence would be properly respected, and that it would be able to enjoy unmolested the rights inherent in its external sovereignty.

It is, therefore, in the interest of every state that all the other states should comply faithfully with all those international obligations that safeguard its sovereignty. A state cannot, however, expect such compliance unless it reciprocally fulfills in a scrupulous manner all its obligations to other states. Those obligations are derived from principles of customary law developed painstakingly over several centuries, and from hundreds of treaties that are concluded every year.

Consequently, states have accepted a considerable body of law limiting their sovereign right of acting as they please. Those restrictions on sovereignty are usually explained as deriving from consent or autolimitation, but it can be easily demonstrated that in some cases states have been considered as bound by certain rules of international law despite the lack of satisfactory proof that these rules were expressly or impliedly accepted by them. On the other hand, new rules cannot ordinarily be imposed upon a state, without its consent, by the will of other states. In this way a balance has been achieved between the needs of the international society and the desire of states to protect their sovereignty to the maximum possible extent.

Nonsovereign States.—The 19th-century distinction between fully sovereign states and several categories of less sovereign units lost its importance under the law of the United Nations. Emphasis was placed not on legal differences among colonies, protected states, protectorates, and states under the suzerainty of another state, but on the practical distinction between self-governing and nonself-governing territories. Nonself-governing territories became under the charter of the UN "a sacred trust," and the states administering them promised to develop them towards self-government. Some of these territories were placed under a UN trusteeship that resulted in a closer supervision of their administration by the UN and in their speedier progress toward self-government or independence. Once a territory achieved self-government, as defined in resolutions of the general assembly, supervision by the UN ceased, even though independent status was not reached. This happened, for instance, when Puerto Rico achieved the status of a self-governing commonwealth in 1953.

Conclusion.—The concept of absolute, unlimited sovereignty did not last long after its adoption, either domestically or internationally. The growth of the democratic form of government imposed important limitations upon the power of the sovereign and of the ruling classes. The increase in the interdependence of states restricted the principle that might is right in international affairs. The peoples of the world have recognized that there can be no peace without law, and that there can be no law without some limitations on sovereignty. They have started, therefore, to pool their sovereignties to the extent needed to maintain peace; and sovereignty is being increasingly exercised on behalf of the peoples of the world not only by national governments but also by organs of the world community. Thus the theory of divided sovereignty, first developed in federal states, has become applicable also in the international sphere.

See INTERNATIONAL LAW, PUBLIC; SPHERE OF INFLUENCE. *See* also references under "Sovereignty" in the Index.

BIBLIOGRAPHY.—The literature on sovereignty is abundant; almost every book on political science, government, international law, and legal theory deals with this problem. A few of the modern special studies are: C. E. Merriam, *History of the Theory of Sovereignty Since Rousseau* (1900), H. Kelsen, *Das Problem der Souveränität und die Theorie des Völkerrechts* (1920); L. Duguit, *Souveraineté et Liberté* (1922); H. J. Laski, *Authority in the Modern State*, ch. i (1919); H. Krabbe, *The Modern Idea of the State*, Eng. trans. by G. H. Sabine and W. J. Shepard (1922); J. Dickinson, "A Working Theory of Sovereignty," *Political Science Quarterly*, vol. xlii–xliii (1927–28); W. Sukiennicki, *La Souveraineté des États en Droit International Moderne* (1927); J. Mattern, *Concepts of State, Sovereignty and International Law* (1928); G. W. Keeton, *National Sovereignty and International Order* (1939); J. Marshall, *Swords and Symbols: the Technique of Sovereignty* (1939); M. S. Korowicz, *La Souveraineté des États et l'Avenir du Droit International* (1945); J. Maritain, "The Concept of Sovereignty," *American Political Science Review*, vol. xliv (1950); K. W B Middleton, "Sovereignty in Theory and Practice," *Juridical Review*, vol. lxiv (1952); J. Calvez, *Droit International et Souveraineté en U.R.S.S.* (1953); E. N. van Kleffens, "Sovereignty in International Law," *Académie de Droit International, Recueil des Cours*, vol. lxxxii (1953); J. L. Brierly, *The Law of Nations*, 6th ed. rev. by C. H. M. Waldock, ch. i–ii (1963). (L. B. SN.)

SOVETSK (formerly TILSIT), a town in Kaliningrad Oblast' of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the left bank of the Neman River about 35 mi. (56 km.) from its mouth. Pop. (1959) 31,941. A castle was founded there by the Teutonic Knights, and the associated settlement became a town in 1552. In 1807 Napoleon and the emperor Alexander I signed the Treaty of Tilsit on a raft moored in the river. The town passed with northern East Prussia to the U.S.S.R. in 1945 and was renamed Sovetsk. It is an important rail junction, with lines to Kaliningrad, Chernyakhovsk, Poland, Klaipeda, and Riga. The chief industry of Sovetsk and of its twin town, Neman, a few miles upstream, is papermaking. (R. A. F.)

SOVETSKAYA GAVAN, a town and seaport of Khabarovsk Kray of the Russian Soviet Federated Socialist Republic, U.S.S.R., stands on the mainland, on the southeastern shore of a deep, narrow gulf of the Tatar Strait (the strait separating the mainland from the island of Sakhalin, just north of Japan). Pop. (1959) 50,421. Although the bay, which forms one of the best natural harbours of the Soviet Far East, was discovered in 1853 by Lieut. N. K. Boshnyak, its first development began only on the eve of World War I, and town status was not achieved until 1941. Great impetus was given when it became the terminus of a railway from Komsomolsk, and it is now the second Russian port of the Pacific coast. Fishing, fish canning, and timber working are important. (R. A. F.)

SOVIET (Russian, "council"), a policy-making and administrative agency established at various levels of government in the Union of Soviet Socialist Republics under the direction of the Communist Party.

Soviets were organized in 1905 by Marxist-stimulated intellectuals in Russian cities as strike committees to coordinate worker opposition to tsarist policies. Suppressed after the 1905 revolution, soviets of workers and soldiers were reestablished by left-wing political leaders to influence governmental policy on the tsar's abdication in March 1917. Lenin demanded that they assume power when he returned from exile in April 1917. A national federation of soviets was created in June 1917, but moderate leftists held control until November. After seizure of power Lenin turned to a congress of deputies of local soviets to provide a semblance of popular support for Bolshevik leadership. By the spring of 1918, he had control, and thereafter the various soviets became unchallenged instruments of Bolshevik policy.

Structurally, soviets were large assemblies elected at village, city, county, province, republic, and all-union levels. From 1917 to 1936 the franchise was denied employers of labour for profit, private merchants, *rentiers*, priests, and monks.

The 1936 constitution opened the franchise to all on secret ballot, but in practice the one-candidate election was continued, permitting citizens only to accept or reject candidates nominated by professional groups under Communist Party direction. Soviets chose executive committees, called councils of ministers, at republic and all-union levels, to supervise administrative departments.

Local soviets operated public housing, local transportation and industry, schools, and health facilities. Republic and all-union soviets approved economic plans and budgets. No critical debate or adverse votes occurred because deputies were largely Communist Party members bound to resolve disputes privately. *See* also UNION OF SOVIET SOCIALIST REPUBLICS: *Administration and Social Conditions.*

BIBLIOGRAPHY.—George Barr Carson, *Electoral Practices in the U.S.S.R.* (1955); John N. Hazard, *The Soviet System of Government* (1957); Merle Fainsod, *How Russia Is Ruled* (1963). (J. N. H.)

SOVIET LAW is a system of public order devised by Communist politicians after the Russian Revolution of 1917 and emulated subsequently in Asian and eastern European states where Communist parties became dominant. Although it was classified by some legal scholars as a subdivision of the Romanist or civil-law system, its adherents claimed independent status for it. This claim was recognized in practice after World War II by election to the International Court of Justice of judges representative of the Soviet legal system in accordance with requirements of the International Court's statute (art. 9) that it include judges representative of the main forms of civilization and of the principal legal systems of the world.

HISTORY

Origin.—Since Karl Marx and Friedrich Engels prescribed no pattern of law for a state organized to achieve the aims set by their *Communist Manifesto* of 1848, it was left for Lenin and his colleagues in the Russian Communist Party after the Bolshevik Revolution of 1917 to improvise a legal system. General guides provided by Marx and Engels were that law be regarded as an instrument of the state, not as a limitation upon those who make policy, and that it enunciate rules of public order facilitating the transition to Socialism and ultimately Communism. Its major tasks were to be two: (1) elimination of the political power of the *bourgeoisie* by depriving them of their ownership of productive resources; and (2) education of citizens in the disciplined pattern of life claimed by Communists to be requisite to achievement of the social order they desired. Completion of these two tasks would, in the Communist view, assure abundant production suitable to realization of the Communist aim of distribution according to need and also the self-discipline of citizens necessary to make coercion unnecessary in preserving order. The police, the army, and courts would become unnecessary, and law would in Marxist terminology "wither away." Citizens would perform social obligations, expressed in morals and unsanctioned administrative regulations, out of conviction of their desirability rather than out of fear of punishment for their violation.

The new principles were embodied in the first constitution of the Russian Soviet Federated Socialist Republic, promulgated July 10, 1918 (art. 9), as follows:

> The basic task placed during the present transitional moment on the constitution of the R.S.F.S.R. is the establishment of the dictatorship of the city and village proletariat and of the poorest peasantry in the form of a powerful all-Russian Soviet authority with the objective of complete suppression of the *bourgeoisie,* the exploitation of man by man and the installation of socialism, under which there will be neither division into classes nor a State authority.

Lacking any precise pattern for a legal system designed to achieve the purposes stated in the first constitution, the new government issued only a few decrees designed to establish a framework for a new society and a primitive institutional structure to enforce it. The decrees deprived private individuals of the ownership of land, banks, insurance companies, merchant fleets, and large-scale industry in implementation of the policy of expropriation of the *bourgeoisie;* created restrictions on the employers of labour; and secularized marriage and divorce. The enforcement instruments, called "people's courts," operated like tribunals in a tribal society, without benefit of professional prosecutor or bar.

To permit regulation of the multitudinous social relationships for which no new law was prescribed, judges were directed at the start to apply Russian imperial laws, but only to the extent that they had not been revoked by the Revolution and were not contrary to the revolutionary conscience of the judges. Perhaps because Lenin and his colleagues were too busy to do otherwise, or perhaps because they preferred to let develop in practice a new legal order, no all-inclusive systematized body of law was prescribed until 1922. Local judges were guided during the first years, apart from the basic decrees indicated, only by suggestions issuing from the people's commissar of justice in the form of instructions, procedural manuals, and journal articles praising some court decisions as the proper application of revolutionary conscience and denouncing others as improper.

Political enemies of the Communists were not brought before the courts but were condemned by political bodies created at the same time as the people's courts and called revolutionary tribunals, or they were arrested and imprisoned without public hearings by the political police (Cheka). Lenin and his jurists professed only temporary resort to treatment of opponents outside the regular court system. Nevertheless, the conduct of revolutionary tribunals and of the Cheka greatly influenced the course of Soviet law long after their formal abolition in 1922, for they stimulated lack of respect for formality and strict adherence to law even among those charged with enforcing and administering the law.

New instrumentalities emerged within the Ministry of Internal Affairs (MVD), under the personal dictatorship of Lenin's successor, Joseph Stalin, to rid him of opponents. Following the latter's death in March 1953 the Communist Party's first secretary, N. S. Khrushchev, disclosed that many innocent persons had been convicted and sentenced to long prison terms and even to death by special boards of the MVD. The Soviet method of preserving order was revealed officially to have been dual, with courts on one side dealing publicly with nonpolitical offenses and social disputes, and administrative boards and the security police punishing secretly on the other side those whom Stalin chose to call his enemies.

Codification.—Reintroduction of private enterprise in limited areas and subject to strict controls followed the declaration, in 1921, of the New Economic Policy (*see* UNION OF SOVIET SOCIALIST REPUBLICS: *History*) aimed to speed reconstruction to overcome the devastating effects of world and civil war. This reintroduction necessitated, in the Communist view, stabilization of law if capitalists were to be induced to invest. Lenin's commissar of justice, D. I. Kurskii, considered these measures retrogressive but necessary, while N. V. Krylenko, later to become federal commissar of justice, called them the natural evolution of five years of Soviet law. Whatever their origin, stabilization efforts took form in the first systematic codification, and a complex institutional framework of courts was constructed.

A judiciary act of Oct. 31, 1922, established within the Russian republic a three-stepped system of general courts with civil and criminal jurisdiction.

1. The name people's court was retained for the lowest level, but the institution was less primitive than that of 1917. Its bench comprised a full-time judge appointed annually and two lay judges selected for service of a few days in rotation from a panel of intelligent and politically trustworthy citizens. Lay judges shared responsibility with the professional judge in deciding issues of law and fact.

2. Appeals from this court lay to a new provincial court evolved from the Congress of People's Judges which, under the prior rules, had gathered periodically in each province to survey the work of the local courts. The provincial courts also received jurisdiction to try offenders against the security of the regime and other serious civil and criminal cases. As an appellate court the provincial court presented a bench of three professional judges. As a trial court it was composed like the people's court, except that the lay judges were selected from a panel of more experienced and politically sound citizens.

3. To coordinate policy in all provinces there was created the Supreme Court of the republic, evolving from a control department established earlier in the Commissariat of Justice. This court was authorized not only to hear appeals from cases tried in provincial courts but also to discipline lower courts, issue rulings interpreting the codes, and even to try cases of an unusually important character.

To provide specialized treatment of crimes relating to military matters and to the disruption of transport, then deemed critical to the success of the regime, special courts which had evolved earlier were continued by the judiciary act and subordinated to the Supreme Court. An office of public prosecutor and a college of defenders to perform the functions of a bar were established to aid the judges, following experimentation with institutions designed to conduct prosecution and defense on a nonprofessional basis.

Substantive law and procedure to be applied by the new courts were established by codes enacted throughout 1922 and early in 1923, these being criminal, civil, family, land, and labour codes and also codes of criminal and civil procedure. Drafters of these codes followed patterns essentially similar to those of the Romanist states of the European continent.

Patterns of legal institutions and substantive and procedural law enacted by the R.S.F.S.R. were copied with little variation in the other Soviet-type republics that had emerged in peripheral regions of what had been the Russian Empire. Federation of the Soviet republics into the Union of Soviet Socialist Republics on Dec. 30, 1922, caused no change in the judicial system except the establishment, as a coordinator of all practice, of a supreme court of the U.S.S.R. The first federal constitution, adopted provisionally on July 6, 1923, and permanently on Jan. 31, 1924, extended authority to the federal government to enact general principles of law to be followed by the republics in maintaining their codes. While this development might have caused a change in the Soviet legal system, it did not; the first federal judiciary act of Oct. 29, 1924, confirmed the system of courts then existing in each of the republics, and no change was made in substantive or procedural law. The sole innovation was that the military and transport courts were placed under sole control of the Supreme Court of the U.S.S.R.; thus was created a self-contained system of federal courts, with lower branches functioning throughout the republics but beyond the reach of republic officials. A sharp line was thus drawn between courts dealing with matters vital to the security of the regime and those concerned with social disputes.

The pattern established for Soviet courts and codes of law in 1922 and 1923 continued until Stalin's death, although the second federal constitution of Dec. 5, 1936, transferred authority to the federal level to enact codes of law to replace those of the republics. Although a second federal judiciary act was issued on Aug. 16, 1938, new codes were not enacted. On Feb. 11, 1957, the constitution was amended to restore the original relationship between republics and federation, namely that they enact their own codes but conform to general principles established federally. The first of the new general principles were enacted for criminal law and procedure on Dec. 25, 1958; for civil law and procedure on Dec. 8, 1961. Others were enacted at intervals thereafter, and republic codes were revised subsequently. The major innovation of the 1958 enactments was the requirement that punishment be ordered only by a court in accordance with the rules of the procedural code. This provision was heralded by Soviet jurists as preventing return to extralegal procedures such as those exercised by the MVD during Stalin's dictatorship.

CHARACTER OF SOVIET LAW

Source.—Montesquieu's concept of separation of powers as embodied in the Constitution of the United States of America was rejected by the founders of the Soviet legal system. The legislature was made supreme over the executive and the judiciary. In principle only the legislature's enactments were a source of law.

Practice made the executive a source of law, however, for the Council of Ministers often acted without convening the legislature. The second federal constitution of 1936 attempted to stop this practice in violation of the constitution of 1923, but the executive soon again usurped legislative authority. The post-Stalin reforms again sought to restore monopoly lawmaking power to the legislature. In practice it was not the legislature in full assembly but its derivative body, the presidium, elected from its membership to legislate during intervals between full sessions, that created the day-to-day changes in law. Although subsequent ratification from the full body was sought, revocation of presidium action was virtually impossible because of passage of time. The presidium thus became the most active source of law, except for the state economic plan, the budget, and enactment of general principles for republic codes.

Communist Party orders were in the technical sense not a source of law, for to be binding on courts these had to be enacted by state agencies. Still, the party provided the initiative for legislative action. Some vital laws, such as those reforming agriculture or industry, were given preeminent status by signature by the party secretary as well as by the chairman of the presidium when they were published as law.

Having rejected judicial lawmaking with assertion of the legislature as the sole lawgiver, Soviet jurists accepted no concept of judicial precedent like that of the Anglo-U.S. common law. Nevertheless, the Supreme Court of the U.S.S.R., in exercise of its authority to enunciate principles of interpretation based upon examination by its full body of a course of judicial practice, proclaimed such far-reaching principles as to make of these a source of law. Even individual decisions of the Supreme Court, although lacking the authoritative force of the general rulings, were given careful attention by lower courts, since they were published as guides. The Soviet legal system in practice adopted much the same approach to judicial decisions as have Romanist systems generally.

The Supreme Court of the U.S.S.R., being subordinate to the legislature, had no right to declare a law unconstitutional nor to create any limitation upon the legislature or the executive on the ground that some higher or general principles of law had been violated.

Control over the judiciary was preserved for the legislature at all levels above the people's court by giving it appointive and recall power, through the soviet operating at the court's level. Only people's judges were elected by the people under provisions introduced by the 1936 constitution, although elections were not begun until after World War II. Elections followed the pattern established for all representative bodies in the U.S.S.R. by Stalin, namely the one-candidate ballot, the candidate being chosen by local professional groups with the guidance of the Communist Party members in the groups. In practice, judicial decisions had to accord generally with Communist Party policy or the judge risked recall.

In light of the foregoing, the constitutional provision that "judges are independent and subject only to the law" (art. 112) had to be understood as denying to local officials the right of intervention in the formation of an individual decision on personal grounds, but as not prohibiting Communist Party intervention when such intervention was properly formulated in criticism of a line of decisions out of keeping with party wishes.

Judges were not deprived entirely of initiative by the rule of

legislative supremacy or by the leading role played by the Communist Party. In both the civil and criminal codes adopted in 1922 there was authorization to depart from the rigid provisions of the code under special circumstances. The civil code opened with art. 1, reading, "Civil rights are protected by law except in cases where they are exercised contrary to their social and economic purpose." This rule permitted judges to apply their concept of state policy, when it was not specifically stated, to achieve an appropriate result in a given situation, but the decision created no rule of law for future cases. The article was used frequently to justify judicial refusal to protect the rights of capitalists during the period of reconstruction (1922–28), but after 1930 it fell into disuse. Soviet jurists argued for its deletion from the code as unrelated to a stable society from which all capitalist elements had been eradicated, but it was retained in the civil-law general principles of 1961 (art. 5).

The criminal code of 1922, and as revised in 1926, permitted the judge to exercise discretion not only in withholding conviction where a crime had clearly been committed, provided that he found no social danger, but also in punishing a citizen who had committed no crime defined as such by the code. The formula for such judicial initiative was that the act be found socially dangerous in a manner analogous to another act specifically defined as a crime. This provision, which led to unpredictability, was eliminated from the criminal law by the general principles adopted in December 1958.

Substantive Provisions.—Soviet jurists claimed uniqueness for their system because of its constitutionally defined function of promoting socialism. The end, in their view, transformed the means.

Civil Rights.—Civil rights as defined in the U.S.S.R. Constitution of 1936 (ch. x) provided the primary example of the Soviet approach. Guarantees of freedom of speech, press, and assembly were limited to such exercise as might strengthen Socialism (art. 125). Judicial practice indicated that speech evoking hostility to Soviet authority on grounds of disagreement with the elements of social order declared by the constitution as fundamental to Socialism—such as state ownership of the productive resources, collectivization of agriculture, the one-party political system, and equality of races—was punished as a crime against the state if intent to disrupt the system was found. In the practice of the security police prior to the death of Stalin, disruptive intent was presumed from social origin (as a prerevolutionary landowner, industrialist, merchant, army officer, or priest, and sometimes in children of such persons).

State Enterprise.—Among the various fundamentals of Socialism, primary place was given to state ownership of productive property, and law codes reflected this emphasis in various provisions. The criminal code established more severe penalties for theft or destruction of state property than for that of personal property. The labour code was concerned with state employment. The provisions of the civil code of 1922 governing incorporation of private businesses became obsolete with the passing of the New Economic Policy, and even those provisions relating to sales, contracts, property damage, and inheritance governed only a small segment of social activity because the volume of personal property relationships was sharply reduced. Nothing monopolized by the state could be sold or leased by private persons or even by state enterprises and cooperatives, with limited exception for land allocated to peasant households disrupted temporarily by war or industrial employment.

A major new area of law that accompanied Socialism was the law governing state enterprise. Its task was to provide the order and accountability required for performance of the national economic plan. After experimentation for a period with direct operation by ministerial departments, state enterprise passed in 1923 to newly created public corporations, having juridical status with power to sue and be sued. State property was assigned to each enterprise, and business-type accounting procedures were introduced to assure strict accountability. The function of law in the accountability process was performed by an adaptation of the principle of contract. Interrelations of public corporations took the form of written agreements to buy, sell, and perform services, these agreements being executed in performance of the planned objective of each corporation.

For scarce commodities, such as aluminum, copper, and steel, the plan was detailed, establishing the precise corporations to produce and consume the materials. Contracts in such cases became documentation of detail in implementation of planned relationships; management had no choice but to execute the contract. With regard to less critical commodities, the plan provided no such detail of corporate interrelationship; management was permitted to seek out parties producing needed raw materials and to enter into relations with distributing units prepared to sell its products to ultimate consumers. With the administrative reform of 1957, which abolished many centralized ministries and substituted regional economic councils to guide the economy, the area of relatively free choice in the establishment of productive relationships among public corporations was enlarged, and contracts took on greater meaning as the documentation of relationships established with some exercise of free will by the parties. Nevertheless, the aim of plan performance was dominant, and a party evidencing reluctance to make a contract with another could be taken before a state agency for determination of the necessity of execution.

The state agency designated to hear precontract disputes as well as those arising over failure of performance was a specialized court designated as "state arbitration." Created on various levels to facilitate quick response to demand for a decision, its arbitrators utilized the law of the civil code in reaching a decision, but more than money damages was required. Since the primary purpose for the contract relationship was execution of the plan, arbitrators were directed to assure specific performance, if possible, and only if that were out of the question, to order the payment of damages; such payment set in balance the profit and loss statement of the injured party and attracted the attention of state inspectors to the offending party by upsetting its profit and loss statement.

Personal Property and Personal Injury.—Provisions of the civil code relating to personal property matters of private citizens, as well as to personal injury actions, provided no surprises for those familiar with the Romanist provisions of the civil codes of the European continent. There was nothing Socialist about these provisions, although Soviet codifiers sought to incorporate some hardship provisions to avoid exploitation of the poorer party to a bargain. Such a provision, already in use even in western Europe, was the one prohibiting enforcement of a contract if one party in negotiating it had taken advantage of the "extreme want" of the other. Likewise, uneconomic maintenance of a dwelling could result in its confiscation by municipal authorities if the owner had resources to make repairs and had been notified of the necessity of repair at an appropriate time.

Labour.—The provisions of labour law reflected state ownership of productive resources, especially after 1936, when the second federal constitution prohibited private employment from which a private citizen anticipated a profit as entrepreneur. Labour law then came to resemble civil-service law in many non-Soviet-type states, its distinguishing feature being that it was the law of all employment rather than of a small segment. Statutes established wages and hours in accordance with a system of standard classification of all jobs. Employers were not permitted to depart from the standards after 1938.

The labour code defined circumstances under which dismissal was permitted (art. 47), such as incompetence, malfeasance, termination of a position, or unauthorized absence for a given period of time. In regard to unauthorized absence, the period was changed as discipline became stricter with the approach of World War II, until finally dismissal was eliminated as a sanction and by law of June 26, 1940, workmen were subjected to criminal prosecution if absent without authorization for more than 20 minutes. After Stalin's death the provision was repealed by law of April 25, 1956. A labour grievance procedure comprising tribunals on which representatives of management and the labour union shared equally in the decision provided a preliminary opportunity for an aggrieved employee to contest management action. A dissatisfied employee might proceed to court in appeal of the tribunal's deci-

sion unless he was a supervisor or in a position requiring security check.

Collective agreements between management and labour unions fell into disuse in the mid-1930s, when wages and hours became fixed by statute, but they were reintroduced in new form in 1947. They then became instruments to stimulate production by restating workers' obligations and assuring distribution, in accordance with the workers' desire, of the bonuses granted by statute to employees for exceeding planned quotas or reducing costs of production. Labour unions were not deprived specifically of the right to lead strikes, but Communist Party leaders in 1928 declared strikes unthinkable as being against the workers' own state.

Land.—Land use was governed by the land code of 1922, by supplementary decrees, such as the Model Charter for the Collective Farm as enacted initially in 1930 and subsequently amended, and by specialized decrees on the use of land by public corporations and individual houseowners. In accord with Socialist tenets, title to land remained in the state and only the use was allocated. Use passed in perpetuity to the individual houseowner, peasant household, collective farm, or public corporation, subject to withdrawal in the event of misuse or requirement for other state activity such as railroad or highway construction. Safeguards against arbitrary withdrawal of farmlands were instituted, it being the desire of policy makers to preserve the psychological advantage provided by capitalism in encouraging maintenance and improvement of land held in perpetuity. No sale or leasing of perpetual right to use was permitted, except as already indicated.

Family.—Family law provided for secularization of marriage and divorce. Early law permitted registration of divorce at the request of one spouse and without notification to the other, but this procedure was replaced in 1944 by a stricter system thought to be in keeping with Engels' dictum that Socialism would see the introduction of the truly monogamous family. Thereafter, the people's court was obligated to attempt reconciliation, and only in the event of failure could a party proceed to the provincial court, which was required to satisfy itself that restoration of the marriage relationship was impossible before granting the divorce. (*See* also MARRIAGE, LAW OF: *Europe;* DIVORCE: *Europe.*)

Procedural Law.—Romanist influence was strong in the codes of criminal and civil procedure, which copied the main features of the French codes, although retaining more possibilities of control by the state (*see* FRENCH LAW).

Emphasis in criminal procedure was placed on investigation before trial, as in French practice. The examiner had to establish more than a prima-facie case required of the grand jury in the traditional common-law system. The suspect had to be given an opportunity to testify and to produce witnesses and evidence in his defense. The examiner became a preliminary judge authorized to bring the case to trial only if he became convinced of guilt, so that the public trial consisted primarily of verifying the examiner's work rather than hearing the defendant's case for the first time. Contrast with the French model was provided by the fact that the examiner was not a magistrate but subordinate to the office of prosecutor, and until 1958 no attorney was permitted the defendant during the preliminary investigation. Even thereafter only juveniles were given an absolute right to an attorney.

Trial procedure followed the French model closely in that the judges were instructed to establish their intimate conviction of guilt. To do so they might go beyond evidence presented by the prosecution and defense. The court might call witnesses, seek its own experts, examine material evidence, and visit places connected with the crime. The judges did most of the interrogating in the courtroom, relying on the prosecutor and the defense only to bring out points overlooked. No rule of evidence of any kind bound the court, except the rule of relevance.

Civil procedure required observance of rules of evidence only for certain formalities established by the civil code, such as those for documents like contracts or wills. Otherwise, the judges were as free in civil as in criminal matters to seek or hear what they desired to establish their intimate conviction. Soviet jurists claimed an innovation in the prosecutor's authority to intervene in a civil case at any stage thought necessary to the interests of society. This sometimes brought in the prosecutor to protect a party whose attorney appeared incompetent.

Appellate courts finding violation of procedural requirements, of substantive law, or unconvincing evidence were required to remand for new trial. Revision of a sentence or of a decision without retrial in a court of original jurisdiction was permitted only when the original court had no jurisdiction, where an amnesty had been improperly applied, or where there had been no basis for trial at all. A penalty could not be increased on defendant's appeal.

A singular feature of Soviet procedure was the authority given to Supreme Court presidents and to the prosecutors of the republics and of the U.S.S.R. to protest a decision, whether civil or criminal, after it had become final on appeal or in absence of appeal. This procedure permitted reopening of convictions or even of acquittals if the highest legal authorities found grave error in their periodical audit of inferior-court activities. Prosecutors frequently requested such reopenings, even of convictions obtained by inferior prosecutors. This created the opportunity of petition to superior prosecutors or to court presidents after all remedies as of right had been exhausted. At the same time it gave state authorities a second chance, not given the convict or the party to a civil suit, to change an otherwise final decision.

See also UNION OF SOVIET SOCIALIST REPUBLICS: *Administration and Social Conditions.*

PEOPLE'S DEMOCRACIES

Application of the pattern of law evolved within the U.S.S.R. to other areas began when Outer Mongolia declared establishment of its "people's government" on July 11, 1921. Mongolia was then a nomadic society without industry. Nationalization of productive resources had little effect on nomadic life, an exception to nationalization being made in the area of primary concern, namely the private ownership of cattle herds. Nevertheless, the base was laid for following Soviet legal patterns in the first constitution of the Mongolian People's Republic promulgated on Nov. 26, 1924. Thereafter, state-owned industry was developed after Soviet patterns, and following adoption in the U.S.S.R. of the second federal constitution, a second Mongolian constitution was promulgated on June 30, 1940. Its provisions on courts and prosecutors were like those of the U.S.S.R., and the subsequently adopted law on land use of Feb. 6, 1942, provided the same formula of state ownership of land allocated for use gratis and in perpetuity to the nomadic herdsmen for their privately owned cattle. A labour law of Feb. 14, 1941, followed the Soviet pattern, as did a social insurance law of June 22, 1942.

The major opportunity to install the Soviet legal system outside the U.S.S.R. came only after World War II, when Soviet-type governments were created for European states occupied by the Soviet army in pursuit of retreating German forces. The system was also established in China, North Korea, and North Vietnam as military victories placed the Communists in power. In each country there was some variation, related to the degree of economic development, from the Soviet pattern evolved in the U.S.S.R.

Eastern European states permitted retention of limited forms of private enterprise in industry and agriculture reminiscent of the economic pattern of the U.S.S.R. during the 1920s. In China, capitalists, except for landlords, were permitted for a time to function widely, though subject to strict state controls and heavy taxation. In North Vietnam even private landowning was permitted except for large estates, which were divided and distributed among tenants and poor peasants. Law in these communities reflected the presence of capitalist elements in the economies. Yet in all the people's democracies the banks, large-scale industry, insurance, mining, and transport were nationalized and monopolized by the state. Law relating to this sector of the economy, termed the socialized sector, followed the U.S.S.R. model.

Throughout all the states that became Communist, law was given the functions previously established in the U.S.S.R. Thus, the constitution of the Hungarian People's Republic of Aug. 20, 1949, declared (art. 41):

> Courts of the Hungarian People's Republic punish enemies of the working people, defend and secure the state, economic and social

structure of the people's democracy, its offices and the rights of the toilers, educate the toilers in the spirit of observance of the rules of socialist intercourse.

Regulations governing the organization of the people's courts in the People's Republic of China, promulgated Sept. 3, 1951, specified that courts were "to consolidate the people's democratic dictatorship, uphold the new democratic social order and safeguard the fruit of the people's revolution." The People's Republic of Bulgaria was even more outspoken than the U.S.S.R. in clarifying the function of law; in its constitution of Dec. 4, 1947, it granted the right to citizens to create associations, but (art. 87) only "if they are not directed against state and public order established by the present constitution," and then stated:

It is forbidden and will be punished by law to form organizations having as their purpose taking from the Bulgarian people the rights and freedoms won by the people's uprising of Sept. 9, 1944, and guaranteed by the present constitution, or to limit these rights and freedoms, to place under threat the national independence and state sovereignty of the country or to propagate open or concealed fascist or antidemocratic ideology, or to facilitate imperialist aggression, and also to participate in these organizations.

The Chinese Communist Party leader, Mao Tse-tung, in his 1949 speech "On People's Democratic Dictatorship" indicated that he had adopted in effect what had been Stalin's interpretation, enunciated in 1930, of the Marxist concept of the withering away of the state. Mao declared that there could be no withering in China as long as there was opposition to the state. The constitution of the People's Republic of China, promulgated on Sept. 20, 1954, nearly five years after proclamation of the people's republic on Oct. 1, 1949, fixed as final the system of courts evolved in emulation of those of the U.S.S.R. There were to be lay assessors, full-time professional judges, Soviet-type prosecutors, and three levels of courts; "judges were independent and subject only to the law," as they had been declared to be in the U.S.S.R. constitution.

Judicial patterns of the same type were established in the constitutions of the other people's democracies: the Democratic People's Republic of Vietnam (Nov. 8, 1946); the People's Republic of Albania (March 14, 1946); the Rumanian People's Republic (April 13, 1948, and Sept. 24, 1952); the Republic of Czechoslovakia (May 9, 1948); the Democratic People's Republic of Korea (Sept. 8, 1948); the German Democratic Republic (Oct. 7, 1949); and the Polish People's Republic (July 22, 1952).

The applicable law was stated in a small number of specific decrees in each republic, but otherwise determination of law was left to the judges. The Chinese government, among its first actions, abrogated all legal codes of the prior government. In October 1950 the Chinese premier, Chou En-lai, listing the documents to serve as guides to the judges, included not only specialized decrees of the government but also the Common Program of the Chinese Communist Party.

The provisional regulations on people's courts of 1951 (art. 4) read, "Where no regulation is applicable, the policy of the Central People's Government will be adhered to." A law codification commission was established in the spring of 1950, but no basic codes emerged, only specialized statutes such as the land reform law of June 28, 1950; the law on marriage, April 13, 1950; the law on questions of employment, July 25, 1952; and the organic law for the court system, Sept. 21, 1954.

The Soviet-type states in eastern Europe moved more rapidly toward codification than had China, and in contrast to China confirmed the existence of prewar civil codes, although with modifications. Poland adopted a General Statute on Civil Law with 119 articles on July 18, 1950. As in the Soviet civil code of 1922, the codifiers sought to avoid possible injury to working-class elements; art. 3 provided: "It is forbidden to use one's rights to violate the fundamentals of socialist intercourse of the people's government." A Polish family code was adopted on June 27, 1950, and a decree on planned contracts on April 26, 1950.

Prior to the expulsion of the Yugoslav Communist Party from the Cominform on June 28, 1948, Yugoslavia, under the constitution of Jan. 31, 1946, had followed the Soviet legal pattern. Under Tito, Yugoslavia accepted the concept of monopoly of power in a Communist Party and the necessity of state ownership of productive resources, but refused to accept dictation from the Soviet Union on every detail, especially on centralization of industrial control and the speed of collectivization. The Yugoslav government experimented with bringing workmen into the managerial process of public corporations and in broadened autonomy for local government councils. Yugoslav legal philosophers rejected the view that the state could not begin to wither away until full achievement of Communism; they wanted to start relaxation of coercion immediately. Stalin's attempts to unseat Tito failed, and on Stalin's death his successors attempted reconciliation for a time, indicating acceptance of the idea that there could be many roads to Socialism.

Stalin's death in 1953 and his repudiation by the Communist Party of the Soviet Union at its 20th Congress in 1956 opened a new era in the law of Communist-led states. The People's Democracies sought to be relieved of Soviet tutelage. Risings in Hungary and Poland forced the Soviet leaders to acquiesce in the evolution of divergent legal systems. Polish jurists, including Communists, demanded humanistic attitudes toward legal procedure, and other east European states followed their lead. Even Soviet legal philosophers in 1958 endorsed a new humanism and declared it always to have been fundamental to Lenin's thought.

Only Chinese Communists resisted the trend toward strengthened legal procedures, denouncing Soviet policies of humanism and legality in an open letter published in 1963. Since 1957 Chinese leaders had abandoned their policy of copying Soviet legal patterns and had ceased efforts to draft codes of law, declaring that they wanted flexibility in court. They evolved a legal system seemingly inspired by relics of traditional Chinese attitudes as modified by their own guerrilla experience prior to 1949 and by study of Soviet practices during the period of War Communism before 1921.

Concurrently with revival of concern for humanistic values, Communist states other than China revised their constitutions and law codes to give formal support to their claim to have passed beyond capitalism to Socialism. Mongolia and Czechoslovakia declared themselves Socialist republics in constitutions adopted July 6 and 11, 1960, respectively. North Vietnam introduced its new constitution on Dec. 31, 1959. Yugoslavia promulgated its revision on April 7, 1963. New civil codes designed for Socialist economies were introduced in Poland and Czechoslovakia in 1964, the latter incorporating new terminology departing sharply from Romanist tradition.

Several features, however, remained unchanged in the law of these states even after the introduction of variations. Most notable was continuation of Lenin's principle that no basis remained after the Russian Revolution for the Romanist division of law into public and private spheres. To Lenin the Soviet state's concern with every detail of social intercourse required maintenance of the state's right and opportunity of intervention in any matter at any time. Consequently he held that in the Soviet system all law must be public, that is, reflect the state's vital concern in the social relationships governed by law. This attitude of state direction of all matters, sometimes called *dirigisme* by French scholars, remained characteristic of Soviet law.

The industrialization characteristic of the 20th century required legal systems in all states to accept increasing state intervention in social relationships. Employers were no longer left free to execute labour contracts as they wished; property owners were subjected to increasing restraints on use, and in some countries were subjected even to the loss of farms or dwellings if these were not used productively; nationalization in banking, industry, and transport was widespread; compulsory insurance against injury to third parties became commonplace, as did social insurance against industrial accidents; and secularization of marriage and divorce nearly everywhere prevented citizens from conducting their marital relationships solely within religious rules.

Soviet law, in its reflection of the state's pervading presence in human relationships, appeared to be but an extreme extension of trends common to legal systems everywhere. Yet many legal philosophers concluded that Soviet law differed in more than degree from other systems. While its techniques were those of the Romanist system, its difference in degree was interpreted as a dif-

ference in quality, justifying the conclusion that Soviet law was a unique system requiring separate categorization. *See* also UNION OF SOVIET SOCIALIST REPUBLICS: *The Judicial System.*

BIBLIOGRAPHY.—John N. Hazard, *Settling Disputes in Soviet Society: the Formative Years of Legal Institutions* (1960) and *Law and Social Change in the U.S.S.R.* (1953); Peter S. H. Tang, *Communist China Today* (1957); Vladimir Gsovski and Kazimierz Grzybowski, *Government, Laws and Courts in the Soviet Union and Eastern Europe,* 2 vol. (1959); Charles P. McVicker, *Titoism: Pattern for International Communism* (1957); George C. Guins, *Soviet Law and Soviet Society* (1954); Harold J. Berman and Miroslav Kerner (eds.), *Soviet Military Law and Administration* (1955); Boris A. Konstantinovsky, *Soviet Law in Action: the Recollected Cases of a Soviet Lawyer* (1953); R. R. Betts (ed.), *Central and South East Europe, 1945–1948* (1950); Harold J. Berman, *Justice in Russia: an Interpretation of Soviet Law* (1950); Vladimir Gsovski, *Soviet Civil Law: Private Rights and Their Background Under the Soviet Regime,* 2 vol. (1948–49); Rudolf Schlesinger, *Soviet Legal Theory: Its Social Background and Development* (1945); V. I. Lenin *et al., Soviet Legal Philosophy,* trans. by Hugh W. Babb (1951); Andrei Y. Vyshinskii, *The Law of the Soviet State* (1948); Hans Kelsen, *The Communist Theory of Law* (1955); Andrei I. Denisov and M. Kirichenko, *Soviet State Law* (1960); Petr S. Romanshkin (ed.), *Fundamentals of Soviet Law* (1962); Glen G. Morgan, *Soviet Administrative Legality* (1962); Kazimierz Grzybowski, *Soviet Legal Institutions* (1962); John N. Hazard and Isaac Shapiro, *The Soviet Legal System* (1962); Harold J. Berman, *Justice in the U.S.S.R.* (1963); George Feifer, *Justice in Moscow* (1964); Albert P. Blaustein (ed.), *Fundamental Legal Documents of Communist China* (1962). (J. N. H.)

SOVIET UNION: *see* UNION OF SOVIET SOCIALIST REPUBLICS.

SOWER, CHRISTOPHER (1695–1758), printer and pietist leader of the Pennsylvania Germans, was born in Ladenburg, near Heidelberg, Ger., and lived for a number of years in Laasphe, duchy of Wittgenstein. He migrated with his wife and son Christopher (1721–84) to Germantown, Pa., in 1724, residing there until his death, except for a period of farming in Lancaster County, 1726–30. An artisan skilled in many crafts, and profoundly religious, he found his true career in 1738 as the first successful printer to the numerous Germans in colonial America. He imported press and type from Germany, but made his own ink and, occasionally, his own paper. The output from his large and prolific press included a newspaper, *Der Hoch-Deutsch Pensylvanische Geschicht-Schreiber,* retitled *Pensylvanische Berichte* in 1748, an almanac, *Der Hoch-Deutsch Americanische Calender,* and more than 150 other imprints, mostly religious and in German, ranging from broadsides to the first European-language colonial Bible (1743).

Though a separatist, he was sympathetic in Pennsylvania to the Church of the Brethren, founded in Laasphe while he lived there. His pietism was evident in all his activities. He urged the politically apathetic Pennsylvania-Germans to go to the polls to maintain pacifist and oath-denying Quakers in power. He worked for Christlike relations with the Indians, legislation to stop mistreatment of immigrants, dissemination of medical knowledge, building of hospitals, and maintenance of the German culture. He opposed defense measures, compulsory Anglicization of Germans, slavery, higher education, and traditional religions, with their forms, sectarianism, and professional clergy. He deeply appreciated the freedom and opportunities he found in Pennsylvania, despite the enmity of men such as Benjamin Franklin, who had fundamentally different principles and resented his great power among the Germans. Both his press and his principles were continued by his son Christopher after his death in 1758.

See F. Reichmann, *Christopher Sower, Sr., 1694–1758, Printer in Germantown; an Annotated Bibliography* (1943); E. W. Hocker, *The Sower Printing House of Colonial Times* (1948). (W. R. SL.)

SOWERBY, LEO (1895–), U.S. composer, teacher, and organist, was born at Grand Rapids, Mich., May 1, 1895. He studied in Chicago, where in 1917 he presented a program of his orchestral works and songs and in 1921 was named the first recipient of the American Grand Prix de Rome. He taught composition and theory at the American Conservatory of Music in Chicago from 1925 to 1962 and was organist at St. James's Church (now the Cathedral of St. James) there from 1927 to 1962. In 1962 he was appointed director of the College of Church Music in Washington, D.C.

Sowerby's early works were influenced by Max Reger and Grieg. His piano pieces, *The Irish Washerwoman* and *Money Musk,* later orchestrated, are in the manner of Percy Grainger. In *Synconata* (1924) and *Monotony* (1925) he experimented with jazz. His numerous vocal works include cantatas, services, anthems, and *Canticle of the Sun* (performed 1945), which won a Pulitzer Prize. His orchestral works include four symphonies. Besides concertos for piano, cello, and organ, Sowerby wrote many chamber works, including sonatas for wind instruments and piano.

See B. C. Tuthill, "Leo Sowerby," *The Musical Quarterly* (April, 1938). (N. SY.)

SOWING: *see* PLANTING MACHINERY.

SOYBEAN, also known in some countries as the soja bean and the soya bean, is an annual, summer leguminous plant (*Glycine soja* or *Glycine max*) native to eastern Asia and extensively cultivated in the United States. Cultivated in China and Japan long before written history, it is from the standpoint of uses and value the most important legume grown in those countries. Many records of soybean culture in China, some going back as far as 2207 B.C., give advice on soil preference, time of planting, varieties for different purposes and numerous uses indicating that the soybean was among the early crops grown by man. The crop was considered the most important cultivated legume and one of the five sacred grains (rice, soybeans, wheat, barley and millet) essential to the existence of Chinese civilization. In a Chinese dictionary dating from about the beginning of the Christian era the soybean is called *sou,* which very probably was the source of the names soi, soy, soya and soja.

Europeans first knew of the soybean through Engelbert Kaempfer, a German botanist who spent three years (1690 to 1693) in Japan. Kaempfer gave the Japanese name *daidsu mame,* and described it as an erect bean with the pod of a lupine and the seeds of a large pea. Although he discussed the many food products prepared from the soybean by the Japanese, little interest was taken in the crop.

In 1875 a great impetus was given to soybean culture in Europe through the experiments of Friedrich Haberlandt of Austria who grew the seed of several varieties obtained at the Vienna exposition in 1873. Seeds of the four varieties that matured were distributed widely in Austria-Hungary and other countries of Europe and gave promising results. Haberlandt published the results of his investigations in 1878, stimulating further research on the crop. While records show packets of seed being received in France as early as 1739 from missionaries in China, no record of successful culture of the soybean was established in that country until 1879, when a Chinese variety matured satisfactorily at Marseilles. In 1880 Vilmorin-Andrieux and Company introduced into France one of the yellow-seeded varieties tested by Haberlandt, which proved to be well adapted to French conditions. Soybeans were grown as early as 1790 in the Royal Botanic Gardens, Kew, Eng., but the crop is not well adapted to the English climate and there was no commercial production. Cultivation of the soybean in Italy dates from about 1840 but it has not become an important crop in any part of the country.

J. HORACE MCFARLAND COMPANY

PODS AND LEAVES OF THE SOYBEAN

The first mention of the soybean in the United States was by James Mease in 1804, who noted that it did well in the climate of Pennsylvania and recommended that it be cultivated. The Perry expedition to Japan in 1854 brought back two varieties of "soja bean"; for several years after that, the soybean was referred to as the Japan pea. In 1882, the North Carolina agricultural experiment station grew a yellow-seeded soybean, probably the Mammoth Yellow, which remained a leading variety in that area for many years. In 1889 W. P. Brooks of the Massachusetts agricultural experiment sta-

tion brought with him from Japan a number of varieties. The next year C. C. Georgeson secured three lots of soybeans from Japan and grew them at the Kansas agricultural experiment station. Since then most of the agricultural experiment stations in the United States have conducted extensive tests with the soybean crop.

In 1898 the U.S. department of agriculture began introducing large numbers of soybean varieties, mostly from Asian countries with a few from Europe and other parts of the world, and since then has grown and evaluated about 10,000 varieties of foreign origin. Probably among the first varieties secured from Europe were some used by Haberlandt in his early trials in Austria. Beginning in 1936 the U.S. department of agriculture cooperated with the agricultural experiment stations of states and Canadian provinces in developing improved soybean varieties through hybridization and selection. These improved varieties have higher yield and oil content and growth habits more suitable to modern farming methods than the Asian varieties. Increase of acreage and production has been closely correlated with the introduction of these varieties and their improvement through breeding and selection. The improved varieties developed through cooperative research have also contributed very substantially to the increased industrial value of the crop.

Areas of Production.—*Asia and Australia.*—The soybean is grown intensively in China including Manchuria. It is the cash crop of the Manchurian farmer and occupies about 25% of the total cultivated area of that region. In the second half of the 20th century China (with Manchuria) was producing nearly 340,000,000 bu. a year or about 40% of the world total. Other important soybean-producing countries of Asia are Indonesia, Japan, Korea, Formosa (Taiwan), Thailand, and India. The soybean is cultivated in the Philippines and grown successfully in parts of Australia and New Zealand.

Europe and Africa.—The climate of Europe is not generally suited to soybean culture although there is limited production in Italy, Yugoslavia, Rumania, and southern U.S.S.R. There is some soybean culture in Africa, mainly in Tanzania, the Congo, Nigeria, and the Republic of South Africa.

North and South America.—In the Western hemisphere, the United States was the leading soybean-producing country, accounting for nearly 60% of world production. Approximately 83% of the U.S. soybean production is in the north central states. Other areas of heavy production are in the Arkansas and Mississippi delta area of the Mississippi River and along the middle Atlantic coast. Soybeans became a major crop in the United States, with production increasing from 5,000,000 bu. in the mid-1920s to more than 700,000,000 bu. in the 1960s. Another important soybean-producing area in North America was in the province of Ontario, Can. In South America, Brazil and Argentina were becoming important soybean-producing countries.

Plant Characteristics.—The soybean is an erect, branching plant, resembling in its early growth ordinary field and navy beans. Varieties range in maturity from early to late, and in height from less than one foot to five or six feet. The root system constitutes about 10% of the weight of the plant but is well branched and often extends to a depth of four or five feet. The characteristic deep roots are responsible in part for the great resistance to drought offered by soybeans. Except for the pair of simple leaves on the first node above the cotyledons, the leaves are trifoliate and borne one to a node. They vary widely in shape, size, colour, and degree of persistence (leaf remaining attached to the stem), though they usually turn yellow as the pods ripen and are shed by the time the plant is ready to harvest. Nearly all varieties are pubescent; *i.e.*, the stems, leaves and pods are covered with fine tawny (brown) or gray hairs.

Several yellow-seeded varieties in the orient are entirely glabrous, that is, lack pubescence; these usually are smaller, shorter, and yield less seed than most pubescent varieties, and when grown in North America are subject to considerable injury from leafhoppers. In the orient, the glabrous varieties have been found highly resistant to attacks by the pod borer, a serious insect pest, which as yet has not been found in North America.

In the axil of each leaf is a bud that may develop into a branch or flower. The flowers are small and inconspicuous and are either white or different shades of purple. They contain both male and female parts and are almost completely self-fertilized although in the field there is usually around 1% cross-fertilization.

The pods, usually containing two or three seeds, range in colour from light straw or tan through shades of gray and brown to nearly black. Those of the wild soybean and a few hay-type varieties dehisce (open and scatter their seed) readily as the seed matures, although most commercial varieties have been selected for the ability to hold their seed for two or more weeks after maturity.

The seeds are usually yellow, green, brown or black but may be bicoloured. No truly white or red seeds are known in soybeans. The most common bicoloured patterns are green or yellow with a saddlelike patch of black or brown extending down on each side of the hilum or seed scar. Some seeds are brown and black with the patterns concentrically arranged. The cotyledons of soybeans may be yellow or green. Varieties with yellow seed coats and yellow cotyledons are preferred for industrial processing. Seed size varies greatly from the 30,000 seeds per pound of the wild bean to 1,200 seeds per pound for some vegetable varieties. Most commercial varieties range in seed size from 2,500 to 3,500 seeds per pound.

Cultivation.—The soybean will succeed on nearly all types of soils but does best on fertile or sandy loams. In general, soil requirements are about the same as those for corn or cotton. Like other legumes, soybeans obtain part of their nitrogen requirement from the air through the action of certain nodule-forming bacteria living on the roots of the plant. Where soybeans have not been grown previously it is advisable to inoculate the seed with the soybean bacteria just before sowing. (*See* NITROGEN, FIXATION OF.) In the United States soybeans are usually planted during May or early June in rows 30 to 40 in. apart. Early cultivations to kill small weeds or to break a soil crust are usually done with a rotary hoe or harrow starting soon after the plants emerge. Later cultivations are done with regular row equipment used for corn or cotton. Chemical weed control is dependent to some extent on weather conditions.

Approximately 50 diseases are known to attack soybeans, but only a few have caused serious economic losses. Varieties resistant to leaf diseases have been developed, and diseases can be controlled to some extent by planting resistant varieties and following proper cultural practices. Several insects attack soybeans and cause severe losses, but good control may be obtained by timely dusting or spraying with a number of different insecticides. (*See* SPRAYS AND DUSTS IN AGRICULTURE.)

When soybean varieties are grown in their area of best adaptation they mature during September or October, utilizing the full growing season. Flowering and ripening in soybeans are controlled by length of day and night (photoperiod) in which the plants are growing. Soybeans are called short-day plants because under normal conditions they will not start blooming or maturing until the day length is shorter than a certain minimum value which is specific for each variety. This is the reason that a soybean variety is adapted to a rather narrow belt of latitude; and why, for example, in the northern hemisphere, if a variety is planted north of its proper latitude, blooming and ripening will be delayed in the longer summer days and the plant may be caught by frost. Conversely if a variety adapted in the north is moved to a lower latitude it will bloom and mature too early and will not take advantage of the full growing season.

Most soybeans are harvested by a combine, which threshes as it reaps, when the plants are mature, the leaves have fallen and the seed has dried to a moisture content of about 13%, at which time the crop can be stored safely.

Uses.—*Food and Food Products.*—In Asian countries the soybean is grown primarily for the seed, which is used largely in the preparation of hundreds of fresh, fermented, and dried food products. These various food preparations not only give flavour to the oriental diet but supply to a considerable extent the necessary protein in the diet. This has been especially true inland, where seafood is not readily obtainable. Nutritionally the protein of the soybean is similar to that of animal protein—even the amino acid analyses of soybean protein and casein are remarkably similar.

Sprouts grown from soybean seed are eaten by the Chinese as a green vegetable throughout the year, either in salads or as a cooked vegetable. They are a good source of vitamin C. Soybean "milk" (soybean flour suspended in water) is said to have been made long before the Christian era. It is used in the fresh state and also in the manufacture of bean curd or vegetable cheese *teou fu* or *tōfu* used so extensively in the orient. When soybean milk is boiled a film called *yuba* forms on the surface. This film is rich in protein and has been a popular food in China and Japan since ancient times. In cooking, *yuba* is either fried or used in soups. *Nattō* and *miso*, prepared by steaming and fermenting soybeans, are two staple foods in Japan. Red *miso* contains a large amount of salt and will keep for several years.

Soy sauce (*shōyū*), a fermented product made from soybeans and parched wheat, is a very salty dark-brown liquid used extensively both in the orient and the west as a seasoning. In China and Japan soybeans are boiled in syrup and eaten as candied beans and also soaked in salt water and roasted as a nut.

In the U.S. soybeans are used as a green vegetable and are preserved by canning or freezing. The mature or dry soybeans are baked or boiled in a manner similar to navy beans. For this purpose special vegetable varieties have been developed that are bland or nutty in flavour.

The soybean was first used in the U.S. as a forage crop and for green manure. When properly harvested and cured the soybean plant makes excellent hay. Soybeans are especially valuable for plowing under before planting certain crops such as sugar cane. The first commercial processing of soybean seed to obtain oil and cake was in 1911 when a small hydraulic-press mill in Seattle, Wash., processed a shipment of beans from Manchuria. In 1915 a cotton-oil mill in Elizabeth City, N.C., using screw-press equipment, made a successful start in processing locally grown beans. In 1917 a plant in Chicago Heights, Ill., began crushing experiments using both hydraulic and screw-press equipment, although the supply of soybeans was not adequate at that time to keep the mill in operation. A corn-processing company in Decatur, Ill., began crushing soybeans in 1922. Several other processing companies built soybean mills in Decatur in the late 1920s and 1930s, and the city soon became recognized as the leading soybean-processing centre in the U.S. It was not until 1935 that the acreage of beans for processing equaled that for forage but thereafter the volume of processing increased very rapidly.

Processing of the beans for oil and meal by hydraulic pressing or by expeller or continuous pressing (screw presses) left from 4 to 5% oil in the press cake. Most processing plants in the U.S. and Europe changed to hydrocarbon solvent extraction, which leaves around 1% oil in the meal. With modern extraction equipment a bushel (60 lb.) of soybeans produces about 11 lb. of crude oil and 47 lb. of meal.

Soybean oil contains from 80 to 85% unsaturated glycerides and is classed as a semidrying oil. With an iodine number (an index of drying rate) of 130 to 135 it lies between rapid-drying oils such as linseed and tung, and the nondrying oils such as peanut oil. Methods of refining have been developed that make soybean oil very good for edible products such as shortening, margarine and salad oils. The generally low fatty-acid content of soybean oil results in a low refining loss and high yield of refined oil. Cooking oils and salad oils may be prepared by refining, deodorizing, and bleaching soybean oil. By controlled hydrogenation that partly solidifies the oil, an excellent solid fat can be prepared for use in the manufacture of shortening and margarine. (*See* OILS, FATS AND WAXES.)

Crude soybean oil contains from 1 to 3% lecithin—an antioxidant, emulsifier, and softener used in food manufacturing and many other industries.

Soybean meal remaining after the oil has been extracted is a valuable product with many uses, particularly as a high-protein feed for livestock and poultry. In the second half of the 20th century about 95% of the soybean meal consumed in the U.S. was used for feed, supplying over 6,000,000 tons or 66% of the total vegetable protein meal consumed by livestock and poultry (*see* ENSILAGE; FEEDS, ANIMAL). Soybean flour, grits and flakes are used in bread, doughnuts, cakes and cookies; as emulsifiers and binders in sausages and related meat products; in breakfast foods, salad dressings, low-starch health foods, macaroni and noodles, confectionery products, ice creams and as whip toppings. Soybean flour proved very successful in meeting food shortages during emergencies. Cereal proteins are low in lysine content, and soybean protein supplies the deficiency. The addition of 5% soybean protein to wheat flour makes a remarkable improvement in nutritive value. It has been stated that soybean flour is one of the cheapest foods available to man when judged by the amount of protein, minerals, vitamins and energy obtainable per unit of cost. Soybean meal also is used in the brewing industry to provide nutrients for the growing yeast and to improve the flavour and body of beer. While whole soybeans are generally used in the preparation of soy sauce by fermentation, soybean meal is used by many manufacturers in its preparation by the more rapid acid hydrolysis method. Monosodium glutamate, an important condiment in the orient and in the U.S., may be prepared by acid hydrolysis of soybean meal. Glutamic acid is an essential constituent of soy sauce.

Uses in Industry.—Although food manufacture absorbs about 85% of the soybean oil consumed in the U.S., the oil has many applications in industry, one being in the production of high-grade industrial enamels, particularly white enamels for kitchen appliances and coloured enamels for automobile finishes. It is used extensively in the production of varnishes and alkyd resin paints and also is used in inks and stains, sealing and caulking compounds, linoleum and oilcloth, pharmaceuticals, cosmetics, synthetic rubber and core oils. The protein in soybean meal is readily soluble in water and when properly processed is similar to casein in viscosity, adhesive strength and many other characteristics. Soybean protein is used in many industrial applications such as adhesives, paper coatings, water-thinned paints, plastics, textile fibres, fire foam stabilizers, printing inks, fillers, core binders, stickers for agricultural sprays, linoleum, emulsifying agents, paper and textile sizing and leather finishes. In the plywood industry soybean glues represent the largest tonnage of any type of adhesive used. Industrial research is rapidly widening the industrial uses of the soybean and its products.

See also BEAN.

BIBLIOGRAPHY.—C. V. Piper and W. J. Morse, *The Soybean* (1923); K. S. Markley and W. G. Goss, *Soybean Chemistry and Technology* (1944); K. S. Markley (ed.), *Soybeans and Soybean Products*, 2 vol. (1950–51); A. G. Norman (ed.), *The Soybean* (1963). *See* also *Bulletins* of the U.S. Department of Agriculture and State Agricultural Experiment Stations. (J. L. CR.)

SOZOMEN (SALAMANES HERMEIAS SOZOMENOS) (fl. *c.* A.D. 425), church historian whose work supplements that of his contemporary Socrates, was born of a Christian family at Bethelia near Gaza and was brought up in Palestine, of which he shows considerable knowledge. Later, not before 425, he worked as a lawyer in Constantinople. He compiled a history of the church running parallel to that of Socrates (*q.v.*) and generally similar to it; in fact, it is largely based upon Socrates' work, though without acknowledgment, and presumably was written in a competitive spirit. In nine books, it is dedicated to the emperor Theodosius II, was begun in the 440s, and was intended to cover A.D. 324 to 439 but stops in 422 with occasional references to later events. It has been variously suggested that Sozomen died before completing it, that the conclusion has been lost, or that either he or the emperor suppressed the conclusion to avoid reference to the empress Eudocia, who had fallen from favour.

Sozomen normally follows Socrates closely, but he frequently consulted the documents used by Socrates and sometimes cites them more fully. He also drew upon additional sources, such as further works of Athanasius and, for the political detail of book ix, the collections of his contemporary the pagan historian Olympiodorus of Thebes. He is more favourable to monks (by whom he was perhaps educated) and gives more of their history, and he sides definitely with Chrysostom. Though Sozomen was somewhat less critical than Socrates, his history is a valuable supplement to, and sometimes corrects, the earlier one and, like it, was used by Theodore the Lector for his *Tripartite History* (early 6th cen-

tury). His style is better than that of Socrates, who was reacting sharply from the verbosity and artifice of their historian-predecessor Eusebius of Caesarea.

BIBLIOGRAPHY.—Edition by J. Bidez and G. C. Hansen (1960) in *Griechische christliche Schriftsteller;* H. Wace and P. Schaff (eds.), *Nicene and Post-Nicene Fathers*, Eng. trans. by C. D. Hartranft, 2nd series, vol. 2 (1891). *See* also G. Schoo, *Die Quellen des Kirchenhistorikers Sozomenos* (1911). (S. L. G.)

SPA, a town in the province of Liège, Belg., 21 mi. (34 km.) SE of Liège, famous for its springs of mineral water, known locally as *pouhons*, which have caused the name spa to be given to all such health resorts. The springs are said to be beneficial to sufferers from heart diseases. Pop. (1961) 9,055. The town became fashionable in the 18th century, when the coloured wood carvings (*bois de Spa*) were also popular. In 1807 much of the town was burned down, and the principal buildings, the casino and the baths, are modern. The surrounding Ardennes scenery is picturesque. In 1918 the emperor William II made his headquarters at Spa, and in 1920 the conference to discuss the terms of the Treaty of Versailles was held there (*see* SPA, CONFERENCE OF).

SPA, CONFERENCE OF, an important meeting of the Supreme Council of the Allied Powers in the aftermath of World War I. It took place at Spa, in Belgium, from July 5 to July 16, 1920. The problems of German disarmament and reparations and the Polish demand for Allied assistance against Soviet Russia were the main items on its agenda. Great Britain was represented by David Lloyd George (prime minister) and Lord Curzon (foreign secretary); France by Alexandre Millerand (prime minister); Italy by Count Carlo Sforza (foreign minister). They were joined by Léon Delacroix, the Belgian prime minister. Moreover, for the first time since the signature of the Treaty of Versailles, a German delegation was invited to present its views. This delegation comprised Konstantin Fehrenbach (chancellor of the *Reich*), Otto Karl Gessler (defense minister), Gen. Hans von Seeckt (commander of the *Reichswehr*), and others.

The Germans made a formal request that the figure of 100,000 men, which was the limit of the German Army allowed by the Treaty of Versailles, might continue to be exceeded, on the ground that it was impossible for the government to keep order with such a small force. Lloyd George explained the Allies' anxiety over the fact that Germany still had an army of 200,000 men, with 50,000 machine guns and 12,000 guns. It was decided that Germany should reduce the strength of the *Reichswehr* to the treaty figure of 100,000 by Jan. 1, 1921, and withdraw arms from semimilitary formations. On July 9 the Germans signed an agreement embodying these stipulations.

The problem of the reparations (*q.v.*) due from Germany under the treaty remained unsolved, but on July 16 Germany agreed to deliver 2,000,000 tons of coal monthly to France, Italy, and Belgium by way of reparation. The Allies agreed among themselves as to the distribution of the still uncertain total of German reparations: France was to receive 52%, the British Empire 22%, Italy 10%, Belgium 8%, and Serbia 5%.

The question of "war criminals" was allowed to lapse by the conference. As the Netherlands government had on Jan. 23 refused to deliver the ex-emperor William II to the Allies, they decided not to ask the German government to hand other accused persons over to them. Instead, they asked the Germans themselves to proceed with trials and punishment; but the German government was unwilling to do so.

Since the Russo-Polish War (*q.v.*) was going badly for Poland, Wladyslaw Grabski, the Polish prime minister, went to Spa on July 6, 1920, to appeal urgently for moral and material assistance. The Allies decided to intervene diplomatically in Moscow, the British government taking the initiative. Lloyd George promised to send a note to the Soviet government requesting an armistice along a line described later as the Curzon Line (*see* POLAND: *History; Independence Restored*); but at the same time he pressed Poland (1) to renounce Lwow (Lvov) and the whole of eastern Galicia, (2) to leave Wilno (Vilnius) to Lithuania, (3) to agree to a partition of Teschen Silesia by the Allies without plebiscite, and (4) to accept, in the Free City of Danzig, a Harbour and Waterways Board, which had not been stipulated by the Treaty of Versailles. On July 10 Grabski signed an agreement embodying these conditions; and on July 11 Curzon started a telegraphic correspondence with G. V. Chicherin, the Soviet people's commissar for foreign affairs. Grabski left Spa convinced that at least arms and munitions in considerable numbers would be sent to Poland, but in fact no such help arrived before the Poles themselves had won their victory. (K. SM.)

SPAAK, PAUL HENRI (1899–), the foremost Belgian statesman of the decades following World War II, a champion of the North Atlantic Treaty Organization (NATO) and of European cooperation. He was born at Schaerbeek (Greater Brussels) on Jan. 25, 1899, the son of the poet and dramatist Paul Spaak and his wife Marie (née Jeanson). In 1915, during World War I, he tried to join the Belgian Army on the Yser River but was arrested by the Germans. Released at the end of the war, he studied law in Brussels and was called to the bar there.

Elected to the Chamber of Deputies as a Socialist in 1932, Spaak became minister of transport, posts, and telegraphs (March 1935–May 1936) and then foreign minister (June 1936–May 1938) in Paul van Zeeland's cabinets. As foreign minister he conducted the negotiations to release Belgium from the Pact of Locarno (*q.v.*), believing that Belgium could remain neutral in the event of war between Germany and the Western Powers. He became Belgium's first Socialist prime minister in May 1938 but resigned in February 1939. Foreign minister again from September 1939, in Hubert Pierlot's coalition, he went into exile with the government in May 1940 after the Belgian Army's capitulation to the Germans during World War II. In London, he laid the foundations of Benelux (*q.v.*) in September 1944. He returned to Brussels that same month and remained foreign minister in the coalition governments of 1945–47 under Achille Van Acker and Camille Huysmans. On Jan. 10, 1946, Spaak was elected president of the first session of the General Assembly of the United Nations.

From March 1947 to August 1949, Spaak was prime minister, with a Christian Social-Socialist coalition: he brought Benelux into operation (1948) and signed the North Atlantic Treaty (1949). In the summer of 1950 he took an active part in the political outcry that led to the abdication of King Leopold III a year later.

Spaak was elected president of the Council of Europe's Consultative Assembly in May 1951 and president of the European Coal and Steel Community's General Assembly in September 1952. He was Belgian foreign minister again in Van Acker's Socialist-Liberal coalition from April 1954 to May 1957, when he resigned in order to work as chairman of the Atlantic Council and secretary-general of NATO. He returned to Belgian politics, as leader of the Socialist Party, in 1961 and became deputy prime minister and foreign minister in a Christian Social-Socialist coalition. He opposed the transformation of the European Economic Community into a political one before the admission of Great Britain to it.

SPAATZ, CARL (1891–), U.S. Air Force officer, the leading U.S. combat air commander of World War II and first chief of staff of the independent United States Air Force, was born at Boyertown, Pa., on June 28, 1891. After a distinguished career as a combat pilot during World War I, Spaatz acquired extensive staff and command experience between 1919 and 1942. As commander of the 8th Air Force in England during World War II, he initiated the American daylight bombing offensive against Germany in 1942. In January 1944 he became commander of the U.S. Strategic Air Forces in Europe and directed the daylight strategic bombing of Germany from both England and Italy until the end of the war in Europe. He moved to the Pacific in July 1945 in time to direct the final strategic bombing of Japan, which culminated in the dropping of atomic bombs on Hiroshima and Nagasaki. Spaatz became chief of staff of the newly independent Air Force in September 1947, but, finding the administrative burden not to his liking, he retired in 1948. (A. GG.)

SPACE EXPLORATION. Space is not empty—not even those regions that lie between the stars and the planets. Throughout the vast reaches of space, matter (largely hydrogen) is scattered at the extremely low density of perhaps one atom per cubic centimetre in interstellar space and about ten particles per cubic

TABLE I.—*Regions of Space: Earth's Atmosphere and Near Space*

Distance from Earth's centre (Earth radii)	Altitude (km.)	Environmental characteristics						Operational aspects	
		Dynamic	Chemical	Electrical	Thermal	Meteors	Gravitational	Technical	Medical
1–1.0024	0–15	Winds Turbulence	20.9% O_2, 78% N_2 0.9% Ar, 0.03% CO_2, variable amount H_2O (volume %)	Thunderstorms	Average temperature gradient about −6.5° C./km.	Very rare	Terrestrial	Air-breathing engines	Ceiling for use of an oxygen mask: 3 km.: O_2 required 7.5 km.: ceiling for operational crew (U.S. air force) 11 km.: ceiling for all personnel (U.S. navy) 12 km.: ceiling for passengers at rest (U.S. air force) 13 km.: ceiling for pressure breathing (U.S. navy)
1.0024–1.0048	15–30	Lower region: jet streams	Ozone appears; maximum concentration at about 25 km.	None	Approximately isothermal atmosphere about −50° C.	Very rare	Terrestrial	30 km. is limit for most air breathing engines; hypersonic ram jets may extend beyond Slow-down region of ICBM nose-[illegible]	10–25 km · pressurized cabin 25 km : sealed cabin; beginning of space equivalence
1.0048–1.0126	30–80	Turbulence in region of negative temperature gradient	Above 50 km., atomic oxygen appears	D layer (daytime) NO^+ predominant ion	30 to 50 km.: +2 to 3° C /km. 50 to 80 km.: −1.5 to −4° C./km. Temp. at 75 km.: −73° C.	Most meteors disappear Accumulation of meteoritic dust	Terrestrial	Slip-flow region Slow down region of entry gliders from space Radicals for propulsion	Space equivalence Aerodynamic heating Cosmic radiation (maximum atmospheric secondaries)
1.0126–1.063	80–400	Hydromagnetic effects on dynamics Pressure about 10^{-6} atmosphere at 100 km.	O_2 dissociated Slight N_2 dissociation Airglow processes Aurorae	E, F_1, F_2 layers O^+, NO^+, O_2^+ major ions present	Temperature increases rapidly with altitude 100 km.: −53° C. 150 km.: +672° C. 300 km.: +1,170° C.	Meteors visible below about 100 km.	Terrestrial Perturbation of satellite orbits due to Earth's oblateness	Free molecule flow region Lower satellite orbit region; departure orbit for lunar and interplanetary probes	Space equivalence
1.063–1.157	400–1,000	Pressure about 10^{-12} atmosphere at 500 km. Lower edge of magnetosphere over equator	Helium becomes significant	O^+ ion predominates He^+ present Lowest portion of Van Allen radiation belts	Temperature becomes isothermal with altitude, but is subject to large diurnal variations	Micrometeorites and meteorites	Same as above	Manned space stations Instrumented satellites	Space conditions Corpuscular radiation
1.157–1.5	1,000–3,200	Some particles escape from Earth's atmosphere Magnetosphere	Hydrogen is predominant	H^+ present Charged particles travel along magnetic force lines Van Allen radiation belts	Diurnal effects predominate	Micrometeorites and meteorites	Same as above	Instrumented satellites	Space conditions Increasingly intense corpuscular radiation

centimetre in interplanetary space. Space is permeated by gravitational fields, traversed by an extremely broad spectrum of electromagnetic radiation and by cosmic rays, and is occupied by magnetic fields of unknown intensities and distributions. Although the scientist has deduced much about space from observations made at the Earth's surface, only with the advent of the artificial satellite and the deep-space probe has he been able to discover at first hand the great complexity of space phenomena.

This article is concerned with the exploration of space by those methods made possible by the large modern rocket. For the purposes of this discussion, the term space should be regarded generally as meaning those reaches of the universe lying beyond the Earth's sensible atmosphere. This definition of space is imprecise, but efforts to sharpen it become extremely complicated. For the biologist outer space means one thing, for the legal expert another, and for the astronomer something still different. The lawyer, who is concerned with the question of where existing international agreements concerning aviation should be considered to apply and where new agreements concerning astronautics should be developed, has an extremely difficult task when he attempts to develop a precise definition of space. Many of the applications of space technology apply primarily to activities on the surface of the Earth—for example, in the making of weather observations, in communications and in navigation. Even in the area of scientific exploration it becomes impossible for the scientist to separate the study of near space from the science of the Earth's atmosphere and ionosphere, and satellite investigations of the Earth's gravitational field by means of artificial satellites lead to important results in geodesy and in the study of the internal structure of the Earth. Tables I and II list some of the properties and characteristics of various regions of space.

The rocket in one form or another furnishes the only practicable means for transportation through space. Late in the 19th century the Russian mathematician Konstantin Tsiolkovski called attention to the usefulness of rockets for this purpose. Robert Goddard, a U.S. physicist, was motivated in his pioneering work on rockets, during the 1920s and 1930s, by the desire to send research instruments into the Earth's upper atmosphere. A German group, led by Hermann Oberth in the 1920s, looked to interplanetary travel as one of its principal objectives.

As these early pioneers clearly foresaw, space activities did not, indeed could not, come into being until the rocket had been further developed. Although the beginning of the space age is associated with the launching of the first Soviet Sputnik on Oct. 4, 1957, the groundwork for this event was laid by the high-altitude rocket exploration of the preceding decade. Beginning in 1945 with the launching of a WAC Corporal rocket, the United States undertook a series of scientific experiments in the Earth's upper atmosphere. For some years captured German V-2 rockets furnished the principal means of carrying on this upper-air research. With the development of the Aerobee, Viking and other rockets more suited to upper-atmosphere exploration, the work with sounding rockets accelerated, reaching a peak of activity during the International Geophysical Year (1957–58). Upper-air re-

search in the Soviet Union followed a similar chronology.

In both the U.S. and the Soviet Union the development of missile technology was essential to the larger satellite and space probe applications. As preparations for the International Geophysical Year (IGY) progressed, the possibility of launching artificial Earth satellites for scientific investigations became more and more apparent. As a result, the planning committee for the IGY in 1954 passed a formal resolution calling attention to the desirability of using artificial satellites in the IGY program. Both the United States and the Soviet Union responded with announcements that they would prepare scientific satellites for launching during the IGY. The first Soviet Sputniks and the early Explorer and Vanguard satellites of the U.S. were the direct outcome.

Although the possibilities offered by artificial satellites had been apparent to scientists and engineers working in these fields, general recognition of their potentialities did not take place until the first satellite was in orbit. After the launching of the first satellites, the U.S. created the National Aeronautics and Space administration (NASA) by the Space act of 1958. Under this act NASA took over the personnel, facilities and aeronautical responsibilities of the former National Advisory Committee for Aeronautics and also assumed responsibility for the civilian space activities. NASA began functioning on Oct. 1, 1958.

Thereafter, U.S. space activities expanded tremendously. The space budget grew rapidly from its initial total of $331,000,000 in 1959 to about $5,000,000,000 by the mid-1960s. The program encompassed research efforts in aeronautics, space science, space applications, manned space flight and research on advanced technology. In addition, the U.S. department of defense developed a program for the application of space techniques to defense uses, and by the mid-1960s the budget for these purposes was about $1,600,000,000. Also, the Atomic Energy commission was cooperating with NASA in the development of nuclear techniques for space, while the department of commerce was playing an ever-increasing role in the use of satellites in meteorology and geodesy.

This article is divided into the following sections:

In reading these sections the following definitions and concepts will be useful. Rockets used to launch spacecraft into orbit are referred to as space vehicles or launch vehicles. Most of them are multi-stage vehicles. It is customary to refer to the first stage as the booster stage and to the remaining parts of the rocket as upper stages. The term spacecraft is used in a general sense to describe the assemblage and its payload that is lifted into space and injected into an orbit by a launch vehicle. If the launching rocket with its payload follows a vertical or nearly vertical trajectory, and remains within one Earth radius of the ground, then the combination is normally referred to as a sounding rocket. When the spacecraft is injected into a closed orbit about the Earth, it becomes an artificial satellite of the Earth. The orbit is described by four factors: apogee and perigee, the points farthest from and nearest to Earth, given in miles or kilometres; the inclination of the plane of the orbit to the plane of the Earth's equator, given in degrees; and the period, or time required for one circuit of the orbit. Spacecraft that are injected into trajectories leading to escape from the Earth are called space probes. These may be classified as lunar, Venus, Mars, deep-space, etc.

The sounding rocket, artificial satellite and deep-space probe provide the scientist with a powerful means for research in a number of areas. Among the important subjects of investigation by space techniques are: the space environment; Sun-Earth relations; investigation of the Moon; investigation of the planets; determination of astronomical constants; investigation into the nature of the universe; and life in space.

Each of these areas can be investigated along lines that were impossible prior to the advent of the rocket. Of particular importance is the opportunity in astronomy to observe celestial objects and phenomena in the gamma ray, X-ray, ultraviolet, infrared and radio wavelengths that cannot penetrate the Earth's atmosphere to reach observatories on the ground. Especially exciting is the possibility that space biology may contribute significantly to efforts to develop a comprehensive biological theory. This collection of researches carried out in space or by means of space techniques is referred to as space science.

The scientific applications and technological advances made during the first five years of the space age were truly remarkable, but perhaps the most important aspect of this early phase of the space age was the establishment of a new capability in human endeavour. Out of these early activities man could develop the knowledge and techniques; trained and experienced personnel; assembly, launch, tracking and telemetering facilities; industrial complexes and teams; flight "hardware," and the ability to operate in space that go to make up a total space capability. With this capability, man can extend his domain beyond the earth as he once did over the land and the seas, into the air and into the ocean depths. (H. E. NL.)

I. ASTRODYNAMICS

Astrodynamics is the branch of astronautics that deals with the mathematical theory of the motions of spacecraft. These motions may be separated into two groups: the translations of the centre of mass of the spacecraft, and the rotations of the spacecraft about its centre of mass. The translations describe the flight path or trajectory or orbit of the spacecraft; the rotations describe its wobbling or angular motion.

In each of the two groups there are three motions. The translational motions are (1) surge, a motion generally along the velocity vector or direction of travel of the spacecraft; (2) heave, a motion perpendicular to the surge and generally in the direction of the gravity vector; and (3) swerve, a motion perpendicular to the surge and the heave. The rotational motions are (1) roll, a spinning motion around the velocity vector of the spacecraft, *i.e.*, spin about the surge direction; (2) pitch, an angular motion about the swerve direction; and (3) yaw, an angular motion about the heave direction. Thus, a spacecraft has a total of six motions which must be completely accounted for if its performance is to be known accurately, or if its position and attitude in space are critical to accomplishing its basic mission.

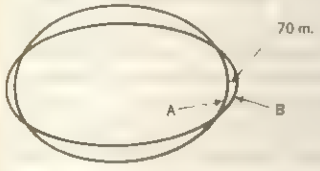


FIG. 1.—EQUATORIAL BULGE IS 21.4 KM. DISCREPANCY BETWEEN OBSERVED SHAPE (A) AND SHAPE OF HYDROSTATIC EQUILIBRIUM (B) IS 70 M.

The theory and applied mathematical methods of astrodynamics depend heavily on the methods of celestial mechanics, exterior ballistics and the flight dynamics of aircraft and missiles.

A. Two-Body Motion

Sir Isaac Newton stated that a sphere, such as the Earth, which is homogeneous in concentric layers, attracts an exterior point mass, such as a spacecraft, as if all of the mass of the sphere were concentrated at its centre. Furthermore, his universal law of gravitation states that any two particles of mass in the universe attract each other with a force that varies directly as the product of their masses and inversely as the square of the distance between them. Thus, a statement of a problem concerning the translational motions of a spacecraft takes the form of a set of equations of motion, which are ordinary differential equations of the second order. A proper understanding of astrodynamics, therefore, requires a knowledge of this branch of mathematics.

For the case of two bodies, such as a spacecraft and a simple spherical Earth, a solution to the differential equations of motion is known. Two types of translational motion are possible. The first is an elliptical or circular orbit of the spacecraft around the Earth. The second is a flight path away from the Earth, a path that may be hyperbolic or parabolic in shape.

While these simple solutions are quite useful in many cases, they are not correct, strictly speaking, since the Earth is not a simple spherical body. It is known, for example, that the Earth has a bulge at its equator. (*See* fig. 1.) This bulge, although amounting to only 21.4 km., seriously affects the motion of a spacecraft. For example, a spacecraft in a nominally circular polar orbit experiences a greater attraction as it passes over the equator and thus it is pulled toward the Earth. Also, its velocity is increased. When the spacecraft passes over the poles, it feels a lesser attraction and therefore it moves out farther from the Earth and its velocity decreases. The bulge at the equator, therefore, has caused a surging and heaving motion of the spacecraft as well as a variation in its altitude. Also, for a spacecraft in a 500-mi. orbit, the perigee altitude variation is approximately 5 mi. in two months.

For a spacecraft in an inclined orbit the bulge also causes the major axis of the orbit to rotate in the orbital plane and the orbital plane to rotate in space about the Earth's spin axis. The size of these two rotations depends upon the exact inclination of the orbital plane to the equator of the Earth. At an inclination of 63.4° there is no rotation of the major axis. For angles less than 63.4° the major axis rotates in the same direction as the satellite motion around the earth. For angles greater than 63.4° the major axis rotates in the opposite sense. There is no rotation

TABLE II.—*Regions of Space: Orbital Space*

Region	Distance*	Environmental conditions	Astronautic aspects
Terrestrial space (Ehricke) Magnetosphere	1–10 r⊕	Atmospheric effects on orbit in inner terrestrial space. Electrical and magnetic effects caused by Earth; especially the high-intensity radiation belt with maximum at 1.5 and 3.5 r⊕, outer limit 10–15 r⊕ on sun side Gravitational dominance of terrestrial perturbations over lunar perturbations	Manned space stations appear safe in region between relevant atmosphere (450 km.) and beginning of intense corpuscular radiation (650 km.); or beyond radiation belt (>15 r⊕) Instrumented satellites: meteorological: 1.1–6.5 r⊕ observational: 3-6.5 r⊕ (6.5 r⊕ = 24 hr. orbit) communication: 3-6.5 r⊕
Cislunar space (Ehricke) Magnetosphere (on antisolar side of Earth)	10–50 r⊕	Equivalence and finally dominance of lunar perturbations over terrestrial perturbations. Gravitational pull of Moon alternately joins and opposes Earth's pull Solar perturbation may be effective on cislunar transfer orbits Tail of Earth's magnetosphere on antisolar side	Essentially research satellites to explore environmental conditions in cislunar space Region is traversed by lunar and interplanetary vehicles Region of chemical and nuclear heat-exchanger rocket drives
Translunar space (Ehricke)	70 to ~300 r⊕	Outer region of geocentric space Lunar and terrestrial gravitational forces never act in opposite direction	No firm technical concepts yet on utilization of translunar space
Region in which the Moon can hold a satellite	40–45 r☾	Lunar gravity field dominant Special environmental conditions near the Moon, not yet known in detail	Lunar satellites Lunar base
Inner solar system	0–1.6 A.U.†	Central force field of the Sun Intense radiation Zodiacal dust cloud	Region of nuclear heat-exchanger drives for fast missions, low-thrust drives for large payload missions
Ecosphere (Strughold)	~0.5–1.5 A.U.†	Liquid water belt Oxygen belt Biotemperature belt	Natural biotechnical region of solar system
Outer solar system	1.6–38 A.U.†	Central force field of the Sun Strong perturbation potential of major planets	Low-thrust drives (electrostatic or electrodynamic)
Region in which respective planet can hold a satellite	Generally greater than 150 planet radii (except Mercury)	Gravitationally, a local anomaly in heliocentric space Probably special electrical, magnetic and radiation conditions not yet known in detail	Region in which artificial satellites can be established
Region in which the planet rather than the Sun should be considered as centre body for an arriving or departing body	Venus: ~100 r♀ Earth: ~144 r⊕ Moon: ~34 r☾ Mars: 170 r♂ Jupiter: 690 r♃	Gravitationally defined as the region in which the Sun is a lesser perturbative force in the planetocentric co-ordinate system than the planet is in the heliocentric co-ordinate system. A co-ordinate system is the frame of reference in which the orbit of the spacecraft (or comet, etc.) is measured. Its centre is in the centre of the planet or the Sun, respectively, for the two above mentioned co-ordinate systems	Region in which transfer from heliocentric to planetocentric orbits (capture) or the reverse (escape) takes place. The transfer orbit between a closed planetocentric (satellite) orbit and a heliocentric (elliptic) orbit is a hyperbola. The limit of the activity sphere is reached when the hyperbolic orbit merges with the heliocentric orbit

*r⊕ = Earth radius; r☾ = Moon radius; r♀ = Venus radius; r♂ = Mars radius; r♃ = Jupiter radius; each measured from the centre of the respective planet.
†1 A.U. = 1 astronomical unit = mean distance from sun to earth, measured from the centre of the sun.

or precession of the orbital plane when the inclination is 90° (polar orbit). Maximum precession, however, occurs for an equatorial orbit.

1. Effect of Asymmetrical Bodies.—Previous work in celestial mechanics on the motion of the innermost satellite of Jupiter and the innermost satellite of Saturn has been helpful to the astrodynamicist in understanding the effect of the bulge on the orbits of near-Earth spacecraft. By actually observing the variations in the orbits of near-Earth satellites the astrodynamicist has discovered that the Earth has, in fact, an extremely complicated shape. Not only is there a lack of symmetry between the northern and southern hemispheres but there is also an asymmetrical distribution of mass about the Earth's spin axis. The latter means that the equator of the Earth is not circular but is elliptical in shape (approximately 100 m.). It is therefore necessary to represent the gravity field of the earth in terms of a potential expressed in spherical harmonics, which include the zonal harmonics (wrinkles in latitude), the sectorial harmonics (wrinkles in longitude), and the tesseral harmonics (wrinkles in both latitude and longitude). This potential may be expressed as

$$U = -\frac{GM}{r}\sum_{n=0}^{\infty}\sum_{m=0}^{n}\left(\frac{a_e}{r}\right)^n J_{nm}\, P_{nm}\,(\sin\phi)\,\cos m\,(\lambda - \lambda_{nm})$$

where

$$P_n(\sin\phi) = \frac{1}{2^n\, n!}\,\frac{d^n\,(\sin^2\phi - 1)^n}{d\,(\sin\phi)^n}$$

$$P_{nm}(\sin\phi) = \cos^m\phi\,\frac{d^m\,P_n(\sin\phi)}{d\,(\sin\phi)^m}$$

In these equations, G = universal gravitational constant; M = mass of Earth; r = distance from origin of co-ordinates to computation point; a_e = equatorial radius of Earth; ϕ = geocentric latitude; and λ = geocentric longitude.

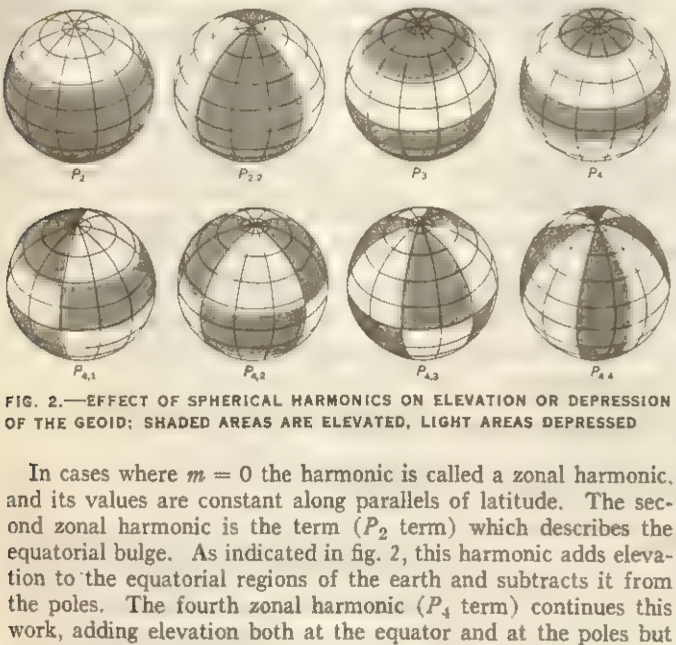


FIG. 2.—EFFECT OF SPHERICAL HARMONICS ON ELEVATION OR DEPRESSION OF THE GEOID: SHADED AREAS ARE ELEVATED, LIGHT AREAS DEPRESSED

In cases where $m = 0$ the harmonic is called a zonal harmonic, and its values are constant along parallels of latitude. The second zonal harmonic is the term (P_2 term) which describes the equatorial bulge. As indicated in fig. 2, this harmonic adds elevation to the equatorial regions of the earth and subtracts it from the poles. The fourth zonal harmonic (P_4 term) continues this work, adding elevation both at the equator and at the poles but slightly depressing the middle latitudes. The proper combination of the second and fourth zonal harmonics will produce a figure that coincides with an ellipsoid within 1 m. Almost all common representations of the earth's figure consider only the second and fourth zonal harmonics. Higher-order even zonal harmonics add to the wrinkles on the earth, remaining symmetric with respect to the equator. The third order zonal harmonic (P_3 term) is the so-called pear-shape term. This term, as with any odd order harmonic, has an opposite effect in the northern and southern hemispheres.

Where $m = n$ the harmonic is called a sectorial harmonic and has values of consistent sign along lines of constant longitude; however, these values change in magnitude as the latitude changes. In addition, for any given latitude the value varies in magnitude and direction as the longitude changes. The $P_{2,2}$ term is the major term indicating that the earth does not have rotational symmetry, and is referred to as the elliptic equator term.

In those cases where $m \neq 0$ and $m \neq n$, the harmonic is called a tesseral harmonic and maintains a value of consistent sign within a tessera, changing sign and magnitude as functions of both latitude and longitude. A tessera is a curvilinear rectangle, in this case bounded by lines of constant longitude and latitude.

Although a satellite may be tracked to an accuracy of 0.1 mi., the lack of knowledge of the gravity field permits prediction of position to an accuracy of only 0.25 mi. per 6 hr. or 1 mi. per day. Thus, even for two relatively simple bodies like a spacecraft and the Earth, the translational motions are complicated because of the complexity of the gravitational force causing them. This same situation also exists for other bodies in space. The Moon, to take an example, is known to be lopsided, and therefore a spacecraft orbiting the Moon will experience similar variations in its orbit.

2. Additional Forces.—In addition to the gravitational force, a spacecraft in its flight around the Earth experiences a drag force due to impact with molecules of air. Furthermore, radiation pressure and a particle wind, originating in the Sun, exert some pressure on the spacecraft and affect the orbit. Additional forces can arise due to the magnetic field of the Earth, as its lines of force are cut by the satellite. All of these forces also must be included in the differential equations for the motion of a spacecraft.

The basic effects of some of these forces are known. The air drag, for example, causes an elliptic orbit eventually to become circular and a circular orbit to decrease in size, with the result that the satellite reenters the atmosphere and either burns up or strikes the Earth. Solar radiation pressures can cause large changes in the altitude of perigee.

B. Three-Body Problem

The simple two-body theory when applied by Newton to the motion of the planets about the Sun verified the excellent observations of Tycho Brahe and the generalizations of Johannes Kepler, except that small variations or perturbations in the orbits were noted. These perturbations in the orbit of one planet about the Sun occurred whenever another planet came near. In a similar manner the orbit of the Moon about the Earth demonstrated appreciable perturbations due to the gravity field of the Sun. It was clear, therefore, that all bodies in space influence all other bodies to some extent, and thus many complicated gravity fields must be considered simultaneously. Even for the relatively simple case of only three bodies, there is no general solution for the equations of motion (*see* Celestial Mechanics).

The situation for a spacecraft is particularly complicated since during some portions of its flight it may be under the strong influence of different bodies and during other portions under weak but simultaneous influences of many bodies. Although the various external forces and gravity fields acting upon a spacecraft may be known approximately and the appropriate equations for the translational motion of the satellite may be set down, no general solutions to the equations are available. Thus the translational motions of the spacecraft would be unknown, were it not for the development of the high-speed computer. The application of the computer by the ballistics scientist to solve projectile trajectory problems has provided the astrodynamicist with an extremely valuable tool that enables him to integrate numerically the complicated equations of motion and, therefore, to arrive at a knowledge of the complete translational motions of a spacecraft. This direct numerical integration is called the special perturbations technique.

If one has a considerable knowledge of the position of the spacecraft as a function of time, it is possible to use the high-speed computer to find special curves which can best represent the observed motions. This process is known as the general perturbations technique.

C. Spacecraft Rotations

Spacecraft injected into orbit without rotation and subject to no external torques will not rotate; the roll, pitch and yaw will all be zero. For spacecraft injected with rotation, the total angular momentum remains a constant if no external torques are present. This characteristic of conservation of angular momentum for spacecraft is fundamental to understanding their rotational performance.

A rotating spacecraft is often subject to small external damping torques, such as magnetic torques arising from the magnetic field of the Earth and air damping torques. There may also be small internal damping torques arising from the movement of components, such as wires, fluids and other masses. The effect of these torques is to cause the spacecraft to change its axis of rotation until it is finally rotating about its axis of maximum moment of inertia. The angular momentum is unchanged when only internal damping is involved. Therefore, spacecraft which are spin-stabilized are designed not as elongated bodies like classical projectiles but rather as somewhat stubby shapes that will fly stably about their initial rolling axis.

Often it is necessary to control actively the angular motion of spacecraft. Again, the fundamental underlying theory is the conservation of angular momentum. The primary control devices are generally small jets to provide torques about the three spacecraft axes. Sometimes internal inertia wheels are used to change the distribution of the angular momentum about each axis.

A special case of the passive control of the angular position of a spacecraft is the use of the gravity gradient torque. This technique is based on the fact that an elongated body in a near-Earth orbit experiences a significant torque due to the difference in the gravity force acting on different parts of spacecraft located at different distances from the centre of the Earth. This torque causes the spacecraft to point its long axis toward the Earth. If, in addition, there exists some damping torque, the spacecraft will be stably Earth-oriented. The gravity stabilization technique is quite reliable because of its simplicity, and there is no limit to its lifetime. It is of special interest for spacecraft used in communications, meteorology, reconnaissance and mapping of the Earth and other planets.

D. Applications

Because the translational motions of a near-Earth satellite are greatly influenced by the complex gravity field of the Earth, and also because there is a need to know the exact locations of satellite tracking stations with respect to each other and with respect to the centre of mass of the Earth, it was recognized that a satellite could be used to determine these quantities. This knowledge, in turn, would make possible both a better understanding of the translational motions of a satellite and an accurate prediction of the orbits of future satellites. Thus, a special geodetic satellite,

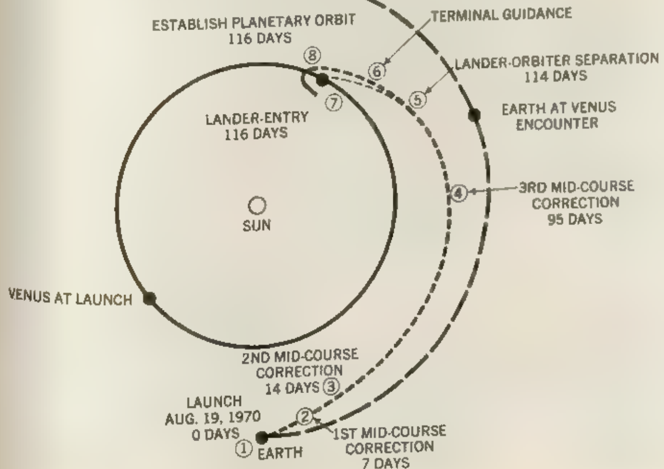


FIG. 3.—TYPICAL TRAJECTORY TO VENUS WITH MANEUVERING POINTS INDICATED

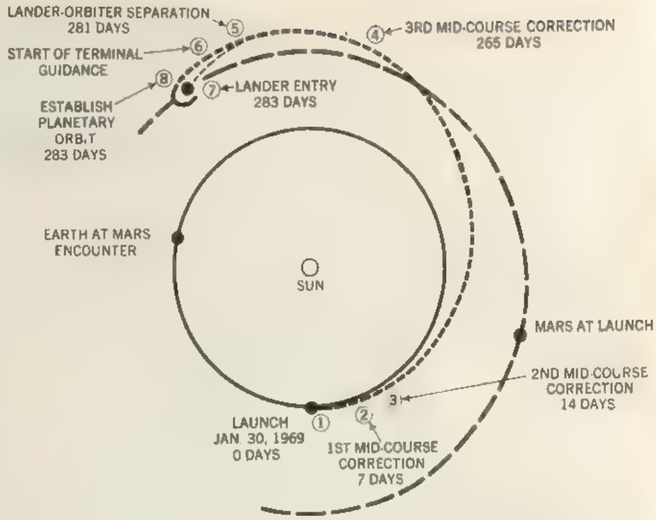


FIG. 4.—TYPICAL TRAJECTORY TO MARS WITH MANEUVERING POINTS INDICATED

ANNA (army, navy, NASA and air force), was conceived and placed into orbit. Various precision tracking techniques, such as optical tracking, doppler tracking, transponder tracking, interferometric tracking and laser tracking, were used to determine: (1) the gravity field of the Earth; and (2) the location of the various tracking stations in geocentric co-ordinates. This was a case, therefore, of a special satellite being used to determine basic astrodynamical information which in turn could be employed to enable a better prediction of spacecraft motion.

Not only is the gravity field of the Earth of interest but also there is considerable interest in determining the gravity fields of the Moon and other planets to enable accurate satellite flight operations in their vicinities. Consider, for example, the motion of a spacecraft traveling from the Earth past Venus. Relative to the Earth the spacecraft travels on a hyperbolic orbit and initially the Earth's gravitational field predominates. However, as the spacecraft moves outward, the Moon's gravitational field becomes dominant over that of the Earth. After passing beyond the Earth-Moon system, the influence of the gravity field of the Sun becomes the primary controlling factor. During its long voyage to Venus (approximately 100 days), the spacecraft follows an elliptical orbit relative to the Sun, subject to small perturbations produced by other planets in the solar system. However, as the spacecraft approaches Venus the gravity field of Venus takes on major importance. Although there is no solution for the translational motions of a spacecraft on such a voyage, the equations of motion may be numerically integrated on a computer and thus the flight path may be obtained (*see* fig. 3 and 4).

In addition, it is necessary to consider the launching energy required to reach a planet when a favourable opportunity is available. As indicated in fig. 5, the favourable opportunities for Mars occur about every 26 months and for Venus about every 19 months. The launching energy, however, is different at each opportunity because of the difference in the geometrical situation. Even at a favourable opportunity there is only a short period (called a launch window) when the energy remains within feasible bounds. It has been shown mathematically that the lowest energy requirement results when it is possible to use what is known as a Hohmann transfer flight path. This simple space maneuver is illustrated in fig. 6. To transfer from the smaller circular orbit to the larger, sufficient velocity is first added in a direction along the orbit to produce a new elliptical orbit with an apogee equal to the radius of the larger orbit. After the satellite has passed along this elliptical orbit and has reached the larger orbit, it is again necessary to add velocity along the orbit to produce the circular orbit at that altitude.

There are also minimum energy transfers for non-coplanar orbits. However, the additional energies are generally quite large. Some of the energy limitations in launching out of the

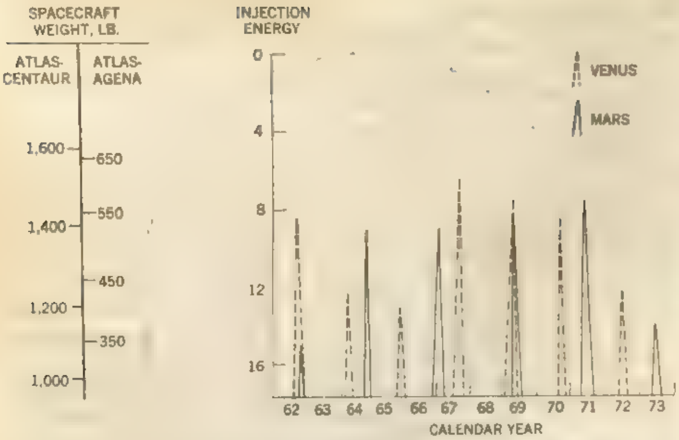


FIG. 5.—PLANETARY ENERGY REQUIREMENTS. RELATIVE ENERGIES IN EXCESS OF ESCAPE ENERGY ARE SHOWN FOR MARS AND VENUS FLIGHTS AT MOST OPPORTUNE TIMES

ecliptic plane are shown in fig. 7. The rays represent inclination to ecliptic plane. The circles represent distance from the Sun in astronomical units (A.U.). For example, a 250-lb. spacecraft launched in the ecliptic plane (0°) can almost reach the Sun (0.08 A.U.). However, if this same spacecraft is launched 37° out of the ecliptic plane it only travels halfway to the Sun.

A voyage of a spacecraft represents a significant breakthrough in the science of astrodynamics. Heretofore, astronomers had only the planets and their motions as a basis for understanding the various gravity fields and the fundamental distances in the solar system. Now man is able to use the spacecraft and their motions in conjunction with those of the planets to obtain a better knowledge of many of the physical constants governing the motion of the planets. For example, a spacecraft flying past the Moon can be used to evaluate both the distance to the Moon and the lunar mass. Similarly, a spacecraft passing near Venus gives trajectory data that can be used to obtain an entirely new estimate of the mass of Venus and the distance from the Earth to Venus. This information, in turn, enables us to obtain a new value for the astronomical unit, which is the fundamental factor for understanding all distances in the solar system and beyond. As a result of the flight of Mariner 2 to Venus and radar observations of the planet, the astronomical unit is estimated to be 149,599,244 km., accurate to within ± 278 km. (*see* Table III).

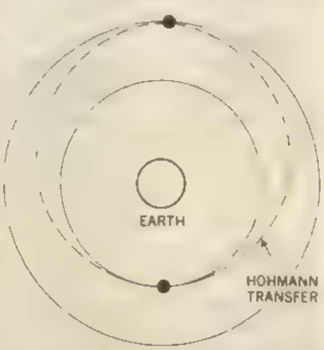


FIG. 6.—SIMPLE SPACE MANEUVER, THE HOHMANN TRANSFER, IN WHICH AN ELLIPTICAL TRANSFER PATH IS EMPLOYED, WITH ENERGY IMPULSES AT PERIGEE AND APOGEE

1. Angular Motion and Controls.—Some examples of the angular motion of spacecraft are as follows: In the case of the Transit navigation satellite it was desired that the satellite not spin in its orbit. However, in order to launch the satellite it was necessary to spin-stabilize the last stage of the launch vehicle. As a result the satellite was placed into orbit spinning. To stop the spin small weights were placed at the satellite's equator. When released, these weights payed out on strings and applied a counterrotating torque to the satellite. In this way most of the angular momentum of the spacecraft was transferred to the weights. The weights were then released into space carrying their share of the spacecraft's angular momentum with them, and the satellite was left with a very small spin.

TABLE III.—*Values for the Astronomical Unit*

Mariner 2	149,599,244 ± 278 km.
Goldstone*	149,598,845 ± 250
Millstone*	149,597,850 ± 400
Jodrell Bank*	149,599,800 ± 5,000
U.S.S.R.*	149,599,500 ± 800

*Radar determinations

The specifications for the Tiros weather satellite called for a spacecraft which was spin-stabilized and from which pictures of the clouds could be taken, but which would not pitch or yaw. Therefore, internal viscous damping was provided; the viscous torque caused the wobbling motion to damp out and gave a stable rolling motion about the axis of maximum moment of inertia. Such passive angular control has been extremely successful.

Tiros also furnishes an excellent example of a simple technique for changing the orientation of the spin axis of a spacecraft. When a change is desired, an electromagnet in the satellite is activated. This magnet then tries to align itself with the local magnetic lines of force of the Earth and in doing so applies a torque that causes the spacecraft to precess, thereby changing the orientation in space of its spin axis. The electromagnet is turned off once the spacecraft has reached the desired new orientation.

In the case of the Orbiting Astronomical observatory (OAO), inertia wheels will be used to vary the distribution of the angular momentum of the spacecraft about each of its three principal axes. By redistributing this momentum from the spacecraft to and from these spinning wheels, it is possible to control the pointing accuracy of the telescope to fractions of a second of arc.

2. In-Space Maneuvering.—Maneuvers in space are quite different from those on earth. In space a body moves in a curved line, the curvature depending upon the instantaneous location vector and velocity vector of the spacecraft. For example, to apply a trajectory correction (*i.e.*, a mid-course correction) to the Mariner 2 spacecraft, it was necessary to know the future time at which the correction would be made. Furthermore, it was necessary to know accurately the location and velocity of the spacecraft at that time. The change in the velocity vector to place the spacecraft in the new trajectory had to be predicted so that the spacecraft would just miss Venus, which was also moving on a curved path. In addition, the angular orientation of the spacecraft in space had to be known and had to be changed to a new orientation so that the corrective force would be applied in the proper direction. Without going into the details of computation, timing and execution, the complexity of such a maneuver is evident (*see* fig. 5 and 6). Mariner 2 originally was on a trajectory that would have missed Venus by 233,000 mi. After the mid-course correction it missed Venus by 21,600 mi., which was on the safe side and only 11,600 mi. off the intended "bull's-eye miss" of 10,000 mi. Similar techniques were used to establish the Mariner 4 trajectory for the fly-by of Mars in 1965 and to correct the lunar impact or landing positions of the Rangers 7, 8 and 9, and of Surveyor 1. *See* also AERODYNAMICS; BALLISTICS: *Exterior Ballistics*; GEODESY. (Jo. D. N.)

II. LAUNCH VEHICLES

Launch vehicles are powered heavier-than-air craft that are used to place a secondary craft or payload into ballistic trajectory. They are closely akin to their predecessors, the guided missiles or military rockets, which were fashioned to deliver warheads

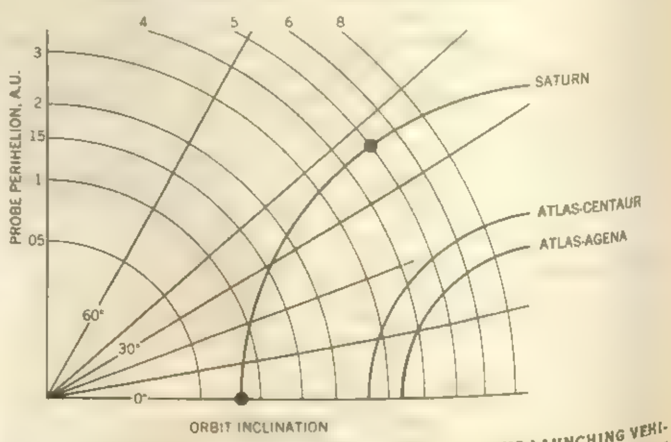


FIG. 7.—NECESSARY BOOSTER PERFORMANCE FOR VARIOUS LAUNCHING VEHICLES TO LAUNCH A 250-LB. SPACECRAFT TOWARD THE SUN AT DIFFERENT ANGLES OUT OF THE EARTH'S ECLIPTIC PLANE

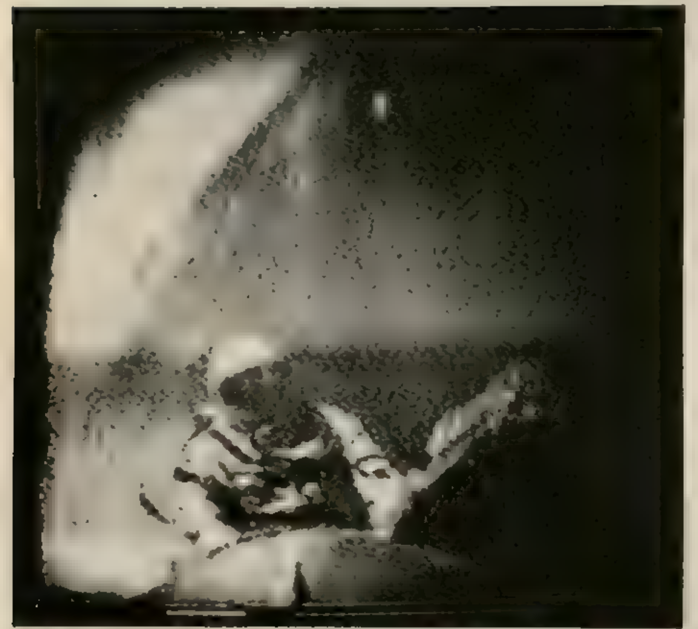
Astronaut Edward H. White as he maneuvered in space more than 100 mi. above the surface of the earth. The 25-ft. "umbilical" line which provided him with oxygen and tethered him to the spacecraft may be seen in this photograph

White photographed from inside the spacecraft by McDivitt. The visor of White's helmet was gold-coated to protect him from unfiltered solar rays; in one hand he held the self-propulsion device which enabled him to maneuver while floating in space

THE FLIGHT OF GEMINI 4

The flight of Gemini 4 in June 1965 lasted 97 hours and 56 minutes, during which the spacecraft containing United States astronauts James A. McDivitt and Edward H. White II made 66 orbits of the earth. During the third orbit, as Gemini 4 swept toward the Pacific coast of the United States, astronaut White emerged from the spacecraft and floated in space for about 22 minutes. Photographs of the first man to float in space, Soviet cosmonaut Alexei Leonov, may be found on the following plate

BOTH PHOTOGRAPHS BY COURTESY OF NASA



Soviet cosmonaut Alexei A. Leonov emerging from the two-man Voskhod 2 capsule. The Russian spacecraft transmitted to earth live television pictures of Leonov's venture into space. This photograph and that at right above were made directly from a television screen on the ground

Cosmonaut Leonov pushing off into space, attached to the capsule by a 16-ft. "umbilical" line which supplied him with oxygen. Unlike U.S. astronaut White, Leonov did not have a self-propulsion device

Leonov, free of the capsule, floating upside down at the end of his tether; at right may be seen the open hatch of Voskhod 2. Leonov remained in space for about ten minutes

THE FLIGHT OF VOSKHOD 2, MARCH 1965

The Soviet cosmonaut floating in space 107 mi. above the earth. During their flight, cosmonauts Leonov and Pavel I. Belyayev completed 17 orbits of the earth

ALL PHOTOGRAPHS FROM SOVFOTO

against military targets. Launch vehicles, however, must be capable of greater flexibility to perform a variety of tasks.

1. History.—Scientific uses for rockets were suggested by Konstantin Tsiolkovski of Russia in his treatise on space travel (1903). Robert Goddard of the United States launched the first rocket vehicle employing a liquid-fueled rocket engine on March 16, 1926. The fast technological growth in aeronautics during the 1930s and 1940s was effectively adapted to rocket technology to produce the V-2 in Germany (1938–45), the first supersonic aircraft, X-1 in the United States (1947), and a series of ballistic missiles. The perfected state of guidance and control systems was a most essential part of this rapid utilization of rocket power. (*See* also ROCKETS AND GUIDED MISSILES.)

2. Configurations.—Launch vehicles usually consist of one or more rocket-powered stages utilizing precise guidance systems in conjunction with control systems that actuate steering mechanisms. Small sounding rockets and upper stages of larger vehicles are sometimes stabilized in a fixed orientation by spin imparted from auxiliary power sources or by canted aerodynamic fins.

The airframe of a launch vehicle is closely integrated with the propellant tanks, which make up most of the vehicle's volume. Propulsion is most commonly supplied from rocket motors operating on the reaction forces produced from the high-pressure combustion of liquid fuel and liquid oxidizer combinations or from solid combustibles. Propulsion produced from the combustion of a fuel with the oxygen of atmospheric air, such as the ramjet, is not suitable for application to space launch vehicles.

3. Guidance.—Guidance systems are composed of several basic elements which provide: (1) continuous flight information for the determination of vehicle position, velocity and angular orientation; and (2) computations based on this information for appropriate control commands. Guidance elements can be located both in the vehicle and on the ground. When all guidance elements are in the vehicle the system is referred to as "all-inertial," and guidance accuracy depends upon the precision of inertial sensors that provide the reference for measurement of position and orientation. These systems, being insensitive to external influences (*e.g.*, radio jamming), are usually employed in military missiles. Nonmilitary vehicles more commonly employ ground-based equipment in conjunction with the airborne elements in order to increase flexibility, performance and accuracy.

4. Launch Facilities.—Final launch vehicle flight preparation is accomplished at a range which provides checkout, propellant, photographic, calibration, tracking, telemetry, timing and safety services. Ranges in use in the 1960s included: (United States) Eastern test range, Cape Kennedy, Fla.; Western test range, Lompoc, Calif.; White Sands missile range, N.M.; Wallops station, Wallops island, Va.; (France) missile centre and test range near Arcachon, southwest of Bordeaux; Colomb-Béchar, Alg.; Levant Island range, south of Toulon; Papeete range in the South Pacific, Tuamotu archipelago; (U.S.S.R.) Kapustin Yar, southeast of Volgograd; Aral'sk on the Aral sea; Franz Josef Land in the Arctic; Mirny in Antarctica; (Canada) Fort Churchill; (Japan) Akita rocket range, in northwestern Honshu; Kagoshima space centre, southern Kyushu; (Australia) Woomera.

5. Operation.—Vehicles are usually launched vertically and so maneuvered in roll, pitch and yaw as to produce the proper final trajectory. Multi-stage vehicles can conveniently make use of coasting periods or parking orbits to change the flight direction in pitch. Trajectory analysis is a highly specialized art requiring extensive use of trained personnel and automatic computation.

6. Vehicle and Spacecraft Communication.—Range-based facilities usually provide for the required vehicle tracking and data acquisition (telemetry) associated with launches. Vehicles spending long periods of time in powered flight or in parking orbits frequently need tracking and data services from bases located away from the launch sites. These bases are linked to a common information centre by a radio and landline network.

The demands for data and tracking information over extended periods of time are met by several established networks. NASA organized three such networks for nonmilitary missions: the Earth Satellite network, the Manned Space Flight network and the Deep Space network. Responsibility for the development, implementation and operation of the major tracking and data acquisition networks for the manned and unmanned satellite programs was assigned to the Goddard Space Flight centre, Greenbelt, Md. The Jet Propulsion laboratory, Pasadena, Calif., is responsible for the Deep Space network. Stations for the Earth Satellite network were established at 14 locations on 5 continents. The Smithsonian Astrophysical observatory, Cambridge, Mass., operates a worldwide optical network in support of certain scientific satellites.

Primary unmanned satellite tracking is based to a large extent upon interferometric systems operating in a radio frequency band of 136–137 mc. Telemetry data are acquired by means of high-gain antenna systems located at data readout stations. The telemetry information is usually recorded on magnetic tape.

Manned spacecraft tracking is done by radar because of the

TABLE IV.—*Summary of U.S. Launch Vehicles*

Vehicle	Stage or Engines	Propellant	Thrust (1,000 lb.)	Diameter* ft	Height† ft	Payload (lb.) 300 NM‡	Escape	Mars Venus
Scout	1 Algol II-B 2 Castor II 3 Antares X 259 4 Altair X 248	Solid Solid Solid Solid	86§ 64 23 3	3.3	72	265	—	—
Delta	1. Thor (SLV-2) 2 AJ10 118 3. Altair X-258	LOX/RP‖ IRFNA UDMH¶ Solid	170§ 7 7 3	8	90	880	150	120
TAD (Thrust augmented Delta)	1 Thor (SLV 2) plus three XM 33♀ 2 AJ10 118 3 Altair X 248	LOX RP Solid IRFNA UDMH Solid	170§ 162§ 7 8 3	8	90	1,185	210	160
Thor Agena	1 Thor (SLV-2) 2 Agena	LOX/RP‖ IRFNA UDMH	170§ 16	8	76	1,600⁰	—	—
TAT (Thrust augmented Thor-Agena)	1 Thor SLV 2 plus three XM 33♀ 2. Agena	LOX RP Solid IRFNA/UDMH	170§ 162§ 16	8	76	2,200	—	—
Atlas Agena	1. Atlas (booster and sustainer) 1a. Atlas (sustainer only) 2. Agena	LOX/RP‖ LOX RP IRFNA/UDMH	388§ 80 16	10	91	5,950	950	600
Atlas Centaur	1 Atlas (booster and sustainer) 1a. Atlas (sustainer only) 2. Centaur (Two RL-10)	LOX RP LOX RP LOX/LH°	388§ 80 30	10	100	8,500	2,300	1,300
Atlas		LOX RP	388§	10	72	—	—	—
Titan II-GLV	1 XLR 87 2 XLR 91	Storable Storable	430§ 100	10	90	approx 7,000	—	—
Titan III-C	1. 5 segment (120 in diam)♀ 2. YLR-87 3 YLR 91 4 AJ10 138	Solid Storable Storable Storable	2,400§ 430 100 16	10	123	23,000	5,390	4,500
Saturn IB	1 S-IB (Eight H-1) 2 S-IVB (One J-2)	LOX RP LOX LH°	1,500§ 200	21 6	142	35,000	—	—
Saturn V	1. S-IC (Five F-1) 2. S-II (Five J-2) 3 S IVB (One J-2)	LOX/RP‖ LOX/LH° LOX/LH°	7,500§ 1,000 200	33	281	240,000	95,000	70,000

*Maximum, excluding fins. †Less spacecraft. ‡300 naut. mi. circular orbit. §At sea level. ‖Liquid oxygen and kerosene. ¶Inhibited red fuming nitric acid and unsymmetrical dimethylhydrazine. ♀"Strapon" units. ⁰Polar orbit. °Liquid oxygen and liquid hydrogen.

usually low orbits associated with these missions. The Deep Space network consists of three instrumentation stations with large steerable "dish" antennas spaced around the world to provide continuous surveillance of lunar and planetary spacecraft.

7. Launch Vehicle Characteristics.—A single-stage vehicle operating under conditions of zero gravity and drag will acquire a velocity gain in the direction of thrust according to the classic rocket equation, $\Delta v = Ig \cdot \ln (m_i/m)$, where Δv is the ideal stage velocity increase; I is the specific impulse of the propellant combination used in the rocket motor and is equal to rocket thrust divided by propellant flow; g is gravitational acceleration (standard); m_i is the stage mass at initiation of rocket thrust; and m is the stage mass at rocket thrust termination.

The quantity Ig, possessing the units of velocity, is the "effective exhaust velocity" of the rocket. It serves as a measure of propellant effectiveness. Characteristic values for Ig extend from 7,000 to 14,000 ft. per second; values in the 9,000–10,000 range are commonly attained in practice. The masses m_i and m contain the inert or tare weights of the payload or spacecraft. Elimination of the latter results in m_i/m being the mass ratio of the stage, this term then being a measure of the structural efficiency of the stage designs. Values for m_i/m of 5 to 10 are common. Lower values are associated with rigid propellant tank designs and the higher numbers with pressure-stabilized designs.

Because of the logarithmic character of the rocket equation, it is convenient to use a logarithmic scale for the payload on these performance charts. For multi-stage vehicles the ideal velocity is the sum of the ideal velocities of the stages. Thus,

$$V_i = \{Ig \cdot \ln (m_i/m)\}_1 + \{Ig \cdot \ln (m_i/m)\}_2 + \dots$$

In actual practice, velocity is lost because of gravitational and aerodynamic forces. These losses are heavily dependent upon the characteristics of the vehicle and the particular trajectory flown, and the computation of them is complicated. (R. B. M.)

III. SPACECRAFT

Tables V and IX list some of the spacecraft successfully launched since the beginning of the space age in 1957. Over the years the number and rate of successes increased steadily, the percentage for the U.S. rising from 29% in 1958 to about 90%.

A. Function

The purpose of the spacecraft is to carry a cargo into space, and sometimes to bring the cargo back to Earth. The cargo may be instruments or it may be man with his instruments. In this section only automated spacecraft will be discussed; *see* also *Manned Space Flight*, below.

The spacecraft must provide the environment and services required for the instrumentation ("instrumentation" here refers to the equipment associated with experiments, measurements, etc., to be made by the spacecraft; *e.g.*, radiation sensors, television cameras and micrometeor counters). With respect to environment, it is necessary to maintain the proper temperature, pressure, vibration level, rigidity and thermoconductivity. With respect to services, it is necessary to supply: (1) a structural platform; (2) power, both electrical and propulsive; (3) communications; (4) a diagnostic monitoring system for determining the condition of the spacecraft; (5) a command centre for channeling various command signals to the various types of instrumentation aboard; and (6) a control centre for guidance and navigation of the spacecraft.

In general the environment and services required by instrumentation are quite similar from one spacecraft to another. However, the instrumentation itself may differ widely in character and function. At one extreme may be a scientific satellite carrying delicate and sophisticated instrumentation for precision measurement and at the other extreme may be a simple transmitter, as in the first Vanguard satellite.

B. Typical Spacecraft

Some typical spacecraft will be described in order to illustrate their diversity and some of their design details.

1. Atmospheric Structure Satellite.—The Atmospheric Structure satellite (Explorer 17) was a 405-lb. spacecraft carrying instruments to measure the density, composition, pressure and temperature of the atmosphere. The spacecraft was a hermetically sealed sphere designed to be nearly leakproof, thus minimizing contamination of the atmosphere of space. For example, at apogee the internal satellite pressure was 10^{13} times greater than the "hard vacuum" of space on the outside; any gas leakage from the spacecraft would interfere with measurements of the tenuous space environment.

Electrical energy to the spacecraft was supplied by approximately 150 lb. of silver-zinc chemical batteries. With all instrumentation operating, the satellite consumed 110 w. of power and could operate at this load for about 70 hr. Since there was no provision for data storage aboard the spacecraft, and since satellite "turn-ons" were to last about 5 min., several hundred data "passes" could be obtained. A reasonable useful lifetime of the satellite was computed at between two and three months (the actual lifetime was 99 days).

When injected into orbit, the satellite was spin-stabilized at 1.5 cycles per second. An inclination of 58° from the equator gave it a coverage of the Earth's surface to within several degrees of the arctic and antarctic regions. The orbital period was 90 min. The spacecraft carried an aspect system to determine the orientation of its spin axis in relation to a reference plane; this information was needed to allow the experimenters to interpret their data correctly. A four-component aspect system was used, comprised of a digital-fan sensor, a slit-fan sensor, a Sun-Moon switch and two infrared Earth sensors. The system was designed to use the Earth and either the Sun or the Moon as references against which to locate the satellite spin axis. A digital-fan sensor uses an optical masking technique and six photodiodes to measure the angle of a light source viewed by the sensor. The slit-fan sensor, located in the same housing as the digital fan, is used to measure azimuth angles accurately.

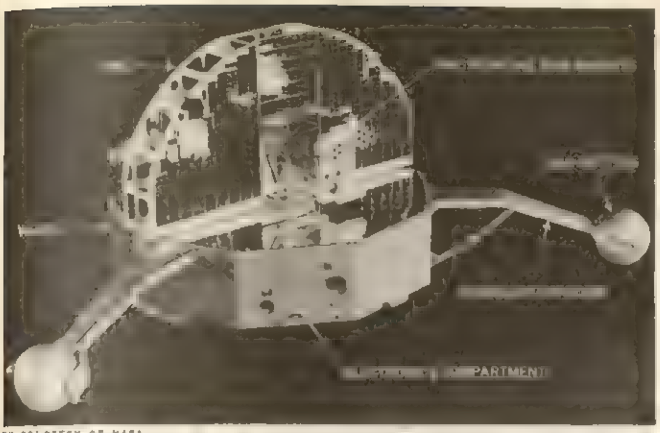
The Sun-Moon switch reduced the gain of the amplifier when sunlight was viewed and restored full gain when the Moon was the reference object. The two Earth sensors had a pencil beam of approximately 2° solid angle and were mounted in the same housing. One sensor had an elevation of 22° north of the equator of the satellite, and the other pointed 22° south of the equator. These Earth sensors operated in the 7–15 micron infrared wavelength region and detected the discontinuity at the Earth's horizon. Prior to launch, steps were taken to insure that the satellite was stable about its predetermined axis of rotation. The battery and electronics components were distributed to ensure maximum moment of inertia about this axis.

The Atmospheric Structure satellite was the first scientific Earth satellite containing a pulse-code modulation (PCM/FM) telemetry system. Previous satellites used the pulse-frequency modulation (PFM/AM) system. The solid-state PCM system provided an output power of 500 mw. and was capable of supplying 40 separate channels of information in digital form.

Provisions were made to permit selection of experiments while the satellite was in orbit. Two command receivers were used, each capable of two independent commands. The number of allowable combinations permitted resolution of possible interference between the various detectors. While equipment "turn-on" was controlled from the ground, "turn-off" of all equipment was automatic and produced from on-board instrumentation, ensuring that power would not be wasted when the satellite passed out of the range of the command receiver.

2. Interplanetary Explorer.—The Interplanetary Explorer (Explorer 18) is a 138-lb. spacecraft placed into a highly elliptical orbit: an apogee altitude of 122,800 mi., a perigee of 120 mi. and an inclination to the equator of 33°, corresponding to an orbital period of approximately four days. The spacecraft is designed to study interplanetary radiation and magnetic fields in the vicinity of the Earth and Moon. Explorers 21 and 26 are also Interplanetary Explorers.

Explorer 18 is an octagon-walled platform, 28 in. across and 12 in. deep, fabricated from nylon honeycomb and fibreglass. A



BY COURTESY OF NASA

FIG. 8.—ORBITING SOLAR OBSERVATORY, FIRST LAUNCHED SUCCESSFULLY IN MARCH 1962

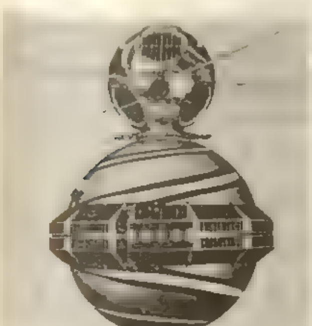
rubidium-vapour magnetometer is located on a 71-in. boom on top of the platform and two fluxgate magnetometers on 14-ft. booms at the sides. Four spring-loaded solar-cell paddles extend from the main structure to supply the required power.

The satellite transmits data continuously during its operating time in orbit on the 136.110 mc. frequency; signals from the 1.5-lb. 4-w. transmitter can be received by NASA stations at distances of 123,000 mi. Data are recorded 100% of the time for approximately the first six months of flight, and then for 25% during the remainder of the satellite's lifetime. The telemetry system is designed so that the recorded data can be processed 16 times faster than real time.

3. Orbiting Solar Observatory.—The Orbiting Solar observatory (OSO) is a 500-lb. spacecraft designed to accommodate interchangeable experiments for studying the Sun. A weight of about 250 lb. is allotted for the total experiment payload.

The OSO consists of two sections: a nine-sided wheel composed of wedge-shaped compartments and a semi-circular section, mounted above the wheel, called the sail. When the spacecraft is in orbit, three arms extend from the wheel and slow its rotation to about 30 r.p.m.; the arms can be used as antennas. The wheel diameter is 44 in. Five of the nine wedge sections, each with a volume of approximately 1,000 cu.in. and a maximum weight capacity of 30 lb., are available for experiments. When in orbit, the wheel rotates much like a gyroscope, each section facing the Sun every two seconds.

The sail, joined to the wheel by a rotating shaft, is designed so that it aims always at the Sun. The sail carries an array of solar cells to supply power and recharge batteries that produce about 16 w.; of this, 9 w. is available for power experiments. In addition, the sail carries two compartments for experiments which must point at the Sun; each is 4 × 8 × 36 in. and is able to carry approximately 50 lb. of equipment, but the two may be combined to give double the space and weight capacity.



BY COURTESY OF U.S. NAVY

FIG. 9.—TRANSIT 2A/NRL SATELLITE LAUNCHED JUNE 22, 1960, FROM CAPE CANAVERAL (NOW CAPE KENNEDY), FLA.

Data from each experiment are stored sequentially in digital form on a tape recorder operating continuously throughout each 95-min. orbit. The information from the recorder is read out at high speed on command from a ground station. The first OSO was launched in 1962, the second early in 1965 and the third in mid-1965.

4. Transit.—Transit (fig. 9.) is a satellite weighing approximately 135 lb. It is designed to provide navigational data to military and civilian ships. The spacecraft has an octagonal body of fibreglass and honeycomb plastic, 12 in. high and 18 in. across. Four panels containing 18,000 solar cells are mounted on the sides, and the power these produce is stored in nickel-cadmium batteries. The spacecraft employs gravity-gradient stabilization, utilizing a single self-erecting boom equipped with a long fine-damping spring with a small mass on the end. The primary instrumentation consists of two ultrastable oscillators operating at 150 and 400 Mc.

5. Mariner.—The Mariner spacecraft (Mariner *2*) weighed 447 lb. and was launched by an Atlas-Agena vehicle in a 180,000,000 mi., 109-day trajectory past Venus. The spacecraft (fig. 10) was 5 ft. in diameter at the base and 9 ft. 11 in. high. In the cruise position, with solar panels and high-gain antenna extended, it was 16.5 ft. across and 11 ft. 11 in. high.

The design was a variation of the hexagonal concept used for the Ranger series. The base housed a liquid-fuel rocket motor for trajectory correction and six modules containing the attitude control system, electronic circuitry for the scientific experiments, power supply, battery and charger, data encoder and command subsystem, digital computer and sequencer, and radio transmitter and receiver. Sun sensors and attitude control jets were mounted on the exterior of the base.

A tubular superstructure extended upward from the base and scientific experiments were attached to this framework. An omnidirectional antenna was mounted at the peak of the superstructure. A parabolic, high-gain antenna was hinge-mounted below the base, and two solar panels also were hinged to the base; these folded up alongside the spacecraft during launch, parking orbit and injection phases, and unfolded like butterfly wings when the craft was in space. A command antenna for receiving transmissions from Earth was mounted on one of the panels.

Two-way communication aboard the Mariner *2* was supplied by the receiver/transmitter, two transmitting antennas (the omnidirectional and high-gain antennas) and the command antenna for receiving instructions from Earth. The spacecraft transmitting power was only 3 w. An Earth sensor was mounted on the antenna yoke near the rim of the high-gain, dish-shaped antenna to search for and keep the antenna pointed at the Earth.

Stabilization of the spacecraft for yaw, pitch and roll was provided by ten cold-gas jets mounted in four locations and fed from two titanium bottles containing 4.3 lb. of nitrogen at 3,500 p.s.i. The jets were linked by logic circuitry to three gyroscopes in the attitude-control subsystem, to the Earth sensor on the parabolic antenna, and to six Sun sensors mounted on the spacecraft frame and on the backs of the two solar panels.

On July 14, 1965, seven and a half months after its launch, Mariner 4, a successor of Mariner *2*, flew past Mars to become the first spacecraft to make a successful journey to that planet. After its Mars encounter, Mariner 4 continued to return data on the interplanetary environment until Oct. 1, 1965. In 1967 it will pass within about 29,000,000 mi. of the Earth, and during that time it should be possible to again obtain data from the spacecraft.

All Mariner spacecraft carried a set of instruments designed to measure the fields, radiation and particles of interplanetary space. There were about twice the number of nuclear particle detectors installed on Mariner 4 as could be accommodated by Mariner 2. The vector magnetic fields, measured by a flux-gate magnetometer on Mariner 2, were detected by a more sensitive molecular resonance magnetometer on Mariner 4. The instrument for measuring the micrometeoroids of interplanetary space on Mariner 4 has been similarly improved. The large quantity of precise orbital data obtained during months of tracking Mariner 4 as it moved away from the Earth provided a basis for significant improvements in man's knowledge of certain fundamental physical constants of the solar system.

Most of the Mariner interplanetary instruments and investigations are useful in exploring the target planet. In addition, the spacecraft carried special instrumentation for planetary measurements. Since Venus is enshrouded in clouds and because of the payload weight limitations imposed on the Mariner 2 design, the planetary instruments on this spacecraft were limited to radiometers capable of spatially resolving the temperatures of Venus. The planetary instrument on Mariner 4 was a television camera, which obtained 22 pictures of the Martian surface. An extension

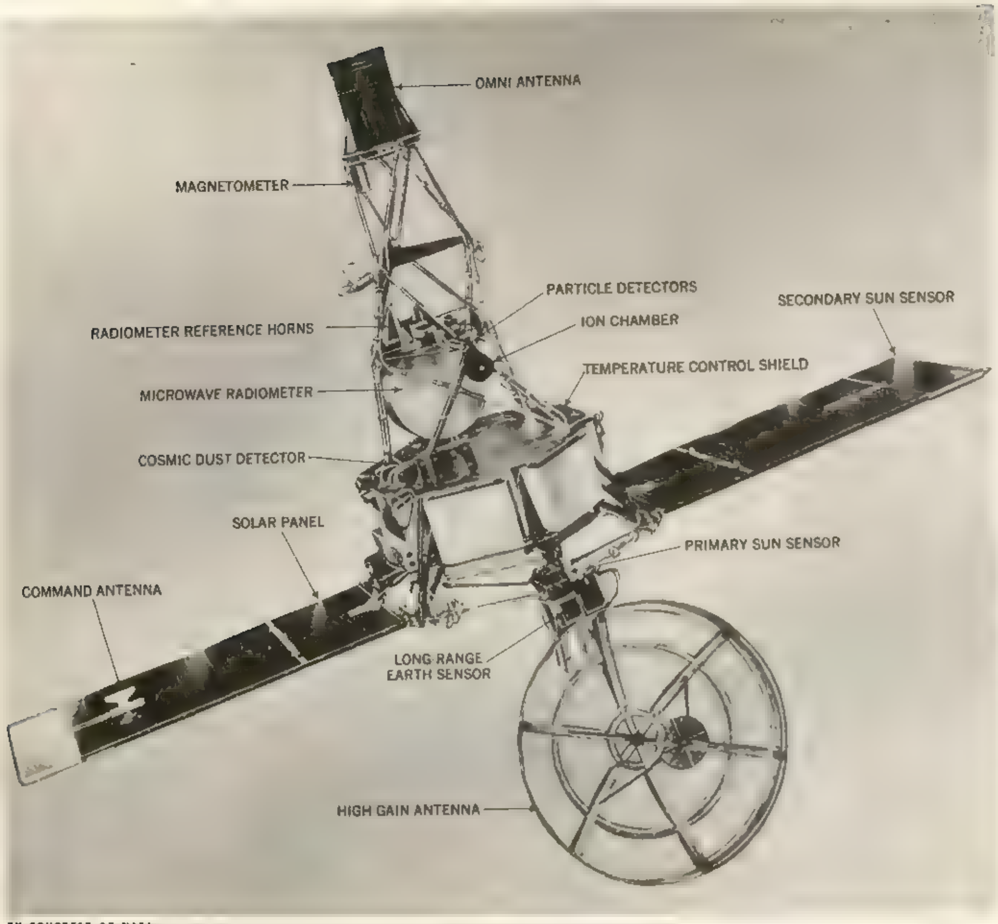
TABLE V.—*Advances in Space Exploration by Unmanned Artificial Satellites and Space Probes*

Name	Nation	Launched	Launch vehicle	Perigee/Apogee (mi.)	Period (min.)	Lifetime*	Weight in orbit† (lb.)	Remarks
Sputnik 1	U.S.S.R.	1957 Oct. 4	n.a.‡	143.5/584	96.17	1958	(184)	Establishment of first artificial satellite
Sputnik 2	U.S.S.R.	Nov. 3	n.a.‡	128/1,056	103.7	April 14, 1958	1,120	First inhabited space capsule (dog, "Laika")
Explorer 1	U.S.	1958 Jan. 31	Jupiter-C	224/1,585	114.7	4–10 yr.	30.8	Discovery of inner (Van Allen) radiation belt
Vanguard 1	U.S.	March 17	Vanguard	405/2,462	134.3	Indefinite	3	First proof that Earth is pear-shaped
Sputnik 3	U.S.S.R.	May 15	n.a.‡	133/1,178	105.8	April 6, 1960	(2,926)	First instrumented multipurpose space laboratory (atmospheric pressure, composition; positive ion concentration, primary cosmic ray data, etc.)
Score	U.S.	Dec. 18	Atlas	115/915	101.5	Jan. 21, 1959	8,800	First transmission of voice messages from space
Luna 1	U.S.S.R.	1959 Jan. 2	n.a.‡	91.5/214.7‖	450¶ (about Sun)	...	3,245	First space probe to become artificial planetoid; radio contact to 371,000 mi.
Vanguard 2	U.S.	Feb. 17	Vanguard	347/2,064	125.9	...	22	Returned first cloud-cover photographs; transmitted for 18 days
Explorer 6	U.S.	Aug. 7	Thor-Able	157/26,366	768	July 1961	143	Discovered Earth's electrical current system; returned first television photograph of Earth; first use of solar-cell paddles for power
Luna 2	U.S.S.R.	Sept. 12	n.a.‡	(Moon impact)	...	33 hr. 32 min.	3,324	First spacecraft on Moon; discovery that Moon is enveloped by layer of low-energy ionized gas
Vanguard 3	U.S.	Sept. 18	Vanguard	317/2,329	130.2	...	100	Mapped Earth's magnetic field, returned micrometeorite and radiation data
Luna 3	U.S.S.R.	Oct. 4	n.a.‡	25,500/294,250	15.3¶ (about Earth)	April 1960	958.7	First circumnavigation of Moon; first pictures of opposite side of Moon
Explorer 7	U.S.	Oct. 13	Juno II	346/676	101.2	...	92	Transmitted data on magnetic fields and storms, solar flares, radiation belt
Pioneer 5	U.S.	1960 March 11	Thor-Able	(Solar orbit)	312¶	...	95	Interplanetary probe; measured size of solar system, returned data on magnetic field and solar wind out to 22,500,000 mi.
Tiros 1	U.S.	April 1	Thor-Able	428.7/465.9	99.19	75 yr.	320	Study of TV systems for weather satellites
Transit 2A	U.S.	June 22	Thor-Able-Star	382/657	101.66	50 yr.	1,587	First successful dual-satellite launch
Discoverer 13	U.S.	Aug. 10	Thor-Agena A	161/436	94.1	Nov. 14, 1960	(300)	First successful recovery of capsule from orbit
Echo 1	U.S.	Aug. 12	Thor-Delta	945/1,049	118.24	...	241.4	(Aluminum-coated plastic sphere) First successful voice and picture transmission by reflection off satellite; first measurements of satellite perturbations caused by solar radiation pressure
Discoverer 14	U.S.	Aug. 18	Thor-Agena A	116/502	94.54	Sept. 16, 1960	(300)	First successful air recovery of orbital capsule
Sputnik 5	U.S.S.R.	Aug. 19	n.a.‡	190/211	90.68	Aug. 20, 1960	10,120	First controlled landing of animals in space vehicle (two dogs, six mice, insects); animal behaviour in orbit monitored by TV and radio
Explorer 8	U.S.	Nov. 3	Juno II	258/1,423	112.75	19 yr.	101	Collection of data for ionospheric studies; static electricity accumulation on satellite surface; impact of micrometeorites
Tiros 2	U.S.	Nov. 23	Thor-Delta	387/453	98.27	75 yr.	330	Radiation and optical mapping of Earth; TV pictures received, but wide-angle pictures of poor quality
Discoverer 18	U.S.	Dec. 7	Thor-Agena B	154/459	94.1	April 2, 1961	(300)	Reentry test, exposure of human eye tissue, bone marrow and blood cells, algae, spores, etc., to corpuscular radiation, air-recovery successful
Sputnik 8/Venus Probe 1	U.S.S.R.	1961 Feb. 12	n.a.‡	123/198 66.7/94.6‖	89.7 300¶	Feb. 25, 1961 (Sputnik 8)	(1,419) (V.P. 1)	First hyperbolic escape of probe to measure interplanetary magnetic fields and solar radiation in space and near Venus, radio contact lost about 4,700,000 mi. from Earth
Explorer 9	U.S.	Feb. 16	Scout	395/1,605	118.3	...	15	Atmospheric density data provided by optical tracking of 12-ft. balloon
Sputnik 9	U.S.S.R.	March 9	n.a.‡	115/155	...	March 9, 1961	10,340	Habitable vehicle with dog, guinea pigs, mice, insects and plant seeds; reentry and recovery successful
Sputnik 10	U.S.S.R.	March 25	n.a.‡	111/153	88.4	March 26, 1961	10,330	Cabin with dog "Zvezdochka" recovered after 17 orbits
Transit 4A	U.S.	June 29	Thor-Able Star	534/623	103.7	...	175	All-weather global navigation satellite; first to use nuclear auxiliary power system
MA-5	U.S.	Nov. 29	Atlas D	99.6/147.5	88.5	Nov. 29, 1961	2,900	First U.S. orbital flight carrying animal (chimpanzee "Enos")
Tiros 4	U.S.	1962 Feb. 8	Delta	441/525	100.4	Indefinite	285	Returned total of 32,593 cloud-cover photographs up to June 10, 1962
OSO-1	U.S.	March 7	Delta	344/370	96.2	...	458	First orbiting solar observatory, transmitting solar flare, radiation data
Cosmos 1	U.S.S.R.	March 16	n.a.‡	135/609	96.35	May 25, 1962	n.a.‡	Unmanned spacecraft to study conditions of prolonged flight and effects of Van Allen belts on long-range radio communications
Ariel	U.K.-U.S.	April 26	Delta	242/754	100.9	...	32	First international co-operative satellite
Tiros 5	U.S.	June 19	Delta	367/604	100.5	...	286	Weather satellite; returned 57,875 cloud-cover photographs
Telstar 1	U.S.	July 10	Delta	593.4/3,501.8	157.8	...	(170)	First transatlantic relay of television signals (July 10); first colour TV relay (July 16); tests of broad-band microwave communications in space

TABLE V.—*Advances in Space Exploration by Unmanned Artificial Satellites and Space Probes (continued)*

Name	Nation	Launched	Launch vehicle	Perigee/Apogee (mi.)	Period (min.)	Lifetime*	Weight in orbit† (lb.)	Remarks
Mariner 2	U.S.	Aug. 27	Atlas-Agena B	65.5/113.8‖	345.9¶		(447)	Geophysical studies of Venus, passing planet at 21,594 mi.; measurements of atmospheric and surface temperatures, magnetic field, planetary density, radio contact to 54,300,000 mi.
Alouette 1	Can.-U.S.	Sept. 28	Thor-Agena B	620/638	105.4		320	Canadian research satellite for ionospheric sounding
Mars 1	U.S.S.R.	Nov. 1	n.a.‡	(Solar orbit)	519¶	...	1,965	Mars probe; transmissions ceased after 66,000,000 mi.
		1963						
Telstar 2	U.S.	May 7	Delta	604/6,713	225	..	175	Active-repeater communications satellite, transmitting on 136,050 Mc.
(None)	U.S.	May 9	Atlas-Agena B	2,254/2,295	166.6	Indefinite	n.a.‡	Ejected 50-lb. package of copper dipole needles for experiments in reflecting high-frequency radio signals over long Earth distances
Syncom 2	U.S.	July 26	Delta	22,132/22,823	1,436	...	86	Active repeater communications satellite in synchronous orbit, initially over Brazil; communications tests successful
		1964						
(None)	U.S.	Jan. 29	Saturn I	164/471	94.8	< 1 yr.	37,700	Heaviest weight in orbit to date; thrust at lift-off 1,500,000 lb.
Elektron 1 Elektron 2	U.S.S.R.	Jan. 30	n.a.‡	252/4,410 285/42,352	169 1,360	... ...	n.a.‡ n.a.‡	Both spacecraft launched into different orbits by one launch vehicle to study separated regions of space
Ranger 7	U.S.	July 28	Atlas-Agena B	(Moon impact)	...	July 31, 1964	806	Transmitted more than 4,000 still television photographs of Moon's surface; last photographs taken at about 1,000 ft. above surface
OGO-1	U.S.	Sept. 5	Atlas-Agena B	173/92,721	3,839	.	1,073	Carries 20 co-ordinated space experiments
Explorer 22	U.S.	Oct. 9	Scout	551/1,673	104.8	...	116	First satellite to carry successful laser experiment
Explorer 24 Explorer 25	U.S.	Nov. 21	Scout	329/1,551	116	...	19 90	With Explorer 25, first simultaneous launch of two spacecraft by NASA
Mariner 4	U.S.	Nov. 28	Atlas-Agena D	(Mars probe)	...	Indefinite	575	Photographed surface of Mars and studied the Martian atmosphere
Zond 2	U.S.S.R.	Nov. 30	n.a.‡	(Mars probe)	...	...	n.a.‡	Scheduled to pass Mars about July 1965, but communication equipment failed
San Marco	Italy-U.S.	Dec. 15	Scout	128/510	95	...	254	First launch of a satellite by a team of nationals from a country other than U.S. or U.S.S.R.
		1965						
Pegasus 1	U.S.	Feb. 16	Saturn I	308/463	97.1	...	33,200	2,000 sq.ft. micrometeoroid sensor area
Ranger 8	U.S.	Feb. 17	Atlas-Agena B	(Moon impact)	...	Feb. 20, 1965	..	Transmitted 7,137 still photographs of Moon's surface; covered much wider area than Ranger 7
Ranger 9	U.S.	March 21	Atlas-Agena B	(Moon impact)	...	March 24, 1965	...	Photographs of Moon transmitted over commercial television
Early Bird 1	U.S.	April 6	Thrust-augmented Delta	22,251/22,270	1,437		87	First commercial satellite launched by NASA for the Comsat Corp.
Tiros 10 (OT-1)	U.S.	July 2	Thrust augmented Delta	...	...	...	290	First U.S. Weather Bureau funded spacecraft; spin-stabilized configuration with two TV cameras, near-perfect Sun-synchronous orbit
Zond 3	U.S.S.R.	July 18	n.a.‡	(Space probe)	...	...	n.a.‡	Photographed far side of Moon while on unspecified deep-space mission
Pegasus 3	U.S.	July 30	Saturn I	...	...	...	3,200	Micrometeoroid studies; last of current Pegasus program; final launch of Saturn I vehicle program with 10 out of 10 successes
Venera 2	U.S.S.R.	Nov. 12	n.a.‡	(Solar orbit)	.	...	..	First successful Soviet planetary probe; passed within 14,900 mi. of Venus on Feb. 27, 1966
Venera 3	U.S.S.R.	Nov. 16	n.a.‡	(Venus impact)	...	March 1, 1966	n.a.‡	First spacecraft to impact on another planet (Venus); Failed to return data on descent to surface because of communications failure
French 1A	France	Dec. 6	Scout	(Solar orbit)	...	.	135	Study of very-low-frequency wavefield in the magnetosphere and irregularities of ionosphere; craft designed, constructed and tested by the Centre National d'Etudes Spatiales in France
		1966						
Luna 9	U.S.S.R.	Jan. 31	n.a.‡	(Moon landing)	...	Feb. 3, 1966	220 (instrumented payload)	First successful lunar "soft landing"; television system with 360° field provided several pictures of surface; other sensors indicated low level of radioactivity
Cosmos 110	U.S.S.R.	Feb. 22	n.a.‡	...	...	March 16, 1966	...	Carried two dogs (one a control) for biomedical research; capsule successfully recovered after 22 days in orbit, passing through radiation zones; live dogs "arrived in Moscow in the same containers in which they had lived in the spaceship"
Luna 10	U.S.S.R.	March 31	n.a.‡	(Moon orbit)	...	...	540	First spacecraft placed in lunar orbit
Surveyor 1	U.S.	May 30	Atlas-Centaur	(Moon landing)	...	July 1966	2,194 at separation, 596 landing	First U.S. soft landing on Moon; successful test of landing system required for subsequent manned landing; spacecraft transmitted photographs of Moon's surface

*Dates indicate time of destruction during reentry or successful reentry of satellite; other lifetimes are estimated
†Figures in parentheses indicate weight of spacecraft or capsule only; other figures show total weight in orbit, including booster stage, etc.
‡n.a. not announced
‖Perihelion and aphelion of artificial planet in miles × 1,000,000. ¶Days. ♀Period, Venus Probe 1, in days.
Source: Based on data compiled by National Aeronautics and Space Administration.



BY COURTESY OF NASA

FIG. 10.—A MODEL OF THE MARINER 2 SPACECRAFT WHICH WAS LAUNCHED AUG. 27, 1962, ON A COURSE WHICH CARRIED IT TO WITHIN 21,600 MI. OF VENUS. ITS MISSION WAS SUCCESSFULLY COMPLETED DEC. 14, 1962, WITH THE RADIOING OF MESSAGES BACK TO THE EARTH. THE MODEL IS SHOWN WITH ITS SOLAR PANELS EXTENDED IN CRUISE POSITION

in the application of precision tracking data obtained in the region of the planet provided more accurate knowledge of the mass and atmosphere of Mars.

Both spacecraft utilized two celestial references for three-axis stabilization—for Mariner 2, the Sun and the Earth, and for Mariner 4, the Sun and Canopus. Both spacecraft also contained a mid-course guidance rocket and obtained their electrical energy from silicon cells. Mariner 4 was larger and heavier than Mariner 2. With solar panels extended and solar pressure vanes opened, it spanned more than 22 ft.—about 6 ft. more than Mariner 2. In the dimension from the top of the omnidirectional antenna to the bottom of the base structure, both spacecraft measured about 10 ft. Mariner 4 weighed 575 lb., approximately 125 lb. more than Mariner 2, and contained almost 140,000 individual parts. The most obvious differences between the two spacecraft were in the number of solar panels, the shape of the base structure, the construction of the omnidirectional antenna masts and the locations of the directional antennas and planetary instruments.

The basic structure of Mariner 4 was a 30-lb. octagonal magnesium framework with seven electronics compartments and a mid-course rocket propulsion system around the perimeter. The compartments themselves provided structural support to the spacecraft. Four solar panels, each 71.4 in. long and 35.5 in. wide, were attached to the top, or sunward side, of the octagon.

Because the Mars mission carried the spacecraft away from the Sun, additional solar cell area had to be provided for Mariner 4. About 28,200 semiconductor cells were mounted on the four panels, and—except during launch and maneuvers—they provided the primary source of electrical power to the spacecraft. The attitude-control gas jets and the active solar vanes, which stabilized the spacecraft relative to the direction of the Sun, were mounted on the ends of the panels. The electrical energy originating in the Mariner 4 solar panels flowed into a pair of power regulators, either one of which could assume the full electrical load.

The interior of the octagon contained gas bottles and regulators for Mariner's dual-attitude control gas system. The propellant tank for the liquid-fuel mid-course motor is supported by a cantilever arrangement inside the octagonal cavity, with the rocket nozzle protruding through one of the eight sides. Two sets of attitude-control jets consisting of six jets each, which controlled the spacecraft on three axes, were mounted on the ends of the solar panels near the pressure vane actuators.

The high-gain antenna was attached to the spacecraft by an eight-legged superstructure atop the octagon. Its aluminum honeycomb dish reflector is an ellipse, 46 by 21.2 in., and is parabolic in cross section. The antenna, which weighed only 4½ lb., was in a fixed position. The low-gain omni-antenna was mounted on the end of a circular aluminum tube, 3.88 in. in diameter and extending 88 in. from the top of the octagonal structure. The tube acted as a wave guide for the low-gain antenna.

The Canopus star tracker assembly was located in the shade of the spacecraft on the lower ring structure of the octagon for a clear field of view. Sun sensors were located on both the top and bottom surfaces of the spacecraft body in order to provide spherical coverage.

6. Orbiting Geophysical Observatory.—In 1964, the first Orbiting Geophysical observatory, OGO-1, was launched into an elliptical orbit about the Earth. This type of spacecraft is larger than the Explorer class, weighing somewhat more than 1,000 lb. and designed to accommodate up to 50 individual experiments.

It is designed to be stabilized by motor-driven inertial flywheels and gas jets to keep one face toward the Earth, maintained so that one axis in the spacecraft is always along the line to the centre of the Earth. The solar paddles were designed to face continuously toward the Sun to ensure maximum power. The payload from an OGO consists primarily of experiments selected to make co-ordinated measurements and observations of many aspects of the Earth environment simultaneously in order to supply a clearer picture of the relationships between them and their dependence upon solar activity. Two types of OGO's are scheduled to be used. The EGO, or Eccentric Orbiting Geophysical observatory, such as OGO-1, is planned to have a highly elliptical orbit reaching beyond the Earth's magnetic field or magnetosphere into the near-Earth interplanetary space.

The POGO, or Polar Orbiting Geophysical observatory, will be placed in a low-altitude orbit over the poles for studies primarily of the upper atmosphere and the ionosphere, and for monitoring incoming radiations.

The OGO is essentially a boxlike structure, 6 × 3 × 3 ft. in size, of aluminum sandwich panels, that houses the scientific instrumentation and the equipment needed for power, telemetry, communications and attitude control. Projecting from the box are antennas for communications, booms for experiments requiring separation from the spacecraft and the other experiments, a boom

and containers for experiments that require alignment in the orbital direction, two paddles of 75 sq.ft. combined area that carry the solar cells, an antenna for the radio astronomy experiment and experiments that must point to the Sun. In orbit OGO-1 resembles a giant insect with a 20-ft. wing spread. OGO-2 and 3 launched in Oct. 1965 and June 1966 were basically similar.

7. Ranger.—Ranger 7, an 806-lb. spacecraft, was launched toward the moon by an Atlas-Agena B rocket on July 28, 1964. The mission of Ranger was to take photographs of the moon before impacting on its surface. The tubular-shaped craft was equipped with two "wings" that extended from the bottom of the main tower structure at approximately right angles; these wings, constructed so that they would face the sun during flight, contained panels of solar cells that converted the sunlight into the electricity needed to power the craft's equipment. Six television cameras mounted in the tower of Ranger focused on the moon through an aperture in the centre of the tower. At the top of the tower was an omnidirectional antenna used during launch and in a mid-course maneuver. Facing away from the moon and toward the earth was a dish-shaped antenna that transmitted the photographs.

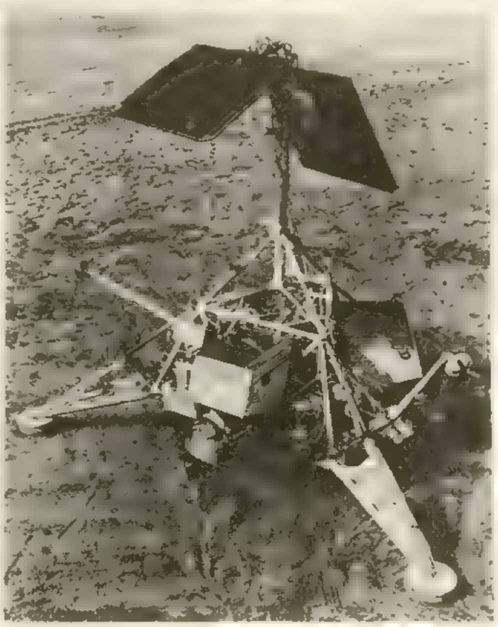
During the last 17 minutes of its flight Ranger returned more than 4,000 high-resolution photographs of the moon's surface, the last at an altitude of about 1,600 ft. Rangers 8 and 9 later successfully performed similar missions.

8. Surveyor.—The Surveyor 1 spacecraft was launched on May 30, 1966, from Cape Kennedy on a mission to land intact ("soft-land") on the Moon's surface. Approximately 16 hr. after launch, a successful mid-course correction maneuver was executed, moving the landing point to an area north of the crater Flamsteed in the Ocean of Storms. Because telemetry indicated that one of the two omnidirectional antennas may not have fully deployed, a terminal maneuver was used which assured correct Earthward orientation of the deployed antenna during descent. The spacecraft properly executed all commands and landed successfully on June 2, 1966. Preliminary data indicates that it touched down at a vertical velocity of approximately 10 ft. per second.

During its first 12 days of operation on the Moon, Surveyor 1 transmitted more than 10,000 pictures. They indicated that the spacecraft had landed on a dark, relatively smooth, bare surface, encircled by hills and low mountains and that the terrain within 1 to 2 km. surrounding the landing site is a gently rolling surface studded with craters with diameters ranging from a few centimeters to several hundred meters and littered with fragmental debris ranging from less than 1 mm. to more than 1 m.

The configuration of the Surveyor spacecraft is shown in fig. 11. The basic structure, which provides mounting surfaces and attachments for the power, communications, propulsion and flight control systems, as well as a platform for the payload packages, is constructed of thin-walled aluminum tubing with the members interconnected to form a triangle. A shock-absorbing landing leg is attached by a hinge to each of the three lower corners of the structure. The legs were folded into a nose fairing during launch. As additional protection against landing loads, blocks of crushable aluminum honeycomb were placed on the bottom of each corner of the frame. A vertical mast with mechanisms that position the high-gain planar array antenna and solar panel was mounted on top of the structure. The basic frame weighed less than 60 lb. and the installation hardware weighed 23 lb.

Surveyor 1 was about 10 ft. high, and its tripod landing gear fitted just within a circle 14 ft. in circumference. Weighing 2,194 lb. at launch, the spacecraft had a landing weight of 596 lb. after depletion of propellants and other consumables and after jettison of the altitude-marking radar and the main retrorocket casing.



BY COURTESY OF NASA

FIG. 11.—(LEFT) MODEL OF U.S. SPACECRAFT SURVEYOR 1 (WITHOUT MAIN RETROROCKET), WHICH MADE A SUCCESSFUL LUNAR LANDING IN JUNE 1966. SPACECRAFT IS 10 FT. HIGH AND WEIGHS 596 LB. WITHOUT FUEL. (ABOVE) PHOTOGRAPH OF THE MOON'S SURFACE TAKEN FROM SURVEYOR 1 SHOWS CRATER AND ROCK 6 IN. HIGH

Two thermally controlled compartments, their temperatures held within acceptable limits by a careful arrangement of absorptive and reflective paints, by conductive heat paths and thermal switches and, on the dark side, by small electric heaters, were provided aboard the spacecraft. One compartment, its temperature maintained between 40° and 125° F., housed communications and most power-supply electronics. The other, held between 0° and 125° F., housed the command and signal processing functions. The solar panel, consisting of a series-parallel array of 792 solar-cell modules approximating 9 sq.ft. in area, supplied up to 85 w. of power during the flight and on the lunar surface. Rechargeable silver-zinc batteries were used for energy storage and to accommodate peak loads.

Inertial references for detecting and controlling the attitude of the spacecraft during cruise flight, as well as during mid-course and terminal maneuvers, were provided by a Canopus star tracker, a Sun sensor and rate gyros on three axes. During cruise flight, attitude control was exercised by a subsystem employing small cold-gas thrusters. Control during descent was exercised initially by the autopilot, vernier engines and by an altitude-marking radar that began the firing of the main retrorocket. Subsequently the spacecraft was steered and decelerated by velocity-measuring and altimeter radars that, in conjunction with the on-board analog computer, autopilot and vernier engines, provided automatic closed-loop guidance for the final descent.

Two propulsion systems were used by Surveyor 1. The main retrorocket employed a solid propellant in a spherical steel case. Its thrust was of the order of 8,000 to 10,000 lb., depending upon temperature. The second, or vernier, propulsion system used hypergolic liquid propellants. Three throttleable thrust chambers were used, each capable of delivering from 30 to 104 lb. of thrust as required by the flight-control subsystem. One chamber could be swiveled for roll control. The vernier engines were used for the mid-course maneuver and in the terminal landing sequence.

For communications, Surveyor 1 carried two transmitters, two receivers, two omnidirectional antennas and the planar-array high-gain antenna, used for transmission of the 600-line television pictures. Tracking and engineering data were transmitted continuously on a frequency of 2,295 Mc. at a power of 10 w. The radio link also incorporated decoders that addressed commands received from Earth to the proper subsystems aboard the spacecraft, and signal processors that conditioned various data signals for transmission back to Earth.

The survey television camera aboard the spacecraft transmitted

200- and 600-line pictures of the lunar surface on command from Earth. The vidicon tube and its shutter, diaphragm and optics were mounted almost vertically and were surmounted by a mirror that could be adjusted both in azimuth and elevation by stepping motors. Provision was also made for remotely inserting filters in the optical system, to permit colorimetric evaluation of individual pictures. The focal-plane shutter of the camera, normally providing an exposure time of 150 milliseconds, could also be held open for longer intervals on Earth command. A lens of variable focal length was used, covering a field of 6.4° by 6.4° at its maximum focal length of 3.9 in. and a field of 25.4° by 25.4° in its wide-angle focal length of 0.98 in. The focus distance of the lens could also be commanded to cover a range from 4 ft. to infinity; and the diaphragm, though adjusted automatically, could also be set by command from Earth.

In addition, more than 100 items of engineering instrumentation were carried by Surveyor 1, and their readings were telemetered to Earth. They included temperature sensors, strain gages, accelerometers and position-indicating devices. The weight of this engineering payload, including an auxiliary battery, was 63.5 lb.

C. Spacecraft Design

The paramount requirements in spacecraft design are reliability and long life; the spacecraft must work, and it must work for a long time. The total cost of a particular development program depends directly on the ability of the designer to fulfill these requirements. A satellite that works for a long time collecting desired data clearly obviates the need for additional spacecraft to collect the same data, and therefore saves additional fabrication and launching costs. Reliable long life, therefore, is the keystone to success in spacecraft technology, not only for near-Earth orbits but especially for spacecraft journeying to other planets. A trip to Mars, for example, takes almost a year, and during that time the spacecraft must be capable of performing guidance and navigation functions as well as scientific measurements. If it is to land when it finally reaches Mars, it must then be capable of operating on the surface. A Martian year is about two Earth years. Therefore, the requirement for long life, completely unattended by human hands, is the great challenge that faces the designer of automated spacecraft.

The first essential for success is simplicity, a feature exhibited by some of the early satellites. They tended to be spherical so that the drag forces would be nearly constant and the resultant orbits could be more readily computed and tracked. They had simple power supplies such as storage batteries or small banks of solar cells to utilize the energy coming from the Sun. Their transmitters and associated electronic circuitry were elementary. That this simplicity of design was successful is evident in the case of the Vanguard satellite, which operated continuously in orbit for over five years.

D. Spacecraft Testing

The importance of skilled workmanship and strict quality control in spacecraft construction is obvious. However, it is well known that reliability greatly decreases as the number of components and complexity of design increase. To achieve success in complex satellites, it has been necessary to develop extensive facilities for testing spacecraft before they are launched.

Fundamentally, successful spacecraft performance depends upon: (1) simple, sound design; (2) highly skilled workmanship; (3) strict quality control; (4) mechanical acceleration and vibration tests; (5) thermal and vacuum tests; (6) radiation tests; (7) guidance and control tests; and (8) complete in-flight monitoring of performance.

1. Acceleration and Vibration Tests.—Every satellite must endure and survive the rugged accelerations and vibrations imposed during launch, so testing facilities that simulate these stresses have been developed. The launching accelerations are simulated by a centrifuge. The accelerations during launch can be programmed, and the spacecraft can be mechanically, structurally and elastically examined in detail under simulated conditions.

A vibration machine is used to pinpoint those resonances that may produce structural failure or distortion. Just as wind-induced resonances can destroy a large suspension bridge, so small vibrations can cause solar paddles, antennas and the main structural components of a satellite to fail. Vibration tests have been extremely important in uncovering weak points in mechanical design that otherwise might not be discovered. About 30% of the spacecraft failures occur during the vibration test.

2. Pressure and Temperature Tests.—An important hazard to spacecraft performance is the low pressure in outer space. Some components, for example, might blow up when placed in a vacuum; some materials might simply evaporate or lose their normal characteristics. The vacuum of space is considerably lower than vacuums readily achievable on Earth, and, therefore, it has been necessary to improve vacuum apparatus. In the mid-1960s vacuum chambers capable of housing complete spacecraft could attain pressures of 10^{-6} mm. of mercury and sometimes as low as 10^{-8}. For small components lower pressures were available.

Temperature presents another problem, for a spacecraft receives heat by radiation directly from the Sun, without the shielding effect provided by the Earth's atmosphere. The amount of heat depends upon the thermoconductivity and the radiation characteristics of the satellite. The temperature of the spacecraft depends upon the heat it absorbs and, in particular, on the heat it radiates away. It is necessary to achieve a good thermal balance in order to minimize extremes of temperature, and it is also desirable to achieve slow changes in temperatures, when changes are unavoidable. Temperature has a direct effect on the lifetime of electronic components; in general, the higher the temperature the poorer their reliability and the shorter their lifetime. Sudden changes in temperature are also serious because they introduce thermoworking and thermofatigue, phenomena that can lead to failure in components and circuitry. The temperature, the radiation and the thermal balance characteristics of spacecraft are thoroughly explored on Earth in thermal chambers, usually combined with vacuum chambers. Heat from simulated "suns" is directed at the satellite. The walls of the test chambers are cooled to the black-body temperatures of space by refrigerants. The satellite itself may be oriented in a systematic manner with respect to the "sun" in order to simulate its flight plan in space. Thermocouples distributed throughout the satellite measure the transient flow of heat and internal distribution of temperatures under various working conditions. In this way "hot spots" may be detected and eliminated.

3. Radiation Testing.—Another problem is that of the intense radiation in space. Radiations coming from the Sun, as well as from the cosmos, pose a special hazard to the performance and life of electronic components, as well as to the characteristics of some structural materials. Solar cells, transistors, resistors, etc., are all subject to serious damage from radiation, particularly during severe solar storms. Artificial radiation also occurs in space as a result of high-altitude explosions of nuclear bombs. Facilities were being constructed in the 1960s to test satellite components and complete spacecraft in strong radiation environments.

4. Guidance and Control Testing.—If spacecraft are to be properly controlled in space, there must be testing facilities on Earth to first prove out the entire guidance and control servomechanism. An excellent approach is to mount the complete spacecraft so that it is free in all three modes of angular motion (roll, pitch and yaw). In this way the servomechanism can be programmed and the spacecraft can be put through various maneuvers. At the same time the system can be placed in a thermo-vacuum chamber and can be operated for long periods to determine its lifetime in the space environment. In this way any deficiencies in the various components of the control system can be discovered and corrected. Complex guidance systems present a major challenge with respect to reliability and long life. Passive attitude-control systems are preferred because of their simplicity. However, in many cases performance requirements dictate direct and immediate control in all three modes of motion or require successive positioning in precise fashion.

For example, the OSO spacecraft must be capable of being pointed in a given direction with an accuracy of one minute of arc; for the OAO satellite the requirement is measured in seconds of arc. In such cases, complex command and control systems are required and these systems, therefore, are of vital concern in determining satellite reliability and performance. The best approach to an answer lies in the long-life guidance and control tests carried out on the ground.

Of equal importance, however, are diagnostic instrumentation and telemetry during the actual space flight, because if reliability and lifetime are to be improved it is necessary to know exactly why past satellites have failed. For this reason some weight and power must be allocated for diagnostic instrumentation which will continuously monitor the performance of the spacecraft, its circuitry, temperature, voltages and currents, and general performance. Thus, in the actual space environment the design and performance of the spacecraft can be monitored and analyzed, and corrective modifications can be made before future launchings.

(Jo. D. N.; R. J. Gu.)

IV. SPACE SCIENCE: PHYSICAL SCIENCES

Science is the expression of man's desire to explore and understand his natural environment. Space science is the extension of that exploration to those new regions of space and to those new phenomena which have become accessible to man as a result of the development of sounding rockets, satellites and deep-space probes. Before this, man was limited in his explorations to the Earth and its atmosphere up to an altitude of about 40 km. He was limited to the study of the natural phenomena which either occurred in this region of space or which emitted electromagnetic or particulate radiation which could penetrate the Earth's atmosphere to this region. He could observe the stars and galaxies in those colours of light which were not absorbed by the atmosphere, and could only speculate about the appearance of the planets and the stars in the ultraviolet and infrared light which is absorbed by the atmosphere. The development of the sounding rocket, the satellite, and the deep-space probe enormously increased the region of space which man can explore and the phenomena he can observe.

In order to understand the factors that must be taken into account in space science and to provide a proper perspective for this article it is necessary to remember that the Earth is immersed in the "atmosphere" of a small star, the Sun, and that the Sun controls the properties of the space occupied by the planets and provides the energy for the processes which take place on the Earth and other planets.

A. Solar Phenomena

The physical properties of the Sun are given in the article SUN. This section will cover primarily those solar phenomena of importance to space science or those which can only be studied by space techniques.

The Sun emits electromagnetic radiation in all wavelengths, from gamma rays to long wavelength radio waves. It also emits charged particles—protons, alpha particles, the positively charged nuclei of heavier atoms, and presumably electrons, although these have not yet been detected. The total power radiated by the Sun, 3.8×10^{33} ergs per second, is nearly constant, varying by less than 1%. However, certain regions of the electromagnetic spectrum, as well as the energies and intensities of the charged particles, are highly variable, both in time and in the location on the Sun from which they are emitted.

Fig. 12 is a graph of the amount of energy that the Sun radiates at each wavelength. The Sun radiates 99% of its energy in wavelengths between 3000 and 40,000 Å. The atmosphere is transparent to this portion of the solar spectrum (with the exceptions of the various absorption bands at 6884, 7621, 9419, 11,350, 14,000 and 18,000 Å shown cross-hatched in fig. 12) and consequently it has been studied extensively at ground-based observatories. The spectral distribution in this portion of the spectrum is similar to that of a 6,000° K. black body, and the radiation comes almost entirely from the photosphere of the Sun. This region of the solar spectrum is of importance to space science because it is primarily responsible for the temperature of objects in space. The solar energy incident on a planet determines the temperature of the planet. Mercury, being closest to the Sun, has a maximum temperature of about 410° C. (770° F.) whereas Jupiter, which is 5 astronomical units from the Sun, has a temperature of −138° C. (−216° F.). As a spacecraft moves closer to the Sun, more solar energy falls on it and its temperature rises.

The shorter wavelengths, below 2900 Å, which are absorbed by the atmosphere can only be studied with instruments on sounding rockets and satellites. Although the amount of energy involved is small, it is an interesting and important region of the solar spectrum. Below 2000 Å the character of the spectral-irradiance curve changes. It is no longer similar to that from a 6,000° K. black body; instead, there is a continuum similar to that of a 4,000° K. black body with a number of very intense emission lines. The intensity of these lines varies with solar activity. The radiation in this part of the spectrum comes not from the photosphere but from the chromosphere and corona of the Sun. Fig. 13 is an enlargement of that portion of the spectral-irradiance curve from 0 to 2000 Å, showing the more important emission lines and the continuum.

The region from about 1000 to 2000 Å is sometimes referred to as the vacuum ultraviolet, the region from 100 to 1000 Å as the extreme ultraviolet (EUV) and the region below 100 Å as the X-ray portion of the spectrum.

The Lyman alpha line of hydrogen (1216 Å) is the most prominent emission line in this region of the spectrum. The first measurements of this line were made with sounding rockets in 1949, and its intensity was continuously monitored on the first Orbiting Solar observatory (OSO-1) in March, April and May, 1962. All of these measurements indicate that the intensity of 6×10^{-7} w. per square centimetre is very nearly constant over a solar cycle. The OSO-1 observations indicated a systematic variation of a few per cent correlated with the number of active regions on the visible disk of the Sun.

Systematic observations of the relative intensities of the 304 Å (HeII), 284 Å (FeXV) and the 335 Å (FeXVI) lines were also made on OSO-1. The intensities of these lines varied in a systematic way with sunspot number and with the flux of 2,800 Mc. radio noise from the Sun. The intensities of the iron lines increased by a much larger factor than the intensity of the helium line, the differences presumably arising from the fact that these lines are produced at different levels in the solar atmosphere.

The solar corona also emits X-rays in the 0–100 Å region. Fig. 14 shows the solar spectral irradiance in the region from 0 to 100 Å for the normal corona and during a solar flare. The spectral-irradiance curve of the normal corona has a shape similar to that of a black body at 500,000° K.; however, the total energy radiated in this region of the spectrum is many orders of magnitude below

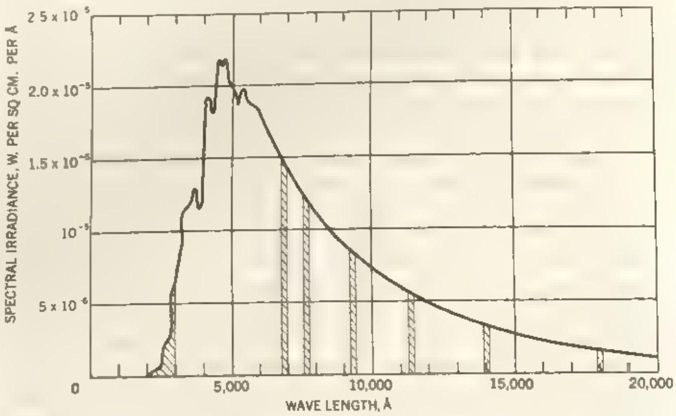


FIG. 12.—DISTRIBUTION OF ENERGY IN THE SOLAR RADIATION INCIDENT ON THE EARTH'S UPPER ATMOSPHERE. NARROW SHADED LINES ARE REGIONS OF STRONG ATMOSPHERIC ABSORPTION; SHADED AREA BELOW 3000 Å IS PORTION OF THE ULTRAVIOLET SPECTRUM ABSORBED BY THE ATMOSPHERE

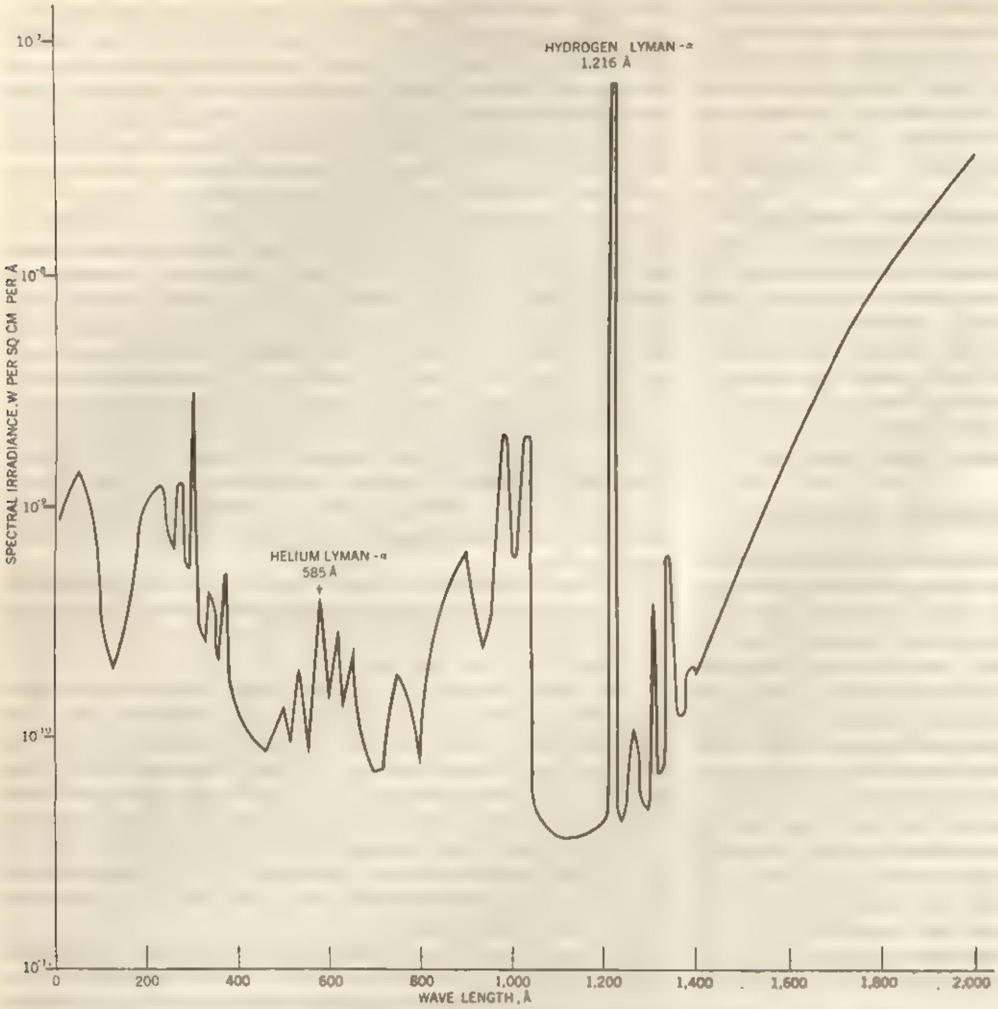


FIG. 13.—SOLAR ULTRAVIOLET SPECTRAL IRRADIANCE ABOVE THE EARTH'S ATMOSPHERE (*see* TEXT)

that which a body the size of the Sun would radiate at this temperature. At shorter wavelengths, 8–20 Å, the intensity varies markedly with solar activity. A typical value in this region is $0.1\text{–}2 \times 10^{-8}$ w. per square centimetre; for the region below 8 Å, 10^{-13} to 10^{-15} w. per square centimetre. During solar flares the shortest wavelength observed is normally about 1 Å although on a few occasions 0.1 Å X-rays have been observed, and on very rare occasions X-rays of 0.02 Å have been noted.

The values quoted for the fluxes of the various regions in the ultraviolet and X-ray spectrums are preliminary and in some cases are based on fleeting measurements made by sounding rockets. The absolute values will undoubtedly change somewhat as systematic, long-term measurements are made in orbiting observatories similar to OSO-1. The study of the spatial and temporal variations of this portion of the solar spectrum should contribute to the solutions to some of the puzzling problems of the chromosphere, the corona and solar flares discussed in the article SUN.

In addition to their importance in leading to an understanding of solar phenomena, the radiations emitted in the ultraviolet and X-ray region of the spectrum are equally important in the understanding of the physics of the upper atmosphere. Fig. 15 shows the altitudes at which the various portions of the spectrum are absorbed by the atmosphere. The properties of a particular altitude region of the atmosphere are determined by the radiation which is absorbed there. The interaction between solar radiation and the atmosphere will be discussed below under *Interplanetary Phenomena.*

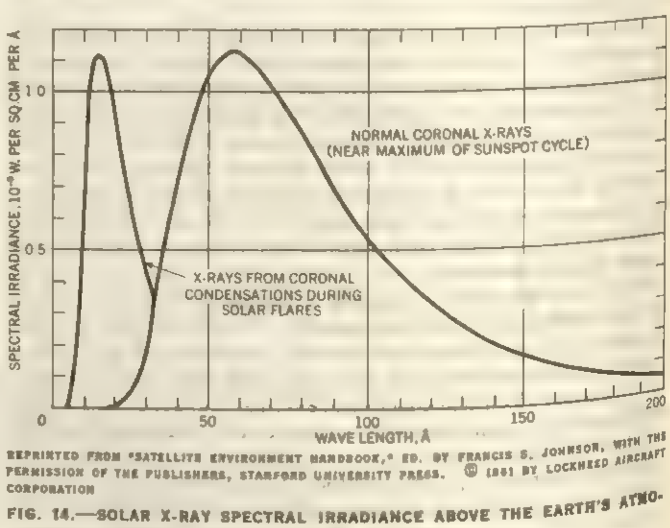
The Sun also continuously emits a flux of low-energy charged particles generally referred to as the "solar wind." During large solar flares the Sun emits large numbers of high-energy protons or solar cosmic rays. The solar wind and solar cosmic rays are best observed in interplanetary space, and the solar wind in particular controls many of the properties of interplanetary space.

B. INTERPLANETARY PHENOMENA

Interplanetary space is defined as that region in the solar system whose properties are primarily determined by solar phenomena rather than by the presence of other bodies such as the planets. The Earth and its magnetic field, for instance, influence a region of space extending out about 70,000 km. on the daylight side of the Earth and about 300,000 km. on the night side. Outside this region the properties of space are determined by solar phenomena.

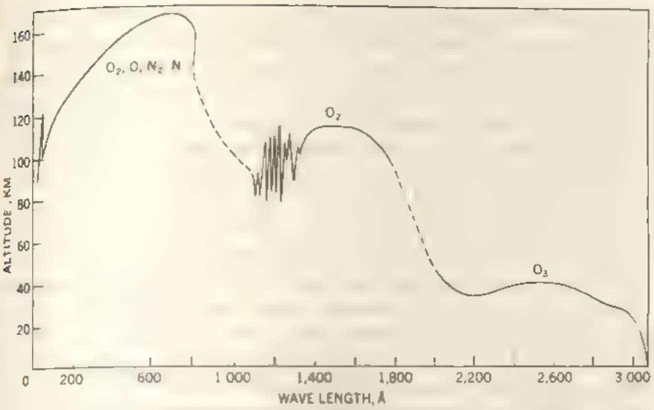
Four interplanetary phenomena will be discussed: the solar wind; the interplanetary magnetic field; solar and galactic cosmic rays; and micrometeorites. These phenomena determine the conditions in interplanetary space and vary appreciably with both time and position in space. In order to understand and describe their behaviour it is necessary to make measurements in interplanetary space.

The solar wind is a moving solar plasma (a plasma consists of a gas of electrically charged particles). Although the individual particles of a plasma are charged, the charges are both positive and negative, and the net charge on any finite volume of plasma is neutral. The solar plasma consists primarily of protons and, presumably, electrons, although the electrons have not yet been measured. The protons in this plasma have energies of about 1 Kev (thousand electron volts), which corresponds to a velocity of about 500 km. per second. Occasionally there is a period of activity on the Sun followed by changes in the behaviour of the solar wind. The number of the particles increases (for example, from less than one per cubic centimetre to more than ten) and their velocities also increase.



REPRINTED FROM "SATELLITE ENVIRONMENT HANDBOOK," ED. BY FRANCIS S. JOHNSON, WITH THE PERMISSION OF THE PUBLISHERS, STANFORD UNIVERSITY PRESS. © 1961 BY LOCKHEED AIRCRAFT CORPORATION

FIG. 14.—SOLAR X-RAY SPECTRAL IRRADIANCE ABOVE THE EARTH'S ATMOSPHERE (*see* TEXT)



FROM J. A. RATCLIFFE, "PHYSICS OF THE UPPER ATMOSPHERE"

FIG. 15.—PENETRATION OF SOLAR ULTRAVIOLET RADIATION INTO THE ATMOSPHERE. ABSORPTION ABOVE 2000 Å IS DUE PRINCIPALLY TO OZONE. BETWEEN 850 AND 2000 Å TO MOLECULAR OXYGEN AND BELOW 850 Å TO ALL CONSTITUENTS

The interplanetary magnetic field has a value of about 5×10^{-15} gauss. The direction of the field lies approximately in the plane of the ecliptic and at an angle of 40° to 50° with the Earth-Sun line. The magnetic field is carried outward from the Sun by the solar wind; hence, it fluctuates with solar activity. Although the field is very weak, it influences the trajectories of cosmic rays since it exerts a continuous force on any moving charged particle.

Galactic cosmic rays are thought to originate outside the solar system. Their intensity shows only a very small variation with solar activity and does not vary with direction. Their properties are discussed in detail in the article COSMIC RAYS. After a major solar flare the intensity of protons in the energy range from about 1 to several hundred Mev (million electron volts) increases by several orders of magnitude in the first few hours and then gradually returns to normal over a period of several days. During the first hour after a flare, while the intensity is increasing, the number of particles observed depends upon the direction of observation. It must also depend upon the position of observation in space, although no direct observation of such a variation has been made. The protons responsible for the increase in intensity are called solar cosmic rays. Except for their large variation in intensity and position they are identical with galactic cosmic rays. In addition to protons there are the positively charged nuclei of heavier atoms—helium, nitrogen, oxygen, and so on up the atomic scale. The relative abundances of these nuclei in the solar cosmic rays differ from those in the galactic cosmic rays and reflect the relative abundances of the elements in the Sun.

About two particles will strike a sphere of 1 sq.cm. cross-sectional area in 1 sec. from the normal galactic cosmic-ray flux. About 100,000 particles will strike a similar sphere in 1 sec. at the peak of a major solar flare.

Meteoroids are a class of astronomical objects which travel in eccentric orbits about the Sun. A meteoroid may strike the Earth's atmosphere and produce an observable effect called a meteor or "shooting star." The solid body which constitutes a meteoroid is called a meteorite and may vary in mass from 10^{-14} g. to several thousand kilograms. The very light meteorites, with mass of less than about 10^{-6} g., are called micrometeorites or "cosmic dust" and appear to be concentrated in the vicinity of planets. Measurements made by Pioneer 1 and Mariner 2, when compared with measurements made by satellites in the vicinity of the Earth, showed that the flux of dust at great distances from the Earth was almost 1,000 times smaller. The material in space is dominated by particles with a mass of about 10^{-9} g. About one particle of mass 10^{-10} g. strikes an area of 1 sq.m. in 1 sec. near the Earth. The integrated accretion rate over the entire surface of the Earth is about 10,000 tons per day.

C. THE EARTH

The Earth, of course, is the planet which has been studied most extensively. It is convenient to divide material in the realm of space science under four major headings: Geodesy, the study of the solid earth; the Upper Atmosphere; the Ionosphere; and the Magnetosphere.

1. Geodesy.—A satellite is a powerful tool for studying the gravitational field of the Earth and for locating points on the surface of the Earth. The position of a satellite in orbit can be determined within limits of about 10 or 20 m. By photographing a passive satellite or a satellite carrying a flashing light it is possible to determine its location relative to the fixed stars. Then, by determining the orbit precisely, studying variations in the orbit and correcting for those variations resulting from atmospheric drag and solar radiation pressure, it is possible to determine the gravitational potential and the exact shape of the Earth. From a knowledge of the gravitational potential it is possible to determine the variation of density within the Earth and to draw some inferences about the internal strength of the Earth.

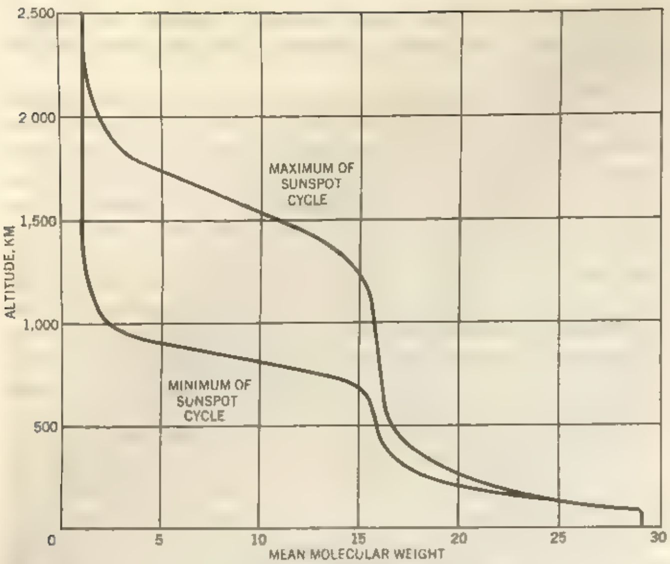
Simultaneous observations of a satellite can be used to determine precisely the relative locations of the observing stations on the surface of the earth. In this way triangulation measurements can be extended over great distances. (*See* also GEODESY.)

2. Upper Atmosphere.—As altitude increases, the first significant departure from the sea-level composition of the atmosphere occurs at about 30 km., where significant amounts of ozone (O_3) are observed. The upper atmosphere will be considered here as beginning at this altitude and extending up to about 200 km. Above this point, the properties of the atmosphere are determined by ions and free electrons and will be discussed in the section entitled *Ionosphere*. The upper atmosphere is studied primarily with sounding rockets; balloons are not feasible above 40 km. and satellites cannot be flown below about 200 km.

In order to describe and understand the atmosphere at a particular point, it is necessary to know its composition, temperature, pressure and density. All of these quantities vary with altitude and latitude, the time of day, the season and the solar cycle. An official publication, *A United States Standard Atmosphere*, gives these quantities as a function of altitude for an idealized middle-latitude atmosphere averaged over the yearly and solar-cycle variations.

The variation in temperature with altitude is largely controlled by the absorption of solar radiation. The rapid increase in temperature starting at about 100 km. and extending to about 300 km. marks the region in which ultraviolet radiation is absorbed. No temperature increase is noted above about 300 km. because no large amount of energy is absorbed in this region. Fig. 15 is a plot of the altitude at which about $\frac{2}{3}$ of the photons of a particular wavelength have been absorbed. The longer wavelengths (1000 to 2000 Å) are absorbed below about 50 km. by the ozone in the Earth's atmosphere. The wavelengths between 1000 and 2000 Å are absorbed by the oxygen in the atmosphere in the region between 50 and 150 km. The extreme ultraviolet radiation (500 to 1000 Å) is absorbed in the region between about 140 and 160 km.

Solar radiation is absorbed in the atmosphere by means of several processes, the two most important of which are photoionization and dissociation. In photoionization a neutral atom or molecule absorbs a photon and emits an electron, thereby forming an electron-ion pair. The electron will have an energy determined by the amount of energy required to liberate it from the atom and the amount of photon energy absorbed. In general, the electron will be emitted with an energy greater than the thermal energy of the ambient neutral particles. The electron then loses this energy by one of three processes: (1) inelastic collision with neutral particles; (2) elastic collision with ambient electrons; or (3) elastic collision with neutral particles and ions. The electron may remain a free electron or it may attach itself to a neutral molecule, thereby forming a negative ion. At altitudes below 60 km. electrons rapidly attach themselves to air molecules to form negative ions, the primary effect being an increase in conductivity of the air. At altitudes above 60 km. during the day, the density of the air is so low that the probability of attachment is small; the rate of production of electrons is high and thus there are an appreciable number of free electrons. At night the production



REPRINTED FROM "SATELLITE ENVIRONMENT HANDBOOK," ED BY FRANCIS S. JOHNSON, WITH THE PERMISSION OF THE PUBLISHERS, STANFORD UNIVERSITY PRESS. © 1961 BY LOCKHEED AIRCRAFT CORPORATION

FIG. 16.—VARIATION OF MEAN MOLECULAR WEIGHT OF THE ATMOSPHERE WITH ALTITUDE, SHOWING VALUES AT THE TWO EXTREMES OF THE SUNSPOT CYCLE

of electrons is very low, and those that are produced attach themselves to molecules, forming negative ions. Above 120 km. the density is so low that the electrons produced during the day remain free electrons during the night.

In the second major process, dissociation, a neutral molecule absorbs a photon and separates into two atoms. An example of this is the dissociation of the O_2 molecule into atomic oxygen. At very low altitudes, around 30 km., the neutral atoms of oxygen attach themselves to an O_2 molecule forming ozone, O_3. At higher altitudes, above 90 km., the density is low enough so that the probability of attachment to an O_2 molecule is small compared to the rate of dissociation and there is a large amount of atomic oxygen. The net effect is that there is a variation of the mean molecular weight of air with altitude. Fig. 16 shows the mean molecular weight for the two extremes of the solar cycle. At low altitudes, below about 90 km., the atmosphere has a mean molecular weight corresponding to the normal proportions of molecular oxygen and nitrogen. From 90 km. upward, the mean molecular weight drops rapidly as the oxygen molecules dissociate. The sharp drop in the mean molecular weight at 800 km. at solar minimum and 1,600 km. at solar maximum is due to the presence of the lighter constituents in the atmosphere. In this region the atmosphere is in diffusive equilibrium; therefore, the lighter components tend to separate out and become the dominant constituents at higher altitudes. Measurements on Explorer 8 showed that there is a large amount of helium ions present at about 1,600 km. At the very high altitudes, above 2,500 km., the mean molecular weight corresponds to that of atomic hydrogen.

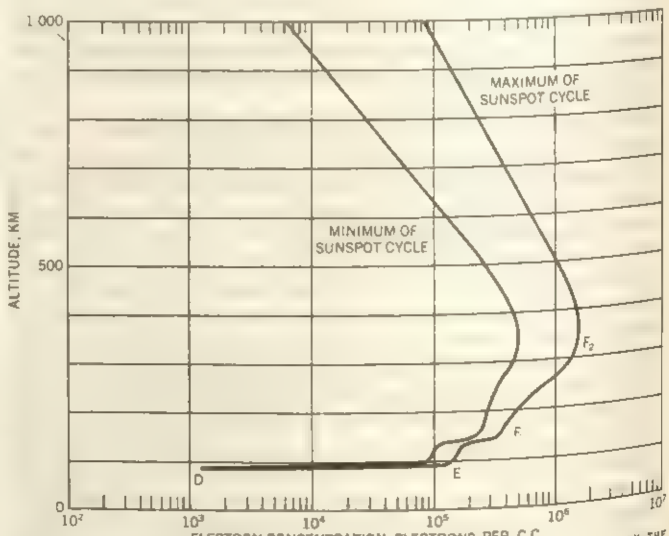
3. Ionosphere.—One of the major energy-absorption processes in the atmosphere is the production of electron-ion pairs. At altitudes below about 90 km. electrons freed by this process are promptly attached to another molecule, forming a negative ion. Above 90 km. the electrons persist for a considerable length of time: in the regions around 200 km. the high concentrations of electrons persist even through the night. The ionosphere was one of the first regions of the upper atmosphere to be studied, because it reflects low-frequency radio waves that can be produced and detected with elementary equipment. A radio wave propagating vertically from the surface of the earth will be returned from a level in the atmosphere where the electron density and the frequency of the radio wave are related by the formula $N_e = 1.24 \times 10^4 f^2$ electrons per cubic centimetre, where f is the frequency in megacycles per second.

Experimentally it is found that the electron density increases with altitude up to about 350 km. Therefore, the higher the frequency used to probe the ionosphere, the higher the altitude from which the signal is returned. The ionosphere has been studied by sending a variable-frequency radio signal vertically upward and measuring the length of time required for the echo to return. This time is then a measure of the altitude at which the wave is reflected. At altitudes above about 300 km. and a frequency above 10 Mc. no echo is returned. At higher frequencies the wave is propagated upward; that is, the ionosphere becomes "transparent."

During the early studies of the ionosphere it was observed that as the probing frequency was increased it would be reflected at a particular level up to a certain limit; then the altitude of reflection would change discontinuously to a higher level, where a wave of higher frequency would be reflected. Thus, it was thought that there were distinct layers, each of higher electron density than the one below. These observations led to the characterization of certain distinct ionospheric layers or regions, the so-called D, E, F_1 and F_2 regions. However, this concept of sharp layers of ionization with little or no free electrons in the intervening regions proved to be wrong as shown in fig. 17. The regions or layers are really locations where the electron concentration reaches a peak, remains relatively constant for some distance, and then again increases.

The D Region.—The D region extends from 60 km. to about 90 km. Photoionization is produced by the Lyman alpha radiation of hydrogen and the primary ions formed are NO^+. Fig. 15 shows that there is a "window" at the frequency of the Lyman alpha line which makes it possible for this radiation to penetrate to the 60 km. altitude. In addition to the normal D region ionization produced in this way, high-energy solar X-rays emitted at times of solar flares can increase the electron concentration in the D region, and can extend appreciably downward the lowest altitude levels of ionization. Also, protons from solar flares can cause an enhancement of the electron concentrations over the polar regions of the Earth. When such an event as this takes place the increased ionization in the D region completely blacks out radio transmission. The increased ionization is produced by protons in the energy range from a few Mev up to several hundred Mev.

The E Region.—The altitude range from about 90 to 140 km. is designated the E region. It has been established that the "soft" solar X-rays are responsible for the photoionization which occurs in this range. The dominant ionic species in this region are O_2^+ and NO^+. The electron density in the E region is of the order of 10^5 electrons per cubic centimetre at noon during sunspot minimum and is increased by about 50% during sunspot maximum. Electron concentration is highest near local noon and falls off roughly symmetrically with time on either side of noon.



REPRINTED FROM "SATELLITE ENVIRONMENT HANDBOOK," ED. BY FRANCIS S. JOHNSON WITH THE PERMISSION OF THE PUBLISHERS, STANFORD UNIVERSITY PRESS. © 1961 BY LOCKHEED AIRCRAFT CORPORATION

FIG. 17.—VARIATION OF ELECTRON DENSITY WITH ALTITUDE, SHOWN AT THE TWO EXTREMES OF THE SUNSPOT CYCLE

The F_1 Region.—There is no clear resolution between the F_1 region and the E and F_2 regions. The F_1 region is generally and rather arbitrarily defined as the altitude range from 140 to 200 km. The helium 304 line in the solar spectrum is responsible for the photoionization in this region. The photons of the helium 304 line have an energy of 41 electron volts. The energy required to free an electron is of the order of 13 electron volts, and therefore the electrons released in this region have a high energy and share their energy, through the mechanism of elastic collision, with the ambient electrons rather than with the ambient ions. Therefore, the electron temperature is several hundred degrees higher than the ion temperature during the daytime. Electron density in this region is of the order of 5×10^5 electrons per cubic centimetre at noon during sunspot minimum and 2×10^6 electrons per cubic centimetre during sunspot maximum. This region is not usually detected at night, when the electron density falls below 10^4 electrons per cubic centimetre.

The F_2 Region.—The F_2 region is not too well understood. It is thought that the helium 304 line from the Sun is also responsible for ionization there. The reason a second ionization peak is formed in this region is that the recombination rate decreases with altitude more rapidly than does the ionization rate, so that larger ion concentrations occur at higher altitudes.

Above the region of the second peak the electron concentration decreases as shown in fig. 17. The ionosphere in this region is studied by instruments carried on satellites, the first direct measurements being made by the early Sputniks. Alouette 1, a Canadian satellite, carried a swept-frequency "topside" sounder, an instrument similar to the conventional Ionosonde used on the ground. This satellite transmitted a radio signal in the 1–30 Mc. range which was reflected at a depth in the ionosphere given by the formula quoted earlier. In this case, since the transmitter was above the ionosphere, the higher the frequency of the transmitted signal the deeper it penetrated, until finally, at a certain critical frequency, it passed completely through the ionosphere and was reflected from the ground.

4. Magnetosphere.—The magnetosphere is defined as the region of space surrounding the Earth, influenced by the Earth's magnetic field, which begins at an altitude of about 400 km. and extends to about 10 Earth radii (64,000 km.) on the sunlit side of the Earth and to perhaps 40 Earth radii (256,000 km.) on the dark side. It is the region in space in which charged particles oscillating from the northern to the southern hemisphere have a lifetime that is long compared to the time required for a single oscillation. It extends out to a boundary which is sharp and well defined but whose location in space varies with solar activity and with the angle to the Earth-Sun line.

At the boundary the magnetic field changes direction sharply and decreases in strength. The intensity of the trapped radiation also decreases sharply. The transition region or boundary is only about 70 km. thick. Inside the boundary the magnetic field agrees with that predicted for a dipole magnetic field. Outside, the field is turbulent for a distance of several Earth radii until the conditions of interplanetary space are reached.

A charged particle moving in a magnetic field experiences a force at right angles to its direction of motion and to the magnetic field. As a result of this force the particle will spiral along a magnetic line of force. In a uniform magnetic field the particle would maintain a constant velocity and a constant "pitch angle" between its velocity vector and the magnetic field. If, however, it is in a region in which the strength of the field is changing, the pitch angle and the size of the spiral will change.

It is believed that the protons in the inner region arise from the decay of neutrons produced by the interaction of solar and galactic cosmic rays with air nuclei.

D. The Moon and Planets

The discussions in the articles Moon and Planets are based for the most part upon data from ground observations and theoretical analyses. The program of systematic exploration of the Moon and planets with unmanned and manned spacecraft did not begin until the 1960s. Soviet scientists measured the strength of the lunar magnetic field at the Moon's surface with Luna 2 and photographs of the far side of the Moon were obtained with Luna 3 (1959) and Zond 3 (1965). Ranger 7 in July 1964, together with Ranger 8 in Feb. 1965 and Ranger 9 in March 1965, provided more than 15,000 high-resolution photographs of the lunar surface. These photographs showed the lunar surface in the *maria* to be peppered with craters of various sizes and to have a smooth undulating contour. In 1966 the Soviet Luna 9 and the U.S. Surveyor 1 spacecraft "soft landed" on the Moon and provided close-up photographs of the lunar surface. The Surveyor 1 pictures showed the surface to be composed of granular material that ranged widely in size; coarse blocks of rock and smaller fragments contrast with fine particles too small to be resolved by a picture resolution of ½ mm.

The first measurements in the vicinity of a planet were made by Mariner 2, which passed Venus at a distance of 34,760 km. Magnetometers and radiation detectors on board the spacecraft showed no change in the magnetic field and no increase in the radiation level. This implies that Venus has no radiation belt, at least none extending out to 35,000 km., and that any field that it may have does not exceed the interplanetary magnetic field at that altitude. Two experiments to measure the planet's temperature gave a surface temperature of about 427° C. (801° F.).

In July 1965 Mariner 4 passed Mars at a distance of 9,868 km. (6,118 mi.) and transmitted 22 photographs (19 of them useful) of about 1% of the planet's surface to the Earth. The pictures revealed a barren, crater-pocked surface that resembled that of the Moon. Mariner 4 also discovered that the Martian atmosphere was thinner than had been expected, and found no evidence of a magnetic field or of a surrounding belt of radiation.

E. Astronomical Phenomena

The atmosphere absorbs the ultraviolet light from the stars just as it does from the Sun. However, while the Sun emits the bulk of its energy in the visible region of the spectrum, there are stars which emit most of their energy in the ultraviolet portion. In addition to absorbing a major portion of the electromagnetic spectrum, the atmosphere also distorts and scatters a large fraction of the light. For these reasons extremely accurate pointing of large telescopes for long periods of time is required.

The ultraviolet spectra of stars have been measured with instruments on sounding rockets. Preliminary results indicate that intensity in the ultraviolet is below that which would be predicted from our present understanding of stellar processes.

Explorer 11 carried an experiment to measure the flux of high-energy gamma rays. This experiment gave a very low flux of gamma rays and gave no evidence of any discrete sources.

Experiments with sounding rockets have given evidence for discrete sources of X-rays. If subsequent experiments confirm these early results the measurements of the X-ray flux will help to understand stellar processes. (J. E. N.)

V. SPACE SCIENCE: LIFE SCIENCES

Bioastronautics, a term used to describe the space life sciences, includes acquisition of knowledge in space biology and application of this knowledge to manned space exploration.

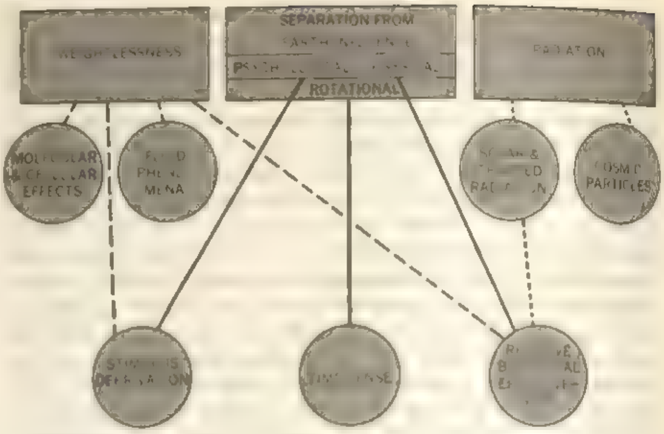
Of the technologies which transform basic scientific knowledge into practical use, biologistics is perhaps the broadest. It includes all knowledge, techniques, equipment and supplies required for life support and crew protection, such as environmental control systems, space suits, food, water and waste management devices. This article will treat bioastronautics in two broad categories. The first, concerned with the search for fundamental scientific knowledge, will be called space biology. The second will describe applied bioastronautics in terms of major unsolved technical problems and the special knowledge, procedures and equipment needed to support astronauts in space.

A. Origin of Bioastronautics

Although men have for many centuries speculated on the probable biological effects of the space environment, the first serious investigations of space hazards resulted from a near-disaster in

1862 when two British balloonists, having reached an altitude of about 29,000 ft., narrowly escaped death from acute oxygen deprivation or hypoxia. The French physiologist Paul Bert, stimulated by reports of this experience, began studying respiratory physiology and the biological effects of barometric pressure.

Bert had the first low-pressure chamber suitable for studies on human subjects. His classic observations formed the basis for aviation medicine, a biomedical specialty which developed slowly until the fourth and fifth decades of the 20th century, when revolutionary advances in aeronautics resulted in a broad and vigorous program of research. One of the first units devoted to biomedical research exclusively in space problems was the department of space medicine, established at the U.S. air force School of Aviation Medicine in 1949 by Harry G. Armstrong, an aviation medicine research pioneer. The rapidly growing body of knowledge in aviation medicine, together with technical developments in aircrew protective equipment, by the time of Sputnik 1 and Sputnik 2 formed the scientific and technological base for bioastronautics in the space age.



BY COURTESY OF NASA

FIG. 18.—BIOLOGICAL EFFECTS OF THE SPACE ENVIRONMENT, SHOWING INTERRELATIONSHIP OF THE VARIOUS EFFECTS (*see* TEXT AND TABLE VI)

B. Space Biology

1. Environmental Biology.—Underlying the many problems in the adaptation of man to space flight are fundamental considerations of the relationship between any living organism and its environment. In a broad sense, the definition of life and its fundamental characteristics are established by this relationship.

Space flight offers an unprecedented opportunity for studying the influence of environment on life characteristics and processes; namely, to determine what effects may proceed from the absence of the normal gravitational force and rotation of the Earth and the presence of ionizing radiation in levels much greater than those on Earth. The absence of the normal gravitational force to which Earth organisms are adapted through evolution may be expected to have effects on the systems in Table VI, with interactions as shown in fig. 18.

Table VI.—*Biological Processes for Study in Space*

A. The role of gravity in:
- (1) Plant morphogenesis
- (2) Animal morphogenesis
- (3) General physiology of cells:
  - (a) Cell nutrition
  - (b) Protein and nucleic acid syntheses

B. Effects of weightlessness (lack of gravitational stimulus) on:
- (1) Animal co-ordinative processes (*i.e.*, orienting reflexes)
- (2) Plant and animal fluid transport (*i.e.*, blood and lymph flow)
- (3) Mammalian physiological processes (*i.e.*, homeostatic mechanisms)

C. Space radiation effects:
- (1) Biological effectiveness of particles having energy not available from accelerators
- (2) Possible synergism between ionizing radiation and other factors in orbital flight (*i.e.*, weightlessness, altered rhythmic processes)

D. Effects of removal from the Earth's rotation on diurnal cycles, time-dependent phenomena, growth of spiral structures

Experimental systems were under development in the 1960s in the biosatellite project of NASA; it was planned to put recoverable spacecraft in orbit for periods of up to 30 days, carrying a wide range of biological experiments involving organisms from bacteria to subhuman primates, to seek answers to the problems outlined in Table VI and fig. 18.

2. Exobiology.—Exobiology is the science of extraterrestrial life. The most revolutionary finding anticipated in the exploration of space would be the discovery of extraterrestrial life or evidence of its preexistence. Aside from its popular impact, this discovery and subsequent investigation would have deep bearing on such topics as the origin of life, organic evolution and even the evolution of the universe.

One of the most widely accepted characteristics of life is the ability to derive energy and matter from the environment and to utilize them for growth and reproduction. Chemical studies on the synthesis of carbon compounds associated with terrestrial life suggest that the synthesis of such substances must occur with great frequency in the universe; therefore, the probability of multiple formation of carbon-based life in the universe is very high. (The probability that life may be derived from other chemical systems is not ruled out, but it is much more difficult to evaluate.) Knowledge of terrestrial evolution leads to the fruitful proposition that life, once it appeared, would evolve into a wide variety of forms to exploit the environment. This proposition raises the distinct possibility of sentient, communicating life forms elsewhere in the universe.

Within the narrow confines of the solar system, however, the most probable forms of life outside the earth are microorganisms (the extant theories of evolution all postulate such forms as a starting point of life). Therefore, exobiology is primarily concerned with the recognition of carbon-based microorganisms on planets of the solar system. Of these planets, Mercury may have too high a temperature for ordinary organic syntheses to occur. Venus, from data derived from Mariner 2, appears to be too hot for conventional biochemistry based on an aqueous medium. Jupiter is too remote for study in the present state of technology.

Including the data obtained by Mariner 4, Mars is still by far the most attractive prospect for the presence of life processes. The seasonal appearance of dark areas associated with the apparent thaw of polar caps may mean that some life forms, perhaps vegetation, revive as warmth and water become seasonally available.

The principal U.S. programs directed toward the investigation of life on Mars are the NASA Mariner and Voyager projects. These include "fly-by" and orbiter missions to gather and record data from electromagnetic radiation either emitted or reflected from the planet by means of spectroscopic analysis, television photography, thermal mapping, etc. Landing capsules were planned that would detect life by identifying some of its most universal characteristics (Table VII); physical-chemical methods were proposed to identify organic compounds. A typical life detection device is illustrated in fig. 19.

Table VII.—*Characteristics or Techniques Useful in the Search for Extraterrestrial Life*

1. Biologically significant elements and diatomic combinations
2. Carbon compounds; organic compounds; optically active compounds
3. Water in liquid form in sufficient amount to permit activity; theoretical substitutes for water
4. Temperatures less than 100° C.
5. Proteins, nucleic acids and amino acids
6. Special staining or dye reactions typical of macromolecules
7. Growth and metabolism
8. Carbon-dioxide assimilation (photosynthetic and autotrophic bacteria)
9. Enzymes or enzyme activity
10. Structure and organization typical of living systems
11. Gross observations by means of near-scan, high-resolution television
12. Bioelectrical phenomena

C. Problems of Man in Space

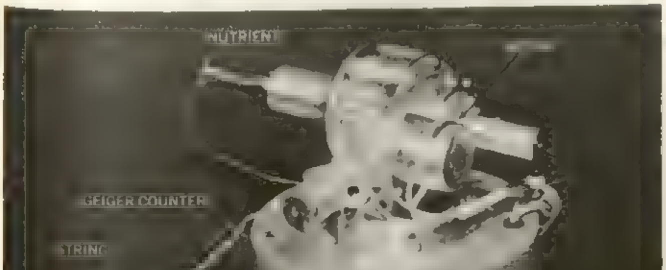
1. Biological Space.—A general scientific definition of space appears in the introduction to this article. In bioastronautics, the definition of space is necessarily associated with its relationship to man. For instance, the unadapted human being becomes partially incapacitated from hypoxia when exposed to altitudes above 10,000 ft. Equipped with pressure-breathing oxygen apparatus,

an aviator can function effectively up to altitudes of about 45,000 ft. However, at any altitude above 18,000 to 20,000 ft., and particularly above 30,000 ft., there is some risk of an attack of the "bends," or decompression sickness. At altitudes above 45,000 ft. an artificial pressure environment is essential to prevent severe decompression effects and to permit adequate oxygenation.

Where space begins is not determinable from a bioastronautics point of view. From a physiological standpoint, space equivalence in terms of atmospheric pressure may be said to begin at about 65,000 ft. The radiobiologist interested in manned space flight recognizes the vicinity of 90,000 ft. as an altitude at which space radiation begins to be significant.

2. Hazards and Stresses of Space Flight.—The environmental hazards of manned space flight include weightlessness, penetrating radiations, "hard" vacuum, temperature extremes, meteorites, and intense, though transient, accelerative and vibratory forces of launch, orbital injection, in-flight maneuvers, reentry and landing. Other factors which may pose problems of safety or effective performance for space travelers include reduced magnetic fields; empty visual fields; unattenuated solar radiation in the ultraviolet, visible and thermal zones; the absence of the customary terrestrial day-night cycle; and the vastness, desolation and silence of space.

Weightlessness.—The accumulating human experience in the "zero-g" weightless state typical of orbital space flight will show whether or not exposure to this state, prolonged for many days, weeks or months, will produce detrimental biological effects. The Mercury and Vostok flights showed that astronauts were able to carry on normally and efficiently as spacecraft operators. Some evidence has accumulated of adverse effects on arterial blood pressure, seen immediately after landing from orbital flights. A very practical question concerns the possibility that prolonged exposure to weightlessness may so de-condition the astronaut that his cardiovascular system may not be capable of compensating for the increased accelerative stresses of reentry, nor for the circulatory demands of standing or moving about during the first few hours or days back on Earth. Moreover, there is lively scientific interest in the possible influence of prolonged weightlessness on various metabolic and regulatory processes, particularly on calcium metabolism and the possibility of skeletal atrophy, kidney stone formation, and predilection for arterial blood clots as a result of elevated blood calcium levels. Similarly, the possible adverse effects on muscular size and strength must be elucidated. Since it is impossible to simulate prolonged weightlessness in Earth laboratories, the answers to numerous questions about its biological and behavioural effects must come from experience in actual space flight, whether from orbiting animal-carrying satellites or spacecraft with human crews.



BY COURTESY OF NASA

FIG. 19.—RADIOISOTOPE BIOCHEMICAL PROBE KNOWN AS "GULLIVER." LIFE-DETECTION DEVICE IS 5½ INCHES IN DIAMETER AT THE BASE AND HAS THREE "STICKY STRINGS," FIRED IN DIFFERENT DIRECTIONS BY BULLETS, THEN REELED BACK, CARRYING SOIL PARTICLES AND MICROORGANISMS INTO A NUTRIENT BROTH. BACTERIA, IF PRESENT, WILL GROW IN BROTH, AND APPROPRIATE SIGNALS WILL BE SENT TO EARTH

Radiation.—It was estimated that the Apollo manned lunar landing and return mission could be accomplished without exposing the crew to radiation of more than 25 rad (radiation absorbed dose; *see* RADIOLOGY: *Therapeutic Radiology: Dosage*). This includes radiation from all sources—the artificial electron belt, the Van Allen Belt, galactic sources and high-magnitude solar flares. An emergency total dose of 150 rad is regarded as a maximum consistent with mission completion and minimum crew safety. This much radiation, however, if absorbed in as short a time as one hour, might produce some symptoms of acute radiation sickness, such as nausea and malaise; however, control of these symptoms with standard drugs is considered feasible.

Since knowledge of the biological effects of high-energy protons and other high-energy space radiations is insufficient to make precise estimates of the hazard to space travelers, laboratory proton accelerators are being used, in part, for proton biological effects studies. There is a modest hope that at least two types of radiation treatment may evolve to a practical state: antiradiation prophylactic drugs and the use, after radiation exposure, of autologous bone marrow.

More powerful space boosters, capable of orbiting large payloads, will allow for the inclusion in the spacecraft design of considerably more inert radiation shielding than can be carried in, for example, the Gemini spacecraft. In the mid-1960s there was optimism regarding the development and perfection of a magnetic means of deflecting charged, high-energy particles from spacecraft.

High-Intensity Light.—Visible solar radiation in space is so intense that, without protective filters in his glasses, helmet visor or spacecraft windows, the astronaut viewing the Sun directly could suffer temporary blindness or even permanent retinal burns. The form of objects viewed in such unscattered solar light may be difficult to discern accurately because of the high contrast between the illuminated and nonilluminated parts.

Ultraviolet Light.—Unattenuated solar ultraviolet light is another hazard to the space traveler. This is not a serious danger because the space suit will efficiently shield against ultraviolet radiation. There is some concern, however, about the deteriorating effects of ultraviolet radiation on materials used in the construction of spacecraft and space suits.

Temperature Extremes.—The range of temperatures to which an astronaut might be exposed is well illustrated by temperature estimates of the Moon's surface. On the dark side at lunar midnight, which is about 160 hr. after lunar sunset, the surface temperature drops to an estimated −150° C. (−238° F.). Conversely, at lunar midday, it would reach a scorching 180° C. (356° F.). Since these temperature extremes are well beyond the tolerance of an unprotected man, the lunar explorer and the astronaut operating in free space outside the spacecraft require insulated pressure suits with reflective outer layers and built-in, temperature-controlled ventilation.

The Vacuum of Space.—Interplanetary space is an extreme vacuum (with a pressure in the order of 10^{-14} mm. of mercury). Exposed to such a condition, the unprotected human being would have less than 15 sec. of consciousness before the onset of acute hypoxia and immediate, catastrophic decompression symptoms including bends, chokes, palsies and ebullism (the profuse evaporation or "boiling" of body fluids). To protect the astronaut against the vacuum of space, U.S. spacecraft are designed to contain a normal operating pressure of 5 p.s.i.a. (pounds per square inch, absolute) of 100% oxygen, and a minimum emergency pressure of 3.5 p.s.i.a. In addition, the astronaut wears a pressure suit which will maintain an internal pressure of 3.5 p.s.i.a. for as long as may be necessary to correct any environmental control system malfunction or to complete the mission and return to Earth.

Isolation and Sensory Deprivation.—Space flights subject the astronauts to varying degrees of physical and social isolation. Laboratory experiments, as well as the experiences of castaways and explorers, indicate that man, when deprived of the customary sights, sounds, smells, vibrations and thermal and kinesthetic sensations, as well as social interactions, may experience profound psychological disturbances to the point of severely impaired judgment and noneffectiveness. Studies and experience show that communication, as by radio, reduces somewhat the sense of isolation. Space missions with multiple crew members, good communications with Earth stations and demanding work schedules will greatly relieve the relative isolation and reduced level of sensory experience.

Empty Visual Fields.—The so-called empty visual field has been important in aviation because it causes problems of accommodation and visual effectiveness. It has been shown that in the presence of an empty visual field, such as a cloudless sky in broad daylight, subjects cannot relax their visual accommodation sufficiently to assure a visual focus at infinity. Instead, the subject experiences a temporary nearsightedness. This situation interferes radically with effective visual search for targets located at visual infinity. Whether or not this impairment will be important in space operations remains to be proved. It is probable that the stars, planets and Moon will afford some of the needed detail in the distant visual fields to permit accommodation of the eye.

Acceleration.—Throughout a typical space mission the astronaut is subjected to an acceleration environment that differs radically from the customary terrestrial situation. The principal features of this unusual experience are: the intense, short-duration, positive accelerations of the launch and staging phases; the long period of weightlessness in orbit; and the various accelerative forces of orbital maneuvers; all culminating in acute deceleration during retrofire, atmospheric reentry and the landing impact.

Aerospace medical research has determined that, with the subject semireclined and well supported, the predicted accelerations in all proposed manned flights are not only tolerable but compatible with an acceptable degree of mental and visual effectiveness and competence in using the fingers and hands for control manipulations. Atmospheric reentry of a Gemini-type spacecraft from a low Earth orbit, such as 120 mi., is marked by a period of decelerative force which lasts for several minutes and may reach as high as 8 to 10 g. Fig. 20 shows two reentry deceleration curves which have been proved by centrifuge testing to be within human tolerance when the subject is properly positioned and supported.

3. Biologistics.—This subject includes the characteristic problems that arise in designing and equipping manned spacecraft. Space crew requirements, as listed in Table VIII, form the basis for many aspects of spacecraft design and are essential specifications for biologistics. From these requirements are evolved the designs for environmental control systems to regulate cabin temperature, pressure, atmospheric composition and humidity; space suits; acceleration couches and restraining devices; food- and waste-handling equipment; and other equipment needed to sustain the space crew throughout the mission.

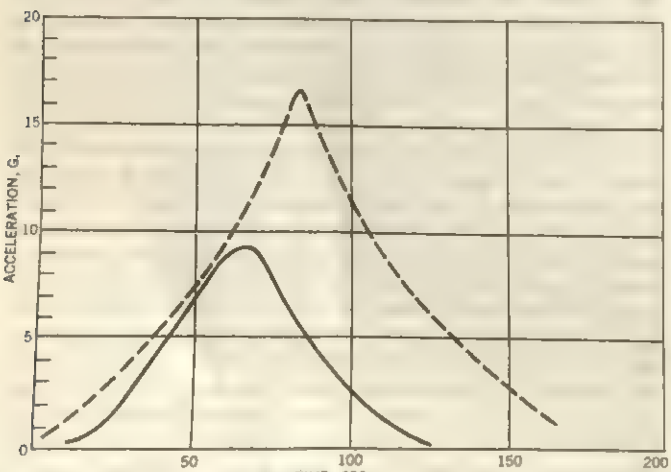


FIG. 20.—TWO EXPERIMENTAL REENTRY DECELERATION CURVES, SHOWING VALUES THAT ARE WITHIN RANGE OF HUMAN TOLERANCE

Manned space flights can be provided with oxygen, food and water for missions of six weeks or more. Partial recovery of water from human wastes and wash water, and regeneration of oxygen from carbon dioxide, would permit somewhat longer missions. However, for long-duration space flights of many weeks or months, in order to keep within the weight and space limitations imposed by the energy requirements for over-all mission completion, it will be necessary to eliminate all but an initial store of consumable supplies and to depend upon a closed ecological system for waste conversion and recovery of life-supporting essentials. In long-duration, earth-orbiting missions it might be feasible to resupply the spacecraft periodically by logistics vehicles. One of the most fascinating, if not the most difficult, technological challenges of the space age is the development of practical closed ecological system concepts in which essentially all waste products, both man- and machine-generated, could be converted for reuse. (O. E. Re.; J. M. Ta.)

TABLE VIII.—*Space Crew Requirements*

Area	Requirements	Item
Metabolic	Oxygen consumption	2.0 lb./man/day
	Carbon-dioxide output	2.3 lb./man/day
	Heat output	11,300 B.T.U./man/day
	Water consumption	6.0 lb./man/day
	Food consumption	2,800 kg.cal./man/day
Environmental	Cabin pressure	5 p.s.i.a. and 100% oxygen or 15 p.s.i.a. mixed oxygen-nitrogen
	Space suit pressure	3.5 p.s.i.a.
	Cabin temperature	75° F. ± 5° (rel. hum. 40%–70%)
	In-flight acceleration	(examples, "eyeballs in" or back to front) 10 g for 1.5 min., 9 g for 2 min., 7.5 g for 3 min.
	Noise levels	80 decibels over-all; 55 db in 600–4,800 cps range

VI. MANNED SPACE FLIGHT

1. Initial Programs.—A goal of both the U.S. and Soviet space programs was the achievement of manned orbital flight. Both nations based their projects on the use of existing ballistic-missile launch vehicles. Because the Soviet ballistic missile had been designed to carry a heavier warhead, their manned spacecraft could also be heavier than its U.S. counterpart; the U.S.S.R. Vostok spacecraft weighed approximately 10,400 lb., while the U.S. Mercury weighed about 3,000 lb.

Soviet Flights.—On April 12, 1961, Yuri Gagarin became the first man to orbit the Earth. In his spacecraft, Vostok 1 (East 1), he circled the Earth once, after a launching near Baikonour, U.S.S.R. On Aug. 6, 1961, Gherman Titov circled the Earth 17 times in Vostok 2, in a 25-hr. flight.

The duration of weightless flight was further extended in a "group flight" by Andrian G. Nikolayev and Pavel Popovich. Nikolayev was launched Aug. 11, 1962, and Popovich was injected into a nearly identical orbit one day later. Both remained in orbit until Aug. 15, Nikolayev completing 64 orbits and Popovich 48. At one point the craft came within 5 mi. of each other but were more than 1,000 mi. apart before reentry.

A second group flight took place in June 1963. On June 14 Valery Bykovsky was launched in Vostok 5. Two days later Valentina Tereshkova, the first woman cosmonaut, was launched in Vostok 6. Both spacecraft returned on June 19, Bykovsky completing 81 orbits and Miss Tereshkova completing 48. In neither group flight was there an attempt at rendezvous.

TABLE IX.—*Manned Space Flights*

Name	Nation	Pilot	Launch date	Weight (lb.)	Perigee/ Apogee (mi.)*	Number of orbits	Time (hr.:min.)
Vostok 1	U.S.S.R.	Yuri A. Gagarin	April 12, 1961	10,419	112/203	1	1:48
"Freedom 7"	U.S.	Alan B. Shepard, Jr.	May 5, 1961	2,855	116†	302‡	:15
"Liberty Bell 7"	U.S.	Virgil I. Grissom	July 21, 1961	2,836	118†	303‡	:16
Vostok 2	U.S.S.R.	Gherman S. Titov	Aug. 6, 1961	10,432	110/152	17	25:11
"Friendship 7"	U.S.	John H. Glenn, Jr.	Feb. 20, 1962	2,987	100/162	3	4:55
"Aurora 7"	U.S.	M. Scott Carpenter	May 24, 1962	2,975	99/167	3	4:56
Vostok 3	U.S.S.R.	Andrian G. Nikolayev	Aug. 11, 1962	10,412	113/146	64	94:22
Vostok 4	U.S.S.R.	Pavel R. Popovich	Aug. 12, 1962	10,425	112/147	48	70:57
"Sigma 7"	U.S.	Walter M. Schirra, Jr.	Oct. 3, 1962	3,029	100/176	6	9:13
"Faith 7"	U.S.	L. Gordon Cooper, Jr.	May 15, 1963	3,033	100/166	22	34:20
Vostok 5	U.S.S.R.	Valery F. Bykovsky	June 14, 1963	10,408	109/146	81	119:06
Vostok 6	U.S.S.R.	Valentina V. Tereshkova	June 16, 1963	10,392	113/143	48	70:50
Voskhod 1	U.S.S.R.	Vladimir M. Komarov Konstantin P. Feoktistov Boris B. Yegorov	Oct. 12, 1964	11,731	110/254	16	24:17
Voskhod 2	U.S.S.R.	Pavel Belyayev Alexei Leonov	March 18, 1965	12,529	107/308	17	26.02
Gemini 3	U.S.	Virgil I. Grissom John W. Young	March 23, 1965	7,111	99/139	3	4:53
Gemini 4	U.S.	James A. McDivitt Edward H. White II	June 3, 1965	7,879	99/184	66	97:56
Gemini 5	U.S.	L. Gordon Cooper, Jr. Charles Conrad, Jr.	Aug. 21, 1965	7,947	101/217	128	190:55
Gemini 7	U.S.	Frank Borman James A. Lovell, Jr.	Dec. 4, 1965	8,076	100/204	220	330:35
Gemini 6	U.S.	Walter M. Schirra, Jr. Thomas P. Stafford	Dec. 15, 1965	7,817	100/193	17	25:51
Gemini 8	U.S.	Neil A. Armstrong David R. Scott	March 16, 1966	8,351	99/186	7	10.41
Gemini 9	U.S.	Thomas P. Stafford Eugene A. Cernan	June 3, 1966	8,268	99/194	48	72:21
Gemini 10	U.S.	John W. Young Michael Collins	July 18, 1966	8,295	99/474	46	70:47
Gemini 11	U.S.	Charles Conrad, Jr. Richard F. Gordon, Jr.	Sept. 12, 1966	8,374	100/851	47	71:17
Gemini 12	U.S.	James A. Lovell, Jr. Edwin E. Aldrin, Jr.	Nov. 11, 1966	8,296	100/187	62	94:35
Soyuz 1	U.S.S.R.	Vladimir M. Komarov	April 23, 1967	n.a.	125/139	18	25:37

*Minimum perigee and maximum apogee, in statute miles. †Maximum altitude, statute miles (ballistic flight). ‡Range, statute miles.

United States Flights.—The first U.S. manned space flight took place on May 5, 1961, when Alan B. Shepard, Jr., flew a short-range ballistic trajectory in a Mercury spacecraft boosted by a Redstone launch vehicle. He reached an altitude of 116.5 mi., a speed of 5,100 m.p.h., and covered a distance of 300 mi. The period of weightless flight was about 5 min. On July 21, 1961, Virgil I. Grissom made a nearly identical flight.

The first U.S. orbital flight took place on Feb. 20, 1962, when John H. Glenn, Jr., completed three orbits in his Mercury spacecraft "Friendship 7." M. Scott Carpenter completed a similar flight on May 24, 1962, and Walter M. Schirra, Jr., made a six-orbit flight on Oct. 3, 1962. The last flight in the Mercury series took place on May 15–16, 1963, when L. Gordon Cooper, Jr., circled the Earth 22 times.

2. Gemini.—While the Mercury and Vostok programs demonstrated man's capabilities in space, their scope was too limited to permit the development of operational proficiency in manned space flight. In the United States the Gemini project was designed to continue where Mercury left off.

Objectives.—The Gemini project had two prime objectives: to extend the duration of flight to as long as two weeks, and to develop techniques for performing a rendezvous in space. The two-man Gemini spacecraft resembles the Mercury spacecraft in shape but is somewhat larger; it weighs more than twice as much as Mercury. The Gemini spacecraft is not fitted with the Mercury-type escape rocket; instead, ejection seats are used for emergencies. This is possible because the Gemini booster—the Titan II—uses storable hypergolic propellants which create a smaller explosion hazard than the liquid oxygen and kerosene used in the Atlas.

Rendezvous Maneuvers.—In a Gemini rendezvous mission, the "target" vehicle is placed into orbit first. The target can be an Agena stage or, alternatively, a passive "docking adapter," placed into orbit by an Atlas launch vehicle. (The target vehicle can also be another manned Gemini spacecraft.) Some time after being launched the target again passes over the launch site. For example, if the Agena's orbital period is exactly 90 min., then it will pass over the launch site 16 orbits or 24 hr. later, when the earth has completed exactly one rotation. If the Gemini spacecraft is then launched at the correct time, it could ideally be inserted into an orbit that is identical to that of the Agena, with the spacecraft and target flying close to each other. Under such ideal conditions, the pilots would then make small maneuvers, using their rocket "thrusters," to join the two craft.

For many practical reasons, these ideal conditions cannot be achieved in actual flights. First, it may be difficult to launch the spacecraft at precisely the right moment. A delay of one second allows the target vehicle to fly five miles beyond the rendezvous point. Second, the launch vehicle guidance systems, although extremely accurate, cannot insert the Gemini spacecraft into the exact same orbit as the Agena. (An error of $\frac{1}{10}$ of 1% in cutoff velocity would cause a height error of more than one mile, one-half revolution after insertion.) Under normal conditions, therefore, the moment of launch is extended to a "launch window" of several minutes duration. Catching up with the Agena and other orbit adjustments are accomplished during flight by using propulsion systems on the Gemini or Agena.

If the two craft are in the same orbital plane but are out of phase (*i.e.*, if they pass the same point at different times), the required corrections can be made with a minimal expenditure of fuel. In this case, one craft is placed into a more elongated orbit with a longer period; the other, in its original orbit with the shorter period, will then overtake the first. The orbit of the first is again shortened after the distance between the two craft has been decreased. These changes require relatively little propulsion.

If the two spacecraft are injected into different orbital planes, the corrective maneuver becomes more difficult. For a given plane change α, a velocity increase equal to $2V\sin\frac{\alpha}{2}$ is required, where V is satellite velocity. Thus, for a 5° plane change, a velocity increment of about 2,200 ft. per second would be required, while for a 60° change an increment equal to satellite velocity (about 25,500 ft. per second) is needed. Even with Agena propulsion, only minor plane changes are possible in Gemini.

A complete rendezvous maneuver, therefore, comprises the following: launch within a specified "window," plane change, phase change, rendezvous and docking.

Gemini Flights.—On March 23, 1965, astronauts Virgil ("Gus") Grissom and John W. Young made the first manned flight in a Gemini spacecraft, a three-orbit mission. The second manned Gemini flight, a four-day mission, took place in June 1965, with astronauts James A. McDivitt and Edward H. White II. White left the spacecraft for about 20 min. during the third orbit and maneuvered freely in space at the end of a 25-ft. tether.

BY COURTESY OF NASA

FIG. 21.—MERCURY AND GEMINI (TWO-MAN) SPACECRAFT, SHOWING COMPARATIVE SIZES

In Aug. 1965 L. Gordon Cooper, Jr., and Charles Conrad, Jr., circled the Earth for eight days in the third manned Gemini flight.

Two major Gemini objectives, long-duration flight and rendezvous, were achieved in Dec. 1965. On Dec. 4, astronauts Frank Borman and James Lovell, Jr., began a flight that lasted more than 330 hr., just a few hours less than two weeks. On the eleventh day of the Borman/Lovell Gemini 7 mission, astronauts Walter M. Schirra, Jr., and Thomas Stafford started their Gemini 6 flight, which culminated in a rendezvous with Gemini 7. The spacecraft flew in close formation for many orbits, prior to the return to Earth by Gemini 6 on Dec. 16 and Gemini 7 on Dec. 18.

In March 1966 astronauts Neil Armstrong and David Scott performed a similar rendezvous but with an unmanned Agena. After rendezvous they "docked" with the Agena, firmly joining the two craft. A subsequent short circuit in one of the spacecraft systems forced an early termination of the flight. Nevertheless, the pilots flew a precise reentry, landing within sight of a recovery airplane.

Astronauts Thomas Stafford and Eugene Cernan flew the third Gemini rendezvous mission, Gemini 9, in June 1966. They gained operational proficiency in rendezvous by joining their target vehicle three separate times. Cernan left his spacecraft for more than two hours in an extended extravehicular activity.

In July 1966 Young and Michael Collins flew the complex Gemini 10 mission, which included rendezvous and docking with an Agena; use of Agena propulsion to fly to an altitude of 474 mi.; rendezvous with an Agena left in orbit during the Gemini 8 mission; and extravehicular activity to retrieve a micrometeorite crater collection experiment from the Gemini 8 Agena.

In Sept. 1966 Charles Conrad, Jr., and Richard F. Gordon, Jr., docked the Gemini 11 with an unmanned Agena on the first orbit. Using the power of the Agena, Gemini 11 ascended with it to an altitude of about 850 mi. While outside the spacecraft for about 45 min., Gordon joined the two craft with a tether. Conrad later used the tether to rotate the Gemini 11 around the Agena in a station-keeping exercise that produced artificial gravity. Astronauts James A. Lovell, Jr., and Edwin E. Aldrin flew the last manned Gemini mission, Gemini 12, in Nov. 1966. The flight included rendezvous, docking with an Agena, two extended periods of extravehicular activity, a tethered station-keeping period and a precise landing within 3.8 mi. of the planned landing point.

Gemini Experience.—The experience gained in Gemini included: 1,940 man-hours in space; long-duration flight; precision maneuvers, culminating in rendezvous on 10 occasions, using 7 different methods; station-keeping with and without a tether; 9 exercises in docking with another vehicle; 11 periods of extravehicular activity, totaling 12 hr. and 35 min.; a record altitude of 851 statute miles; and precision-controlled landings, within approximately 7 nautical miles of the aiming point on the last 7 flights. Also, the astronauts successfully performed 40 scientific, technological and medical experiments. The Gemini flights clearly demonstrated that man can easily adapt to the weightless space environment for periods of at least 14 days and then readapt to the gravitational environment of earth.

3. Voskhod and Soyuz.—The first Soviet multi-manned space flight took place on Oct. 12–13, 1964, when Voskhod 1 ("Sunrise 1") orbited the Earth 16 times. Piloted by Vladimir M. Komarov, the craft also carried Konstantin P. Feoktistov, an engineer-scientist, and Boris B. Yegorov, a physician. The crew wore standard clothing rather than pressure suits during this flight; a landing system comprising parachutes and braking rockets permitted the craft and its occupants to land on the ground. Voskhod 2, a two-man vehicle piloted by Pavel Belyayev, was launched on March 18, 1965. During this flight Alexei Leonov, co-pilot, left his spacecraft briefly, becoming the first man to float freely in space.

In April 1967 the U.S.S.R. launched Soyuz 1, described as the largest and most complex spacecraft yet to be flown. However, the pilot, Vladimir Komarov, apparently encountered difficulties early in the flight. During the 18th orbit he attempted to land the craft, but it became tangled in its parachute cords and crashed to earth in central Asia; Komarov was killed, becoming the first known space flier to die during a space flight.

4. Lunar Landing Program.—In May 1961 the U.S. announced plans to land a manned spacecraft on the Moon by the end of the 1960s. The name given to the program was Apollo.

Lunar Landing Techniques.—Several techniques for the lunar mission appeared to be feasible and were studied in detail. These included the "direct method," wherein a single giant launch vehicle (Nova) would launch a spacecraft toward the Moon. Near the Moon the spacecraft would be slowed down by its own propulsion system, called the lunar landing stage. Another propulsion system, the lunar takeoff stage, would launch the spacecraft from the Moon and send it on its return trajectory toward Earth.

A second proposal, called the "earth-orbit rendezvous" method, differed from the direct method in that two somewhat smaller launch vehicles (Saturn V) would separately transport two parts of the spacecraft (*i.e.*, the lunar landing stage as one part, and the lunar takeoff stage and spacecraft proper as the second part) into Earth orbit. There they would be joined and then propelled as a single craft toward the Moon.

A third method, called the "lunar orbit rendezvous" method, would employ a single Saturn V launch vehicle to send the three-man Apollo spacecraft, weighing 90,000 lb., on a trajectory toward the Moon. Approximately 2½ days later, as the craft approached the Moon, its own propulsion system would place it into an orbit around the Moon. In lunar orbit one element of the spacecraft, called the lunar module (LM), with two men aboard, would separate from the mother ship. The two men would then fly the LM toward the Moon, while the third man circled the Moon in the other section of the spacecraft, the "command mod-

ule." After executing a "soft" landing on the lunar surface, the two lunar explorers would make scientific observations and collect geological samples. After about one day, they would take off from the surface and fly toward a rendezvous with the command module in lunar orbit. After the rendezvous, spacecraft propulsion would eject the spacecraft from lunar orbit and send it back toward the Earth. Precise guidance and control on the return trip would allow the spacecraft to enter within the limits of a rather narrow "reentry corridor." This corridor, about 25 mi. high, is limited at the lower side by too thick an atmosphere, which would cause the craft to exceed acceptable deceleration limits, and at the upper by so thin an atmosphere that the craft would not be "captured" and therefore would not reenter. Within the atmosphere the Apollo command module will have a limited degree of maneuverability. Although its lift-to-drag ratio is only about 0.34, the reentry path will be 1,000 mi. long and 200 mi. wide as a result of the high entry velocity.

The first manned Apollo flight, in earth orbit, was scheduled for Feb. 1967. On Jan. 27, 1967, during a countdown rehearsal, a fire broke out inside the spacecraft cabin and spread rapidly in the concentrated oxygen environment at sea level pressure. Astronauts Virgil I. Grissom, Edward H. White II and Roger B. Chaffee attempted to escape from the spacecraft but were unable to leave and lost their lives.

5. Manned Spacecraft Design.—Two factors are of major importance in the design of manned spacecraft: weight and reliability. The need for minimum weight becomes apparent when it is recognized that, for a manned lunar mission, 750 lb. of launch vehicle weight is required for every pound returned to Earth at the end of the mission; or, since the takeoff thrust is about 25% higher than the takeoff weight, an additional 1,000 lb. of thrust is needed for each added pound of spacecraft weight. Reliability is generally assured by providing redundant elements where a failure would critically affect the success of the mission. In manned missions the pilot can significantly enhance reliability by detailed failure analyses and corrective actions.

External Characteristics.—The external shape and materials of a manned spacecraft are governed by the necessity of reentering the Earth's atmosphere at speeds approaching satellite or escape velocity, depending on the type of mission. The tremendous heat generated during reentry is dissipated either by radiation from the craft's surface, or by ablating (evaporating or boiling away) some of the surface material. For example, in the Mercury and Gemini spacecraft (fig. 21), the major portion of the heat is dissipated by ablation from the blunt forward-facing heat shield; heat is also rejected from the conical afterbody, where thin (0.016-in.), high-temperature alloy "shingles" quickly reach an equilibrium temperature and give off heat by radiation.

The blunt shape of the forward face of the Mercury, Gemini and Apollo spacecraft also helps in the heat rejection process. This shape creates a strong shock wave which in turn dissipates much of the craft's kinetic energy; therefore, only a small part of the total energy is given off as heat at the vehicle's surface.

Major Subsystems.—Subsystems for providing a habitable atmosphere, for generating electric power, for determining and maintaining the correct attitude, flight path and velocity, and for communications are located inside the spacecraft.

The environmental control system provides oxygen (or air) at the proper temperature, pressure and humidity, and removes carbon dioxide and other contaminants. In all U.S. spacecraft, oxygen at a pressure of 5 p.s.i.a. is used for breathing purposes; in the U.S.S.R. Vostok spacecraft, a standard mixture of oxygen and nitrogen at a pressure of 14.7 p.s.i.a. is employed. Although the cabin atmosphere itself is habitable, the pilot is further protected by a pressure suit which allows him to survive in case the cabin is punctured by a meteoroid or is otherwise damaged.

Electric power sources include ordinary batteries, solar cells and hydrogen-oxygen fuel cells. For short missions, up to two or three days in duration, ordinary batteries are acceptable; for longer missions, solar cells or fuel cells can generate greater amounts of power with lower weight requirements. Fuel cells also produce potable water as a by-product.

The spacecraft's attitude is stabilized with small rocket motors, or thrusters. The thrusters can be fired automatically or manually by the pilot through the use of a control similar to an airplane control stick. Whichever method is chosen, the necessary information concerning the craft's attitude comes from an autopilot, which uses gyroscopes to measure attitudes and attitude changes. In Earth satellite missions it is important to stabilize the craft's attitude precisely about all axes of rotation just prior to retrofire, so that reentry will be made in the proper direction. In rendezvous missions and in lunar missions, attitude control also must be maintained while making course corrections and during docking maneuvers.

Guidance and navigation systems are employed to find the spacecraft's position and velocity in space, to determine course corrections that must be applied to follow the desired trajectory, and to measure changes in speed and direction while maneuvers are made. A typical guidance system consists of an inertial measurement unit (an array of gyroscopes and accelerometers) used to measure the acceleration vectors relative to a fixed point in space, and a guidance computer to translate acceleration measurements into trajectory data and to determine required course corrections. With this information the pilot then applies thrust in the proper direction.

Communications systems in manned spacecraft include voice receivers and transmitters, telemetry receivers and transmitters for sending measurement data, and radar beacons for transmitting tracking information to ground network stations. (G. M. Lo.)

VII. APPLICATIONS

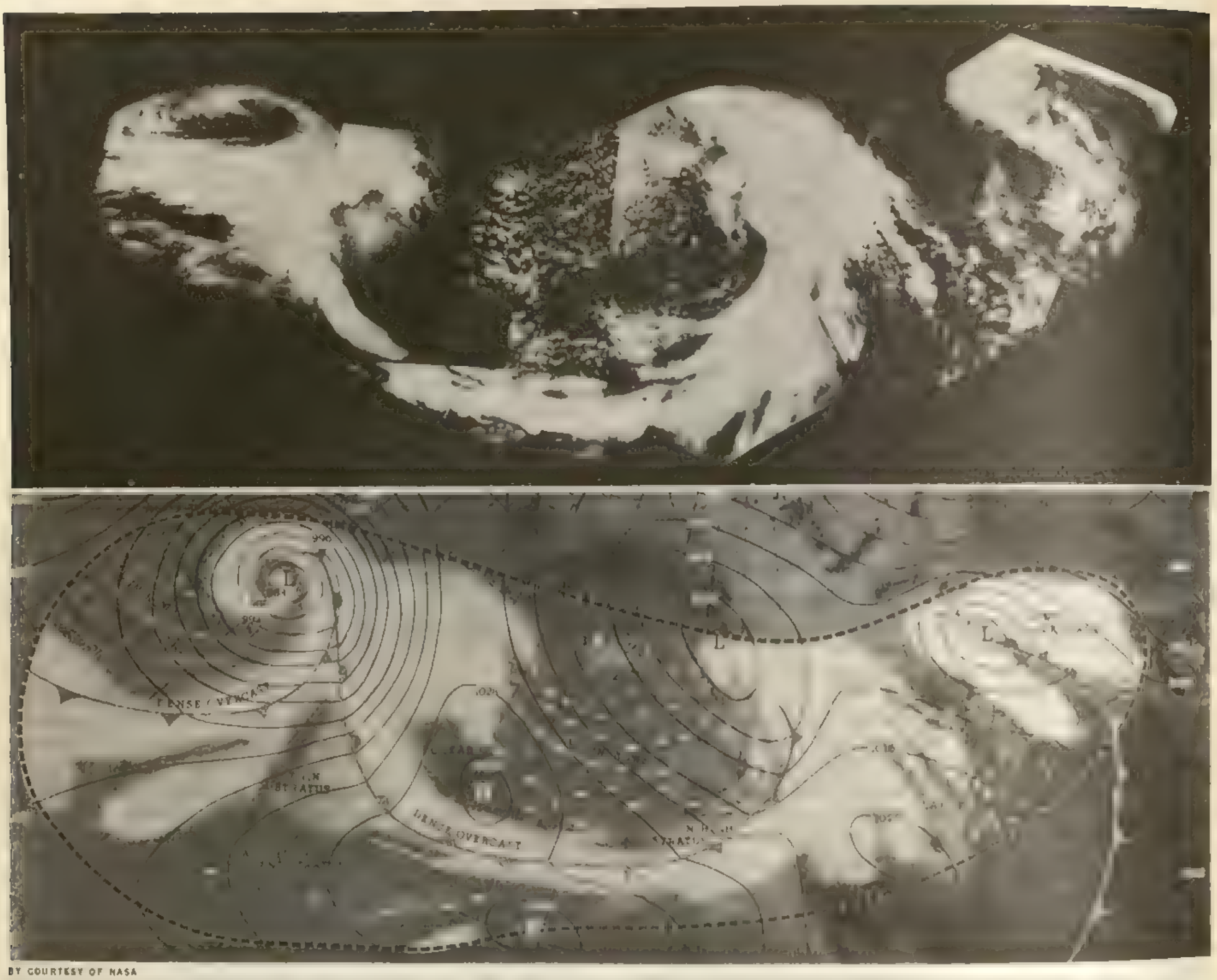
Satellites have several properties which make them uniquely useful: (1) their ability to provide a view of a large portion of the Earth's surface at any instant; (2) their ability to view all of the surface over a period of time if in a nearby polar orbit; and (3) their ability to view continuously the disk of the Earth's surface if in a synchronous orbit. Conversely, a satellite is mutually visible from a large number of widely separated points on the Earth's surface. Results obtained by the first communication, meteorological and geodetic applications satellites pointed the way to other useful applications, including navigation, agriculture and forestry, cartography and geography, and geology and hydrology.

1. Meteorological Satellites.—Satellites such as the Tiros series provide the means to monitor the Earth's cloud cover and temperature distribution. This information is telemetered to data-acquisition stations on the Earth in the forms of television pictures and temperatures as measured with infrared-sensitive radiation detectors. (*See* fig. 22.)

Since weather is affected by conditions over a large portion of the Earth and by the solar energy input to the Earth, it is important that meteorological data be obtained over the entire surface of the Earth. Complete global monitoring was previously impossible because of the many observation stations required and inability to place them in many areas, particularly in the ocean areas where severe weather disturbances originate.

Satellites such as Tiros and Nimbus provided coverage of some of these meteorological events for the first time. Satellite observations have led to the discovery of severe storms, hurricanes and typhoons long before they could be detected by conventional means. The Tiros system has led to the development of the world's first operational weather satellite system. The ESSA 1 and ESSA 2 satellites, operated by the U.S. department of commerce, provide global picture data on a daily basis for use in short-range weather forecasts.

The Nimbus 1 satellite, launched in Aug. 1964, carried an advanced vidicon camera system, a high-resolution infrared radiometer and automatic picture transmission (APT). This satellite sent daytime pictures to simple APT ground stations for local use. Nimbus 2, launched in May 1966, covers 150 APT stations around the clock, including 44 stations in 26 different countries. This satellite has the added capability of transmitting nighttime infrared photographs of cloud cover to specifically modified ground stations. In addition to the utilization of Nimbus data for meteorological research and weather forecasting, the data have been



BY COURTESY OF NASA

FIG. 22.—STORMS AND FRONTS AS SEEN BY TIROS. (TOP) MOSAIC OF TIROS PHOTOGRAPHS; (BOTTOM) WEATHER MAP FOR MAY 20, 1960, PREPARED BY USING TIROS CLOUD DATA

utilized in other areas of research such as glaciology, geology and oceanography.

Meteorological satellites have also led to a new understanding of the tremendously large area of influence of certain weather conditions. There was little doubt that this new capability to observe from above, and over the entire surface of the Earth, would result in better information on which to base weather forecasting and lead to a more thorough understanding of the nature of the weather processes, all of which would make the interpretation of observations more accurate. Meteorological satellites will place greater emphasis on remote sensing of the atmosphere to obtain the information that will make it possible to extend weather forecasting to two to three weeks in advance.

2. Communications Satellites.—Another promising application of satellites, which takes advantage of the property that satellites can be seen simultaneously from widely separated points on the earth, was first proposed in 1945 by Arthur C. Clarke. He suggested the use of a manned satellite station for relaying communications over large distances. Although they were unmanned, the Echo, Telstar, Relay and Syncom satellites demonstrated the utility of this means of communication.

Satellite Coverage.—Satellites such as Telstar and Relay are in intermediate-altitude orbits (3,000 to 6,000 mi. high) and therefore move with respect to the Earth. Thus, to provide continued communications between two widely separated points would require that satellites passing out of range be replaced by other satellites just coming into view. From 30 to 50 satellites would be required at these altitudes to provide essentially continuous communications, if the satellites were distributed randomly with respect to each other in their orbits.

The Syncom satellite was placed in an orbit at an altitude of 22,245 mi. At this altitude the period of rotation of the satellite about the Earth is exactly 24 hr. By selecting an equatorial orbit and a period that is the same as the Earth's rotational period, the satellite remains stationary or fixed relative to the surface of the Earth. Thus a single "synchronous" satellite of this type would be available continuously for communications between any two stations within its area of visibility, and three such satellites, suitably spaced, could provide global coverage with the exception of the extreme polar regions.

Types of Communications Satellites.—Telstar, Relay and Syncom are all "active" satellites; *i.e.*, they carry transmitter-receiver sets to receive, amplify and retransmit signals from the ground, as well as the necessary power supplies. There are also "passive" communications satellites which carry no electronics equipment but merely reflect a signal directed at them back toward the Earth. Examples are the 100-ft. inflatable sphere Echo I which was visible as a rapidly moving "star" in the evening skies, and its successor, Echo 2. Passive satellites require larger receiving antennas on the ground and larger amounts of ground transmitter power than active satellites because they reflect only a small fraction of the incident energy back to the receiving antenna.

Commercial Communications Satellites.—The successful development and demonstration of these satellites paved the way for a new era in commercial communications. The Communication Satellite corporation (Comsat), authorized by the U.S. congress in

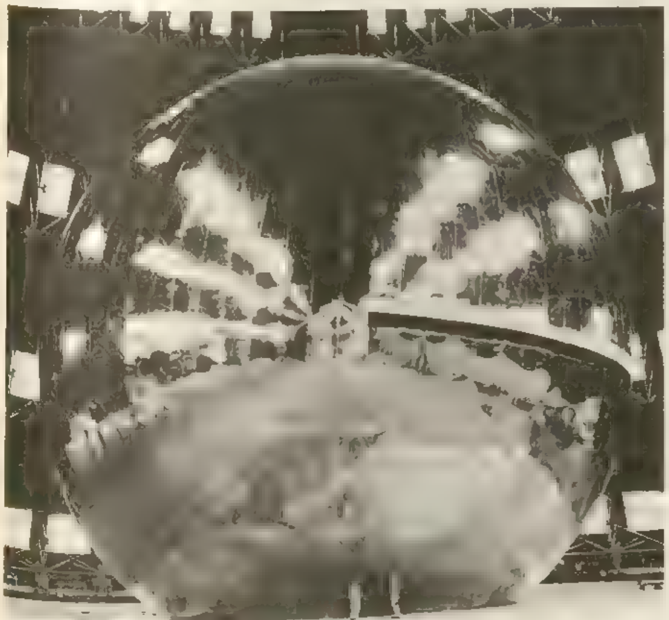
1962, and incorporated in Feb. 1963, was created to establish a global commercial communication satellite system. Comsat acts as the system manager for the International Telecommunication Satellite Consortium ("Intelsat")—the name adopted to describe the participants in the interim international arrangements for the financing and ownership of the space segment of the system.

Comsat's first satellite (Early Bird) was launched by NASA in April 1965 and placed in synchronous orbit over the Atlantic for operation between North America and Europe. Early Bird was placed in commercial operation in June 1965 and provides a basic capability of 240 transatlantic telephone channels. It has also been used for the transmission of television between the two continents.

Comsat was preparing to launch additional satellites in 1966 to support NASA's Apollo program (manned flight to the Moon). These satellites were to also provide circuits for commercial use and were to be placed in synchronous orbit over the Atlantic and Pacific oceans. It was planned that the first launchings of a truly global communication satellite system would begin in 1968.

3. Navigation and Data Collection Satellites.—Advances in space technology have made it possible to envision an Earth-orbiting satellite to achieve a new and improved worldwide, all-weather navigation system for use in sea and air traffic control and for the global collection of data pertinent to traffic routing and safety. Satellites, together with an associated ground station complex and transponders in craft using the system, have the potential capability of providing accurate position information automatically, quickly and reliably to commercial sea and air traffic, to control and routing centres and to air-sea rescue stations. Satellites could also augment this position data with weather forecasts, ocean sea-state reports and other information to ensure safety to both sea and air lanes. Direct long-distance voice communications between traffic controllers and the craft, and distress messages advising controllers of disabled craft location and situation could be provided by a communications channel in the satellite. Research and development work to determine the feasibility of such a satellite was planned for the mid- and late 1960s.

4. Geodetic Satellites.—The first satellite whose major objective was geodesy was ANNA 1-B launched in Oct. 1962. It contained three independent instrumentation systems to provide an increase in the quality of observational data, to permit the comparison of simultaneous measurements and to enable an integrated analysis of all the data. In Aug. 1964, the National Geodetic Satellite Program was initiated with the following objectives:



PICTORIAL PARADE

FIG. 23.—ECHO 1, A HIGH-ALTITUDE REFLECTOR FOR LONG-RANGE COMMUNICATIONS, SHOWN HERE DURING GROUND INFLATION TESTS. THE ALUMINUM-COATED BALLOON, WHICH INFLATES TO 100 FT. IN DIAMETER, WAS LAUNCHED AUG. 12, 1960

WIDE WORLD

FIG. 24.—TELSTAR 1, FIRST PRIVATELY OWNED COMMUNICATIONS SATELLITE, WAS LAUNCHED IN 1962, MAKING POSSIBLE THE TRANSMISSION OF LIVE TELEVISION PROGRAMS BETWEEN THE U.S. AND EUROPE

(1) connect geodetic datums to establish one world datum and adjust all local datums to the common centre-of-mass of the Earth so that positions of geodetic control stations will have a relative accuracy of ± 10 m. or better in an Earth centre-of-mass coordinate system; (2) define the structure of the Earth's gravitational field to 5 parts in 100,000,000 and refine the locations and magnitude of large gravity anomalies; (3) compare and correlate geodetic systems for an integrated analysis; and (4) further refine the positions of tracking stations relative to each other and relative to an improved world geodetic reference system.

To realize these objectives both optical and electronic systems, stationary and portable, are utilized to obtain tracking data from both active and passive satellites. The active geodetic satellites are the Beacon Explorers-B and -C, and GEOS-A and -B. The passive PAGEOS is a 100-ft. diameter sun-reflective balloon. The Beacon Explorers and GEOS-A have already been placed in orbit at altitudes of about 620 mi. PAGEOS was launched in June 1966 into a 2,635-mi. orbit.

Completed studies show that the Earth is an unsymmetrical oblate spheroid with differences in geoid heights of as much as ± 100 m. It is also evident that its shape, size and mass are sufficiently well known for most purposes. Small uncertainties in values for flattening of the Earth and the semi-major axis are significant, however, for space and scientific activities. Better information on the distribution of the Earth's mass and gravity, plus more accurate positions and an improved world geodetic datum, will permit more accurate deep space launches with less need for mid-course and terminal guidance corrections.

Advanced geodetic satellites now being investigated will use new observational systems such as hybrid ground stations, radar altimeters, gravity-gradiometers and other spacecraft-borne sensors. Resulting data will give important support to other scientific disciplines such as geophysics, geology, oceanography and celestial mechanics.

5. Earth Resource Survey Satellites.—It has been demon-

strated by such space programs as Tiros, Nimbus and Gemini that some of the Earth's natural and cultural resources can be surveyed more advantageously from space than by other means. Resources which can be surveyed in this manner include mineral districts, soils and soil moisture, crops, timber stands, certain types of marine life and metropolitan land and transportation networks. Advantages which satellites possess for Earth resource surveys include: rapidity and continuity of observations, reduced data acquisition times, synoptic views for regional syntheses, and greater freedom from weather disturbances.

To maintain world economic developments at their present levels requires continued assessment and utilization of the Earth's resources. Satellite surveys appear to be the only feasible way to complete and keep up-to-date worldwide resource maps within the foreseeable future. The Earth resource satellites of the future are expected to incorporate instrumentation capable of scanning the full electromagnetic spectrum from radio waves through the visible to the infrared and ultraviolet regions. These instruments are expected to reveal for the first time the full potential of the Earth in terms of those natural resources which have direct and real benefit to mankind. (L. JA.)

VIII. INTERNATIONAL SPACE ACTIVITIES

The space age began in an atmosphere of international co-operation. Following suggestions made in Rome in 1954 by the central planning body, Comité Spéciale de l'Année Géophysique Internationale (CSAGI), for the International Geophysical Year, the U.S. and the U.S.S.R. announced plans for artificial scientific satellites. Both countries, plus the United Kingdom, Canada, Japan, Australia and France (unofficially), also organized programs for the launching of small, scientific sounding rockets.

CSAGI subsequently considered resolutions for the exchange of satellite and sounding-rocket information consistent with the open, scientific character of the IGY. The western members sought the widest possible exchange of information relating to experiments and results. Soviet scientists preferred to restrict the resolutions essentially to published scientific results. The fact that the tools, techniques and certain of the research objectives of space research were common to both scientific and military interests was clearly a limiting factor.

Scientists of countries concerned with satellites or sounding-rocket work, including tracking, met with delegates from interested international scientific unions in London in Nov. 1958 to draft a charter for a permanent International Committee for Space Research (COSPAR) to function under the International Council of Scientific Unions. The meeting was convened by Homer E. Newell of the United States. The group formulated a charter and selected H. C. Van de Hulst of the Netherlands as president. Representatives of the U.S.S.R. Academy of Sciences participated in the drafting of the charter, but at the next meeting of COSPAR, in March 1959 in The Hague, they alleged that the charter gave inadequate "representation" to the eastern bloc.

Other delegates strongly opposed overt consideration of political factors in the organization of a nongovernmental scientific body. A compromise provided for two vice-presidents, one each from the U.S. and the U.S.S.R., with additional members of COSPAR's executive committee to be selected from slates proposed by the two countries.

COSPAR subsequently met amicably in Nice, France (1960), Florence, Italy (1961 and 1964), Washington, D.C. (1962), and Warsaw, Pol. (1963). Little change was made in the original IGY patterns for exchange of data and notification of launchings. While other nations went well beyond the IGY-COSPAR conventions for co-operation, the Soviet Union in general continued to publish scientific results often lacking sufficient detail for independent evaluation, listed "look-angles" for satellites at specific locations but failed to provide their orbital elements, and furnished only general information on numbers of sounding rockets launched in various areas, without listing experiments or results. Catalogues of rocket-satellite data deposited in the IGY World Data centre in Moscow were to have been provided at six-month intervals but were not made available after 1960. Only an approximate compliance with the fundamental requirement for information of scientific results can be said to have been obtained.

Beginning in 1962 COSPAR attempted to modify its structure so as to make a positive contribution to the organization of useful space research. Significant progress became evident by 1963 in the planning of synoptic sounding-rocket programs in meteorology, ionospherics and aeronomy.

Another forum, principally for annual symposia on engineering aspects of space research, exists in the International Astronautical federation, which brings together a considerable number of national rocket societies with broad and long-standing interests in the promotion of space exploration.

1. U.S. Programs of Co-operation.—The charter of the National Aeronautics and Space administration provided, among its objectives, for co-operation with other nations in space activities. Co-operation was offered by NASA with officially sponsored, central civilian agencies in other countries, with literal sharing of the requirements for specific projects of scientific validity and mutual interest, without exchange of funds. The program included:

Joint Satellite Projects.—The first international satellites were Ariel and Alouette, the former instrumented in the U.K. and engineered by NASA, the latter prepared entirely in Canada, and both launched successfully in 1962 by NASA. By the end of 1964 a second British and a first Italian satellite had been successfully launched, the latter by an Italian team using U.S. launch facilities. Additional co-operative satellites were planned by British, French and Italian national space councils. Under a 1964 agreement, NASA undertook the launching, in 1967, of two satellites to be prepared by the European Space Research organization (*see* below).

In Nov. and Dec. 1965, a second Canadian and a first French satellite were launched by the U.S. Also, in 1965, a bilateral agreement was negotiated by NASA and the West German ministry of science for a joint scientific satellite project in 1968. A project involving NASA and the French Space committee was agreed upon in June 1966. It provided for a French satellite launched by NASA to work in conjunction with high-altitude balloons, deriving wind and temperature information from them.

Individual Satellite Experiments.—A number of experiments proposed by scientists outside of the U.S. and judged in competition with U.S. proposals were selected by NASA for incorporation in large observatory satellites to be launched over a period of several years.

Sounding Rocket Experiments.—Australia, Canada, India, Italy, Japan, New Zealand, Pakistan, Denmark, Norway, Sweden, France, the Netherlands and Argentina co-operated with NASA in the launching of scientific sounding rockets, from sites in several countries. Responsibility was divided for the essential components: sounding rockets, payloads and ground equipment. Countries which more recently entered joint programs of this type include Brazil, Germany, Spain and the U.K.

Ground-Based Experimental Participation.—More than 40 countries carried out weather observations made with aircraft, balloons or ground instruments, synchronized for correlation purposes with cloud photographs obtained by NASA's Tiros weather satellites; 10 countries established major ground terminals for experimental communications by means of NASA communications satellites.

Tracking and Data-Acquisition Operations.—About two-thirds of NASA stations outside the U.S. are operated wholly or in part by citizens of the host countries, which, in a number of cases, also provide and support the technical staff.

Training.—Three NASA programs offer (1) graduate fellowships at U.S. universities for students sponsored by their central space agencies; (2) training arrangements at NASA centres in support of agreed co-operative projects; and (3) associateships at NASA centres for gifted senior scientists.

2. U.S.–U.S.S.R. Relationships.—United States representatives attempted during the IGY to broaden exchanges between the two major space powers but met with little success. Other bilateral overtures were made privately, and in 1960 NASA publicly offered tracking services by Project Mercury stations for future Soviet manned flights. The offer was not accepted.

In Feb. and March 1962 Pres. John F. Kennedy responded

to routine congratulations from Chairman Nikita S. Khrushchev, on the occasion of the first U.S. manned orbital flight, by proposing specific co-operative projects and designating Hugh L. Dryden, deputy administrator of NASA, as chief negotiator. Khrushchev designated Soviet academician A. A. Blagonravov to meet with Dryden. By June 1962 their discussions produced an agreement calling for co-ordinated weather satellite launchings and data exchange, a complementary satellite survey of the Earth's magnetic field, and demonstration of experimental space communications by means of the U.S. satellite Echo 2. By Aug. 1963 a detailed implementing agreement became effective.

With increasing strength in the U.S. space program and the easing of international tension following the nuclear test ban treaty of 1963, President Kennedy was encouraged to suggest, in his speech to the general assembly of the United Nations in Sept. 1963, that the two countries explore co-operation in their manned lunar programs.

The bilateral agreement of 1962 led, early in 1964, to communication tests by means of Echo 2, conducted through transmissions from a U.K. station to a Soviet station. The Soviet side provided data on the tests. By Oct. 1964 a conventional communications link for exchanging weather data was operating between Washington and Moscow on a shared cost basis. Conventional weather data was being exchanged twice daily but satellite data was not yet exchanged since it was not yet available from the Soviet side.

3. European Regional Co-operation.—In the period 1961–63, 9 countries established a European Space Research organization (ESRO) to develop scientific satellites and employ sounding rockets. By late 1963 plans had been made for a budget of $300,000,000 over an eight-year period, and for the establishment of a data centre in Darmstadt, Ger., a technical centre in Delft, Neth., and a test centre in Italy, as well as administrative headquarters in Paris. ESRO member countries were the United Kingdom, France, West Germany, Spain, Switzerland, Denmark, Sweden, Belgium and the Netherlands. Among the ESRO projects were development of two small satellites to be launched by NASA, two medium satellites to be launched by boosters purchased through NASA and a large astronomical satellite to be launched by the ELDO vehicle (*see* below).

A second agency, the European Launcher Development organization (ELDO), was organized during the same period. Consisting of seven countries, including Australia, ELDO sought to develop a European satellite booster using the British Blue Streak vehicle as a first stage, a French second stage and a German third stage. Belgium, the Netherlands and Italy were the other participants. ELDO's budget called for expenditures of about $200,000,000 over five years. Its launching range was planned at Woomera, Austr. The first stage of the Blue Streak was successfully flight-tested in 1964. In early 1966, rumours of British withdrawal from ELDO precipitated a crisis in European space affairs. It was expected to be resolved through a review of costs to reduce the U.K. share and a reshaping of the program to provide an early capability to launch a synchronous communications satellite.

4. UN Activities.—Following a U.S. proposal, the United Nations established an *ad hoc* committee on the peaceful uses of outer space in 1959. The Soviet Union, the United Arab Republic and India abstained, essentially on issues of political representation and procedure. The committee prepared a report describing the opportunities for useful activity by UN specialized agencies such as the World Meteorological organization and the International Telecommunications union, stating opportunities for the development of space law and seeking to encourage co-operation among nations in peaceful space research. By late 1961 a formula was arrived at which expanded the committee, increasing the representation of the eastern bloc, and provided for the committee to attempt to do its business on the basis of unanimity. Designated the UN Committee on the Peaceful Uses of Outer Space, it established two subcommittees, one on scientific and technical matters, the other on legal matters. It was agreed that the committee's function would be to facilitate co-operation among nations rather than engage in operational activities. The Scientific and Technical subcommittee met in Geneva, Switz., in mid-1962 and recommended the publication of information on national program activities, requested support for specific programs of the international scientific community and provided for possible UN sponsorship of an international sounding-rocket range to be established on the geomagnetic equator. India offered to establish such a range, then in preparation through bilateral arrangements with the United States. The parent committee unanimously approved the Scientific and Technical subcommittee's report in September, and the general assembly added its approval later in the fall of 1962. Although the subcommittee reconvened in Geneva in the spring of 1963 and again in 1964, little more was accomplished other than the receipt and preparation of various reports. The international rocket sounding range at Thumba in India did, however, progress well, receiving U.S., French and Soviet contributions in the form of equipment. While reflecting Soviet political inhibitions in part, the absence of further substantive achievement also could be ascribed to the limited scope available at the time for UN activity within its prescribed nonoperational function.

Early in 1966, an *ad hoc* working group met in New York and agreed on a UN International Conference on the Peaceful Uses of Outer Space to meet in 1967. Also in 1966, first the United States and then the Soviet Union proposed treaties to preserve celestial bodies exclusively for peaceful purposes. It was agreed that the Legal subcommittee of the UN space committee would convene in Geneva in July 1966 to attempt to draft such a treaty.

5. Prospects.—Despite political and security difficulties, unexpected progress was made in the first six years of the space age in bringing nations together in useful co-operation of more than token character. The bulk of substantive activity followed from the international co-operative projects of NASA. However, other centres of international collaboration began to appear. The regional organizations in Europe, ELDO and ESRO, mapped ambitious programs. The International Committee on Space Research (COSPAR) began to generate useful operational activities pursuant to recommendations made by certain of its working groups. The first steps taken toward significant co-ordination of effort by the Soviet Union and the United States, however, although giving early promise, tended to lose momentum and lag as the U.S.S.R. responded slowly to U.S. overtures. (A. W. FR.)

IX. SPACE LAW

It was apparent from the outset of the era of space flight that the legal status of outer space must differ essentially from that of air space. It is a universally recognized principle of international law that every nation has complete and exclusive sovereignty over the air space above its territory and therefore possesses the right to exclude aircraft from its air space and to determine unilaterally the conditions of entry and use, unless a different rule has been prescribed by an international agreement to which the nation is a party. The vertical extent of the air space has never been defined by international agreement and there is no agreed limit upon the upward reach of the area within which the underlying nation may exercise exclusive control.

If the rule of exclusive sovereignty which prevails in the air space were to extend to altitudes at which orbital flight can be accomplished, it would be necessary to secure the prior consent of nations over which a satellite would pass. In apparent reliance, however, on a basic difference between the legal status of air space and the legal status of outer space, no nation has sought the prior consent of other nations before sending its satellites aloft. No objection was registered by any other nation.

In addition to the question of the legal status of outer space, the related question of the legal status of celestial bodies was raised by the prospect of exploration and exploitation of the Moon and the planets of the solar system.

Both of these questions were the subject of Resolution 1721 (XVI) adopted unanimously by the general assembly of the United Nations on Dec. 20, 1961. The resolution, which was the first authoritative statement of legal principles applicable to space activities, "commends to States for their guidance" the principles that "international law, including the Charter of the United Nations, applies to outer space and celestial bodies" and that "outer

space and celestial bodies are free for exploration and use by all States in conformity with international law and are not subject to national appropriation."

Resolution 1721 (XVI) also requested the United Nations Committee on the Peaceful Uses of Outer Space "to study and report on the legal problems which may arise from the exploration and use of outer space." During the two years after the adoption of the resolution, a consensus on certain general principles emerged from the deliberations of the committee and its Legal subcommittee. These principles were given formal expression in Resolution 1962 (XVIII), entitled "Declaration of Legal Principles Governing the Activities of States in the Exploration and Use of Outer Space," which was adopted unanimously by the general assembly on Dec. 13, 1963.

In addition to restating the principles enunciated in Resolution 1721 (XVI), Resolution 1962 (XVIII) declared that states bear international responsibility for national activities in outer space, whether carried on by governmental agencies or by nongovernmental entities; that the activities of nongovernmental entities in outer space require authorization and continuing supervision by the state concerned; that each state which launches or procures the launching of an object into outer space, and each state from whose territory or facility an object is launched, is internationally liable for damage caused by such object to a foreign state or to its natural or juridical persons on the Earth, in air space or in outer space; that the state of registry of an object launched into outer space retains jurisdiction and control over the object and its personnel while in outer space; that ownership of objects launched into outer space is not affected by their passage through outer space or by their return to Earth, and there is a duty to return any such object to the state of registry regardless of the place where found; and that personnel engaging in space flight who make a landing on the territory of a foreign state or on the high seas because of accident or emergency shall be safely and promptly returned to the state of registry.

Resolution 1962 (XVIII) recognized that activities in outer space might be carried on by international organizations, and it provided that in such cases responsibility for compliance with the principles set forth in the resolution shall be borne by the international organization and by the states participating in it. Taking account of the possibility of harmful interference being caused by a particular space activity with the activities of other states in the peaceful exploration and use of outer space, the resolution declared that if a state has reason to believe that an outer space activity planned by it or its nationals would have such an adverse effect, it shall undertake appropriate international consultations before proceeding with any such activity.

Although both Resolution 1721 (XVI) and Resolution 1962 (XVIII) declared the principle that outer space is free and is not subject to national appropriation, neither resolution attempted to define at what altitude outer space begins or air space ends. The problem presented by the apparent juxtaposition of two realms governed by antithetical legal regimes was left by the general assembly for future consideration. It appeared that the future development of manned space flight, involving extensive flight during return to earth at altitudes below that at which orbital flight can occur but above that at which the flight of conventional aircraft has taken place, eventually would require a more precise treatment of this question. (J. A. Jo.)

BIBLIOGRAPHY.—Arthur Clarke, *The Exploration of Space* (1959); H. E. Newell, *Express to the Stars* (1961); S. Ramo (ed.), *Peacetime Uses of Outer Space* (1961); Krafft A. Ehricke, *Space Flight* (1962); R. Jastrow (ed.), *Exploration of Space* (1960); R. M. L. Baker and M. W. Makemson, *An Introduction to Astrodynamics* (1960); G. P. Sutton, *Rocket Propulsion Elements* (1956); H. S. Seifert (ed.), *Space Technology* (1959); H. H. Koelle, *Handbook of Astronautical Engineering* (1961); R. B. Morrison and M. J. Ingle (eds.), *Design Data for Aeronautics and Astronautics* (1962); Francis S. Johnson, *Satellite Environment Handbook* (1961); John A. Ratcliffe (ed.), *Physics of the Upper Atmosphere* (1960); John C. Houbolt, "Lunar Rendezvous," *International Science and Technology*, No. 14, p. 62 (Feb. 1963).

SPACE-TIME. The theory of relativity has brought about a fundamental change in the scientific conception of space and time, described in a famous saying of Hermann Minkowski—"From henceforth space in itself and time in itself sink to mere shadows, and only a kind of union of the two preserves an independent existence." This union is called space-time. In this article, the concept itself is discussed first by Albert Einstein. Recent developments are summarized in the section that follows Einstein's contribution. The method of determining the structure of space-time is described in the article RELATIVITY.—Editor.

EINSTEIN'S CONCEPTS

All our thoughts and concepts are called up by sense experiences and have a meaning only in reference to these sense experiences. On the other hand, however, they are products of the spontaneous activity of our minds; they are thus in no wise logical consequences of the contents of these sense experiences. If, therefore, we wish to grasp the essence of a complex of abstract notions we must for the one part investigate the mutual relationships between the concepts and the assertions made about them; for the other, we must investigate how they are related to the experiences.

As far as the way is concerned in which concepts are connected with one another and with the experiences there is no difference of principle between the concept systems of science and those of daily life. The concept systems of science have grown out of those of daily life and have been modified and completed according to the objects and purposes of the science in question.

The more universal a concept is, the more frequently it enters into our thinking; and the more indirect its relation to sense experience, the more difficult it is for us to comprehend its meaning; this is particularly the case with prescientific concepts that we have been accustomed to use since childhood. Consider the concepts referred to in the words "where," "when," "why," "being," to the elucidation of which innumerable volumes of philosophy have been devoted. We fare no better in our speculations than a fish which should strive to become clear as to what is water.

SPACE

In the present article we are concerned with the meaning of "where," that is, of space. It appears that there is no quality contained in our individual primitive sense experiences that may be designated as spatial. Rather, what is spatial appears to be a sort of order of the material objects of experience. The concept "material object" must therefore be available if concepts concerning space are to be possible. It is the logically primary concept. This is easily seen if we analyze the spatial concepts for example, "next to," "touch," and so forth, that is, if we strive to become aware of their equivalents in experience. The concept "object" is a means of taking into account the persistence in time or the continuity, respectively, of certain groups of experience complexes. The existence of objects is thus of a conceptual nature, and the meaning of the concepts of objects depends wholly on their being connected (intuitively) with groups of elementary sense experiences. This connection is the basis of the illusion which makes primitive experience appear to inform us directly about the relation of material bodies (which exist, after all, only insofar as they are thought).

In the sense thus indicated we have (the indirect) experience of the contact of two bodies. We need do no more than call attention to this, as we gain nothing for our present purpose by singling out the individual experiences to which this assertion alludes. Many bodies can be brought into permanent contact with one another in manifold ways. We speak in this sense of the position relationships of bodies (*Lagenbeziehungen*). The general laws of such position relationships are essentially the concern of geometry. This holds, at least, if we do not wish to restrict ourselves to regarding the propositions that occur in this branch of knowledge merely as relationships between empty words that have been set up according to certain principles.

Prescientific Thought.—Now, what is the meaning of the concept "space" which we also encounter in prescientific thought? The concept of space in prescientific thought is characterized by the sentence: "We can think away things but not the space which

they occupy." It is as if, without having had experience of any sort, we had a concept, nay, even a presentation, of space; and as if we ordered our sense experiences with the help of this concept, present a priori. On the other hand, space appears as a physical reality, as a thing which exists independently of our thought, like material objects. Under the influence of this view of space the fundamental concepts of geometry, the point, the straight line, the plane, were even regarded as having a self-evident character. The fundamental principles that deal with these configurations were regarded as being necessarily valid and as having at the same time an objective content. No scruples were felt about ascribing an objective meaning to such statements as "three empirically given bodies (practically infinitely small) lie on one straight line," without demanding a physical definition for such an assertion. This blind faith in evidence and in the immediately real meaning of the concepts and propositions of geometry became uncertain only after non-Euclidean geometry had been introduced.

Reference to the Earth.—If we start from the view that all spatial concepts are related to contact experiences of solid bodies, it is easy to understand how the concept "space" originated, namely, how a thing independent of bodies and yet embodying their position possibilities (*Lagerungsmöglichkeiten*) was posited. If we have a system of bodies in contact and at rest relatively to one another, some can be replaced by others. This property of allowing substitution is interpreted as "available space." Space denotes the property in virtue of which rigid bodies can occupy different positions. The view that space is something with a unity of its own is perhaps due to the circumstance that in prescientific thought all positions of bodies were referred to one body (reference body), namely the earth. In scientific thought the earth is represented by the coordinate system. The assertion that it would be possible to place an unlimited number of bodies next to one another denotes that space is infinite. In prescientific thought the concepts "space" and "time" and "body of reference" are scarcely differentiated at all. A place or point in space is always taken to mean a material point on a body of reference.

Euclidean Geometry.—If we consider Euclidean geometry we clearly discern that it refers to the laws regulating the positions of rigid bodies. It turns to account the ingenious thought of tracing back all relations concerning bodies and their relative positions to the very simple concept "distance" (*Strecke*). Distance denotes a rigid body on which two material points (marks) have been specified. The concept of the equality of distances (and angles) refers to experiments involving coincidences; the same remarks apply to the theorems on congruence. Now, Euclidean geometry, in the form in which it has been handed down to us from Euclid, uses the fundamental concepts "straight line" and "plane" which do not appear to correspond, or at any rate, not so directly, with experiences concerning the position of rigid bodies. (On this it must be remarked that the concept of the straight line may be reduced to that of the distance. A hint of this is contained in the theorem: "The straight line is the shortest connection between two points." This theorem served well as a definition of the straight line, although the definition played no part in the logical texture of the deductions.) Moreover, geometricians were less concerned with bringing out the relation of their fundamental concepts to experience than with deducing logically the geometrical propositions from a few axioms enunciated at the outset.

Let us outline briefly how perhaps the basis of Euclidean geometry may be gained from the concepts of distance. We start from the equality of distances (axiom of the equality of distances). Suppose that of two unequal distances one is always greater than the other. The same axioms are to hold for the inequality of distances as hold for the inequality of numbers. Three distances $\overline{AB'}$, $\overline{BC'}$, $\overline{CA'}$ may, if CA' be suitably chosen, have their marks BB', CC', AA' superposed on one another in such a way that a triangle ABC results. The distance CA' has an upper limit for which this construction is still just possible. The points A, (BB') and C then lie in a "straight line" (definition). This leads to the concepts: producing a distance by an amount equal to itself; dividing a distance into equal parts; expressing a distance in terms of a number by means of a measuring rod (definition of the space interval between two points).

When the concept of the interval between two points or the length of a distance has been gained in this way we require only the following axiom (Pythagoras' theorem) in order to arrive at Euclidean geometry analytically. To every point of space (body of reference) three numbers (coordinates) x, y, z may be assigned —and conversely—in such a way that for each pair of points A (x_1, y_1, z_1) and B (x_2, y_2, z_2) the theorem holds:

$$\text{measure number } AB = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

All further concepts and propositions of Euclidean geometry can then be built up purely logically on this basis, in particular also the propositions about the straight line and the plane. These remarks are not, of course, intended to replace the strictly axiomatic construction of Euclidean geometry. We merely wish to indicate plausibly how all conceptions of geometry may be traced back to that of distance. We might equally well have epitomized the whole basis of Euclidean geometry in the last theorem above. The relation to the foundations of experience would then be furnished by means of a supplementary theorem. The coordinate may and *must* be chosen so that two pairs of points separated by equal intervals, as calculated by the help of Pythagoras' theorem, may be made to coincide with one and the same suitably chosen distance (on a solid). The concepts and propositions of Euclidean geometry may be derived from Pythagoras' proposition without the introduction of rigid bodies; but these concepts and propositions would not then have contents that could be tested. They are not "true" propositions but only logically correct propositions of purely formal content.

Difficulties.—A serious difficulty is encountered in the above represented interpretation of geometry in that the rigid body of experience does not correspond *exactly* with the geometrical body. There are no absolutely definite marks, and, moreover, temperature, pressure, and other circumstances modify the laws relating to position. It is also to be recollected that the structural constituents of matter (such as atom and electron, *qq.v.*) assumed by physics are not in principle commensurate with rigid bodies, but that nevertheless the concepts of geometry are applied to them and to their parts. For this reason consistent thinkers have been disinclined to allow real contents of facts (*reale Tatsachenbestände*) to correspond to geometry alone. They considered it preferable to allow the content of experience (*Erfahrungsbestände*) to correspond to geometry and physics conjointly.

This view is certainly less open to attack than the one represented above; as opposed to the atomic theory it is the only one that can be consistently carried through. Nevertheless it would not be advisable to give up the first view, from which geometry derives its origin. This connection is essentially founded on the belief that the ideal rigid body is an abstraction that is well rooted in the laws of nature.

Foundations of Geometry.—We come now to the question: what is a priori certain or necessary, respectively in geometry (doctrine of space) or its foundations? Formerly we thought everything; nowadays we think—nothing. Already the distance concept is logically arbitrary; there need be no things that correspond to it, even approximately. Something similar may be said of the concepts straight line, plane, of three-dimensionality and of the validity of Pythagoras' theorem. Even the continuum doctrine is in no wise given with the nature of human thought, so that from the epistemological point of view no greater authority attaches to the purely topological relations than to the others (*see* Geometry; Knowledge, Theory of).

Earlier Physical Concepts.—We have yet to deal with those modifications in the space concept which have accompanied the advent of the theory of relativity. For this purpose we must consider the space concept of the earlier physics from a point of view different from that above. If we apply the theorem of Pythagoras to infinitely near points, it reads

$$ds^2 = dx^2 + dy^2 + dz^2$$

where ds denotes the measurable interval between them. For

an empirically given ds the coordinate system is not yet fully determined for every combination of points by this equation. Besides being translated, a coordinate system may also be rotated. This signifies analytically: The relations of Euclidean geometry are covariant with respect to linear orthogonal transformations of the coordinates.

In applying Euclidean geometry to prerelativistic mechanics a further indeterminateness enters through the choice of the coordinate system: the state of motion of the coordinate system is arbitrary to a certain degree, namely, in that substitutions of the coordinates of the form

$$x' = x - vt$$
$$y' = y$$
$$z' = z$$

also appear possible. On the other hand, earlier mechanics did not allow coordinate systems to be applied of which the states of motion were different from those expressed in these equations. In this sense we speak of "inertial systems." In these favoured inertial systems we are confronted with a new property of space as far as geometrical relations are concerned. Regarded more accurately, this is not a property of space alone but of the four-dimensional continuum consisting of time and space conjointly.

Appearance of Time.—At this point time enters explicitly into our discussion for the first time. In their applications space (place) and time always occur together. Every event that happens in the world is determined by the space coordinates x, y, z, and the time coordinate t. Thus the physical description was four-dimensional right from the beginning. But this four-dimensional continuum seemed to resolve itself into the three-dimensional continuum of space and the one-dimensional continuum of time. This apparent resolution owed its origin to the illusion that the meaning of the concept "simultaneity" is self-evident, and this illusion arises from the fact that we receive news of near events almost instantaneously owing to the agency of light.

This faith in the absolute significance of simultaneity was destroyed by the law regulating the propagation of light in empty space or, respectively, by the Maxwell-Lorentz electrodynamics. Two infinitely near points can be connected by means of a light signal if the relation

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2 = 0$$

holds for them. It further follows that ds has a value which, for arbitrarily chosen infinitely near space-time points, is independent of the particular inertial system selected. In agreement with this we find that for passing from one inertial system to another, linear equations of transformation hold which do not in general leave the time values of the events unchanged. It thus became manifest that the four-dimensional continuum of space cannot be split up into a time continuum and a space continuum except in an arbitrary way. This invariant quantity ds may be measured by means of measuring rods and clocks.

Four-dimensional Geometry.—On the invariant ds a four-dimensional geometry may be built up which is in a large measure analogous to Euclidean geometry in three dimensions. In this way physics becomes a sort of statics in a four-dimensional continuum. Apart from the difference in the number of dimensions the latter continuum is distinguished from that of Euclidean geometry in that ds^2 may be greater or less than zero. Corresponding to this we differentiate between timelike and spacelike line elements. The boundary between them is marked out by the element of the "light-cone" $ds^2 = 0$ which starts out from every point. If we consider only elements which belong to the same time value, we have

$$-ds^2 = dx^2 + dy^2 + dz^2$$

These elements ds may have real counterparts in distances at rest and, as before, Euclidean geometry holds for these elements.

Effect of Relativity, Special and General.—This is the modification which the doctrine of space and time has undergone through the restricted theory of relativity. The doctrine of space has been still further modified by the general theory of relativity, because this theory denies that the three-dimensional spatial section of the space-time continuum is Euclidean in character. Therefore it asserts that Euclidean geometry does not hold for the relative positions of bodies that are continuously in contact.

For the empirical law of the equality of inertial and gravitational mass led us to interpret the state of the continuum, insofar as it manifests itself with reference to a noninertial system, as a gravitational field, and to treat noninertial systems as equivalent to inertial systems. Referred to such a system, which is connected with the inertial system by a nonlinear transformation of the coordinates, the metrical invariant ds^2 assumes the general form:

$$ds^2 = \sum_{\mu\nu} g_{\mu\nu} dx_\mu dx_\nu$$

where the $g_{\mu\nu}$'s are functions of the coordinates and where the sum is to be taken over the indices for all combinations 11, 12, . . . 44. The variability of the $g_{\mu\nu}$'s is equivalent to the existence of a gravitational field. If the gravitational field is sufficiently general it is not possible at all to find an inertial system, that is, a coordinate system with reference to which ds^2 may be expressed in the simple form given above:

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$$

but in this case, too, there is in the infinitesimal neighbourhood of a space-time point a local system of reference for which the last-mentioned simple form for ds holds. This state of the facts leads to a type of geometry which Riemann's genius created more than half a century before the advent of the General Theory of Relativity of which Riemann divined the high importance for physics.

Riemann's Geometry.—Riemann's geometry of an n-dimensional space bears the same relation to Euclidean geometry of an n-dimensional space as the general geometry of curved surfaces bears to the geometry of the plane. For the infinitesimal neighbourhood of a point on a curved surface there is a local coordinate system in which the distance ds between two infinitely near points is given by the equation

$$ds^2 = dx^2 + dy^2$$

For any arbitrary (Gaussian) coordinate system, however, an expression of the form

$$ds^2 = g_{11}dx^2 + 2g_{12}dx_1dx_2 = g_{22}dx_2^2$$

holds in a finite region of the curved surface. If the $g_{\mu\nu}$'s are given as functions of x_1 and x_2 the surface is then fully determined geometrically. For from this formula we can calculate for every combination of two infinitely near points on the surface the length ds of the minute rod connecting them; and with the help of this formula all networks that can be constructed on the surface with these little rods can be calculated. In particular, the "curvature" at every point of the surface can be calculated; this is the quantity that expresses to what extent and in what way the laws regulating the positions of the minute rods in the immediate vicinity of the point under consideration deviate from those of the geometry of the plane.

This theory of surfaces by Gauss has been extended by Riemann to continua of any arbitrary number of dimensions and has thus paved the way for the general theory of relativity. For it was shown above that corresponding to two infinitely near space-time points there is a number ds which can be obtained by measurement with rigid measuring rods and clocks (in the case of timelike elements, indeed, with a clock alone). This quantity occurs in the mathematical theory in place of the length of the minute rods in three-dimensional geometry. The curves for which $\int ds$ has stationary values determine the paths of material points and rays of light in the gravitational field, and the "curvature" of space is dependent on the matter distributed over space (*see* RIEMANNIAN GEOMETRY).

Just as in Euclidean geometry the space concept refers to the position possibilities of rigid bodies, so in the General Theory of Relativity the space-time concept refers to the behaviour of rigid

bodies and clocks. The space-time continuum, however, differs from the space continuum in that the laws regulating the behaviour of these objects (clocks and measuring rods) depend on where they happen to be. The continuum (or the quantities that describe it) enters explicitly into the laws of nature, and conversely these properties of the continuum are determined by physical factors. The relations that connect space and time can no longer be kept distinct from physics proper. Nothing certain is known of what the properties of the space-time continuum may be as a whole. Through the general theory of relativity, however, the view that the continuum is infinite in its timelike extent but finite in its spacelike extent has gained in probability.

Time

The physical time concept answers to the time concept of the extrascientific mind. Now, the latter has its root in the time order of the experiences of the individual, and this order we must accept as something primarily given. One experiences the moment "now," or, expressed more accurately, the present sense experience (*Sinnen-Erlebnis*) combined with the recollection of (earlier) sense experiences. That is why the sense experiences seem to form a series, namely the time series indicated by "earlier" and "later." The experience series is thought of as a one-dimensional continuum. Experience series can repeat themselves and can then be recognized. They can also be repeated inexactly, wherein some events are replaced by others without the character of the repetition becoming lost for us. In this way we form the time concept as a one-dimensional frame which can be filled in by experiences in various ways. The same series of experiences answer to the same subjective time intervals.

The transition from this "subjective" time (*Ich-Zeit*) to the time concept of prescientific thought is connected with the formation of the idea that there is a real external world independent of the subject. In this sense the (objective) event is made to correspond with the subjective experience. In the same sense there is attributed to the "subjective" time of the experience a "time" of the corresponding "objective" event. In contrast with experiences, external events and their order in time claim validity for all subjects.

This process of objectification would encounter no difficulties were the time order of the experiences corresponding to a series of external events the same for all individuals. In the case of the immediate visual perceptions of our daily lives, this correspondence is exact. That is why the idea that there is an objective time order became established to an extraordinary extent. In working out the idea of an objective world of external events in greater detail, it was found necessary to make events and experiences depend on each other in a more complicated way. This was at first done by means of rules and modes of thought instinctively gained, in which the conception of space plays a particularly prominent part. This process of refinement leads ultimately to natural science.

The measurement of time is effected by means of clocks. A clock is a thing which automatically passes in succession through a (practically) equal series of events (period). The number of periods (clock time) elapsed serves as a measure of time. The meaning of this definition is at once clear if the event occurs in the immediate vicinity of the clock in space; for all observers then observe the same clock time simultaneously with the event (by means of the eye) independently of their position. Until the theory of relativity was propounded it was assumed that the conception of simultaneity had an absolute objective meaning also for events separated in space.

This assumption was demolished by the discovery of the law of propagation of light. For if the velocity of light in empty space is to be a quantity that is independent of the choice (or, respectively, of the state of motion) of the inertial system to which it is referred, no absolute meaning can be assigned to the conception of the simultaneity of events that occur at points separated by a distance in space (*see* MICHELSON-MORLEY EXPERIMENT). Rather, a special time must be allocated to every inertial system. If no coordinate system (inertial system) is used as a basis of reference there is no sense in asserting that events at different points in space occur simultaneously. It is in consequence of this that space and time are welded together into a uniform four-dimensional continuum. (A. EI.)

RECENT DEVELOPMENTS

The preceding section of this article remains essentially as it was originally written by Albert Einstein.

The more recent developments in the structure of space-time have taken three different directions. First, cosmologists have attempted to describe the properties of space-time in the large. By this is meant that instead of asking what the structure of space-time is, for instance, in the neighbourhood of a star or in the solar system or even in the neighbourhood of a galaxy, the goal is to describe the gross properties of the universe as a whole in terms of the gross geometrical structure of space-time. Second, unified field theories were developed, the aim of which was to show that not only gravitational forces, but also electromagnetic (and nuclear) forces are manifestations of the geometry of space-time. Third, modern developments in the quantum theory of the electromagnetic field suggested that a certain structure should be attributed to those portions of space-time which in the past were considered to be empty. The detailed description of this structure forms the modern quantum theory of the so-called vacuum.

A brief description of the new results must necessarily be incomplete. According to the theory of relativity, the geometrical structure of space-time determines the physical processes; hence, a detailed description of this structure would be equivalent to characterizing all the physical processes that can take place. This would entail rather more than any short discussion.

Cosmological Theories.—Briefly, here is what cosmological theories indicate about space-time. They emphasize four essential features of space-time: (1) the distribution of matter in space-time; (2) the time dependence of the geometry; (3) whether space-time is open or closed; (4) whether space-time is singly connected or multiply connected. The last two points refer to the geometry of space-time in the large. To understand the distinction implied in point (3) observe that a sphere (*q.v.*) represents a curved space that is closed, while the surface of a paraboloid (*q.v.*) is open. To appreciate the idea of multiple-connectedness in point (4), note that the surface of a sphere is called a singly connected, two-dimensional space while the surface of a doughnut (torus) is classed as a doubly connected two-dimensional space. (Roughly speaking, if the space as viewed from a higher dimensional space has one or more "holes" in it, then it is multiply connected.)

Point (1) is answered by the so-called cosmological principle, which states that on a large scale all positions in space are equivalent; irregularities are only local. Thus on a large scale the universe is homogeneous. This still tells nothing about point (2), the development of the universe in time. With this postulate there are two classes of theories. The steady-state theory assumes that the universe should be stationary; *i.e.*, its broad aspects should not change in time. The other theories assume the existence of a cosmic time that measures the evolution of the universe. The limiting case of both classes is the static universe. However, about 1930 E. Hubble at the Mount Wilson Observatory discovered that the nebulae in the observable part of the universe are receding from earth. This observation automatically rules out the possibility that space-time should be static as was originally proposed by Albert Einstein (1917). The evolving universes can be of the following kinds: (1) In the very early initial stages of the evolution, the universe was exceedingly small and contained exceedingly high densities of matter; that is to say, it has developed from a point source. (2) The universe was originally a static Einstein universe; from this stage it is now developing by expansion (Lemaître-Eddington universe). (3) The universe, originally contracted from an extremely large initial size, reached a minimum size and is expanding again. The space-time structure so described can still be open or closed; singly connected or multiply connected.

From these different theories, the second was the most commonly accepted. However, since 1940 interest in the other the-

ories revived. The point-source model (1) was invoked by George Gamow, Edward Teller (1939), and others to explain not only the origin of the universe but the generation of the elements composing the universe as well (*see* COSMOGONY; COSMOLOGY).

The steady-state theory was put forward by Hermann Bondi and T. Gold (1948) and by Fred Hoyle (1948). According to this theory the large-scale structure of space-time is independent of the position of the observer and of the time of observation. This postulate is called the "perfect cosmological principle," in contradistinction to the above-mentioned "cosmological principle" which postulates homogeneity only as far as space is concerned. However, with this assumption a difficulty arises. Notwithstanding the homogeneity in space and time, it is necessary to incorporate the observed fact that the universe is expanding and carrying matter away from the neighbourhood of any observer. Thus on the one hand an observer should see on the average the same amount of matter around him, clustered into nebulae; and on the other hand he should see these nebulae, with the matter contained in them, moving away. If this be true the thinning out of matter by expansion must be replaced by a continuous creation of matter out of nothing. This process cannot be observed per se (the amounts are too small) but it gives rise to a number of predictions of things that are observable, including the distribution of age among nebulae.

For many years there was no significant observational evidence to discriminate among the different cosmological theories. In 1960 an analysis (A. Sandage) of the ages of a number of galactic and globular clusters in terms of the relationship between their colour (red shift) and their visual magnitude suggested that the steady-state model may be inconsistent with observed data (*see* NEBULA: *Red Shifts and Expansion of the Universe*). From 1961 to 1965 two new classes of celestial objects were discovered: the quasi-stellar radio sources (quasars) and the quasi-stellar galaxies (so-called blue objects) which resemble the quasars in many optical properties, but do not emit at radio frequencies (*see* QUASI-STELLAR RADIO SOURCES). In addition, in 1965 cosmic radio-noise emission was observed. Analysis of these experimental data suggested that the actual distribution of the quasi-stellar objects is incompatible with the steady-state theory; and that the cosmic noise can be interpreted as red-shifted radiation that initially emerged from a primordial fireball out of which the universe developed. This interpretation also discriminates against the steady-state theory; however, the results were not sufficient to tell whether the universe is open or closed.

Less work has been done on the connectivity of space-time. John A. Wheeler has put forward the interesting suggestion that space-time is multiply connected, to such an extent that if it could be viewed from a higher-dimensional space it would be foamlike, riddled with holes. The "bubbles" in the foam are used in his theory as apparent sources and sinks of the electric field.

Unified Field Theories.—In the General Theory of Relativity all forces that produce acceleration are gravitational. Moreover, these gravitational forces stem from the geometry of space-time. This, however, does not mean that the only interactions described by the theory are those commonly called gravitational. In the General Theory all the usual forces (*e.g.*, electromagnetic, elastic) produce accelerations by modifying the geometry of space-time which in turn produces a gravitational field. Mathematically this arises in the following fashion: For every interaction, whether it be gravitational, electromagnetic, and so on, there is a corresponding energy-momentum tensor; and the geometry of space-time is determined by this energy-momentum tensor (*see* FORCE; ANALYSIS: *Analysis and Space*). However, the General Theory alone cannot give the laws for these tensors; *i.e.*, cannot show how the electromagnetic fields modify the geometry of space-time. The unified field theories are attempts to accomplish this as far as electromagnetic forces are concerned, conceiving these laws to be geometrical restrictions on space-time.

This modifies the geometry of space-time profoundly. It makes the geometry either higher dimensional or non-Riemannian. The modifications brought about this way are too abstract and technical to be discussed here.

An entirely different approach was proposed by Wheeler, who preserved the Riemannian geometry of space-time by retaining the source-free equation of the electromagnetic field and the Einstein equation for the gravitational field, with the electromagnetic field energy as its source (*see* GRAVITATION: *Theories of Gravitation*). However, the charges, which are usually thought of as sources of the electromagnetic field, do not appear here at all as particles, but are seen as the manifestation of the multiply connected nature of space-time.

Modern Quantum Theory of Fields.—This area of study has also left its mark on the development of space-time concepts. Quantum theory denies the possibility of the simultaneous measurement of certain quantities with arbitrary accuracy. For example, it asserts that the momentum (mass times velocity) and position of a particle cannot be measured simultaneously with arbitrary accuracy. These rules are expressed by the Heisenberg uncertainty relations (*see* UNCERTAINTY PRINCIPLE). In general, three main directions can be pursued.

1. The question can arise of how far the uncertainty relations will be reflected in efforts toward specification of space-time. First, is it possible to attach labels to points in space-time that would tell with complete accuracy what is meant by "here" and "now"? Second, even if it were possible to do this labeling locally, can the labeling be spread over the whole of space-time while preserving complete accuracy? The answers to these questions are not yet known.

2. Suppose the first question were answered in the affirmative, thus allowing use of the concept of space-time as it stands. It is immediately apparent that the uncertainty relations would affect the General Theory. For, according to the Heisenberg relations, it is not possible to measure the position and momentum of a particle simultaneously and precisely. On the other hand, if the gravitational field of the body can be measured, the General Theory permits just that. Thus, the General Theory must be modified to make it consistent with quantum theory. This can be accomplished for weak gravitational fields, with the following results. Space-time is still characterized by the $g_{\mu\nu}$ tensor (*see* above: *Effect of Relativity, Special and General*) but this $g_{\mu\nu}$ tensor cannot be known in general with perfect precision; instead, what is known is the probability that in a given physical situation it will have certain values. Thus the discussion returns to the questions raised in point 1 above.

3. If no gravitational fields are present the General Theory considers space-time to be empty and its metric to be flat. If there is any physical process going on there will be a gravitational field generated by it, and space-time will become curved. According to the modern quantum theory of fields, however, even empty space is not completely void. This empty space (called the vacuum) is endowed with potentialities. Electrons, positrons, photons, and other elementary particles can be spontaneously created there even if only to be annihilated immediately (*see* PARTICLES, ELEMENTARY). These virtual physical processes have an observable influence on such phenomena as the emission of light by actual atoms. How will the existence of these virtual processes influence the description of space-time? Is there any gravitational field associated with these processes? Does a gravitational field influence these virtual processes? It is hard to tell what the answers will be. *See* also QUANTUM MECHANICS and references to "Space-Time" in the Index.

BIBLIOGRAPHY.—A. Einstein, *The Meaning of Relativity*, 5th rev. ed. (1956); E. Schrödinger, *Space-Time Structure* (1950), *Expanding Universes* (1956); H. Bondi, *Cosmology*, 2nd ed. (1960); P. G. Bergmann, *Introduction to the Theory of Relativity* (1942); J. A. Wheeler, *Geometrodynamics* . . . (1962); *Reviews of Modern Physics*, vol. 29, no. 3 (1957); L. Infeld (ed.), *International Conference on the Theory of Gravitation* (1964). For experimental evidence *see* papers in the *Astrophysical Journal* by A. R. Sandage, 133:355 (1961), 141:1560 (1965), by R. H. Dicke *et al.*, 142:414, and by A. A. Penzias and R. W. Wilson, p. 419 (1965); also M. Schmidt in *Nature*, 197:1040 (1963). (N. L. B.)

SPAGHETTI: *see* MACARONI.

SPAIN (ESPAÑA), a state (nominally a kingdom under the 1947 Law of Succession, see *Administration and Social Conditions*, below) occupying more than three-fourths of the second largest

peninsula in Europe (after the Scandinavian), termed IBERIA by the Greeks, HISPANIA by the Romans. Portugal, Andorra, and Gibraltar occupy the remainder of the peninsula. The Spanish provinces comprise the mainland portion of the Spanish state, the Balearic Islands, and the Canary Islands. Total area: 194,883 sq.mi. (504,748 sq.km.). Also included in the Spanish state, although they do not constitute provinces, are certain places under Spanish sovereignty along the Mediterranean coast of Morocco (Ceuta, Alhucemas, Melilla [*qq.v.*], the Chafarinas Islands, and Peñón de Vélez de la Gomera). Total area: 12.4 sq.mi. (32.1 sq.km.). The colonial possessions of Spain comprise Spanish West Africa (the two overseas provinces of Ifni and Spanish Sahara [*qq.v.*]) and Equatorial Guinea (*q.v.;* formerly Spanish Guinea, consisting of the overseas provinces of Río Muni and Fernando Po).

This article is divided into the following sections and subsections:

I. PHYSICAL GEOGRAPHY

1. General Features.—The Iberian Peninsula may be described as more continental in dimensions and more African in character than the other two great Mediterranean peninsulas, the Italian and the Greek. From Madrid, just south of which is the reputed geometrical centre of the peninsula (at Cerro de los Ángeles), the rail distances to Cádiz (424 mi.; 682 km.) on the south coast, Valencia (294 mi.; 473 km.) on the east, and Santander (302 mi.; 486 km.) on the north give some idea of the dimensions of the peninsula. Spain includes the greater portion, but it seems legitimate to discuss at some points the physical characteristics of the peninsula as a whole.

Although seven-eighths of the peninsula's borders are washed by seas, its massive form of 250,000 sq.mi. (647,500 sq.km.) is dominated by the great block of the Meseta (Spanish, Meseta Central, "central tableland") at an average elevation of 2,000–2,500 ft. (610–760 m.) above sea level. Mountain walls forming the rims of the Meseta effectively isolate the extensive interior from the coast and on the western side have provided Europe's oldest frontier, between Portugal and Spain, for more than eight centuries. Thus, the monotony of its wide central plateaus and the diversity of its peripheral mountain ranges have emphasized its continental character.

The narrow width of the Strait of Gibraltar suggests that the peninsula shares many North African affinities. The vast plateaus have their geographical counterpart in the High Plateaus of Algeria and a geological similarity with the Moroccan Meseta. The great spreads of pediments (erosional surface features associated with arid climates, termed *rañas* in southern Spain) brightly stained by the drought; the badlands of the Ebro Basin, Upper Tagus (Tejo or Tajo), and the Segura; and the relics of inland drainage deepen this similarity. The dwarf palm so characteristic of North Africa is found along the Spanish littoral as far north as Tarragona, while the pseudosteppe lands of southeast Spain have a similar cover of esparto and other plants.

But the peninsula has European characteristics also. The landscapes of the rainy northern and northwestern regions are similar to those of other Atlantic communities, such as Brittany and Ireland. Comparable customs reflect a common origin, in what has been termed the "Atlantic civilization." Much of the settlement of the Iberian Peninsula has moreover been profoundly influenced by the Christian reconquest and the 500-year struggle with Islam. It may be, however, that due to the peninsula's peculiar position and relief it has never successfully served as a bridge between North Africa and Europe. It has suffered the worst of both worlds, somewhat remote from European developments and yet not unified in one culture. Its backward look fostered regionalism and localism so that the province and its *comarca* subdivision have much geographical significance.

2. Geology.—Broadly, the geological outcrops consist of Paleozoic crystalline rocks in the western Meseta, Tertiary clays and coarser detrital deposits in the central basins, and Mesozoic folded limestones in the east and north. The Meseta, which has been tilted westward, occupies four-sevenths of the peninsula. Some of its bordering mountains represent the epicontinental margins of Secondary seas, others Secondary geosynclines. Two Tertiary tectonic depressions of contrasted evolution flank the Meseta: the Ebro trough to the north, and the Baetic Basin to the south (now deeply eroded by the bed of the Guadalquivir). A third depression is infilled by the lower valley of the Tagus and the Sado in Portugal. The Alpine systems of the Pyrenees and the central basins of the Meseta have been infilled by Tertiary and Quaternary deposits. Thus in the paleography of the peninsula, three units have been recognizable since Secondary times: the Paleozoic block of the Meseta; the epicontinental margins, especially to the east; and the geosynclines to the north and southeast.

The Meseta is a fragment only of the great Hercynian continent which extended from North Africa across Europe; it has existed as a peneplain since the close of the Carboniferous Period. It consists of Paleozoic rocks which range from Cambrian to Carboniferous. The great thicknesses of Cambrian slates in the Asturias, Galicia, Extremadura (Estremadura), and Portugal indicate relatively uniform marine conditions at the time of their deposition. Evidence of a pre-Paleozoic or Huronian mountain system is uncertain. The evidence of Caledonian orogenesis between the Cambrian and Silurian epochs is more certain, especially in the folds of the Sierra Morena. Silurian deposits reflect the folded relief with thick conglomerates and sands covering the borders of the southwestern massif and the isolated block under the Ebro trough. During the Devonian, deep geosynclinal condi-

MAJOR POLITICAL AND PHYSICAL FEATURES OF SPAIN. INSET MAP SHOWS LAND USE AND MINING SITES

tions developed over the Pyrenees and the northwest. The Lower Carboniferous was also marine, markedly detrital, and then, near the Hercynian block of northeast Spain, moderately thick in the south and southwest and very thick in the Asturias. Thus marine sediments (Dinantian) alternated with continental (Culm) facies. The Culm conglomerates witness the first Hercynian movements but in the Cantabrians (Cordillera Cantábrica) the very thick Mountain Limestones indicate that folding there was much later. The Middle Carboniferous is everywhere detrital with coal seams, discordant in the northeast, but concordant and more widespread in the north. The Upper Carboniferous is continental and discordant everywhere, indicating the effects of the Hercynian uplift.

Apart from this sequence of Paleozoic strata on the borders of the Meseta, the other characteristic is the widespread occurrence of extrusive rocks and denuded batholiths of granites. According to C. Neiva there are five important magmatic phases in the Paleozoic of Portugal, the granites resulting from the solidifying of eruptive granites of various ages. These begin with basic and end with acidic rocks. The Hercynian folds gave place to the great batholiths now exposed by erosion, especially in the Pyrenees and the centre of the northwest of the peninsula where the Paleozoic geosynclines were deepest. Everywhere the effects of Hercynian orogenesis are visible in the northwest-southeast trending Armorican folds.

The second paleogeographic unit of the peninsula, the epicontinental margins of the Meseta, reflect the long period of orogenic tranquillity during the Secondary and early Tertiary eras, when marine transgressions laid down thick, fine sediments. Three subdivisions are recognizable: the wide platform of shallow depth which bordered the Meseta from south of Oviedo to about Albacete; the shallow sedimentary basin of the northeast; and the deep geosyncline which occurs between Gibraltar and the Balearic Islands.

Oscillations of land and sea laid down the three-fold sequence of Bunter sandstones, Muschelkalk limestone, and lagoonal Keuper during the Triassic. Gentle folding, in the Jurassic and the Cretaceous, produced oscillations of sea level, the marine transgressions of the Lias, Malm, Barremian-Aptian, and Upper Cretaceous (Cenomanian, Turonian, and Senonian) alternating with the regressions of the Dogger and the Gault. The beginning of the Tertiary marks the commencement of more continental conditions and the first phases of the Alpine movements.

At the end of the Eocene, Alpine movements commenced in the Pyrenees and the Sistema Penibético (Baetic Cordillera), moving outward from the centre of the ranges in several phases according to H. Stille: Upper Eocene-Oligocene; Oligocene-Miocene; Intra Miocene (Helvetian-Tortonian), and Pliocene. During the early Tertiary (Paleocene), flysch and littoral deposits were laid down

in the two geosynclines of the Pyrenees and the sub-Baetic straits between lower Andalusia and the southern Balearic Islands. But during the Miocene all the peninsula was subjected to continental conditions except the sub-Baetic straits which were wider than before. Over the eastern Meseta unconformable beds were deposited by lakes covering Old Castile, New Castile, the Ebro trough, and smaller ones in the Badajoz and lower Tagus districts. Tertiary uplift fractured the Mesetan block, often traverse to the Hercynian alignment, as in the Cordillera Carpetovetónica (Central Cordillera), Montes de Toledo, and Sierra Morena. Since the Pliocene, the basins of the Meseta have initiated the primary courses of the main lines of drainage, followed by the Ebro, Douro (Duero), Tagus, Guadiana, and Guadalquivir rivers.

3. Physiography.—Prolonged erosion, especially since the Pontian stage, has produced typical elements of relief on the Meseta. The typical form is the tableland (*mesa*), consisting of unconformable Miocene strata, laid down by rivers (Tortonian) in brackish ponds (Sarmatian) and in marshes (Pontian). The flat, upper surface of hard limestone forms an eroded capping (*páramo*), protecting the softer, underlying marls that outcrop abruptly on the slopes (*cuestas*). The sands and clays of the lower slopes form plains (*campos, campiñas, llanuras*). Villafranchian pediments (*rañas*) form a thin veneer over the plateaus, especially in Extremadura and southern Portugal. These appear related with the highest river terraces at 330–400 ft. (100–120 m.). The other terraces are Quaternary, apparently climatic, at 200 ft. (60 m.), 100 ft. (30 m.), and 30–40 ft. (10–12 m.) above the major rivers, and related to the terminal moraines. Glaciation however was of minor importance except in the Pyrenees where it has produced glacial sculpture along 186 mi. (300 km.) of the central ranges.

Landforms.—The major surface features of the peninsula may be grouped under three broad headings: the Meseta; its borderlands; the Pyrenees (Pirineos) and Sistema Penibético, together with the troughs of the Ebro and the Guadalquivir.

The crystalline base of the Meseta provides most of the high plateau country of the peninsula, rising from 1,000 to over 3,000 ft. (300–900 m.). Tilted toward the west and southwest, the Meseta carries several of the major rivers, notably the Douro, Tagus, and Guadiana. Two major series of contrasts divide the Meseta regionally. West of a line drawn roughly through Zamora, Salamanca, and Ciudad Real are crystalline outcrops, with thin soils, and much wild country. This is most extensive in Extremadura, where plateaus of 1,100–1,500 ft. (330–460 m.) run 200 mi. (320 km.) from east to west and 90–120 mi. (145–200 km.) from north to south. Only the small Tertiary basins, notably Badajoz and La Serena, have kinder landscapes. In the eastern Meseta, Tertiary sediments, several thousands of feet thick, have been carved out by differential erosion into tablelands (*páramos*), cuestas, isolated buttes (*alcores*) and alluvial plains (*campiñas*). These are more docile landscapes, widely cultivated under cereals. The Cordillera Carpetovetónica, which stretches east-northeast to west-southwest in a series of vertebrae (Cerro de la Estrella, Gata, Peña de Francia, Gredos, and Guadarrama) at heights varying between 5,000 and 8,500 ft. (1,500–2,600 m.), shows a marked contrast on its two flanks. From the north, these block mountains generally rise gradually, but on the south they fall steeply in fault scarps toward the Tagus trough. This explains the second major feature of the Meseta. In Old Castile, the Douro shows little evidence of headward erosion above the gorges on the Portuguese frontier. In contrast, the Tagus flows through a series of troughs, facilitating more vigorous erosion along its tributaries, as seen in La Alcarria. Further differences occur in La Mancha (2,200–2,600 ft.; 670–800 m.) forming a flat tableland, backed to the south by the volcanic cones of Campo de Calatrava. The Sierra Morena (4,500–5,900 ft.; 1,370–1,800 m.) forming the southern edge of the Meseta is fretted by vigorous linear erosion and fractured along the border of the Guadalquivir Basin.

The Portuguese landforms are essentially a western annex of the Meseta, protected from the more energetic orogenic movements that affected the northern and the eastern flanks of the peninsula. Alentejo is similar to Extremadura; Beira and Trás-os-Montes are Castilian in aspect. The most individualized landforms are the folded Mesozoic cuestas of southern Algarve and Extremadura. Galicia is characterized by senile features in the interior and by the effects of youthful, coastal submergence. A vast batholith, it has been fractured into a series of north-south blocks. The Cantabrian Mountains, perhaps the most formidable barrier of the Meseta's borders, have two types of relief: Hercynian landforms in the Asturias and more plastic Mesozoic relief in the east. The Iberian Mountains (Sistema Ibérico) are less of a barrier, a complex of ancient tablelands and folded ranges such as the Sierra de la Demanda, Sierra Cebollera, Montes Universales, and Sierra de Javalambre, which rise from 5,000 to 7,600 ft. (1,500–2,300 m.). A series of step faults breaks up the eastern edge of the Meseta toward the Valencian Plain.

Peripheral to the Meseta is the deceptive symmetry of arrangement in the Ebro depression and Pyrenean ranges to the north, the Guadalquivir Basin and the Baetic ranges to the south, respectively. The Spanish Pyrenees have a three-fold sequence: a central axis of high mountains (8,000–11,000 ft.; 2,400–3,350 m.), flanked by a discontinuous depression that is followed by some of the major valleys and the outer ranges that rise to 4,000 ft. The triangular-shaped trough of the Ebro is shut off from the sea by the Catalan Mountains, composite in style of relief. The Ebro Plain differs in important respects from its Andalusian counterpart. In the southeast, the Sistema Penibético rises steeply from the coast to 5,000–11,000 ft. (1,500–3,350 m.), to culminate in the Sierra Nevada. Inland, the Penibetic trough is followed by a string of valleys and plains, bordering the lower, complex sub-Baetic ranges that lie to the northwest.

Drainage.—Runoff averages about 35% of the total precipitation in the Iberian Peninsula, but it is 20% or less over much of the Meseta. As one-third of the peninsula has less than 20 in. (500 mm.) annual rainfall, the majority of the rivers have only a modest flow, less than 700 cu.m. per sec. The Pyrenees are the most important source of water so that the outflow of the Ebro is greatest. The northern streams also have abundant flow, the Nalón, for example, having more than the Guadiana (*q.v.*) at the Portuguese frontier, with only one-twelfth of its drainage area. In the south, especially in the southeast and in the Sierra Morena, even large catchment areas dry out in summer.

The characteristic regime of much of the peninsula, especially of the interior, is the double spring and autumn maxima. The marked increase of rainfall in spring, coupled with the melting snows of the mountains, accounts for this. In addition to the summer minimum, a secondary one occurs in winter. In eastern Spain, the flow is often complicated by karstic phenomena and the maximum is, in consequence, more delayed. Regimes with a single maximum occur both on the Atlantic and on the southern Mediterranean coasts, though the volume of flow in the latter is much less. Floods are common on all the large rivers, the Ebro and Tagus (*qq.v.*) having records that compare with inundations of other European rivers. But the most disastrous floods have occurred on the east coast with its steep gradients and the dense concentrations of population, especially on the Segura, Júcar (*qq.v.*), and Turia rivers.

4. Climate.—The Iberian Peninsula, lying between latitudes 36° and 44° N, is influenced by the seasonal shift of the westerly cyclonic belt in winter and the subtropical high pressure system in summer. The winter cold is consequent on the continental dimensions and high elevation of the Meseta. In summer, drought is explained by the spread of the Azores anticyclone over the peninsula. The Pyrenees coincide approximately with the jet stream (a narrow belt of intensely strong westerly wind) of the tropopause (the zone of transition between the stratosphere and troposphere, or lower part of the atmosphere) in summer, providing an effective aerological barrier to the more temperate conditions farther north. Polar air dominates Mediterranean meteorology so that in winter the Mediterranean Sea, 3–6° C (6–10° F) warmer than the Atlantic in the same latitude, acts as a catalyst, influencing frontogenesis and cyclogenesis. Spanish weather conditions therefore reflect four climatic controls: continental, Atlantic, Mediterranean, and mountain influences.

The peripheral regions have milder conditions than the continental interior, but the Mediterranean littoral is generally warmer and drier than the Atlantic coasts. The most marked contrast is between the northwest and the southeast (Santiago 66 in. [1,680 mm.] and Almería 10 in. [250 mm.]).

The humid zone follows approximately the watershed of the north coast from Galicia (excluding southeast Galicia and Bierzo) to the Basque provinces, then along the Pyrenees to northeastern Catalonia. Elsewhere, the orographic rainfall may be high—as in the Serranía de Ronda in the south, 8 in. (200 mm.)—but it is markedly seasonal. At the other extreme, the subdesert region of Murcia and Almería is perennially dry, having less than 1.2 in. (30 mm.) for each month of the year. Between them, the greater part of Spain embraces a dry zone with less than 24 in. (610 mm.), badly distributed seasonally and irregular in amount from year to year. Rainfall is low also in the centre of the river basins, being less than 12 in. (300 mm.) in the Zamora and Salamanca districts of Old Castile and at Saragossa in the mid-Ebro Basin.

In the seasonal distribution of precipitation three systems can be distinguished. On the northwestern and southern oceanic facades, a single winter maximum is marked, related to frontal depressions of polar maritime air. Along the Mediterranean littoral from Alicante to Catalonia, an autumn maximum reflects the importance of Mediterranean fronts at this season. Spring maximum rainfall in the central part of the Meseta indicates the effects of continental conditions having dry, cold anticyclonic weather prevailing in winter but the high pressure system breaking down in spring with marked convectional rainfall. The border areas of the Meseta have double maxima of rainfall; the west has a more pronounced rainfall in spring, however, and the east in autumn.

Areas above 3,000 ft. (900 m.) cover approximately one-seventh of the peninsula, so mountain climates are significant. Apart from the high ranges of the Central Pyrenees, the Sierra Nevada, and isolated summits in the Cantabrians and Cordillera Carpetovetónica that rise above 6,500 ft. (1,900 m.) and have prolonged snow cover, the majority of the mountains lie between 3,000 and 6,250 ft. (900–1,900 m.). These have 15–60 days of snow, with moderate to heavy rainfall depending upon their orientation, local winds, etc. The peripheral mountains have lower annual ranges of temperature and 40–60 in. (1,000–1,500 mm.) of rainfall, while the interior mountains have higher ranges and only 25–40 in. (630–1,000 mm.) annually.

5. Vegetation and Soils.—The rich variety of plant associations in Spain is the consequence of the varied climates and relief. Winter cold limits the olive to south of the Cordillera Carpetovetónica, except in the Ebro Valley and in small sheltered basins in the northwest. The Guadalupe-Toledo Mountains, the borders of La Mancha, and the Sierra Nevada are also excluded. North and northwest of this true Mediterranean limit, mesophytic flora of mixed deciduous species are characteristic. Two northern zones may be distinguished: a maritime, more humid zone where the common oak is important, mingling on the higher slopes with the sweet chestnut and the beech; and a transitional interior zone where the Pyrenean and Portuguese oaks mingle with the evergreen oak. Man has introduced a number of other species, notably the maritime pine and in recent decades the blue gum, especially into Santander. The sweet chestnut (as a valued food in the past) has been greatly spread, especially in the Basque provinces; the beech has competed successfully against the fir on the higher slopes, but is excluded climatically from Galicia.

South and east of these deciduous oak zones, the peninsula was formerly widely covered with the Mediterranean climax of the holm or evergreen oak. Apart from some of the mountains of Catalonia and a few other areas, this has been largely destroyed, vast areas of the Meseta being covered with regressive associations of the mint family, especially thyme, lavender, and rosemary. The pea family is represented by the broom heaths while in the northern mountains are heather and erica heaths. The plants that occur frequently in the woodland scrub also form dominant associations, such as the various rockroses (especially *Cistus ladaniferus*), boxwood, and the kermes oak. In the more humid, siliceous soils, the cork oak covers wide areas of Cádiz and Málaga provinces, and also of northern Catalonia. Pines constitute about one-third of Spanish forests, notably the stone pine in Old Castile, the Aleppo pine in the east, and the Scotch, Corsican, and mountain pines in the mountains. The zonal sequence becomes impoverished south of the Cordillera Carpetovetónica, so that there is an upward rise of the evergreen oak climax and elimination of the colder, continental species (except in the Sierra Nevada).

Along the east coast patches of dwarf palms are common and the carob is characteristic. In the southeast, esparto and Spanish broom steppes cover extensive areas, associated with salt-tolerant species in saline tracts. In addition to drought, the cold winters of Teruel and the Ebro Valley account for the occurrence there of Spanish cedar or *sabinar* steppes. Both P. Reyes, who estimated the Spanish steppes grossly at 28,000 sq.mi. (72,500 sq.km.), and E. H. del Villar, who denied their importance, have overstated their cases. In La Mancha, Almería, Murcia, Valencia, Los Monegros, and around Valladolid are extensive steppes but their limits have been confused with the Labiatae cover.

Soils.—The soils of the peninsula have been classified by del Villar as "acid-humic" in the northwest and northern Portugal, "siallitic (sialic)" over the western half of the peninsula, and largely "calcareous" to the east. The acid-humic soils are leached, associated with humid conditions and mesophytic flora, and modified into podsolized soils by the presence of pine woods. Siallitic soils are of three broad types: the humid soils of the north found on both siliceous and calcareous rocks under mesophytic flora; the intermediate type found on the borders of Old Castile with mixed deciduous flora; and the most widespread xerosiallitic soils occurring only on siliceous rocks and in dry climates, notably in the southwest. Most of eastern Spain, the Ebro Basin, and plains of Old Castile have intermediate or dry calcareous soils. These range in colour from the bright red earths of the east coast and south, through the chestnut soils of the interior, to the tawny or gray soils of the arid southeast. In the areas where cultivation has been attempted beyond its proper limits, an extreme type of soil, *calvero*, covers at least half the surface; eroded soil caps (regoliths or loose rock fragments partly covered by soil) are discontinuous and the basic rock is conspicuous on the surface.

Saline soils are widespread in the Ebro trough, especially in Lérida and also in the southeast. Patches occur also in the Campos de Criptana and Daimiel and on a wide scale in Las Marismas of the lower Guadalquivir. The black earths of Andalusia are a legacy of the Quaternary marshes which formerly constituted the Campiña.

6. Animal Life.—Its close links with Africa give Spain more African species than exist in the other Mediterranean peninsulas while the Pyrenean barrier and extent of the country explains the number of indigenous species. There are two clearly distinguished groups of fauna. The larger one is typical of the Mediterranean region, confined to the drier eastern and southern Spain, with many African and indigenous forms. The second group associated with the central European forests is found in the north. There are also alpine forms in the higher mountains, including various deer, mountain goats, and birds such as the eagle and vulture. Indigenous to the country are a type of red deer (*Cervus elaphus hispanicus*), an ibex (*Capra pyrenaica*), a squirrel of the Sierra de Segura (*Sciurus vulgaris segurae*), a hare (*Lepus granatensis*), and a water vole (*Arvicola terrestris sapidus*). Among the native birds, the azure-winged magpie (*Cyanopolius cooki*) is abundant in central Spain and the Sierra Morena, while a partridge (*Perdix perdix hispaniensis*) lives in the Pyrenees. The red partridge is common on the steppe lands and the great bustard feeds in flocks on the Meseta cornfields.

In the Pyrenean region the fauna is similar to that of France. The chamois, Pyrenean mountain goat, squirrel, and mole are characteristic. In the Cantabrians, the fauna is mainly European in character and includes, in the Picos de Europa, the chamois and the bear. In the northwest the Galician hare, the mountain goat, and ibex are characteristic.

The Meseta and Andalusia are rich in indigenous species of moles and squirrels. Muskrat, roebuck, and wolf are also found. In the Sierra de Guadarrama there are species of lynx, genet, badger, stone marten, fox, and weasel. African species predominate in the south, and in the southeast there are peculiar species of wolf, jackal, rabbit, and squirrel. There is the Andalusian weasel and the hedgehog, as well as the Egyptian mongoose which is not found elsewhere in Europe. In the Balearic Islands there are peculiar species of rat, dormouse, and genet, but an absence of squirrels and of moles and some other insectivores. There are generally no squirrels in the Mediterranean regions of Spain. At Gibraltar is confined the Barbary ape, thought either to be indigenous or to have been imported by the Romans or Arabs.

The great variety of bird life is explained partly by the variation of relief and climate and partly by the routes of migration followed by African and Northern European species. In Andalusia alone there are more than 400 species of birds, but only 20 are resident throughout the year. It is thus a characteristic of Spain to find only a few of its many birds living there permanently.

Snakes are common, the largest being *Coelopeltis monspessalanus*, but only the viper, which is rare, is deadly to man. There are many varieties of lizard, especially in the south together with the *quelonia* tortoise, and in the mountains the salamander is found. Common freshwater fish are the trout, barbel, tench, and carp.

The insect life is also varied. In Madrid Province alone, more than 350 species of butterfly have been counted, many of them endemic as well as truly African and European species. The marked contrasts between the humid northern zone and the arid centre and east are reflected in the differing species. *See* also PORTUGAL: *Physical Geography.*

II. GEOGRAPHICAL REGIONS

The geographical position of Spain explains much of its character. Tarifa on the Strait of Gibraltar is on the same latitude as both the island of Malta and Susah (Sousse, Tunisia). Cape Roca (Cabo da Roca) in the northwest is almost on the same meridian as Valentia in southwest Ireland. Spain therefore exhibits Atlantic, West European, Mediterranean, Central European, and African affinities. It is as Celtic as France, as Roman as Italy, as German as Central Europe, and even as arabicized as North Africa. The title of the Spanish kings, *Rex Hispaniarium*, indicates, apart from the empire in the New World, that there were "Spains" within the peninsula. The usage of the Roman term *provincia* first appears in 16th-century documents. Some of the provinces mark stages in the reconquest, such as those of Aragon and Old Castile; others distinguish medieval political units such as the kingdom of Navarre, the county of Catalonia, the lordship of Santander, and the principality of the Asturias. In 1811 Napoleon imposed a framework of prefectures, based roughly on drainage basins: 52 provinces were created in 1822, modified later to 47 peninsular provinces and 2 insular provinces.

Various geographers have demarcated the regional divisions of Spain. Dantín Cereceda recognized 17 regions on the basis of the ancient tribal divisions. J. F. Unstead made a division into 22. But whatever detailed divisions are made, the essential regional lineaments are those of the peripheral and central zones and Andalusia.

1. Peripheral Regions.—*The North.*—Three elements, the Atlantic Ocean, the mountain backbone, and the mild and humid climate, distinguish the northern regions from all others. Galicia is the largest region with a northern and western coastline of nearly 375 mi. (600 km.) and a dense population of more than 2,700,000. Its structure has yielded the following units: the west coastal plateaus and rías (drowned coastal river valleys), the central tablelands of Lugo, the central depression drained by the upper Miño (Minho) and Sil, and the high mountains on the east and south to the Portuguese frontier. It is a Celtic landscape of small fields, many churches, and *castros*, or hill settlements, densely scattered. Population has focused on three main areas: on the fishing villages of Rías Bajas with their sheltered, rich fishing grounds; on the Rías Altas with their more open hinterlands of rich agricultural lands, such as Betanzos; and on the valley of the lower Minho with its vineyards and lines of communication. The chief towns are Vigo, La Coruña, and El Ferrol del Caudillo—all ports. The historic towns of Santiago, Lugo, and Orense are all inland and former ecclesiastical centres of note.

The Cantabrians form a formidable barrier 30–60 mi. (50–100 km.) wide along the north coast for about 180 mi. (300 km.). The larger eastern sector comprises the Asturias in which an interior longitudinal depression centres on Oviedo. Because of the coastal cliffs, Asturias has turned its back on the sea, to concentrate on the industrial benefits of coal mining in the Oviedo Basin and its valleys. Only at Gijón is there an important port. But at Avilés a new town has developed with modern steel mills. Pastures, apple orchards, beech forests, and maritime pines form a friendly hinterland to the industrial centres.

The East.—The eastern provinces in Catalonia, Valencia, and Murcia have more cultural unity than physical uniformity. Whereas northeast Catalonia has a humid Mediterranean climate with maize (corn) fields and beechwoods, Murcia in the south is largely steppe. The mountains behind the coastal plains range from the alpine scenery of the Eastern Pyrenees (Pirineos Orientales) to the faulted plateaus and ranges of the Iberian Mountains and to the hill and basin relief of the sub-Baetic. Harnessing of the rivers has created two important economic activities: textile manufactures in the valleys of Catalonia and irrigation in the lower valleys of the Llobregat, Ebro, Turia, Júcar, and Segura. Irrigational development explains the dense lowland concentrations of population.

Of the four Catalan provinces, Lérida belongs more properly to the Ebro Valley, and Barcelona (*q.v.*) is described separately. Gerona in the northeast comprises the trough of El Ampurdán, whose landscape of shelter belts indicates it is in the realm of the cold *tramontana* winds. The southern province of Tarragona is drier, especially in the limestone hills of the Garraf, and noted primarily for its vineyards and olive groves. Tarragona is one of the oldest cities of Spain.

South of the Ebro delta with its ricefields, the low *garrique* (heathland) vegetation of northern Castellón is replaced by a narrow coastal belt of orange groves, on which Castellón de la Plana bases its wealth. From Valencia, with its fertile *vega* (valley land), to the mouth of the Júcar is the richest *huerta* (irrigated land) in Spain, concentrating three-quarters of the provincial population on less than one-tenth of its area. Valencia is the chief manufacturing city and port. In northern Alicante there has also been a concentration of population since Arab times, but the relief is distinct. The mountains are boldly terraced for muscatels, almonds, and figs, and the enclosed basins are densely populated with much irrigation from springs.

The unity of Murcia is the Segura Basin, whose river irrigates the *vega* of Murcia City and of other towns. The relief is more open and spacious, the high plateaus of Hellín and Chinchilla forming an easy approach to Castile. To the south of the plain is Cartagena, a naval base and historic port. Off the coast is the archipelago of the Balearic Islands (*q.v.*), consisting of five inhabited islands and numerous islets.

2. Central Spain.—Stretching from the Pyrenees to the Sierra Morena in Andalusia, Castilian Spain has played a historic role in uniting the diverse regions of the Meseta. In its widest sense it has about 9,300,000 inhabitants, *i.e.*, about one-third of the total population on half of the national territory. It consists of vast plateaus, framed by mountains, none of them spectacular except the southern flank of the Cordillera Carpetovetónica, and especially the Pyrenees. Wheat lands (and in New Castile, vineyards) alternate with immense waste of xerophytic shrubs (especially in Extremadura) or steppe. Light densities of population are characteristic in Old Castile, in New Castile, and in Aragon (*see* Table). But a marked contrast exists between the numerous hamlets of Old Castile and León camouflaged in adobe brick, and the whitewashed, large agricultural villages of New Castile, many of them created by the military orders of the reconquest.

The Spanish ranges of the Pyrenees (*q.v.*) stretch 275 mi. (440 km.) from west to east. The highest summit, Pico de Aneto,

rises to 11,168 ft. (3,404 m.). Whereas the Central Pyrenees widen to nearly 80 mi. (130 km.), they are only 15–18 mi. (24–30 km.) wide in the west and 6 mi. (10 km.) in the Montes Alberes, in the east. The main transverse valleys such as the Aragon, Gállego, Cinca, and Segre isolate the communities into small natural regions. Stock rearing, timber, and hydroelectric power are the main resources.

The Iberian Mountains to the south of the Ebro Valley are much less of a barrier, despite their length of 285 mi. (460 km.) and constant width of 50–60 mi. (80–100 km.). In the central sector is a series of parallel ranges separated by the basins of Teruel and Calatayud, but in the northwest the Sierras de la Demanda and Cebollera and to the southeast the Montes Universales have higher and more impressive scenery rising to 6,200–7,500 ft. (1,900–2,300 m.).

The Ebro depression, framed by the Pyrenees and the Iberian Mountains, forms a triangle running 220 mi. (350 km.) to its apex at the foot of the Cantabrians, and 100 mi. (160 km.) along its base bordering the Catalan Mountains. Lying mostly between 600 and 1,000 ft. (180–300 m.) its relief is erosional, forming a sequence of river valleys and terraces, tablelands, scarps, and residual hills such as the Sierra de Alcubierre. Piedmonts, bordering the trough, have been dissected by the tributaries of the Ebro into a series of extensive steppes, especially on the northern side, such as La Ribera of Navarre, Las Bárdenas, Violada, Hoya de Huesca, Los Monegros, Llanos de Urgel, La Segarra, and Garrigas. They are for the most part sparsely settled, except where irrigation is being developed, as in the Lérida and Violada districts. Large villages are concentrated along the Ebro Valley at intervals of 10–15 mi. (16–25 km.). But, apart from Saragossa, the capital, towns are small, such as the historic centres of Huesca and Tudela.

La Rioja, antechamber of Aragon and famed for its vineyards on the terraces of the Ebro, leads through the Burgos corridor into Old Castile. Although Burgos is a feudal and royal city rich in architecture, the surrounding countryside is relatively poor. The basin of Old Castile extends westward for 110 mi. (180 km.), fronted by the faulted edges of the Cantabrians and mountains of León. Below them are repeated monotonously the piedmonts of Palencia and León. Southward the Tertiary basin stretches for about 140 mi. (225 km.). It continuously repeats the erosional forms of Pontian plateaus and tablelands, fronted by scarps and plains widening into the alluvial valleys of the Douro and especially its northern tributaries, the Arlanzón, Esla, and Pisuerga. The southern valleys are less well watered, smaller, but more numerous, and fray the ramp that rises almost imperceptibly to the Cordillera Carpetovetónica. This extensive area, centred in the chief city of Valladolid, is the breadbasket of Spain. To the west, however, the fertile Tertiary cover with its cereal lands is replaced by thin, schistose soils fit only for stock rearing and a poorer economy. Zamora and Salamanca, the university town, stand at the borders of these contrasted landscapes.

The Cordillera Carpetovetónica, consisting chiefly of the Sierras de Gredos and de Guadarrama, culminates in the Pico de Almanzor (8,504 ft.; 2,592 m.) and Peñalara (7,972 ft.; 2,430 m.) respectively. From the ramp of Old Castile at about 3,300 ft. (1,000 m.) the central ranges merge almost imperceptibly in places. But from the south, the Tagus trough is overlooked by these mountains rising steeply above it. Thus on the south the Cordillera forms an effective climatic limit to some plants, notably the olive. Only a few passes cross the mountains, notably Puerto de Navacerrada (5,833 ft.; 1,778 m.) and Puerto de Somosierra (4,757 ft.; 1,450 m.), and these have emphasized Madrid's nodality.

The southern Meseta is much larger than the northern counterpart and extends about 150 mi. (250 km.) from north to south and about 75 mi. (120 km.) from east to west. Its major units are the Tagus trough, the tablelands of New Castile, the mountains of Guadalupe and Toledo, and the plateaus of Extremadura. The Tagus trough consists of several basins, such as La Vera and La Sagra, while the Tagus Valley is in places narrowly incised in the bordering highlands. The irrigated *campiña* of the Río Henares (upstream from Aranjuez) and the rolling tablelands widen eastward in New Castile. Two main types of landscape are distinguishable. To the north, the headstreams of the Tagus, deeply trenched, explain the dissected scarplands of La Alcarria and Serranía de Guadalajara. Settlements follow the valleys. South and east of Ocaña is La Mancha, with the most extensive area of undissected relief and—in the Campo de San Juan—the flattest plains of the peninsula. The regional name (Arabic Al Mansha, "the dry land") explains its steppe vegetation and its disrupted courses of drainage, sometimes forming salt lakes. Nevertheless about two-thirds of La Mancha is cultivated, with vast areas of cereals and extensive vineyards, especially between Alcázar and Albacete and from the Campo de Criptana to the Río Záncara. Farther west the general altitude of 2,200–2,600 ft. (670–800 m.) decreases and merges into the thin siliceous soils and poor stock-rearing country bordering Extremadura. La Mancha, the land of Don Quixote, has large villages but the historic market towns of Cuenca and Albacete are located eccentric to the region.

To the west and northwest rise the quartzitic mountains of Toledo (5,253 ft.; 1,601 m.) and Guadalupe, through which the Guadiana has cut a defile at Pantano de Cíjara where a great dam has been sited. To the southwest the 70 volcanic cones of the Campo de Calatrava give a distinctive landscape. Westward of Ciudad Real, the plateaus of Extremadura extend about 200 mi. (330 km.) into southern Portugal and 90–120 mi. (145–200 km.) from north to south. Tilted gently to the west, at 1,000–1,600 ft. (300–500 m.), the relief is represented by three elements: the softly undulating relief of the Silurian slates, the domed mountains of granite, and the steep ridges of quartzite. Apart from the fertile plains of Mérida and Badajoz under cereals and with irrigation schemes, vast areas are covered with gum cistus and other waste, with holm woodlands grazed by pigs. In the south, the Sierra de Aracena announces the borders of the Sierra Morena, with its dry valleys and sparse settlements (except in the mining districts of Ríotinto and Almadén).

3. Andalusia.—One of the most distinctive major regions of the peninsula, Andalusia had in the 1960s a population of 6,000,000. Framed by the southern border of the Sierra Morena, the Campo de Montiel, the Murcian Steppes, and elsewhere by the sea, it is divided into two distinct sectors, the Guadalquivir Valley of Lower Andalusia, and the sub-Baetic and the Sistema Penibético of Upper Andalusia. Perhaps no other region of Western Europe has had a richer cultural heritage, dating from the mineral exploitation and town life of the Tartessians and Romans. Of the 42 major towns, 38 were Roman in origin and later intensely arabicized. Compared with the rural peasantry of Castile, hard-working and resourceful, the Andalusian labourer is traditionally urban-minded and pleasure-loving. The economic strength of Andalusia has always been the balanced interests of mountain and plain, with complementary economies; thus Seville and Granada tend to represent distinct interests.

Of the two, the Andalusian Plain is the more important, supporting about 2,750,000 people, with a high population density around Seville. Settlement is markedly nucleated, with approximately 90% of the population living in about 240 large villages and towns. Significantly, the Guadalquivir Valley is the only extensive lowland in the peninsula that penetrates deeply into the interior from the coast. Its apex strikes 200 mi. (320 km.) toward the northeast from the Gulf of Cádiz. From a width of about 100 mi. (160 km.) at its mouth, it narrows to 60 mi. (100 km.) at Seville and 6 mi. (10 km.) at Úbeda, but to the south it is flanked by two other narrow depressions followed by the valleys of the Genil and Guadiana Menor. A marine trough until the late Quaternary, the fine sediments have been carved into smoothly undulating relief, famed for its black soils. These are intensively cropped, with olives on the hill slopes. The marshes of Las Marismas and sand dunes of the Arenas Gordas are the only lands not utilized for agriculture. In Huelva, where cork forests and stock rearing are important, is a major afforestation scheme with eucalyptus forest, planted since 1940. Olive groves are very important in the other provinces; they cover 46% of the cultivated area of Jaén, 32% in Córdoba, and 26% in Seville; many have been planted since the 19th century. Vineyards are significant

in the Condado of Huelva and Montilla in Córdoba. In the Campiña, more than 80% of the nation's cotton is grown. The chief towns are Seville and Huelva (both ports), Córdoba, and Jaén.

Upper Andalusia, though slightly smaller than the plain, is the dominant physical influence, bringing heavy rainfall on its southwestern flanks, and the source of the major tributaries of the Guadalquivir. Although the mountain ranges extend structurally for 500 mi. (800 km.) from east-northeast to west-southwest and are 125 mi. (200 km.) at their maximum width, they are much less compact than the Pyrenees. At their extremities they appear more like isolated mountain ranges emerging from a tangled relief of hills, plateaus, and plains. Termed the Sistema Penibético, the ranges can be divided into three zones: the sub-Baetic Mountains which rise above the Andalusian Plain to 3,000–6,500 ft. (910–2,000 m.) in three series of ranges; the central valleys and depressions traceable for about 110 mi. (180 km.) between Antequera, Granada, and Baza; and the principal chain of Sierra Nevada. Rivers have trenched, deep valleys to the coast, and deforestation of the steep slopes has created new deltas.

In Cádiz as in the Andalusian Plain, latifundia (large landed estates) are an agrarian problem, with 65% of the province comprising estates of more than 620 ac. (250 ha.) yet representing only 1.6% of the total holdings. Family interests may form several great estates, the largest of which are under tree crops, especially olive groves. Specialized cultures such as the vineyards of Jerez de la Frontera and Málaga are in smaller holdings. Vegetables and cereals are irrigated in the intermontane basins of the Genil Valley, Arroyo de la Vega of Granada, Río de Guadix, and others. On the coast at Motril, Málaga, Torrox, Torre del Mar, and Adra sugar cane and bananas are grown. Tourism on the Costa del Sol is an important economic activity. Málaga, Granada, and Cádiz are the chief cities. (J. M. Ho.)

III. THE PEOPLE

1. Ethnic Groups.—In general, modern Spanish stock is characteristically "Mediterranean" with two divisions: Ibero-insular (short, swarthy, and dolichocephalic) and Ibero-saharan (tall, dark, and slender) (*see* IBERIANS). In spite of the continuity of the Mediterranean type from the Neolithic Period, there exist contributions of Nordic, Celtic, Eastern Mediterranean, and Dinaric ancestors. In addition the Basques (*q.v.*) present a peculiar ethnic type of obscure origin. The people of modern Spain are composed of Castilians (including Cantabrians and Andalusians), Galicians (including some Portuguese), Basques, and Catalans (together with related Valencians and Balearics). (*See* also *Archaeology* and *History: Prehistory* below, and EUROPE: *Anthropology* and *Archaeology*.)

2. Languages.—The predominant language is Castilian (Castellano), also called Spanish (Español) since the 19th century, with its dialect forms (Andalusian, Asturian, and Aragonese). It is the official and literary language of Spain. The Basque language is a mysterious survival (probably dating from the Neolithic Period) in the Western Pyrenees (Navarra) and in the Basque provinces to the west. Catalan (with its variants) is widely spoken in Catalonia (northeastern Spain), in Valencia, in the Balearic Islands, in Andorra, and in Roussillon, France; it is allied to Balearic speech and Provençal and differs from Castilian Spanish, though all these languages, including Galician, which is akin to Portuguese, are derived from Latin. (*See* further SPANISH LANGUAGE; BASQUE LANGUAGE; CATALAN; PORTUGUESE LANGUAGE.)

3. Religion.—Roman Catholicism is the established religion of Spain. Following the pronouncements of the second Vatican Council, however, the Organic Law of the State (1966) established the principle of religious liberty for non-Catholics as a legal right. Though since the 18th century the percentage of practising Catholics has been on the decline, the Catholic faith has a profound influence on the life of both the ordinary people and the more educated classes. Civil marriage is allowed only for non-Catholics. There are 11 metropolitan and 54 suffragan sees. The primate of all Spain is the archbishop of Toledo. There are about 30,000 Protestants. Jews are estimated at 5,000 and the first synagogue since their expulsion in 1492 was opened in Madrid in 1959.

The anticlerical measures of the Second Republic (1931) were reversed after the Civil War of 1936–39. On June 7, 1941, the government signed an interim agreement with the Holy See regulating procedure for appointments to the Spanish hierarchy. This was confirmed in the Concordat of Aug. 27, 1953. The dignitaries of the church subscribed to an oath of fealty to the state.

4. Customs and Culture.—Regional characteristics persist (*e.g.*, Castilian, Andalusian, Galician, Basque, and Catalan), accentuated by physical considerations such as the contrast between the wetter and drier regions, and by the varying degrees in which regions have adopted contemporary ideas. Local communities, with the town as the natural unit of society, preserve a certain vitality, weakened perhaps by authoritarian rule. By the 1960s independent social organizations, common in Catalonia, in Valencia, and in the Basque country, were becoming more active and vocal, and tending to weaken the corporative structure of a "uniform" Spanish working society. Industrialization had also created an influential upper social class of businessmen and bankers. The Catholic Church, with its great social influence, was slowly re-

(ABOVE) MICHAL HERON—PHOTO RESEARCHERS INC.; (TOP RIGHT) INGE MORATH—MAGNUM; (BOTTOM RIGHT) HENRI CARTIER-BRESSON—MAGNUM; (BOTTOM CENTRE) KIT ROBBINS—RAPHO GUILLUMETTE

(Above) A street in Barcelona's Gothic quarter; (top right) Plaza Mayor in Astorga, at right the Archbishop's Palace erected by Antonio Gaudí in 1909; (bottom right) gypsies making their annual Rocío pilgrimage to Almonte; (bottom centre) village homes in the hills of Alcalá de Guadaira, built beneath ruins of a 14th-century Moorish castle

sponding to Vatican pressure for social and economic reforms.

Modern ways of life are changing much that is old, though the new is but slowly accepted. Traditional customs persist; the fiesta is one of the principal features of Spanish social life, particularly in the villages, but also in the towns. Besides the religious processions (*e.g.*, Corpus Christi and those of the Virgin Mary under various dedications), the fiestas are held to celebrate saints' days and include pilgrimages, fairs, and carnivals, accompanied by fireworks and bullfighting. The *romería*, or village pilgrimage to the saint's shrine, is held mostly in summer; one of the best known is the Romeria del Rocio, held on Whitsunday at Almonte in Huelva. The *verbena* is a night fair held in cities and towns, notably in Madrid. Seville has its April fair and its Holy Week processions (La Semana Santa); Valencia has its procession on St. Joseph's Eve (March 19) in which the huge *fallas*, or carnival figures, are carried; Pamplona has its festival where young bulls are driven through the streets; Lorca has, remarkably, a carnival-type fiesta on Good Friday.

(LEFT) GEORGE HOLTON—PHOTO RESEARCHERS, IN (ABOVE) HENRI CARTIER-BRESSON MAGNUM

ROMAN CATHOLICISM IS THE ESTABLISHED RELIGION OF SPAIN

(Left) Hooded penitents (*nazarenos*) passing through the streets of Seville during a traditional Holy Week procession; (above) holiday excursion of seminarians in the countryside near Burgos in Old Castile

Bullfighting (*q.v.*) remains a national spectacle, displaying a combination of personal bravery and a special art. Various sports and athletics have become popular. Football (soccer) draws huge crowds, and the Basque game of pelota is played (professionally, as a spectacle) in the Basque country and Castile.

Spanish folk music varies greatly from province to province, as do the folk instruments. Thus, in Andalusia the guitar predominates, in Castile the fife-and-drum, in Galicia and Asturias the bagpipe. The interaction of folk and art music is long-standing and intense. It showed very vividly in the *tonadillas* that began simply as incidental songs introduced into popular plays and developed, during the 18th century, into complete musical spectacles employing a fair degree of stylized folk song. The *zarzuela* is a more elevated form of these humble musical comedies. Since the middle of the 19th century the professional urban popular music of Andalusia, called *cante flamenco*, has exerted enormous influence throughout the peninsula, not only on popular music but also on the national repertory of ballet, opera, and concert.

Handcrafts are widespread. In rural areas there are basket weaving, ceramics and clay modeling, blacksmithing, glass blowing, wood and stone carving, knitting, and lace and embroidery work. In coastal areas there is a rich heritage of the lore and crafts of maritime trade and fishing, with special celebrations connected with the sea. *See* also SPANISH LITERATURE and the section on Spain in the articles PAINTING and RENAISSANCE ARCHITECTURE. (X.)

HENRI CARTIER-BRESSON—MAGNUM

BOYS PLAY AT BULLFIGHTING ON A STREET IN SEGOVIA

IV. ARCHAEOLOGY

The progress of Spanish archaeology is revealed by the increasing number of excavations, museums, and archaeological university posts and periodicals.

The earliest discoveries go back to the Lower Paleolithic. They consist of some implements belonging to the pebble-culture and to the Abbevillian and Acheulean industries. The majority of the beds are in the terraces of certain rivers (Manzanares) and the Cantabrian zone. Discoveries of Mousterian industry are more widespread. In the Upper Paleolithic, the peninsula entered a period of some prosperity and a relatively dense population. The Cantabrian coast, the Levante, and regions around Madrid and Lisbon are the richest areas. Remains of Perigordian and Aurignacian industry occur in the north; distribution of Gravettian is more widespread. Solutrean finds are particularly common in the Cantabrian zone, the Levante, and Catalonia. Magdalenian occurs mostly in the north, but is found sporadically in the Levante (Parpalló cave). These cultures exhibit certain original features in Spain. The most notable is the manufacture by the Spanish Solutreans of a type of barbed and tanged points, anticipating by many millennia those of the Neolithic and giving rise to intriguing problems concerning its possible relationship with the African Aterian. The centre for this variant is the Parpalló cave (Gandía), but it is found throughout the Levante, extending to the south and Portugal.

One of the most admirable achievements of the Upper Paleolithic is in the realm of art. The first known Quaternary painting (*see* ALTAMIRA CAVE) was discovered in Spain; today there are about 35 known sites, from Málaga to Santander, of naturalistic art in the so-called Franco-Cantabrian style. The wealth of wall painting contrasts with the paucity of portable art (with the exception of the Parpalló cave, which has over 5,000 engraved and painted slabs). The so-called East Spanish art, which depicts scenes in rock shelters of animals painted in a naturalistic style and more or less stylized human figures, is believed to belong to the Mesolithic. More than 60 sites are known, from the province of Lérida to that of Almería.

After the Mesolithic, the Neolithic introduced a new element—pottery. The earliest known pottery (handmade) was decorated with impressions of a *Cardium* shell; pottery with incised or relief decoration was also common. In the earliest metal-using period, a Spanish form which reached the British Isles—the bell beaker—had spread halfway across Europe. The settlements and their cemeteries (the most famous is that at Los Millares, in Almería, *c.* 2000 B.C.) offer various objects of adornment, polished stone axes, flint knives and arrowheads, etc. Copper soon appeared—one of its first forms being the dagger. Megalithic building was widespread, especially in the south, west, and the Pyrenean area; it varied from simple polygonal dolmens to complex galleries and domed tombs, of indubitably Oriental origin.

With the Bronze Age, grave goods changed with the decline in the use of stone. Daggers, which were lengthened to form swords, axes, from the flat type to the palstave, arrowheads, necklaces, diadems, bracelets, rings, and bodkins were all of metal. Pottery was plain, predominantly carinated.

The El Argar culture in the southeast is the most characteristic. The western Atlantic zone was very rich in this period because of the abundance of gold, of which there is evidence in the form of jewelry clearly related to that of the British Isles. The tin trade led to the formation between the Atlantic lands of links

which endured until Roman times. In schematic art, the last reflection of Stone Age art, there are some unusual forms of engraving (Gallegan *insculturas*).

In the Balearic Islands, the later Bronze Age presents a Cyclopean architecture with parallels in Malta, Sicily, and Sardinia. In Majorca and Minorca there are settlements, artificial burial caves, and talayots (towers of various kinds). In Minorca there are also *navetas* (graves shaped like an inverted ship) and taulas, whose purpose is unknown. (*See* BALEARIC ISLANDS: *Archaeology*.)

The end of the Bronze Age brought new invasions of Indo-European-speaking peoples, especially the Celts (about the 9th century B.C.). They introduced cremation in their enormous urnfield cemeteries. Pottery was the totally exotic Hallstatt type, or the traditional excised; there were great innovations in armour and utensils. Among the deposits, that of the Huelva estuary (8th century B.C.) is outstanding. The Celts, established in the centre and the west, continued their peculiar culture into Roman times.

On the Mediterranean coasts, the first millennium B.C. witnessed the renaissance of the indigenous people, influenced by colonization: first Phoenician, continued by the Carthaginians, on the south coast and in the island of Ibiza, then Greek, on the south coast and in the northeast (Emporion, Rhode). Many trade objects, jewels, and pottery fill the collections. Thousands of native settlements or cities show a rather rudimentary planning. The enigmatic city of Tartessos, head of a kingdom in Andalusia, gives its name to a variant of this culture. Its archaeological remains are limited to some jewelry and a few inscriptions in an archaic alphabet of eastern origins. The Iberian culture includes cities, sanctuaries, and cemeteries, and in them much stone, bronze, and ceramic plastic art. The "Lady of Elche" (4th century B.C.?) is the most representative work of this group. The animal sculpture offers clear eastern prototypes. Wheel-made pottery with geometric decoration is conspicuous. In some regions stylized plant decoration is very rich (Elche, Archena), but the most important is the human decoration which is found in the southeast and most abundantly at Liria (Valencia). It depicts scenes of war, hunting, fishing, sacrifices, dances, and domestic life. There are also numerous inscriptions in the Iberian alphabet, derived from the Tartessian.

MANSELL/ALINARI

"THE LADY OF ELCHE"

Iberian sculptured bust (probably 4th century B.C.), in the Museo del Prado

In 218 B.C. the Romans entered Spain and began the process of romanization, which was accelerated in the second half of the 1st century B.C. Indigenous elements became buried under the avalanche of foreign cultural, political, juridical, and artistic elements. Study of the evolution of pottery permits a clear chronology. In the second half of the 1st century B.C. the so-called terra sigillata began to spread, made first in Italy, then in Gaul; Spanish manufacture began at the same time. In the 3rd and 4th centuries A.D., manufacturing became more common and stamped ware predominated, continuing until the Visigothic era. A new chapter of Roman archaeology has been opened by underwater exploration; among the many remains of sunken ships are a great number of amphorae in which wine and olive oil were exported.

Roman cities sometimes rise over Iberian towns, but in many cases on the plain nearby (Liria, for example); others are founded entirely afresh (Mérida), or derive from camps (León). The peace which reigned during the first centuries of the empire scattered the population through the countryside, in innumerable villas decorated with mosaics. The cities offer great monuments: temples, theatres (that of Mérida is one of the best in the Roman world), amphitheatres (Italica, near Seville), circuses, and town walls. The same richness and perfection are shown in the aqueducts (Segovia), bridges (Alcantara), triumphal arches, and monumental tombs. In the second half of the 3rd century, with the first invasions, the cities once again enclosed themselves behind strong walls and all life was impoverished.

The spread of Christianity revealed new aspects: sarcophagi, tomb mosaics, and basilica mosaics. The number of known basilicas, showing clear African and Eastern influence, constantly grows. The last chapter of Hispanic archaeology is usually found in the invasion period and early medieval times. (L. P. G.)

V. HISTORY

A. PREHISTORY

Paleolithic.—Lower Paleolithic industries are well represented in Spain, especially in the central Meseta north and west of Madrid and in Cantabria, where conditions in the penultimate interglacial period appear to have been particularly favourable to flourishing hunting communities. The Meseta group is somewhat earlier than the Cantabrian and includes industries of Clactonian type, but in general Acheulean is the first widespread industry in Spain and is found from Cantabria to Andalusia. In the east, however, industries before the Mousterian are rare. In the southern provinces, Clactonian is well though sparsely represented; other southern stations are almost entirely Mousterian.

In the central Meseta and in Cantabria the widespread Mousterian culture is continued into a mixed Levalloisian-Micoquian. Classical Levalloisian and Tayacian are found at a number of central, southern, and eastern sites. In the east industries are French in type and compare with classical Mousterian of Charente and the Dordogne. Mousterian man is represented by bones of Neanderthal type.

In the Upper Paleolithic the Cantabrian-Biscay region formed a continuous cultural area with south-central France. While the classical Aurignacian (French Middle Aurignacian) is not represented outside the Cantabrian zone, the transition from Mousterian to Aurignacian is found not there but elsewhere. Later Aurignacian with Gravettian is well represented in southern and eastern Spain and is the earliest known industry over most of this region; it includes types which have a certain parallelism with the Capsian of North Africa and the Grimaldian of southeast France and northern Italy. Solutrean occupied Cantabria and Catalonia and extended down the east coast as far as Valencia.

Magdalenian industry is particularly well represented in Cantabria. Throughout the region typical Magdalenian carved bonework, harpoons, and animal remains attest to the temporary onset of cold conditions. Magdalenian remains are also found in Catalonia and at the important cave site of Parpalló, Gandía (Valencia). Doubtless there were scattered Magdalenians throughout most of the peninsula but their influx was not strong enough to upset the widespread late Gravettian people, who in a cultural phase known as Epigravettian remained the basic population at the end of the Upper Paleolithic, with Capsian peoples on the Levantine, and Epipaleolithic-Mesolithic Azilian and Asturian cultures in the north and west seaboards.

For the world-famous Spanish cave paintings, *see* PRIMITIVE ART.

Neolithic and Bronze Age.—After the invention of pottery, Capsian flints continued in the earliest phases of the Spanish Neolithic, which first appeared in eastern and southern Spain in coastal caves and is therefore known as the "Cave Culture." Stratification of these cave deposits is difficult, but grain cultivation and the use of querns certainly reached the Spanish Cave Culture at the same time as similar cultures in North Africa, Liguria, and southern France.

There is, however, little to separate the Neolithic and the earliest metal-using cultures, for Spain was especially rich in metal ores, and a rich copper-using culture soon developed in the southeast. This was centred in the province of Almería, where most of the material comes from megalithic passage graves, probably introduced by an invading Mediterranean people. Undoubtedly there was a pre-metal phase of the Almerian culture, but this is

difficult to define. Two habitation sites of round and oval huts have associated passage graves and have no traces of metal. One of these is El Garcel, a village of wattle and daub houses and grain storage pits. Agriculture (grain, olives, vines) was well developed, but the flint industry is still the primitive Capsian. Pottery is related to that of North Africa, and there is a little stone idol with eastern Mediterranean affinities. Some fragments of copper are, however, known in later phases.

Later village sites in Murcia Province have rectangular houses. A number of dolmen and cave burials and an extension of the Almerian culture into Andalusia also fit into this phase. In these earlier phases burial was in cists and collective round megalithic graves of dolmen type.

The richest phase of the Almerian culture (*c.* 2000–1500 B.C.) is found at Los Millares on a bank of the Andarax at the base of the Sierra Nevada, a triangular site defended by a fosse and ditch and predominantly associated with passage-grave burial. Copper becomes more abundant, used only for small objects such as earrings and pins. During this phase the Almerian culture spread into the southwest and into Portugal, and throughout "beakers" related to those of West and Central Europe are found.

The fourth phase of the Almerian culture (*c.* 1500–1000 B.C.) is fully Bronze Age and is best represented at the small prehistoric township of El Argar. Here there is a complete change in burial (to jar burial) and skull type, obviously indicating the arrival of new people. In one grave were found eight segmented faïence Egyptian beads of a type known to date to *c.* 1400 B.C.; these provided the first fixed chronological point in Iberian prehistory.

The Almerian culture can also be traced into southern Portugal, where there occur a number of passage graves of the second and third phases. In the mountain lands there is a progressive impoverishment, seen in megalithic tombs without vaulting.

The full Bronze Age in Spain was a period of retrogression except in the southeast where around El Argar on the Río de Antas a flourishing culture arose and spread throughout the southwest where it is mainly represented by naturally defensive settlement sites in Almería Province and in Murcia. Copper was still more common than bronze and many molds for metal casting were found; gold and silver were used for beads and other ornaments; metal mining was probably the chief source of wealth. A pair of pottery horns from El Oficio (Almería) suggests eastern Mediterranean connections.

This strong Bronze Age culture of El Argar flourished throughout the second half of the 2nd millennium B.C. and spread north to Catalonia, east to Alicante, south to Andalusia and Portugal; north of the Segura River it becomes less typical, especially in Catalonia; here as in other peripheral regions pottery of Neolithic Cave Culture type persists almost to the Iron Age. In the northeast in the final phases of the Bronze Age new pottery types and bronze ax heads of European type foreshadow increased contacts across the Pyrenees and the drawing of the Iberian Peninsula into the Central European Late Bronze Age cultural orbit.

Atlantic Bronze Age.—The final phases of the Bronze Age culture flourished on the Atlantic coast and in the extreme southwest. New South European and central Mediterranean influences brought to these regions new metal types, above all a notched shield (the Herzsprung shield) which is known in Cyprus and Samos and can be dated *c.* 720. No actual shields survive in southern Spain, but a number are engraved on flat rock pictures, chiefly in Extremadura. Besides shields, these rock pictures show carp's-tongue swords, molded fibulae, combs, mirrors, and wheeled chariots, all of which suggest a local invasion of a Mediterranean people introducing new metal techniques. Close cultural contact between the Atlantic seaboard and Galicia and Brittany and Ireland is particularly strong in bronze and gold work.

Iron Age.—Iron-using Celtic peoples of trans-Pyrenean Hallstatt culture invaded Spain in waves from 900 to 500 B.C. (these cannot be precisely dated). Their first phase of settlement, lasting down to 650 B.C., brought them in concentrated numbers as far south as the Ebro and penetrated lower Aragon. Villages are known, made up of rectangular clay and stone houses facing on to a simple pattern of streets. Late offshoots of this first Hallstatt phase passed to the south through Castellón, and even Almería was influenced by the early Catalonian urnfields. This phase reached the west coast (Alcácer do Sal) and Guadalquivir Valley (Carmona); also, antennae-hilted swords of a type used in the final phases of Central European Hallstatt were engraved on stones in Extremadura, and bronzes of Hallstatt type were found in the Huelva River.

A second major wave between 600 and 400 B.C. carried Hallstatt culture over the central Meseta into Lusitania and the far southwest as well as into Galicia, where in the last four centuries B.C. an evolved culture of scattered forts (*castra*) aligned itself with northern and central Portugal rather than with the west of Spain. Avienus' *Ora Maritima* speaks of Celtic tribes occupying the Meseta in the 6th century, and ancient place names ending in *-briga*, *-dunum*, and *-acum* attest to a Celtic settlement which must certainly have taken place well before the Central European Celts passed about 450 B.C. into the La Tène cultural phase, which did not reach or influence Spain.

Four main areas stand apart from the general pattern of Celtic settlement. The lands of the north coast were settled by the Cantabri (*q.v.*), Astures, and Vascones (modern Basques), who, while partly celticized, were ethnically separate from the Celtic Galicians of the west and such Celtic tribes as the Ilergetes in the east. The south coast between Gibraltar and Málaga was occupied by a mixed race of local Bronze Age peoples with a Phoenician element called Bastulo-Poeni. South Catalonia, part of Aragon, Valencia, Castellón, Granada, Jaén, and Andalusia were held by the non-Celtic Iberians (*q.v.*), a Mediterranean people with Hallstatt burial rites but of strictly unknown ethnic formation who had a written language and excellent pottery of non-Central-European derivation and were heavily influenced by Greek and Phoenician culture. Although many of their metal types were derived from the Urnfield people and their villages were very similar to those in Aragon and Catalonia, their shrines and chambered tombs were decorated with reliefs and free-standing sculpture in an exotic and highly individual style. They were masters at bronze casting and hundreds of small bronze votive statuettes provide a picture of their physical characteristics. Lastly, in the Ebro Valley and the northeastern Meseta the Iberians and Celts fused into a group of tribes known to classical authors as the Celtiberians (*see* CELTIBERIA). This mixed Celtiberian culture spread over most of east and central Spain in the century preceding the Roman conquest.

B. Early Colonization

Tartessos.—The rich mining culture of southeast and southwest Spain first became known to the classical world by the voyage of Colaeus of Samos, who sailed there *c.* 630 B.C. and opened trade with the region of Tartessos. Classical sources agree that before the Greeks discovered southern Spain the Phoenicians traded with Tartessos (according to some Greek writers Gades [Cádiz] was founded *c.* 1000 B.C.); consequently, it has been equated with the biblical land of Tarshish, to which the ships of Solomon and Hiram of Tyre made three-year round voyages. Possibly the names of Tartessos and Tarshish can be philologically connected, and there is no doubt that while biblical writers used the word "Tarshish" as a rather general geographical term for lands where minerals were mined, it was nevertheless properly regarded as being at the western end of the Mediterranean. Tartessos itself, according to Strabo, was an island city at the mouth of the Guadalquivir and was destroyed by the Carthaginians about 500 B.C.; its site was unknown in Roman times and has remained so. From the 7th century B.C. the Guadalquivir Valley was occupied by an Iberian tribe called Turdetani, a name whose root might also have some connection with *Tart*essos.

Phoenicians.—Existing archaeological evidence shows that Phoenician seaborne trade reached the estuaries and lower reaches of the Guadalquivir, Guadiana, and Tagus rivers about 600 B.C. A few metal objects in southern Spain attest to local manufacture of Phoenician objects, some of which found their way to the Celtic hill-forts on the west of the Meseta. Phoenician influence in the southeast is less marked, but there seem to have been early

contacts with the region of Valencia and also with the inland Iberians of Tutugi (Galera district, Granada), of Peal de Becerro (Jaén), and of Cástulo (Jaén). Although nothing as early as the date of Solomon and Hiram has been found, there is every likelihood that the Phoenicians obtained ores from the El Argar people long before active trade between the Phoenicians and Tartessos.

Greek Settlement.—From the mid-7th to the late 6th century B.C. Greek contact with eastern Spain was incessant. Besides the commerce with Tartessos, Greek trading settlements were founded on the east coast at Hemeroscopeion (Punta de Ifach), Mainake (possibly Málaga or Almuñécar), and the unlocalized Oinussa, Molybdana, and Heracleia by the Ionian Phocaeans. The Rhodians founded Rhode on the Golfo de Rosas in the northeast in the 7th century. After the Phocaeans founded Massilia (Marseilles) *c.* 600 B.C., this northern corner of Iberia became important; and the first permanent large Greek colony on Iberian soil was founded from Massilia at Emporion (modern Ampurias on the coast of Catalonia). Although, meanwhile, the Phocaean foundations to the south had fallen to the Carthaginians, Ampurias remained independent down to the Roman occupation and did a brisk trade with Punic Ibiza as well as with the Iberian tribes of the interior. There are considerable objects of Greek trade found in the southern and eastern regions of Spain. The cultural impact of the Phocaean Greeks upon the indigenous Celtic tribes of the north and the Iberians of the south is very great and far outweighs that of the Carthaginians.

Carthaginians.—About 520 B.C. the Phoenician settlements came under the chief Phoenician city in the west, Carthage (*q.v.*), and with them access to the mines and a recruiting ground for mercenaries. After the First Punic War (264–241 B.C.) Carthage was left with Gades and a few other coastal settlements in Spain. From 237/236 Hamilcar Barca and his successors Hasdrubal and Hannibal (*qq.v.*) extended Carthaginian rule over many tribes. Hasdrubal founded Carthago Nova (Cartagena) and accepted the Ebro as a northern frontier in a treaty with the Romans in 226/225. Hannibal's attack in 219 on Saguntum (south of the Ebro but allied to Rome) led to the Second Punic War (*see* PUNIC WARS). (WM. C.)

C. THE ROMANS

Victory in the Second Punic War (218–201 B.C.) left Rome with two embryonic Spanish provinces: Hispania Citerior, the eastern coastal strip based on Tarraco (Tarragona) and Carthago Nova; and Hispania Ulterior, comprising the southern coast and the valley of the Baetis (Guadalquivir); the two are often referred to as Hither and Farther Spain, respectively. Thereafter the task of deepening the provinces by advance inland was bitterly pursued. Year by year, with each province under its proconsul, Roman soldiers and Roman money were poured in, until Scipio Africanus' epic capture of Numantia in 133 B.C. allowed advance northward to the Durius (Douro) and westward to the middle Tagus. Administration, though usually opportunist, was not always merely oppressive or self-interested; and although annual tribute was imposed on native communities they received certain advantages—security, peace, safe communications, new opportunities of production and commerce, employment under Roman arms, and acquaintance with Roman civil law. Agriculture, viticulture, and fisheries flourished in the south and east, as the local coin types attest, even if these activities were controlled by Roman capital; trunk roads were constructed (*e.g.*, that from the Pyrenees to Carthago Nova), and coinage developed from the proceeds of the innumerable mines, of which Rome claimed the gold while leasing out the silver, lead, and (probably) copper.

When in 80–72 B.C. the Roman Quintus Sertorius used Spanish supporters against Roman political rivals, he did so as the leader of an already half-romanized army. Julius Caesar's governorship of Hispania Ulterior (61–60 B.C.) saw the frontiers extended up to the mouth of the Durius. Later, as dictator, he initiated a new colonial and municipal policy, bringing Italian settlers to Spain and granting Latin rights to certain indigenous communities. Gades became a fully fledged Roman *municipium*. Roman peace was swiftly followed by Roman political consciousness; language, religion, and public life were henceforth molded and unified by the central Roman model.

Augustus' first task was to undertake, by bloody and savage warfare, the conquest of the northwestern (Asturio-Cantabrian) corner of Spain—not complete until 19 B.C. Now, with the whole peninsula tamed, he effected its total reorganization, though the precise date (in any case earlier than 2 B.C.) is uncertain. Baetica (roughly the Guadalquivir Valley with the southern coast) became a "public" province, without regular troops, under senatorial proconsular administration; Lusitania (a new province roughly coextensive with modern Portugal south of the Durius) and Tarraconensis (comprising all the rest of the peninsula) were imperially administered under propraetors with three legions as a permanent garrison in the latter. Major regional towns or cities became centres of government and local jurisdiction. Emerita Augusta (Mérida), a pattern of Augustus' many newly founded watchdog colonies of veteran soldiers, typifies, by the wealth of its remains today, their imposing character. Colonies and full *municipia*, formed *ex hypothesi* of Roman citizens, were self-governing and subject only to citizens' taxes. Unprivileged communities were administered by the provincial governor and subject still to payment of tribute. Although Augustus was chary in extending the Roman franchise, his work for Spain was enormously beneficial and widely recognized as such. The building of roads and bridges, the encouragement of local enterprise, the unification of sentiment, and the establishment of profound peace, while raising the standard of living everywhere, were not imposed at the cost of rigid conformity. Political, religious, and financial anomalies were tolerated. Rome's principal reward consisted of the vast output (perhaps unequaled elsewhere) of gold, silver, copper, iron, tin, and lead from Spanish mines.

The Augustan model was faithfully preserved. Claudius reduced the garrison to two legions, Vespasian to one. The road complex was developed swiftly. Spanish wine, by its cheapness, and oil, by its quality, disturbed the previous pattern of world markets. Vespasian imposed stricter regulations for the control of all the principal Spanish mines, the revenue of which afterward flowed into the imperial treasury. But early in his reign he recognized a century of astonishing progress by simultaneously promoting about 350 unprivileged communities to Latin rights. Spain could not now deny the presence of the golden age, marked by security, a high and well-diversified culture, and a balanced and critical civilization, which gave Rome the younger Seneca, Lucan, Martial, Quintilian, Columella, and Pomponius Mela as literary or academic ornaments, and the emperors Trajan and Hadrian. The 3rd and 4th centuries saw initiative, if not prosperity, declining. City walls (some hastily built in the 3rd-century crises) crumbled in neglect, through excessive dependence on Rome; and when the barbarians burst the Rhine frontier in A.D. 406–407 and flooded Gaul and Spain, there could be neither resistance nor hope of recovery. (C. H. V. S.; X.)

With the irruption in 409 of the Vandals, the Suebi (Sueves), and the Alani (Alans), Spain entered on a long period of division and confusion. The Vandals and their associates, who plundered far and wide, were not numerous enough to swamp the population. When in 429 Gaiseric led the Vandals and Alani out of Spain into North Africa, his whole horde numbered only 80,000 persons, including old men, women and children, and runaway slaves. The contemporary writers Orosius and Salvian say that many subjects of the empire in Gaul and Spain preferred the barbarians to the tyranny of the imperial tax collectors. The great landowners had, moreover, almost as much to fear from the agrarian insurgents known as Bagaudae as from the barbarians. The most resistance to the barbarians came from the least romanized part, the north; the Basques were independent until 574. Spain was thrown back into its pre-Roman state of division, with the vital difference that the country now possessed the tradition of the Roman law, the municipalities, and one great organization in the Christian church (unfortunately, the barbarians were usually converted to Arian Christianity).

After the fall of the rival emperors, first Constantine, then

BY COURTESY OF THE STIFTSBIBLIOTHEK, ST. GALLEN SWITZERLAND

VISIGOTHIC LEGISLATOR

Representation of a legislator from *Leges Barbarorum*, a Visigothic collection of laws written in A.D. 793

Maximus (*qq.v.*), supported by the Roman army in Spain, the Romans enlisted the help of the Visigoths (*see* GOTHS). Ataulphus (*q.v.*), the successor of Alaric I and husband of Placidia, sister of the emperor Honorius, entered Spain in 415 but was murdered the same year. Wallia was soon elected king and continued Ataulphus' work. He crushed the Alani and drove the Vandals and Suebi into the northwest. But then he handed Spain back to the imperial officials and marched with all his people (418) into Aquitania Secunda, which had been assigned to them by Honorius.

The Romans did not give up hope of recovering Spain, even after the Visigoths (*see* below) were firmly established. In 552 Justinian took advantage of a disputed Visigothic succession to send a force which occupied much of the south coast for the Eastern Roman Empire. The Visigoths regained Corduba in 572 and Carthago *c.* 622; part of Algarve was Byzantine until 629. (X.)

D. THE VISIGOTHS

The Visigothic rulers of Aquitaine began to intervene in Hispanic affairs early in the 5th century but it was not until 494, during the reign (484–507) of Alaric II, that large-scale migration to Spain began. The migrants, who certainly did not number more than 200,000, included peasants as well as warriors and clergy and established their overlordship without serious opposition (497). In the areas where they elected to settle, the Hispano-Roman landowners (*hospites*) were compelled to surrender to the newcomers two-thirds of their cultivated land (*sortes Gothicae*), but there was no general expropriation even in Old Castile, where Visigothic settlement was densest. The political capital of the Visigothic monarchy was transferred from France to Toledo in the reign (551–567) of Athanagild.

Though the invaders were both politically weak and already highly romanized, powerful influences at first prevented any rapid assimilation to the native population. The Visigoths were followers of the Arian heresy and remained subject to their own legal codes, while the Hispano-Romans were Catholics and continued to be governed by Roman law. Intermarriage between the two races was prohibited. Instability was to be the outstanding feature of the state set up by the Visigoths in Spain. Because of the ambitious Visigothic nobility's strict insistence on the principle of elective monarchy, the crown was constantly menaced by political intrigue, and the removal of the sovereign by deposition or assassination was frequent. Leovigild (568–586)—one of the few effective Visigothic kings—finally destroyed (585) the remnant of the Germanic kingdom set up by the Suebi in northwestern Spain about 175 years before and managed to bring the ever-rebellious Basques under his control; he also dealt successfully with a Catholic rebellion against the state religion. But the next king, Reccared (586–601), was baptized a Catholic soon after his accession, and the third Council of Toledo then proclaimed the conversion of the whole kingdom (589).

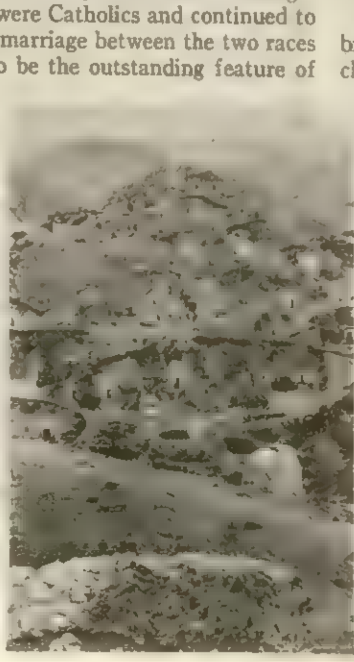
The removal of the religious barrier allowed the assimilation of the two groups to progress more rapidly, and the councils of Toledo, where nobles as well as clergy were represented, now began to dominate the crown. The reign (621–631) of Swintila saw the final expulsion of the Byzantine prefecture set up in southern Spain in Athanagild's time. In the time of Sisenand (631–636), the fourth Council of Toledo completed the subjection of king and government to ecclesiastical authority. An event of great significance for the subsequent history of medieval Spain was the promulgation by Recceswinth (649–672) of the *Liber judiciorum* as the sole legal code for both peoples. This was a new compilation, drawn up by order of the king *c.* 654, in which Germanic customary law predominated over Roman legal traditions.

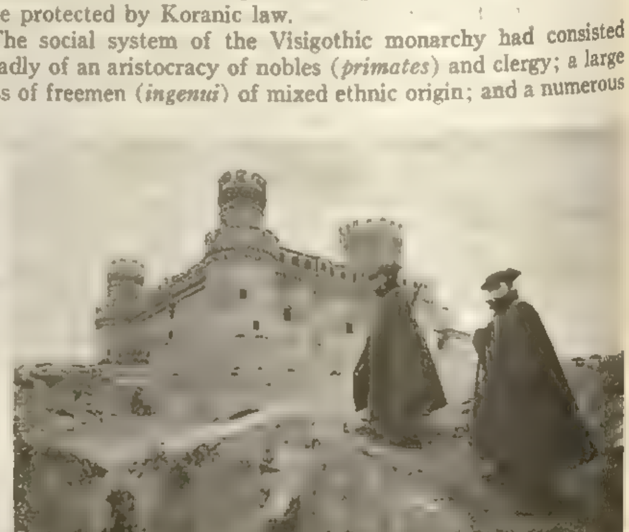
It is probable, however, that the extent of assimilation brought about by Reccared and Recceswinth has been somewhat exaggerated. Even after their reforms only nobles of pure Visigothic descent could be elected to the throne, and the Germanic nobility was thus encouraged to maintain its blood line up to (and after) the final fall of the monarchy.

Wamba (672–680), the last king of any stature, attempted to reorganize the decadent military institutions of his kingdom but was deposed by a trick; and the energies of the next three kings were chiefly concerned with holding their thrones against the intrigues of their political enemies. By the time of Roderick (Roderic; 710–711) the disintegration of the state was complete. Apparently as a result of an invitation from Roderick's opponents, Tarik, the Arab governor of Tangier, crossed the Strait of Gibraltar with an invading Arab and Berber army in 711. Roderick called for national unity against this grave threat, but even then the Visigothic nobles declined to abandon their suicidal partisanship. Defeatism was deliberately encouraged by some of them in the army that Roderick had assembled; in consequence it was routed by Tarik at the Battle of Guadalete, south of Medina Sidonia (July 19–26, 711), the king himself disappearing from history.

The collapse of the Visigothic monarchy in Spain followed at once. A substantial number of members of the large slave population abjured Christianity to improve their social position; most of the peasantry did likewise. The Jews, subjected to fierce persecution since the reign of Sisebut (612–621), welcomed the tolerant invaders as liberators. A number of high Visigothic nobles accepted Arab suzerainty, and the invaders speedily completed the occupation of almost the whole peninsula. Though the Arabs took one-fifth of the land for themselves the Christian communities were left with their religious and legal independence intact and were protected by Koranic law.

The social system of the Visigothic monarchy had consisted broadly of an aristocracy of nobles (*primates*) and clergy; a large class of freemen (*ingenui*) of mixed ethnic origin; and a numerous





(LEFT) INGE MORATH—MAGNUM; (ABOVE) JOHN LEWIS STAGE—PHOTO RESEARCHERS, INC.

VESTIGES OF SPANISH HISTORY

(Left) Cave dwellings near Guadix, inhabited by gypsies; (above) Spanish civil guards approach the Castle of the Mendozas, a 15th-century stronghold, at Manzanares el Real

slave class (*servi*). The *aula regia* (royal household) was made up of palace officials (*palatini*), who held the title of count by virtue of their offices. Provinces were governed by *duces* or *comites* and subdivided into areas ruled by *judices*. This stratified society without clear political ideals proved incapable of creating any sense of national cohesion during the two centuries of Visigothic rule in Spain. Though they were by no means without creative capacity in letters and the visual arts, it was as lawgivers that the Visigoths exercised their greatest influence on the history of Spain. Recceswinth's *Liber judiciorum* remained the basic legal code of Christian Spain for centuries after the disappearance of the last Visigothic king; and, by a curious paradox, memories of the feeble Visigothic state were now to prove strong enough to sustain the small Christian states of the north in their long resistance to Islamic military and cultural power. (P. E. R.)

E. Moorish Spain to 1031

During the reign in Spain of the Visigothic king Witiza (701–709) Arab forces of the caliphate (*q.v.*) had conquered northern Morocco and laid siege to Ceuta, the last remnant of the Byzantine possessions. It seems certain that Julian, the imperial count or governor of Ceuta, incited the Arab viceroy in North Africa, Musa ibn Nusair, to attack Spain, possibly by arrangement with the sons of Witiza. After a successful preliminary raid by the Berber Tarif in 710, Musa ordered his governor at Tangier, Tarik ibn Zaid, to make a descent in force. In 711 Tarik crossed with 7,000 men, mostly Berbers, and, after reinforcement by 5,000 more, awaited the advance of Roderick, Witiza's successor, near the Laguna de la Janda, south of Vejer de la Frontera. There, on July 19, Roderick was defeated and probably killed. The sons and partisans of Witiza, who had withdrawn during the battle, now joined Tarik and encouraged him to advance northward to seize Toledo, while a detached force occupied Córdoba. In June 712 Musa crossed with an army of 18,000, mostly Arabs, captured Seville and Mérida, and dispatched his son Abdul-Aziz to the southwest, before he joined Tarik at Talavera. After wintering at Toledo, he resumed his advance in 714, captured Saragossa, and made with Tarik a two-pronged expedition into León and Galicia before returning to Damascus at the caliph's command. His son Abdul-Aziz, after occupying Portugal, completed the conquest of Granada and Murcia, where the duke Theodomir (called Tudmir by the Arabs) became tributary. Other Arab forces pushed northeast into Septimania; and later governors continued to raid into Gaul, where, although defeated by Charles Martel at Poitiers in 732, they maintained an advanced base at Narbonne until its capture by Pepin III in 751.

The invaders met little opposition on the whole. The sons of Witiza and other great Visigothic families, whether converted to Islam or not, compounded by payment of tribute for extensive domains. The Jews, freed from persecution, were eager allies, and the serfs gained a measure of freedom. Conversions were on a large scale, and the Spanish converts (*muwallads*) became an active and turbulent element in the general Moorish population. The unconverted, called *mustaribs* (*mozárabes*), did not suffer much from interference and, like the Jews, formed prosperous communities in the Muslim cities. The Arabs, in spite of continuous immigration, were too few to colonize the country; they formed the administrative and military cadres, maintained by the allotment of fiefs in the region of Saragossa, the eastern and southern coasts, and the Guadalquivir Valley. The Berbers settled mainly in the centre and in mountainous regions which resembled their native land and favoured their anarchical tendencies. The Berber revolt in Morocco in 740 produced a parallel outbreak in Spain, and Berber risings, large and small, continued for the next two centuries. Among the Arabs, also, the intertribal feud between Kais (Qays) and Kalb (*see* Caliphate) was reflected in Spain, complicated by rivalries between the early immigrants (*baladis*) and the Syrian contingents (*shamis*) which came in from 741 onward. In this patchwork of racial and tribal discords, rebellion was almost endemic, and the resulting lack of unity remained a permanent characteristic of al-Andalus, as Moorish Spain was called (*see* Andalusia).

The Omayyad Dynasty.—In 755 the Omayyad prince Abd-al-Rahman ibn Mu'awiya, escaping from the massacre of his relatives in Syria, landed at Almuñécar, and with Kalbite support captured Córdoba (756). He waged pitiless warfare against internal disorder for more than 30 years and established a centralized power based upon an imperial guard of Berbers and European slaves (Slavonians). The pious Hisham I (788–796) tried to stem the expansion of the Asturian kingdom. A series of *muwallad* revolts, at Saragossa, at Toledo, at Mérida, and in Córdoba, threatened to dislodge his successor, al-Hakam I (796–822), who, however, by his implacable resolution reaffirmed the Omayyad power and laid the foundations of the material and literary culture which developed under the urbane Abd-al-Rahman II (822–852). The peace was again broken under Mohammed I (852–886) by successive *muwallad* and Berber revolts, often with Asturian support, notably those of the "Visigothic" Beni Kasi at Tudela and Saragossa, and of Omar ibn Hafsun in the Sierra de Ronda. General disintegration and civil war among and between Arabs, Berbers, and *muwallads* in all quarters forced his son 'Abdallah (888–912) to adopt a temporizing and defensive strategy.

'Abdallah's grandson Abd-al-Rahman III (912–961) by resolute and far-sighted leadership lifted the kingdom from this morass to an apogee of power, culture, and magnificence. Within ten years he mastered Ibn Hafsun (d. 917) and his sons, and in ten more he reunited Moorish Spain. He proclaimed himself caliph in 929. In spite of occasional reverses, he forced the Christian princes to pay tribute to Córdoba, now a metropolis enjoying immense reputation and prestige in all Europe. He greatly enlarged the Slavonian guard, whose officers acquired increasing power at the expense of the old Arab aristocracy. The reign of his son al-Hakam II (961–976) was no less brilliant. The weakness of Hakam's son Hisham II (976–1013) favoured the rise of a military dictator, the Arab Ibn abi-'Amir, later entitled al-Mansur (Almanzor) (981–1002), under whom, as under his son 'Abd al-Malik al-Muzaffar (1002–08), the military power of the caliphate, now recruited mainly from Morocco, was at its height. But beneath the surface the old antagonisms and social discontents were still alive. In 1009 they broke out in a political crisis which temporarily reduced Moorish Spain to chaos and from which it never fully recovered. While Slavonians and Berbers by turn occupied Córdoba and made and unmade caliphs, the rest of al-Andalus threw off all allegiance to the Omayyads and was dissected among a host of petty rulers. After Córdoba itself had been occupied by the Moroccan Arab house of Hammuda (1016–27), the last Omayyad prince was dethroned in 1031 and replaced by a republic. *See* also Abd-al-Rahman; Córdoba, Caliphate of. (H. A. R. G.; X.)

F. Christian Spain to 1479

The history of Christian Spain during the period 711–1479—a subject long distorted by religiopatriotic myths originating in the time of the Catholic Kings (Ferdinand and Isabella)—is dominated by the interplay of two opposed forces: the idea of an ultimate Christian conquest of the whole peninsula; and the inescapable fact that, historically, culturally, and economically, the Christian and Islamic peoples of Spain had become irretrievably associated with each other. The Christian will to reconquer, based partly on religious feeling, partly on the Visigothic tradition of Spanish unity, and mostly, perhaps, on economic motives, only sporadically took on a militant crusading form before the end of the 11th century. When it did so the economic consequences of the crusading spirit were often scarcely less harmful to the Christian conquerors than to the conquered Moors. Thus Ferdinand III's conquest of Andalusia in the 13th century was undertaken in a spirit of intolerance that brought permanent economic ruin to southern Spain and so severely damaged the economic prospects of the conquering state, Castile. This experience, and the very different results of the conquest of the Moorish kingdom of Valencia by the economically more educated Aragonese and Catalans, taught the Castilians prudence and the fact of the *convivencia* between Moor and Christian reasserted itself. From the middle of the 13th century to 1479 the Castilian attitude to Granada was decidedly ambivalent while Castile, Aragon, and Navarre all ac-

cepted the presence within their frontiers of substantial numbers of Islamic subjects (mudéjares) to whom royal protection was accorded.

The Kingdom of Asturias.—Before the tide of Arab and Berber invasion a number of Visigothic nobles and their retinues withdrew to the mountains of Asturias and there elected a Visigothic prince, Pelayo (*c.* 718–*c.* 737), as their king. The Berber revolt in Morocco and Spain (741) and the consequent withdrawal of the Berber garrison in Galicia led to the attachment of Galicia to the new kingdom by Alfonso I (739–757). Under Alfonso II (791–842) the Asturian capital was moved from Cangas de Onís to Oviedo. Later kings pushed forward from the mountains to the central plain and their armies operated with success far to the south. The discovery, early in the 9th century, of the supposed tomb of St. James at Padrón in Galicia and the establishment of the shrine at Santiago de Compostela made the Asturian kingdom guardian of a shrine of European significance and gave it a symbol of national unity. Alfonso III (866–910), the greatest of the Asturian kings, expanded his domains as far as Coimbra in the southwest and Burgos in the southeast. By now the situation was stable enough for his successor, García I (909–914), to move the capital from the Asturian mountains to the city of León.

The Asturian kings, like their Leonese successors, regarded themselves as direct heirs to the traditions of the fallen Visigothic monarchy; and, for this reason, the historical writings of Alfonso III's reign emphasized their obligation to undertake the reconquest of all Spain. The Asturian administration followed Toledan models as closely as circumstances permitted. Yet the organization, social and political, inherited from the Visigoths was already being changed by the immigration of a free Mozarabic peasantry from the south, while Frankish political ideas also exerted some influence on Asturian society. In general, however, it can be said that the Asturian kings handed on to the kings of León a tradition of isolation, not only from Moorish Spain but also from the rest of Christian Europe, which experience was to show could not be sustained.

The Kingdom of León.—The transfer of the Christian capital from Oviedo to León coincided with the establishment of the caliphate of Córdoba and, until the early 11th century, the balance of power in the peninsula swung strongly in favour of the Moors. That the Leonese kingdom survived at all was the result of the generalship of Ordoño II (914–924) and Ramiro II (*c.* 931–951) who, in unfavourable circumstances, secured important victories against the caliph's armies. Ramiro was, however, unable to prevent the breakaway from his kingdom of the Christian county of Castile. This district, partly recolonized by immigrants from the north, had rejected the Visigothic legal code (*Forum judicum*) in favour of Castilian customary law and showed itself hostile in other ways to the Visigothic traditions of León until, under Count Fernán González (932–970), it achieved *de facto* independence. Henceforth clashes between the innovating tendencies of Castile and the conservative traditions of León were to weaken the Christians. Ordoño III of León (951–956) acknowledged the hegemony of the caliphate, and the later attempts of Bermudo II (984–999) to assert his independence of Córdoba brought disaster on the kingdom. The armies of Ibn abi-'Amir al-Mansur occupied and destroyed León (988) and sacked Santiago de Compostela itself (997), sparing only the Apostle's tomb. Under Alfonso V (999–1028) the capital was repopulated and recovery began. However, early in the reign of Bermudo III (1028–37), a new threat to the independence of León materialized when Sancho III the Great of Navarre, after seizing the county of Castile (1028), occupied the eastern part of the Leonese kingdom. Later he entered the capital (1034) and there assumed the title of "emperor of Spain," Bermudo fleeing to Galicia. León was never wholly to recover its supremacy again.

The Rise of Catalonia and Navarre.—During the reign of Charlemagne the Franks succeeded in capturing Barcelona (801) and in driving the Moors out of the northeastern corner of the peninsula. This area, later known as the Marca Hispánica, was then occupied by a number of Visigothic counts under Frankish suzerainty. In the time of Count Guifredo (d. 898) the county of Barcelona became independent and established its hegemony over the other counties of the Marca Hispánica. Under Borrell II (*c.* 950–992) al-Mansur's armies devastated Catalonia and burned Barcelona (985); but the region recovered during the rule of Ramón Borrell (992–1018). Under Berenguer Ramón I (1018–35) it began to flourish. Because of the influence of the Franks it had already developed political and social institutions different from those of the rest of Christian Spain. The 10th century also saw the dramatic rise of the Basque kingdom of Navarre, which has little history before this period. Sancho I Garcés (905–925) added the small county of Aragon to his dominions but suffered severely at the hands of Abd-al-Rahman III's armies. A period of subordination to the caliphate followed, but Sancho II Garcés (970–*c.* 994) was spared the worst of al-Mansur's attacks and devoted himself to the political organization of his kingdom. There was, however, nothing to indicate that, under Sancho III Garcés, called "the Great" (1005–35), Navarre was to become, for a brief period, the dominant power in Christian Spain.

The Medieval Spanish Empire.—Though Sancho the Great had assumed the imperial title in León, his political ideas were hostile to the old Visigothic-Leonese tradition of peninsular unity, and his empire was largely built on a simple desire for personal aggrandizement. It included Navarre, Castile, León, Sobrarbe, and Ribagorza while Berenguer Ramón I of Barcelona became his vassal. Sancho deliberately encouraged contacts between his dominions and the rest of Christian Europe and probably did much to make the Spanish nobles aware of the feudal aspirations of their class north of the Pyrenees. He regarded his empire as his personal property and distributed it among his four sons in his will. The eldest, García IV (1035–54), inherited an enlarged Navarre. Castile was made a kingdom and given to his second son, Ferdinand; Sobrarbe and Ribagorza were united in a separate kingdom and given to another son, while the small county of Aragon was also erected into a kingdom for his fourth son, Ramiro I (1035–63). Ramiro soon incorporated Sobrarbe and Ribagorza into his new kingdom (1045), and Aragon thus began its independent history.

Sancho's will provided for the return of León to Bermudo III on his death, but this restoration was overthrown by Ferdinand I of Castile (1035–65), who speedily annexed León (1037) and reoccupied (1054) the Castilian territories given in his father's will to his eldest brother, García IV of Navarre. Ferdinand also took the imperial title and by virtue of this claimed suzerainty over his brothers. He achieved great successes against the Moors, pushing his frontiers to the Tagus in the south and the Mondego in the west and making the *taifa* (principality) kings of Badajoz, Seville, and Toledo his tributaries. On his death Ferdinand, as his father had done, divided his dominions among his sons, and there was a period of violent strife before the second of them, the brilliant Alfonso VI (1065–1109), succeeded to the whole inheritance. His reign brought great changes to Spain.

The extensive Moorish kingdom of Toledo was conquered in 1085; it had been in close political relations with Alfonso for some time before this. The conquest, involving the absorption not only of the large and culturally advanced Mozarab and Jewish communities of Toledo but also of great numbers of often highly skilled Muslim artisans and agriculturalists, considerably altered the ethnic and cultural structure of the Castilian kingdom. There was no attempt to expel the newly conquered Moorish populations, who were promised royal protection and the preservation of their religion and customs. To give formal expression to his tolerant policy, Alfonso changed the imperial title by styling himself "emperor of the two religions." It seemed for a brief time as if these changes might produce a genuine Hispano-Moorish civilization under Christian political control. Unfortunately the Almoravids' invasion of Andalusia (1086) soon removed the *taifa* kingdoms from Alfonso's control and replaced the tolerant spirit of the Spanish Moors by one of bigoted African fanaticism. Meanwhile, under the influence of his wife Constance of Burgundy, the emperor allowed French Cluniac monks, to whom the Spanish tradition of *convivencia* was wholly repugnant, to get complete control of the Spanish church. They forced Alfonso to modify his promises

of toleration toward the conquered mudéjares, compelled him to replace the old Mozarabic missal by the Roman, and contrived to have the ancient Visigothic script of Spain replaced by the Carolingian hand in use in the rest of western Europe. These innovations brought into Christian relations with the Moors something of the crusading zeal which the French monks had found deplorably lacking and ensured that medieval Spain would eventually belong much more definitely to the common political and cultural pattern of medieval Europe, though mudéjar influences long affected Spanish life and culture.

Alfonso was followed by his daughter Urraca (1109–26), widow of Count Raymond of Burgundy; her turbulent reign cannot be followed here. She was succeeded by her son, Alfonso VII (1126–57), who gave new meaning to the imperial title by extending his influence over the county of Barcelona and by forcing Ramiro II of Aragon and García V of Navarre to accept his suzerainty. The counts of Toulouse and Montpellier also accepted him as overlord and, at a ceremony at León (1135), attended by his new Christian and Moorish vassals, he assumed the title "emperor of all Spain." He also achieved resounding, if inconclusive, successes against the Moors, and the European prestige of the Spanish imperial court was recognized when Louis VII visited Alfonso at Toledo (1150). Nevertheless separatist tendencies in the peninsula secured another ominous victory when the count of Portugal, Alfonso Henriques, assumed the royal style (1140). On his death, Alfonso VII, too, divided his realm among his sons, leaving Castile to the elder, Sancho III, and León to Ferdinand II. This new partition proved the deathblow to any future revival of the imperial idea originally inherited from the Visigoths. Castile and León now remained separated for 73 years, during which the magnates began to display in a marked degree that anarchical partisanship and resistance to royal authority which were to plague Spanish life for the next three centuries. Some important successes were nevertheless achieved against the Moors during this time. In 1212 Alfonso VIII of Castile (1158–1214), Sancho's son, supported by the armies of Aragon, Navarre, and Portugal, routed the Almohad amir of Morocco, Mohammed al-Nasir, at Las Navas de Tolosa and so removed the last really serious Islamic threat to Christian hegemony in Spain. In an atmosphere of ever-mounting religious fanaticism the way was now open to the conquest of Andalusia.

The Rise of Aragon.—The Pyrenean kingdom of Aragon was brought into existence (1035) by a testamentary act of Sancho the Great of Navarre. During the rest of the 11th century it gradually encroached on Moorish territory north of the Ebro. Saragossa was captured (1118) by Alfonso I (1104–34) who, in spite of his entanglements in the internal affairs of Castile, was able to follow up this important success by occupying a large area southwest of the former Moorish capital. In 1137 Petronilla, the infant daughter of Ramiro II (accession 1134; death 1154), was betrothed to Ramón Berenguer IV, count of Barcelona, to whom Ramiro thereupon handed over the government of Aragon (1137–62). The son of the marriage, which was solemnized in 1150, Alfonso II (1162–96), formally initiated the personal union of both countries under the house of Barcelona which was to last until 1410. In 1179 an important step for the future of Aragon was taken when Alfonso made an agreement with Castile whereby the task of reconquering the Moorish kingdom of Valencia was reserved to the Aragonese crown. In exchange Aragon relinquished any claims to any other Moorish-held territory in the peninsula. During this reign, too, Aragon developed political interests north of the Pyrenees; Provence (1166) and Roussillon (1172) fell to Alfonso by inheritance, but he left Provence to his second son, Alfonso. This preoccupation with French affairs proved fatal to Alfonso's son Peter II (1196–1213) who, having added Montpellier (1204) to his dominions by marriage, became involved in the Albigensian War and was killed at the Battle of Muret (1213). His death put an end to major Aragonese political pretensions north of the Pyrenees. It is to be noted that, from 1137 to 1412, Catalonia was the dominant partner in the association of the two countries. Economically and culturally Catalonia was much in advance of isolationist and conservative Aragon, whose nobility often found itself more in sympathy with the politics of the Castilian magnates than with the bourgeois ideas of the nationalistic Catalans. Catalan was the language of the court of Aragon during the whole of this period.

BY COURTESY OF THE ARCHIVO DE LA CORONA DE ARAGON, BARCELONA

THE DONATION OF FEUDAL RIGHTS THROUGH MARRIAGE. ILLUMINATION FROM THE "LIBER FEUDORUM MAJOR," ARAGON, 1162–96

Early in the reign of Ramiro II, Navarre seceded (1134) from the Aragonese crown to which it had been united since 1076. Sancho VII of Navarre (1194–1234) tried to nominate James I of Aragon as his successor and to reunite the two kingdoms, but the Navarrese preferred Thibaut (Theobald) IV of Champagne as king (1234–53), and Navarre was thereafter ruled mostly by a succession of French princes. The little mountain kingdom nevertheless retained many of its distinctively Spanish institutions.

End of the Reconquest.—The last king of León, Alfonso IX (1188–1230), was succeeded, on his death, by his son, who was already king of Castile. Castile and León were thus finally united. The new sovereign, Ferdinand III the Saint (1217–52), at once embarked on a great series of campaigns to subdue Andalusia. These began with the capture of Córdoba (1236) and culminated in the surrender of Seville (1248). Influenced by the crusading zeal instilled into the Spanish church by the Cluniac and Cistercian orders, Ferdinand at first expelled the Moorish inhabitants of the Andalusian cities en masse but was later forced to modify his policy by the collapse of the Andalusian economy that inevitably ensued. He also assented, chiefly for financial reasons, to the establishment of the new Moorish kingdom of Granada under Castilian suzerainty. The Granadine Moors were forced to create, by hard work, a high degree of productivity in their new state in order to pay to Castile a large annual tribute, which probably became essential to the well-being of Castile's finances. This reviving tendency toward a renewal of the older tolerance toward the Moors was also reflected in the cultural field; Toledo had already become famous throughout Europe as a centre where Moorish philosophy and science were, under the stimulus of successive archbishops, made available in translation to European scholars. During the same period James I of Aragon (1213–76), Peter II's son, completed Aragon's part in the reconquest. After occupying the Balearics (1235) he captured Valencia (1238)—unlike Ferdinand III carefully preserving the agricultural economy of the Moors—and so established the final peninsular frontiers of Aragon. By these Castilian and Aragonese achievements the reconquest was, for all practical purposes, brought to an end. Both countries now gave themselves up to internal social and constitutional struggles.

Castile and León from 1252 to 1479.—The reign of the cultured Alfonso X the Wise (1252–84) was politically ill-fated. Alfonso involved himself in wars with Portugal, Aragon, and Navarre and, having got himself elected to the German throne (1257), devoted much effort and treasure to an unsuccessful attempt to make good his claims against papal opposition. At home, following the death (1275) of his eldest son Ferdinand, a constitutional struggle developed over the rival claims to the succession of Alfonso of La Cerda and of Sancho, respectively the grandson and the second son of the king. This ended in an attempt by a junta of prelates and nobles, supporters of Sancho, to depose Alfonso X (1282) and saw the beginning of an acute phase in the violent clash between the magnates and the crown which was to dominate Castilian politics for 200 years. Alfonso X's legal works, based on the renascent doctrines of Roman and canon law, laid the theoretical foundations for monarchical supremacy; but the king was too inept a politician to succeed in enforcing them. The troubled reign of Sancho IV (1284–95) was complicated by war with Aragon, which supported the succession of the house of

La Cerda; and Castile did not emerge from continuous civil strife until the end of the minority of Sancho's grandson Alfonso XI (1312–50). This Alfonso, aided by the armies of the autonomous municipalities (*concejos*), sternly repressed the magnates and succeeded in establishing a measure of absolutism which found expression in the legal and political reforms promulgated at Alcalá de Henares (1348). He also took up arms successfully against a threat from Morocco, defeating the invaders at Río Salado (1340).

The trend toward strong central government would probably have continued under his son Pedro the Cruel (1350–69) without major trouble, had not Alfonso left a number of powerful bastard sons to contest his legitimate heir's inheritance. The eldest, Henry of Trastámara, posing as the defender of the magnates against royal attempts to limit their ancient privileges, began a struggle which lasted about 19 years to seize his half-brother's throne. Meanwhile Aragonese designs on Murcia—Castile's only outlet to the Mediterranean—and Castilian ambitions in western Aragon plunged the two countries into a ferocious war (1356–66) which became even more embittered when Peter IV of Aragon espoused the Trastámaran cause in Castile. It was not, however, until France intervened actively in support of Peter IV and Henry of Trastámara that Pedro of Castile's victories in Aragon were checked. Both France and England believed that the support of the powerful Castilian fleet might prove decisive in settling the outcome of the Hundred Years' War. Pedro's alliance with England (1362) therefore led to the dispatch of the *compagnies*, the veteran French mercenaries, to Castile, financed by France and the pope (1365). Pedro fled to Gascony and secured the military intervention of Edward the Black Prince, who routed Henry's supporters at Nájera (1367). Pedro was restored but quarreled with his English ally, who then began negotiations with Aragon, Navarre, and Portugal for a quadripartite partition of Castile. In 1369 Henry of Trastámara, again with French help, defeated and killed the Castilian king at Montiel and finally gained full control of the kingdom (1371).

To counteract this French success, John of Gaunt, duke of Lancaster (*q.v.*), married Pedro's daughter and heiress Constance (1371) and thus became pretender to the Castilian throne. He was, however, unable to invade Castile during the reign of his rival Henry II (1369–79), who succeeded in consolidating his position against strong loyalist opposition. For several decades to come a close political and military alliance with France dominated Castilian foreign policy, and Castilian sea power menaced the English south coast. Especially notable was the attack of 1377, in which combined French and Castilian fleets, led by Admiral Fernán Sánchez de Tovar, ravaged the ports from Plymouth to Folkestone, and that of 1380 when they burned Gravesend and threatened London. In spite of the conservative and class-conscious ideas of the Trastámaran monarchy, the constant wars in which they were involved compelled Henry and his son John I (1379–90) repeatedly to seek financial aid from the representatives of the municipalities in the Castilian *Cortes*, who thus secured considerable control over the government. John brought military and financial ruin to his kingdom, partly because of new Lancastrian attempts to conquer Castile and partly because of his own ambitions to add Portugal to his crown. His crushing defeat by the Portuguese at Aljubarrota (1385) was followed by the landing of John of Gaunt in Galicia with a large army (1386). French reinforcements and John of Gaunt's military ineptitude, however, saved the Trastámaran dynasty and the dynastic dispute was finally settled (1388) by the marriage of Catherine of Lancaster, Pedro's granddaughter, to John I's heir, Henry III (1390–1406). Violent outbreaks against the Jews marked Henry's accession, leading to mass forced conversions destined to have a tremendous impact on Castilian culture and society. In Henry's short reign the crown reverted with success to the policies of Alfonso XI and Pedro and reduced the magnates to order by stern measures. The reign of John II (1406–54) saw a continuance of the political disintegration of Castile as a result of a renewal of the struggle of the magnates with the crown. Civil war was now frequently accompanied by armed Aragonese and Navarrese interventions. John, a politically inept lover of the arts, left the government in the hands of his able favourite, Álvaro de Luna. Luna's power was not broken until 1453, when the king, acceding to the demands of the nobles and of the queen, ordered his execution.

Under the impotent Henry IV (1454–74) the Trastámaran dynasty reached its lowest ebb. Civil strife continued unabated and was now accompanied by a collapse of the moral standards of the court. The king had married Joan (Juana) of Portugal (1455), but their daughter and heiress, also called Joan, was alleged to be the child of an adulterous relationship of the queen's. Opponents of the court accordingly favoured the adoption of the king's sister Isabella (Isabel) as heiress to the throne. Joan's marriage to Alfonso V of Portugal and his active support of her claims led, on Henry's death, to five years of civil war before Isabella triumphed and was acknowledged to be queen of Castile (Isabella I, 1474–1504) by the peace of Trujillo (1479). Ten years before she had married Ferdinand, son and heir of John II of Aragon. Thus, on John II's death (also in 1479), the crowns of Castile and Aragon became associated though not yet formally united. The 15th century had so far been an age of political, economic, and moral decay in Castile where the economy remained pastoral and where agriculture and industry had failed to develop, causing credit to be obtainable only from Jewish and *converso* ("New Christian") moneylenders. There was nothing to hint, in 1474, that the central kingdom would, within a few years, become the heart of a world state.

Aragon and Catalonia, 1276–1479.—The conquest of Valencia having ended Aragonese participation in the reconquest, it became necessary for the eastern kingdom to expand outside the peninsula if it was to avoid becoming subordinate to the hegemony of Castile. Under Peter III (1276–85) Aragon therefore began political and military intervention in Italy. As husband of Constance, daughter of the emperor Frederick II's bastard son Manfred (*see* SICILY), Peter accepted the throne of Sicily and promised to free his new subjects from the rule of Charles of Anjou. In pursuit of this aim Aragonese troops and Catalan ships secured substantial successes against French military power and papal diplomacy and made Aragon a major factor in Italian politics (*see* SICILIAN VESPERS). Papal pressure on Peter III's sons Alfonso III (1285–91) and James II (1291–1327) resulted in Aragonese undertakings to abandon Sicily; but these were not accepted by the Sicilians, and James's brother Frederick (Fadrique) was eventually recognized as king of the island of Sicily for life (*see* FREDERICK III). James also secured the grant of Sardinia to Aragon as a papal fief (1297), though the island was not completely subjugated until the 15th century. Aragonese power in the Mediterranean, which went hand-in-hand with Catalan industrial growth, was further strengthened by the activities of the Catalan Company in Asia Minor and Greece, which eventually brought about the establishment of the Catalan duchy of Athens, first under Sicilian and then under Aragonese suzerainty (*see* ALMOGÁVARES). The Aragonese dynasty in Sicily remained closely associated with the Aragonese crown, and both countries were united when Martin I of Aragon (1395–1410) succeeded his son Martin I of Sicily in 1409. The reign of Alfonso V (1416–58) saw a further dramatic extension of Aragonese power in Italy when, after a long struggle, Alfonso finally became the undisputed ruler of the kingdom of Naples (1443) and transferred his capital there.

These events reacted unfavourably on the political situation in the parent kingdom, where Aragonese and Catalan particularism grew apace and reached its climax under John II (1458–79), when Catalonia was in revolt against the crown for 11 years. The discontent of the Catalans dated back to the compromise of Caspe (1412), which had placed Alfonso V's father, the Trastámaran prince Ferdinand I (1412–16), on the Aragonese throne when the male line of the house of Barcelona had become extinct. Since 1164, when Alfonso II, already count of Barcelona, succeeded his mother as ruler of Aragon, Catalonia had been the dominant partner in the Aragonese union. Now the Catalans feared, with some reason, that the Castilian dynasty would favour Aragonese interests at their expense.

The preoccupation of successive kings with Italian affairs weakened the crown at home. In 1283 the nobles and some mu-

nicipalities of Aragon formed an *unión* and extorted from Peter III a General Privilege granting them special privileges and immunities. These were increased substantially by Alfonso III in 1287 and remained in force throughout the reigns of James II and Alfonso IV (1327–36). But in 1348 Peter IV (1336–87), who had strong absolutist leanings, routed the army of the *unión* at Epila and put an end to an intolerable constitutional situation. Peter's reign represents something of a break with the general policy of Aragon during this period. Having annexed the possessions of the Majorcan crown, *i.e.*, the Balearic Islands and Roussillon (1343–44), he challenged the peninsular hegemony of Castile (1356); but, in the ten years' war which followed, a large part of his kingdom was overrun and occupied by the armies of King Pedro of Castile. He saved himself by joining in the French effort to dethrone his Castilian rival and replace him by Henry of Trastámara. Henry promised to cede the whole of eastern Castile to Aragon in exchange for Aragonese support but declined to fulfil his promise when he had secured the Castilian throne. Long years of negotiation for an Anglo-Aragonese alliance against Castile followed, during which Peter IV sought to play off French and English interests in the peninsula for his own ends without finally committing himself to either side. In the last years of his reign the common problems with which the schism in the papacy faced the peninsular kingdoms led him to try to form a peninsular bloc that would be neutral in both religious and military affairs; but French pressure prevented the success of this far-sighted plan. Peter's son John I (1387–95), elder brother and predecessor of Martin I, left the conduct of diplomacy largely to his French wife Yolande de Bar, and Aragonese foreign policy became largely subservient to that of France. The accession of a Castilian prince to the Aragonese throne as Ferdinand I in 1412 had already foreshadowed the union of the two largest peninsular states, and this was made certain in 1469, during the reign of Ferdinand I's second son John II, when Ferdinand, the heir to the throne (born 1452), married the Castilian heiress Isabella (born 1451). The same year which saw Isabella undisputed ruler of Castile (1479) brought her husband to the Aragonese throne as Ferdinand II. The destinies of the two countries were thereafter bound together.

Medieval Institutions: Asturias, León, Castile.—The Asturo-Leonese kingdom inherited from the Visigoths the principle of elective monarchy; but in practice its sovereigns were always chosen from the same family. By the 11th century primogeniture was accepted, though the elective tradition was never formally abandoned. In accordance with their declared policy, the first Asturian kings reestablished at Oviedo the pattern of civil administration which had existed at the Visigothic capital, Toledo. Outlying regions were governed by *comites* in the king's name, and the officers of the royal household were the direct successors of the Visigothic *palatini*. The steward (*maiordomus*) and the standard-bearer (*armiger regis* or *dapifer*) eventually emerged as the chief civil and military officers of the court. By the 13th century the governorship of the larger provinces was usually held by an *adelantado*, who exercised regional military and judicial authority on behalf of the king. In some areas similar duties were carried out by the *merino mayor*. The *merino menor*, who was appointed by either of these two officers, acted as their deputy in matters of criminal jurisdiction reserved to the crown (*voz de rey*). The traditional military organization of the country was overhauled by John I of Castile in 1383 when, following French practice, a constable aided by two marshals took over the duties previously performed in the king's name by his *alférez*. The office of admiral of Castile was created by Ferdinand III. The *cancellarius* rarely appears before the 12th century. By the time of Alfonso X, however, the chancellor was recognized to be the chief civil officer of the household. The *cancillería* (chancery), like the court itself, remained ambulatory until the end of the medieval period. The employment by the king of a privy seal (*sello de la poridad*), independent of the chancery and held in the custody of its own *canciller*, was well established by the end of the 13th century. As elsewhere in Europe the privy seal was evidently used by Alfonso XI and King Pedro to further the direct control of the king over the government since, in the political reaction brought about by the accession of Henry of Trastámara, attempts were made to restrict its use. The treasury, whose officials were largely Jewish, was headed by the *almojarife mayor*, later known as the *tesorero mayor*. King Pedro created several *contadores reales* to supervise the work of local tax collectors.

In theory feudalism did not exist in Castile and León. Vassalage was not tenurial, and lands were not held for military service. The *ricoshombres*, the highest class of nobles in the medieval Iberian kingdoms, held their *honores* (estates), at least theoretically, on life tenure only. In practice, however, feudal grants did occur and from the 10th century onward the magnates attempted to usurp feudal rights for themselves. The *ricoshombres* also possessed *señoríos* (seigneurial lands), often reconquered Moorish territory given to them, with its populations, as a reward for their services. The term *ricohombre* (literally "rich man," but actually reflecting an idea of power rather than of riches) first appears in the *fueros* (laws) of Miranda de Arga, in Navarre (1162), and of Santarém in Portugal (1179). The *Siete Partidas*, the great corpus of law compiled under Alfonso X of Castile and León, describes the *ricoshombres* as the equivalent of counts and barons in other lands. Medieval Spanish writers regard as the distinctive feature of this class, which was noble by lineage, its members' right to enlist men under their private banners and to retain nobles of the second class as vassals. The *ricoshombres* regarded themselves as the natural advisers of the king and were, until the 14th century, ex officio members of the royal council. From their ranks were chosen the governors of the provinces and the military commanders. The status of the Castilian *ricoshombres* was, however, not unassailable. The king could—and sometimes did—deprive them of all their possessions. The right to coin was not granted to them; royal officers could, in theory, enter their territories to deal with certain classes of crime; a limited right of appeal from their courts to the royal courts existed—at least in name. This underlying dependence of the magnates on the crown's goodwill was largely responsible for the continual struggle of the Castilian and Leonese *ricoshombres* to keep the crown weak and to resist all centralizing and absolutist tendencies. The lesser nobles (*infanzones* and *hijos d'algo*) could be vassals of the crown or of a lord. The free peasants paid tribute for permission to till the soil either to the king on royal lands (*realengos*) or to a lord on seigneurial lands. On the *behetrías* the peasants were semifree, having the theoretical right to change their lord at will. This institution was chiefly found in those areas of Castile and León proper which had been repopulated by Mozarabic immigration during the reconquest. The serfs (*solariegos*) were chiefly found on the lands of the magnates, of the military orders, and of the church (*abadengos*). The status of the mudéjares (Muslims living under Christian protection) varied according to political conditions and to locality. The Jewish communities in the towns (*aljamas*) enjoyed royal protection until the end of the 14th century, when an outburst of popular fanaticism led to great massacres of Jews in Seville, Córdoba, Toledo, and other cities (1391) and Jewish communal life was irreparably damaged.

The most notable institution of medieval Castile was the semiautonomous municipality (*concejo*) comprising a town (*villa*) and its surrounding *comarca*. The rise of the *concejos* dates from the 11th century, and they were largely brought into being by the need to attract settlers, by special privileges, to newly reconquered areas. This was achieved by the royal grant of a franchise (*fuero*) giving the *concejo* in perpetuity a large measure of control over its own affairs and its own legal code. The *concejos* raised their own militia and formed military associations with each other (*hermandades*) in defense of their interests. They were invaluable allies of the crown in its struggles with the nobles, particularly during the 14th century, when their deputies (*procuradores*) to the Castilian parliament (*Cortes*) dominated that body. The appearance of the third estate in the *Cortes* dates from the end of the 12th century.

In the later Middle Ages the king depended on the vote of the *procuradores* for the grant of financial aid (*servicio*) necessary to cover civil and military expenditure. The heavy costs of the wars of Henry II and John I increased the dependence of the king

on the third estate. John I was forced to meet the severe criticisms of his policy voiced by the *procuradores* by creating the royal council (1386). At the beginning four representatives of each estate were appointed to the new body, and the king delegated to it (theoretically at least) many of his powers. In the 15th century the *concejos* themselves fell victims to the growth of the royal power and had to accept royal officials known as *corregidores*. A hint of impending decay had already appeared during the reign of Henry III (1390–1406) when, against the wishes of the *Cortes*, the king succeeded in making permanent the *alcabala*, a highly unpopular sales tax of one-tenth, which then became one of the main sources of royal revenue.

Medieval Institutions: Aragon, Catalonia, Valencia.—In spite of their union under the Aragonese crown, these three regions retained their own separate institutions and parliaments. There was a strong tendency toward feudalism in Aragon from the beginning, and grants of fiefs for military service were frequent. After the union with Catalonia, where feudalism was fully developed, the Aragonese nobility intensified its efforts to obtain full feudal rights. The smaller municipalities (*universidades*) were, until the 15th century, less important than in Castile because of the dominant position of Saragossa, Barcelona, and Valencia respectively in the parliamentary representation of the popular arm. The Aragonese nobility consisted of four classes, *ricoshombres de natura*, *ricoshombres de la mesnada* (created by James I), *caballeros*, and *infanzones*. These, like the representatives of the *universidades*, attended parliamentary sessions by right.

The situation of the *ricoshombres* in Aragon was much more secure than that of those in Castile. They formed a separate estate in the Aragonese *Cortes* and had absolute power over their serfs and vassals. In Aragon the *ricoshombres de natura* claimed to be descended from the original reconquerors of the kingdom. The *ricoshombres de la mesnada* were particularly attached to the royal household. In Catalonia there were three classes of magnates (*comtes*, *vescomtes*, and *valvassors*). The Aragonese *Cortes*, which consisted of four estates (*brazos*) because of the separate representation of the greater and lesser nobles, was governed by complicated rules of procedure. Redress preceded supply, and unanimity was in theory required for the approval of any measure. The *Cortes* also claimed the right to declare war. The most important political office in the country was that of the *justicia mayor*. He was arbitrator between king and nobles or people when disputes arose with the crown, and his decisions were virtually final. A permanent parliamentary commission (*diputación del reyno*) existed to deal with finance, observance of the *fueros*, and matters pertaining to the peace of the realm. The organization of the Catalan and Valencian *corts* was somewhat similar, though the former asserted a right to control all legislation. A *diputació general* in Catalonia performed functions like those of the Aragonese permanent parliamentary commission. In spite of the many restrictions on their powers it has been suggested that the Aragonese kings had greater *de facto* authority than the kings of Castile. The appointment of the *justicia* was made by the crown, and only the crown could initiate legislation. Evidence of the growth of the royal bureaucracy was the creation, in the 14th century, of the office of *maestre racional* to supervise all aspects of public finance throughout the Aragonese crown's territories.

A peculiar feature of Aragonese administration was the absence of intermediate authorities between the central government and purely local officials. In Catalonia, land belonging to the king was ruled on his behalf by the *veguers*, with subordinates known as *sots-veguers* or *batlles*; both were crown appointments. The importance of Barcelona gave that city a dominant place in the life of Catalonia. It was ruled by the *consell de cent* though, in practice, this body's authority was normally delegated to a small *consell* of five persons with whom were associated the *veguer* and the *batlle*. Catalan law was based on the *Usatges* (*c.* 1060), the oldest written feudal code. The kingdom of Valencia was governed by the constitution (*Furs*) granted to it by James I in 1239. There were wide variations in the conditions of the unprivileged classes under the Aragonese crown. The legal and economic position of the Catalan serfs (*pagesos de remença*) was worse than that of serfs anywhere else in the peninsula. On the other hand the mudéjares were comparatively well treated in Aragon proper where, until the 15th century, crown and nobles appreciated their economic importance and allowed them freedom of worship and their own tribunals. The Jews, vassals of the king, were favourably treated until the 15th century when, as in Castile, their position rapidly deteriorated. (P. E. R.)

G. The Catholic Kings

The Union of Aragon and Castile.—When Ferdinand II (*q.v.*; 1479–1516, also Ferdinand V of Castile from 1474) succeeded to the crown of Aragon in 1479, the union of Aragon and Castile was finally achieved. Ferdinand and Isabella ruled jointly in both kingdoms and were known as the *Reyes Católicos* ("Catholic Kings"). It was, however, a union of crowns and not of kingdoms. In size, institutions, traditions, and, partly, even in language, the two kingdoms differed greatly. Within the kingdom of Aragon (*q.v.*), Aragon and Valencia each had about 270,000 inhabitants of whom some 20% and more than 30%, respectively, were Moors and Moriscos (*q.v.*; Moors officially converted to Christianity). Catalonia had about 300,000 inhabitants. In each part the powers of the crown were severely limited. The barons ruled their estates like kings, dispensing arbitrary justice over their peasants. In Catalonia they had the right to wage private war. In Aragon, anyone arrested by order of the king could put himself under the jurisdiction of the *justicia*, who held his office for life and was therefore independent of the king's pleasure. It was this highest judge of the kingdom who crowned the kneeling king with the formula, "We who are as good as you swear to you who are no better than we, to accept you as our king and sovereign lord, provided you accept all our liberties and laws; but if not, not." Barcelona, in the later Middle Ages the capital city of a flourishing commercial and territorial empire in the Mediterranean, had declined owing to Italian competition and the ravages of ten years of civil war (1462–72); however, it did retain its privileges and autonomy.

Castile, too, was a poor country. Much of its soil was arid and its agriculture was backward. The armed shepherds of the powerful sheep-owners' guild, the Mesta, drove their flocks over

BY COURTESY OF THE MUSEO DEL PRADO

"THE VIRGIN OF THE CATHOLIC KINGS" BY AN ANONYMOUS HISPANO-FLEMISH MASTER

Ferdinand II, kneeling at left, and Isabella, kneeling at right, were given the title "Catholic Kings" by Pope Innocent VIII

hundreds of miles, from summer to winter pastures, spoiling much cultivated land. Despite the violent hostility of the landowners, the government upheld the Mesta's privileges, since the Mesta paid generously for them, and was supported by the merchants who exported the raw wool to the cloth industry of Flanders. The power of the Mesta, and the harm it did, increased in the 16th century. The impoverished Spanish peasant was unable to buy the manufactures of his urban industries, and thus the Castilian towns remained small and their industries underdeveloped, compared with those of Italy and the Netherlands. But Castile was a large country, with a population of more than 4,500,000. Despite its poverty it was much more powerful and had, potentially, much greater resources than Aragon. The Catholic Kings determined to restore the power of the crown in Castile. Once this was achieved, the liberties of the smaller kingdoms would become a relatively minor problem. (For the details of this achievement *see* ISABELLA I.)

There was also a remarkable development in the administrative machinery of Castile. The old royal council, a council of great nobles advising the king, was transformed into a bureaucratic body for the execution of royal policy, staffed by a prelate, three nobles, and eight or nine lawyers. These lawyers, mostly drawn from the poor hidalgo class (lower nobility), were entirely dependent on the royal will and became willing instruments of a more efficient and powerful central government. The Catholic Kings also set up a Council of Finance (1480, but not fully developed until much later), the Council of the Hermandad (*q.v.;* 1476–98), the Council of the Inquisition (1483), and the Council of the Orders of Knighthood (for the administration of the vast property and patronage of the Orders of Santiago, Calatrava, and Alcántara), and they reorganized the Council of Aragon. Charles V and Philip II were later to continue this work and to add further Councils, notably those of the Indies (1524) and of Italy (1558).

The Spanish Inquisition.—With its large Moorish and Jewish populations, Spain was the only multiracial and multireligious country in Western Europe and much of the individual development of Spanish civilization, in religion, literature, art, and architecture, during the later Middle Ages, was due to this fact. The Jews had served Spain and its monarchs well, providing an active commercial class and an educated elite for many administrative posts. But, inevitably, their wealth had created jealousy, and their heterodoxy hatred, in a population which traditionally saw itself as the defender of Christianity against the infidel. The Catholic Kings, ever good tacticians, profited from this feeling. In 1478 they first obtained a papal bull from Sixtus IV setting up the Inquisition (*q.v.*) to deal with the alleged evil influence of the Jews and *conversos* (converted Jews). Since the Spanish Inquisition was constituted as a royal court, all appointments were made by the crown. Too late, Sixtus IV realized the enormous ecclesiastical powers which he had given away and the moral dangers inherent in an institution whose proceedings were secret and which did not allow appeals to Rome.

With its army of lay familiars, who were exempt from normal jurisdiction and who acted both as bodyguards and as informers to the inquisitors, and with its combination of civil and ecclesiastical powers, the Spanish Inquisition became a formidable weapon in the armoury of royal absolutism. Yet the number of those whom it actually tried and condemned secretly has often been greatly exaggerated. Many good Catholics in Spain opposed its introduction, and the Neapolitans and the Milanese (who prided themselves on their Catholicism and who were supported by the popes) later successfully resisted the attempts by their Spanish rulers to impose the Spanish Inquisition on them (*see* below). Even in Spain itself, only its sumptuous autos-da-fé (*see* AUTO-DA-FÉ) seem to have been popular, not the institution itself.

The first inquisitor general, Tomás de Torquemada (*q.v.*), himself from a *converso* family, at once started a propaganda campaign against the Jews. In 1492, he persuaded the Catholic Kings to expel all Jews who refused to be baptized. Isabella and most of her contemporaries looked upon this expulsion of about 170,000 of her subjects as a pious duty. At the moment when Spain needed all her economic resources to sustain her new European position and her overseas empire (*see* below), Ferdinand and Isabella deprived Spain of its economically most active citizens and laid it open to exploitation by German and Italian financiers.

Conquest of Granada.—The Moorish question was of a different order, for the Moors still ruled their independent kingdom of Granada. The Catholic Kings had to concentrate all their military resources and call on the enthusiastic support of the Castilians to conquer the kingdom in a long and arduous campaign, which ended with the capture of Granada, the capital, in 1492. In this campaign Gonzalo de Córdoba (*q.v.*), the "Great Captain," developed the tactics, training, and organization which made Spanish infantry almost unbeatable for 150 years. The Moors were granted generous terms and religious freedom. But, against the advice of the saintly Hernando de Talavera, archbishop of Granada, who was trying to convert the Moors by precept and education, the queen's confessor, Francisco (later Cardinal) Jiménez de Cisneros (*q.v.*), introduced forced mass conversions. The Moors rebelled (1499–1500) and, after their renewed defeat, they were given the choice of conversion or expulsion (1502). Though many chose conversion, the problem now became virtually insoluble. There were never enough Arab-speaking priests, nor money for education, to make outward conversion a religious reality. The Moriscos remained an alien community, suspicious of, and suspect to, the Old Christians.

Moreover, almost for the first time in history, the two shores of the western straits of the Mediterranean were now under separate and hostile political control. The African Moors raided the Spanish coasts and often received help from the Moriscos. Isabella and Jiménez drew the logical conclusion that North Africa must be conquered. After the queen's death (1504), and largely thanks to the energy of Jiménez, the Spaniards captured Oran and some other ports (1505–10). Then Ferdinand called off the campaign.

Acquisition of Naples.—Ferdinand's foreign policy was dominated by the traditional Aragonese rivalry with France along the Pyrenees and in Italy. Aragon still held Sicily and Sardinia from the much more extensive, medieval Aragonese empire. French intervention in Italy, from 1494, gave Ferdinand his chance. Through his own Machiavellian diplomacy and through the generalship of Gonzalo de Córdoba, he acquired the kingdom of Naples (1503). For the first time the union of Aragon and Castile had shown its power, and Spain now rivaled France as the most powerful state in Europe. Ferdinand had carefully arranged the marriages of his children to strengthen his diplomatic position against France. The unexpected deaths of the two eldest and their children, however, left the succession of Castile after Isabella's death to the third, Joan ("the Mad"), and her husband, Philip the Handsome of Habsburg (*qq.v.*), ruler of the Burgundian Netherlands. As Philip I of Castile, he was supported by a large section of the nobility and forced Ferdinand to recognize his claims. But he died in 1506 and Ferdinand was left as sole ruler. Ferdinand's last great success was the annexation of the Spanish part of the small Pyrenees kingdom of Navarre in 1512.

H. THE HOUSE OF AUSTRIA

Charles I.—Ferdinand died on Jan. 23, 1516, and the crowns of the Spanish kingdoms now devolved on his grandson, Charles I (1516–56; *see* CHARLES V, Holy Roman Emperor), the ruler of the Netherlands and heir to the Austrian Habsburg dominions. This new union had not been planned in Spain and, at first, it was deeply resented. Cardinal Jiménez, the regent until Charles's arrival in Spain, had to battle against the old antagonisms between nobles and towns which were flaring up again when the magnates tried to regain their old power. Although the court at Brussels had been very careful to hold its hand, the Spaniards accused the Netherlanders of greed and place-hunting. When Charles arrived in Spain in September 1517, his supporters were already disillusioned and the country was apprehensive of the rule of a foreigner. Charles himself, young, ugly, and inexperienced, speak-

ing no Spanish and surrounded by Burgundian councillors and courtiers, did not initially make a good impression. The *Cortes* of Castile, Aragon, and Catalonia granted his financial demands but attached to them much pointed advice and criticism.

The Comunero Movement.—On June 28, 1519, Charles was elected Holy Roman Emperor as Charles V and prepared to go to Germany. His chancellor, Mercurino Gattinara, summoned the Castilian *Cortes* to Santiago de Compostela (April 1520) to demand more money, even though the former grant had not yet expired. The towns immediately made difficulties. The Toledans refused to appear; the others demanded the discussion of grievances before supply. By a mixture of bribery and concessions, the government finally induced a majority of the delegates (who had transferred from Santiago to Corunna) to vote the new grant. As Charles set sail (May 20, 1520) the Castilian revolution had already begun.

The towns, led by Toledo, formed a league and set up a revolutionary government. This *Comunero* movement spread rapidly through Castile, and the nobles did nothing to check it. They had not forgiven Charles his foreign councillors, his bestowal of the archbishopric of Toledo on a young Burgundian, Guillaume de Croy, and the appointment of Adrian of Utrecht (later Pope Adrian VI) as regent. Only when the more radical and popular elements in the towns were gaining control of the *Comunero* movement and were spreading it to the nobles' estates did the nobles combine and defeat the *Comunero* army at Villalar (April 23, 1521).

The power of the monarchy was thus restored in Castile, never to be seriously shaken again under the Habsburg kings. The towns kept much of their autonomy; but royal control in the towns was safeguarded by the appointment of royal officials, the *corregidors.* The *Cortes* continued to function, but they now willingly voted the taxes the government asked for, and the king granted or ignored their petitions as he saw fit. The nobles had won the civil war, but they could no longer break their alliance with the monarchy. The monarchy, now that it could manage the *Cortes* and afford a standing army, was the stronger partner. But when, in 1538, Charles V proposed a tax from which the nobles should not be exempt, there were immediate rumblings of revolt. Charles had to give way; but he never summoned the nobility again to the meetings of the *Cortes.* The monarchy had thus won its political victory in Castile only at the cost of letting the nobility contract out of the financial obligations of the state and the empire. The rising burden of taxes fell, therefore, on those least able to bear them, and on the only classes whose activities and investments could have developed the Castilian economy.

The Nobility.—The traditions of the grandees and hidalgos, formed in the centuries of struggle against the Moors, made them even more averse to economic activities than the rest of the European nobility. They invested their money in land, without however improving agriculture, and preferred careers in the army, the church, and the civil service to the ignoble occupations of commerce. In the long run, the economic weakness of Spain, aggravated by its social traditions and its system of taxation, proved a serious handicap in Spain's struggle with her western European rivals. After Villalar, however, the Spanish nobility had come to accept Charles V. His championship of Catholic Christianity against the Muslim Turks and the German heretics appealed to their own traditions of Christian warfare against the Moors. While Charles kept the grandees out of the central government of Spain itself, he had many prizes to offer, in military commands, provincial governorships, and even viceroyalties in Italy and Spanish America. The hidalgos, trained as lawyers at Salamanca, or as theologians at Alcalá de Henares, could look forward to dazzling careers in the king's councils and in the Spanish church. The Spanish upper classes were beginning to accept and enjoy their monarch's position as the greatest ruler in Europe.

Charles V's Foreign Policy.—Spain as a result, however, became involved in interminable wars. The necessity of defending southern Italy from the Turks brought Charles V's empire into collision with the Ottoman Empire and turned into a struggle for the control of the central Mediterranean. Ferdinand's failure to complete the conquest of North Africa now brought a bitter revenge. The corsair leader Khair ed-Din, known as Barbarossa (*q.v.*), had made himself master of Algiers (1529) and acknowledged the suzerainty of the sultan of Istanbul. Thus the purely local problem of the Moorish raids on the Spanish south coast became merged into the much more formidable struggle with the Ottoman Empire. In 1535 Charles captured Tunis; but in 1541 he failed against Algiers. At the end of his reign, the balance of the two great naval powers in the Mediterranean, the Turkish and the Spanish, was still even.

The rival Aragonese and Angevin claims to Naples also brought conflict with France, with whom Charles fought four wars. His armies conquered Milan and reduced most of the still independent Italian states to Spanish satellites (*see* NAPLES, KINGDOM OF; ITALY: *History*). As an increasing part of the burden of these wars fell on Spain, and especially on Castile, so Charles V's empire in Europe became gradually transformed into a Spanish and Castilian empire. In the last half of his reign, Spaniards and Hispano-Italians monopolized all high positions in the empire south of the Alps and began to appear in Germany and the Netherlands.

Spain in the New World.—In 1492 Columbus discovered the West Indies and, in the next half century, the Spaniards conquered huge empires in America (*see* LATIN AMERICA: *Colonial Period*) and made their first settlements in the Far East. From the beginning there were disputes with the Portuguese, who were pioneering the sea route to the East Indies around Africa and were conquering their own colonial empire. The Catholic Kings obtained a series of papal bulls (1493) from the Spanish pope, Alexander VI, and as a result concluded the Treaty of Tordesillas (*q.v.*) with Portugal (1494) to settle their respective claims. Everything west

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

THE CORONATION OF THE EMPEROR CHARLES V

Charles V was crowned Holy Roman emperor by Pope Clement VII at Bologna in 1530. A small portion of the coronation procession is shown above in part of an engraving by N. Hogenberg, Antwerp, 1632; the pope and the emperor are seen beneath the canopy

of an imaginary line 370 leagues to the west of the Cape Verde Islands in the Atlantic was assigned to Spain; everything east to Portugal. The rest of Europe saw no reason to accept the pope's decision, and the result was constant and brutal warfare in the overseas colonies, even when the European governments were officially at peace.

Colonial Policy.—Unlike the other European colonists of that age, the Spaniards were vitally concerned with the moral problems of the conquest, conversion, and government of heathen peoples. If the great majority of the *conquistadores* ruthlessly pursued gold and power, they took with them the Dominican friars, who set themselves to convert and educate the American Indians. The Dominican Bartolomé de las Casas (*q.v.*) fought long battles to modify at least the greatest evils of colonial exploitation. His debates with the theologian Juan Ginés de Sepúlveda and the writings of Francisco de Vitoria (*q.v.*) provide the first systematic discussions of the moral and legal problems of conquest and colonial rule. Their importance lay in their effects on Spanish colonial legislation. The *Leyes Nuevas* (New Laws of the Indies) of 1542 were based largely on the arguments of Las Casas. By about the middle of Charles V's reign, the crown had managed to transform the robber empires of the *conquistadores* into relatively settled colonies whose administration was controlled from Spain.

Trade and Commerce.—The crown insisted that all trade with the colonies should be carried on through Seville and that it should be reserved for Castilians. It was closely regulated by the Casa de Contratación (1503), the house of trade, in Seville. That city itself rapidly became one of the greatest trading centres in Europe, and its population rose, from 25,000 in 1517, to 90,000 in 1594. Yet Castile was unable to supply all the manufactures which the colonists demanded and for which they paid in solid gold and silver. Castile had to import these goods itself and the Castilian monopoly of trade with Spanish America only had the effect of giving the whole of Europe the chance to compete on equal terms for the Spanish trade with the emperor's non-Castilian subjects. In both Mexico and Peru, silver mining became a major industry and rapidly increasing quantities of silver were shipped to Spain, partly as the crown's right to one-fifth (*quinto real*) but, more important, as payment for imports. The average annual quantities rose rapidly from about 1,000,000 pesos in the five-year period 1526–30, to 5,000,000 in 1541–45, and then to the peak of more than 35,000,000 in 1591–95. Prices, especially of agricultural produce, had started to rise in Spain, as in the rest of Europe, long before American silver was imported in considerable quantities. But whatever the ultimate causes of the price revolution of the 16th century, there can be little doubt that American silver greatly aggravated the inflation in Spain in the second half of the 16th century. Very little of this silver seems to have been invested in economic production. Most of it was used to pay for Spanish imports, for the Spanish armies abroad, and to satisfy the Spanish government's German and Italian creditors. Thus Spain, with all the treasure of the New World at its command, remained a poor country.

Philip II.—When Charles V abdicated his various lands (1555–56), Philip II (*q.v.*; 1556–98) succeeded to all his father's dominions except Germany. His empire in Europe, now without the imperial title, was still only a loose union of independent states recognizing the same head. Philip, a great traditionalist, was not the man to inspire his different subjects with a new unifying idea, though he improved the central administration of his empire by the creation of the Council of Italy (1558). But his own Castilian upbringing and preferences increased the tendency toward transforming the empire into a Castilian empire. Moreover the intellectual atmosphere in Spain itself had changed, from the freer and more liberal days of Charles V, when the country was open to the influence of Erasmus and of reforming ideas. All thought of reconciliation with the Protestants had now died. The influence of St. Teresa, and of St. Ignatius Loyola (*qq.v.*), the founder of the Society of Jesus, and, above all, of its new king, Philip II, made Spain the intellectual, as well as the financial and military, spearhead of the Counter-Reformation.

After another inconclusive war with France, Philip concluded the Peace of Câteau-Cambrésis (1559), which confirmed Spanish hegemony in Italy and left the frontiers of the Netherlands intact. He had inherited a debt of some 20,000,000 ducats and, in 1557, his government declared a moratorium on its debts. He still had a major naval war with the Turks on his hands and there were many problems in Spain itself awaiting solution. For the next 20 years Spain remained on the defensive in Western Europe.

The Moriscos.—The most immediate problem was that of the Moriscos of Granada. The attempt to christianize and assimilate them had proceeded much too slowly. The ineptitude and mutual wrangling of the different Spanish authorities in Granada, and the king's own rigid view of religion, drove the Moriscos into a rebellion which was put down only after a long and ferocious campaign (1568–70). At the end of it the Moriscos of Granada were deported, in small groups, to different parts of Castile and settled among the Old Christian population in a last attempt to achieve assimilation. In the absence of systematic education, and in the face of the hostility of the Christian population, this attempt was also doomed to failure.

Portugal and Aragon.—There remained the question of the complete unification of the Iberian Peninsula. In the case of Portugal, Philip's opportunity came when his nephew, King Sebastian of Portugal, in an ill-prepared crusade in Morocco lost his life and a great Portuguese army at the battle of Alcazarquivir (1578). During the short reign (1578–80) of Sebastian's old uncle, the Cardinal King Henry, Philip carefully prepared his ground in Portugal by intrigue and bribery. Nevertheless when Henry died, it was still necessary for Philip to make good his own claims by conquering Portugal in 1580. Philip respected the laws and privileges of his new subjects and left them to administer their own colonial empire. But the union increased, rather than diminished, the old hostility between the Castilians and the Portuguese. When an opportunity arose, in the 17th century, the Portuguese were quick to reassert their independence.

Philip II's last action in the peninsula was against Aragon. It was precipitated by a court intrigue which led to the flight (1590) of the king's secretary, Antonio Pérez (*q.v.*), to Aragon. Since Pérez was unlikely to be convicted in the *justicia's* court there, the king demanded his transfer to the court of the Inquisition. The populace of Saragossa rioted, freed Pérez, and killed the king's special representative (1591). To the Aragonese, it meant the defense of their liberties; to Philip it meant open rebellion. A Castilian army marched into Aragon (1591) and Philip made a number of constitutional changes. The *justicia* was, from now on, removable at royal pleasure; the viceroy could be a Castilian, and majority voting was substituted for the principle of unanimity in the Aragonese *Cortes*. These changes gave the crown the ultimate power of decision in Aragon but preserved the kingdom's autonomy.

Lepanto.—In the Mediterranean the Spanish fleet was inferior to that of the Turks and Philip had to remain on the defensive, even when the Turks were besieging Malta (1565). Their failure to capture the island from the Knights of St. John marked the end of their great offensive. Six years later, the combined Spanish, Venetian, and papal fleets—in alliance the numerical equals of the Turks—virtually annihilated the Turkish fleet at Lepanto (*q.v.*; 1571). The strategic effects of this great victory were negligible. Its moral effects were immense. It confirmed the Spaniards in their chosen role of champions of Christendom and explains much of their continued willingness to support their king's religious and imperial policies, even in the face of ruinous costs and mounting disasters. After Lepanto, however, it became clear that the stalemate in the Mediterranean could not be broken. In 1580 Spain signed a truce with the Porte (Turkish government).

The Armada.—From about 1580 the Spanish government became convinced that the rebellion (1568–1609) and heresy in the Netherlands (*see* NETHERLANDS, THE: *History*) could not be crushed as long as the rebels received help from England and France. Philip began to give financial aid to the League, the ultra-Catholic party in France. From 1586 he began to prepare the invasion of England. The Armada (*q.v.*) which set sail from Lisbon

BY COURTESY OF THE NATIONAL MARITIME MUSEUM, GREENWICH

A PAINTING (VENETIAN SCHOOL) OF THE BATTLE OF LEPANTO, 1571, IN WHICH THE SPANISH, VENETIAN, AND PAPAL FLEETS DEFEATED THE TURKS

in May 1588 numbered 130 ships and nearly 30,000 men, bravely commanded by the duque de Medina Sidonia. But it had been set an impossible task: that of convoying the army of Alessandro Farnese (*q.v.*), duke of Parma, from the Netherlands to England, in the face of a better armed English fleet and without control of a single deepwater Channel port. The defeat of the Armada was inevitable, but not dishonourable.

Spanish intervention in France from 1590 was equally doomed to failure. The duke of Parma, with his Spanish veterans, won great tactical victories. But Spain failed to prevent the succession of Henry of Navarre as Henry IV of France and the collapse of its ally, the League, when Henry became a Catholic (1593). Spain had gambled its own prosperity and the American treasure on establishing its hegemony in Western Europe, and with it the decisive victory of the Catholic Church over the Protestants, and had failed. Shortly before his death, Philip II concluded the Treaty of Vervins (1598) with France, which substantially reestablished the position of 1559. Yet, if Spain had failed in its highest ambitions, at the end of the 16th century it was still the greatest power in Europe. It had brought Christianity to millions and Protestantism had been contained, though not defeated. Spanish monks and mystics had given Catholicism a new content, and the theologians and jurists had created the basis of international law. Spanish literature and art were only now entering their greatest period. Morally and economically there were dark sides to the picture; but to the Spaniards the 16th and early 17th centuries have always been their golden century. (H. G. Ko.)

From the death of Philip II in 1598 to the accession of the house of Bourbon to the Spanish throne in 1700, that is to say during the reigns of Philip III, Philip IV, and Charles II, Spain experienced a period of serious decline, which is often referred to as the decadence of the 17th century. It was not until the last years of Charles II that a very gradual improvement took place.

Philip III.—The reign of Philip II's son, Philip III (*q.v.;* 1598–1621) saw the landed aristocracy established in power, and the rule of the *validos,* or royal favourites; first, to 1618, of the duque de Lerma (*q.v.*) and afterward of Lerma's son, the duque de Uceda. In foreign affairs, the monarchy followed a peaceful policy, ending the war with England by the Treaty of London (1604) and concluding the Twelve Years' Truce with the United Provinces (1609). It also sought to maintain its hegemony in Europe by relying on the influence of dynastic ties with other reigning families. After the assassination (1610) of Henry IV of France friendly relations were established with the French. However, schemes to aid the Balkan peoples under Turkish rule, and the problems arising in Italy, notably Spanish rivalry with Venice, which resulted in 1618 in the famous "conspiracy of Venice," an alleged Spanish plot to overthrow the Venetian Republic, compelled Spain to keep considerable forces in the Mediterranean. To meet the danger from North Africa, Philip formed an alliance with the kings of Cuco, in the Atlas Mountains; the resulting split in the Berber Kingdoms well suited Spanish policy.

This peaceful foreign policy gave an opportunity to the Spanish monarchy to try and strengthen the internal unity of the country by expelling the Moriscos (1609–14). About 300,000 were driven from Spain, well over half of whom were from the kingdoms of Valencia and of Aragon. The expulsion, however, seriously weakened the Spanish economy and brought financial ruin to the middle classes of Valencia and Aragon. They lost not only the interest but also the capital which they had advanced to the landowning Spanish aristocracy, for the latter had depended on the Moriscos for source of labour and for the dues which the Moriscos paid as vassals. On the other hand, internal unity was threatened by the beginnings of the breakdown of the union of the realm which the Catholic Kings had begun by binding the autonomous kingdoms together. In the lands of Aragon, especially Catalonia, the first indications could be seen of the regional discontent that was to lead to revolution in 1640.

Economic Depression.—Signs of an economic crisis, marking the change from the expansion of the 16th century to the depression of the 17th, are reflected in Cervantes' *Don Quixote,* the first part of which appeared in 1605. The first important monetary inflation occurred, prices rose slightly, and trade declined, despite the fact that in 1608 Spanish exports to the Indies reached a peak of 45,000 tons in a single year. By 1610 the depression was under way, a change experienced in Spain a quarter of a century before the rest of Europe, due to the fact that Spain alone received precious metals direct from Spanish America. In an attempt to solve the financial difficulties, Philip III resorted to the issue of copper currency (*vellón*). The first measure of this kind had already been taken in 1599, and was extended in 1602 and 1603. In 1617 the gap between income (5,375,000 ducats) and expenditure (8,235,000) forced the king to seek permission from the *Cortes* to issue enough copper coinage to make a profit of 600,000 ducats. The *Cortes* acceded to this request only on condition that no more *vellón* currency would be issued for at least 20 years. Nevertheless, the king resorted to a new currency debasement shortly before his death in 1621. The economic and financial crisis accentuated the social divisions within Spain, with, on the one hand, a privileged minority identifiable with the landowning aristocracy, and the mass of the poor on the other. This division explains most social ills of the time, notably begging, encouraged by the indiscriminate charity of the church; roguery (*la picardía*), which inspired much memorable literature; and banditry, particularly in the mountainous parts of Catalonia.

Philip IV.—Under Philip IV (*q.v.*; 1621–65), the eldest son of Philip III, the decline of Spain became more marked. Philip's reign in fact proved to be one of the most dramatic periods of Spanish history. Like his father, Philip relied upon favourites, the conde-duque de Olivares and Luis de Haro (*qq.v.*) and relinquished the exercise of power to the aristocracy, who governed by a system of executive councils. In 1621 Olivares renewed the war with the Dutch, the treaty of 1609 having just expired, and this move coincided with the extension of operations in the Thirty Years' War (*q.v.*), with which Spain had been involved since 1618, from Bohemia to the Palatinate and the Rhine.

Olivares' Foreign Policy.—From the very beginning Olivares planned a most ambitious policy designed to impose an Austro-

Spanish hegemony in Europe, in an attempt to recreate the empire of Charles V. The early victories of the Spanish armies at Fleurus (1625) and Breda (1625) were countered by the diplomatic success of the French minister, the duc de Richelieu, in provoking the Wars of the Valtelline (1624–26) and of the Mantuan succession (1628–31), which broke the Austro-Spanish control of northern Italy. In 1634 Spain gained its last spectacular victory of the Thirty Years' War at Nördlingen, for in the following year France intervened directly in the war, which then completely changed its character, particularly after the internal rebellions of 1640 in Catalonia and Portugal (*see* below) laid these territories open to French intrigue. The defeat of the Spanish forces by the French at Rocroi (1643) and at Lens (1648) forced Spain to recognize the independence of the United Province at the Treaty of Münster (January 1648). Eleven years later the Treaty of the Pyrenees (1659) with France deprived Spain of Roussillon and part of Cerdagne on the Pyrenean frontier, and parts of Flanders.

Catalonia.—The increasing financial needs of the crown compelled Olivares to try and forge the group of autonomous kingdoms, brought under one rule by the Catholic Kings, into a centralized state. His plan brought into the open two incompatible forces: on the one hand the system of *fueros*, or local privileges (especially in the kingdom of Aragon and above all in Catalonia); and on the other the attempt of the absolute monarchy to concentrate all power in the hands of the sovereign. Olivares clashed openly in 1626 with the Catalan and Aragonese *Cortes* over his policy of more centralized control of arms and money, and from then until the outbreak of war with France in 1635 the crisis revolved around the ambitions of Olivares and the opposition of the Barcelona *bourgeoisie*, neither party, it seems, wishing utterly to exclude the possibility of some future agreement.

After 1635 Catalonia was subjected to pressure from both France and Castile. The problem then became one of whether this discontented *bourgeoisie*, although opposed to Olivares' centralizing policy, would be able to prevent an alliance between the more extreme Pyrenean nobility and Cardinal Richelieu. In 1638 this latter alliance became more probable, as the balance of power swung toward the nobility with the election of a fervent Catalan patriot, the Canon of Urgel, Pau Clarís, as president of the *generalidad* (autonomous government of Catalonia). At the same time, in Castile the party led by Cardinal Borja and the *protonotario* of the Council of Aragon, Jerónimo de Villanueva, which demanded a more intransigent policy toward Catalonia, gained ground.

When Olivares, bent on extracting financial assistance from the Catalans, decided to launch an offensive against France from Catalonia, the billeting of soldiers there produced a revolt among the peasantry (May 1640). Most of the aristocracy and upper *bourgeoisie* remained faithful to the Spanish crown. The revolutionary solution, however, and the subsequent incorporation of the principality with France (January 1641) were imposed by the peasantry and by the lesser nobility, who reflected the social instability of the times. Catalonia then remained the site of a civil war and a battlefield between Spain and France until the Treaty of the Pyrenees in 1659. During the course of the war a new move between Madrid and the moderates in Barcelona led to the defeat of the extremists with the capture of Barcelona by Don John of Austria in October 1652, and the Catalan revolution ended with the reincorporation of the principality of Catalonia with the crown of Philip IV in January 1653. In October 1640, while the Spanish government was preoccupied with affairs in Catalonia, a successful revolt in Lisbon proclaimed Portugal's independence under the duke of Braganza, who became John IV, though it was not until the Treaty of Lisbon in 1668 that Spain finally recognized Portugal's independence (*see* PORTUGAL: *History*).

The economic and financial distress of the monarchy was aggravated further by the burden of these wars. Philip IV was forced to resort to inflation, especially in the years 1634–35, and so great was the increase in prices that some sort of deflationary measures became essential. From 1656 until the deflation of 1680 (in the reign of Charles II) the country endured a very real monetary crisis, which marked the lowest point of the depression of the 17th century. This fact explains why social distress was even greater than in the reign of Philip III.

Charles II.—Philip IV's son Charles II (*q.v.;* 1665–1700), the last of the Spanish Habsburg kings of Spain, came to the throne as a minor, and his reign began with the ten-year regency of the queen mother, Mariana of Austria. In domestic affairs, the ruin of Olivares' policy had produced a strong reaction in favour of the decentralization of government. This was accompanied by some economic improvement of the peripheral regions of Spain, notably in Catalonia after 1680, in contrast to the position in Castile. This improvement, however, was due partly to the fact that from about 1680 gold from Brazil did much to improve the economic and financial position throughout Western Europe. From 1665 to 1680 several rises in prices resulting from government action had proved disastrous and made essential the deflation of 1680 which resulted in a 45% drop in prices and restored some order to the monetary chaos in Spain. By 1686 the position was becoming more stable and a slow recovery in the economy was in progress. In 1680 the first minister, the duque de Medinaceli, created the Junta de Comercio y Moneda and inaugurated an economic policy based on the mercantilist principles of the French minister Jean Colbert (*q.v.*). Trade increased, especially through the port of Barcelona, and there was some development of Castilian industry. Charles II promulgated an order declaring that commercial activities were compatible with noble birth. The beginnings of the drastic reforms of the economy, and of the administrative system of Spain, that the Bourbons were to carry out in the 18th century, can in fact be seen in these minor reforms.

Wars with France.—In foreign affairs Charles was forced to involve Spain in continuous wars against the imperialism of Louis XIV of France. In 1667 Louis claimed the Spanish Netherlands on behalf of his wife Maria Teresa, the daughter of Philip IV's first marriage and consequently half sister to Charles II. The resulting War of Devolution (*q.v.;* 1667–68) produced a European coalition against France, but by the Peace of Aix-la-Chapelle (1668) Spain ceded to France several places in Flanders. A further war by Louis XIV from 1672 to 1678 (*see* DUTCH WARS), which was ended by the Peace of Nijmegen, deprived Spain of the Franche-Comté and resulted in frontier changes in the Low Countries. However in the War of the Grand Alliance (*q.v.;* 1689–97), despite Spanish defeats, Louis negotiated the generous Peace of Rijswijk, whereby he evacuated his forces from Catalonia and restored to Spain its possessions in Flanders as set down at the Peace of Nijmegen. By this time Louis had no wish to prejudice the Spaniards against him, for the question of the succession to the Spanish throne (to which France had strong claims, *see* below) was becoming acute.

The lack of any issue from Charles II's two marriages, to Marie Louise d'Orléans and Maria Anna of Neuburg, had led to great speculation among the major European powers as to his successor. Rival claims from both France and the emperor Leopold I, and fears by the European powers of dominance by France if the former were successful, resulted in the treaties of 1698 and 1699 for the partition of the Spanish Empire. Anxious to preserve the unity of his empire, Charles II in October 1700, a month before his death, made a will naming as his sole heir Philip, duc d'Anjou (later Philip V of Spain), a younger grandson of Louis XIV. Louis' acceptance of the will, despite his obligations to the partition agreements, and his declaration in December 1700 that Philip was not debarred from the succession to the French throne, led directly to the War of the Spanish Succession (1701–13; for details of the dynastic claims to the succession, of the partition treaties, and of the war itself, *see* SPANISH SUCCESSION, WAR OF THE). (J. RE.)

I. THE BOURBON DYNASTY

Philip V.—In the first half of the 18th century, Spain played an active part in foreign affairs, partly because of the reforms that followed the accession of a French prince, Philip V (*q.v.;* 1700–January 1724 and August 1724–1746). Louis XIV sent Jean Orry (*q.v.*) to inspect the Spanish finances in 1701, and in 1702

Orry was put in charge of military finances. After 1705 the work of reform was considerably helped by the arrival of Michel Amelot as French ambassador, and, even though Amelot was recalled in 1709 and Orry was out of Spain between 1706 and 1713, the reforms which they had initiated did much to revive Spain. The basic object of Orry's reforms was to centralize the financial administration, a task made easier by the fact that the old kingdoms of Aragon and Valencia and the principality of Catalonia, which had declared for the archduke Charles (from 1711 the emperor Charles VI), the Austrian Habsburg claimant to the Spanish throne, were eventually conquered, so that their financial independence could be ended. Orry also worked to simplify the methods of collecting taxes, though it was not until after 1713 that his plan of one tax collector in each province with only one assistant in each area came into operation. When Orry proposed to tax the church, however, he was dismissed (1715). The treasury had been freed of considerable sums which had previously been granted every year as pensions; income from alienated crown property was resumed for one year, and treasure brought by private individuals from the Indies was taxed heavily. Spain was fortunate in that Philip V had no extravagant favourites, such as Philip III and IV had had in Lerma or Olivares, and his shabby court was not so costly as the luxurious court of Philip IV. From 1702 to 1714 prices had risen a little because of bad harvests, plague, and an exceptionally cold winter, but this rise was very slight by comparison with what had happened in the 17th century; from 1714 to 1732 prices were generally stable.

Administrative and political centralization was increased as the districts which had supported the archduke Charles were reconquered. In 1707 the *fueros* of Aragon and Valencia were ended, and municipal government on Castilian lines was introduced in those provinces. The eastern kingdoms ceased to be regarded as separate. The office of viceroy was abolished and in its place the king appointed captains general, one of whose functions was to preside over the *audiencias* (high courts of justice) set up at Saragossa, Valencia, Barcelona, and at Palma de Mallorca in the Balearics. The extinction of the *Cortes* of Aragon and Valencia in 1709 and of Catalonia in 1724 was another sign of the increasing centralization. In achieving these reforms Philip V took little part. His main preoccupation in these years was the war for the succession, and his determination to retain the throne of Spain was loyally supported by his young wife María Luisa of Savoy. On the whole Philip's foreign policy followed the line laid down by Louis XIV. The loyalty of the Spanish court to France was encouraged by the influence of Marie Anne de la Trémoïlle, princesse des Ursins (*q.v.*), who had been sent to Spain by Louis XIV in 1701 to become chief lady of the bedchamber to the queen of Spain.

By the treaties of Utrecht (1713) and of Rastatt (1714), which ended the war, Philip was recognized as king of Spain and of Spanish America, but lost the Spanish possessions in Italy (Milan, the kingdom of Naples, and the Spanish bases on the Tuscan coast), Sardinia, and the Spanish Netherlands to the emperor; Sicily to the king of Savoy; and Gibraltar and Minorca to England. Philip had previously renounced his claim to the French throne. Philip and the emperor, however, refused to make peace with each other and remained technically at war.

Isabella Farnese.—In 1714 María Luisa died and the choice of a second wife for Philip fell on Isabella Farnese (*q.v.*), niece and stepdaughter of the duke of Parma, who was commended to Madame des Ursins as being young, countrified, and completely inexperienced politically. During her journey overland to Spain, Isabella met Maria Anna, the widow of Charles II of Spain, and also had opportunities for considerable conversations with Giulio Alberoni (*q.v.*), the envoy at Madrid of the duke of Parma. Although inexperienced, she quickly realized that she must break the power of the princesse des Ursins. In her first interview with the favourite, Isabella showed her character by exiling her (Dec. 23, 1714). For a few years Alberoni was the dominating influence, though the official chief minister was Cardinal Francesco del Giudice. Alberoni was convinced that Spain had the resources to become a great power. He was also deeply concerned for the small states of Italy, particularly Parma, and wanted to free them from the emperor's influence. This he hoped to do with the help of Spain, if Spanish policy could be directed by the new Farnese queen.

Foreign Policy.—The death of Louis XIV in September 1715 provided Alberoni with an opportunity to carry out his plans. Philip V had already shown signs of wishing to escape from French control. In December 1715 Alberoni persuaded the Spanish court to conclude a commercial treaty with Britain which eradicated the difficulties to British trade that had arisen from the shortcomings of the commercial treaty signed at Utrecht in 1713. This was followed in May 1716 by another agreement with Britain which gave satisfaction for grievances arising out of the Asiento Treaty (*see* ASIENTO) of 1713, permitting Britain to supply Negro slaves to Spanish America. In July 1716 Cardinal Giudice fell from power and Alberoni's influence, though still unofficial, increased, but he was badly discredited when England and the emperor concluded the Treaty of Westminster, 1716, and when this was followed in November 1716 by an alliance between England and France. The Spanish court, fearing that the emperor might try and increase his power in Italy, used the pretext of the arrest in Lombardy of the Spanish Inquisition General, Cardinal Molinés, by imperial authorities, to launch a successful attack in November 1717 on Sardinia, a former Spanish possession. This was followed by an attack on Sicily in July 1718.

For a time it looked as if Spain might again embroil all the powers in war. In August 1718 Great Britain, France, the emperor, and the United Provinces formed the Quadruple Alliance to settle the differences in Southern Europe and if necessary to coerce Spain to accept their terms. Later that month the English destroyed the Spanish fleet off Cape Passaro, Sicily; in 1719 French troops overran the Basque provinces; and finally in 1720 Philip V accepted the terms of the Quadruple Alliance as a basis on which peace could be reestablished between Spain and the empire. The emperor renounced his claim to the Spanish crown, and Philip V renounced his claim to the Spanish territories in Italy, lost to the emperor in 1713. Philip agreed to restore Sardinia to the emperor (who in fact gave it to Savoy) and to relinquish his claims to Sicily, which passed to the emperor. Some issues still undecided were to be settled at a congress to meet at Cambrai, in France. Alberoni had been dismissed in December 1719, but he and Isabella had achieved something. In the Quadruple Alliance the powers acknowledged the claim of Don Carlos (afterward Charles III of Spain), Philip's eldest son by Isabella, to succeed to Parma and Tuscany, territories to which Isabella had dynastic claims.

With Alberoni's dismissal Spain relapsed temporarily into a state of stagnation. Alberoni had carried the administrative reforms of Orry a stage further and simplified the institutions of government. He had abolished the *Despacho*, or small Cabinet Council, and had reduced the powers of the larger *Consejo de Estado*. The old councils of Castile, Finance, War, and the Indies remained, but their influence had declined. Spanish affairs were, in fact, in the hands of three secretaries of state, but in 1720 there was no statesman capable of exploiting the power of one of these posts. The marqués de Grimaldo became the chief of the secretaries of state, but he was not a statesman. Under his guidance Spain in 1721 renewed its alliance with France and resumed cordial relations with England. Isabella hoped that England and France would help to secure for Don Carlos effective possession of the Italian lands, but the dispute between Spain and the emperor was not easily settled and negotiations at the Congress of Cambrai dragged on from 1721 to 1724.

Abdication of Philip.—In January 1724 Philip V caused a sensation by abdicating in favour of Luis, his eldest son by his first wife. Whether he did this to strengthen the claim he intermittently made to the throne of France or because he was tormented by religious scruples, knew himself unfit to govern, and was only putting into effect an earlier vow to abdicate is uncertain. The episode ended swiftly for in August 1724 Luis died of smallpox and Philip yielded to the pressure of Isabella, France, the papacy, and the Council of Castile to resume the crown. The abdication

had had very little effect on Spanish policy, but when it became clear that France would not take active steps to secure the Italian possessions for Don Carlos and when in 1725 the French court returned to Spain the infanta Mariana (b. 1718), who had been sent to France on her betrothal to Louis XV at the time of the alliance in 1721, Isabella and Philip decided to abandon the French alliance and try a direct approach to the emperor. The negotiations with the imperial court at Vienna were entrusted to a remarkable Dutchman Jan Willem Ripperdá, who managed to achieve peace and even an alliance with the emperor in 1725 by the Treaty of Vienna. Ripperdá reported in Spain that the emperor had requested that he should be made chief minister, but he was soon exposed as having misled both the emperor and the rulers of Spain by rash promises and was dismissed in May 1726.

Succession to Parma.—The improved relations with the emperor, however, soon cooled and Spain reverted to a policy of friendship with England and France, signing the Treaty of Seville with them in 1729. On the death in January 1731 of Antonio, the last Farnese duke of Parma, the emperor seized the duchy. The dispute over Parma was finally solved by the diplomacy of England, which in return for accepting the Pragmatic Sanction, guaranteeing the Austrian succession to Charles VI's daughter Maria Theresa, persuaded the emperor in the Treaty of Vienna (1731) to cede Parma to Don Carlos and invest him with the reversion to Tuscany. In 1733 Spain concluded the first Family Compact with France, joining that country in the War of the Polish Succession (*q.v.*). As a result, by the Peace of 1735 (ratified in 1738) Don Carlos was confirmed in his conquest of Naples and Sicily in return for relinquishing Parma and his claim to Tuscany.

War with England.—In 1739 war broke out between Spain and England. Since 1713 there had been considerable causes of friction between them, particularly the large amount of illicit trade carried on by the English with the Spanish colonies on the American mainland and in the West Indies. The Spanish colonial governors, finding it difficult to fit out ships to patrol the coasts, allowed adventurers and privateers to intercept smugglers, and British merchants claimed that their ships were illegally stopped and searched. Although neither the English prime minister Sir Robert Walpole nor the Spanish secretary of state the marqués de Patiño (*q.v.*) wanted war, mercantile clamour in England could not be pacified, and in 1739 the war known as the War of Jenkins' Ear broke out (*see* JENKINS, ROBERT). In 1740 this war merged with the War of the Austrian Succession (*q.v.;* 1740–48), and in 1743 Spain concluded with France the second Family Compact, whereby France undertook to secure Milan, Parma, and Piacenza for Don Philip, Isabella's second son, and to help Spain recover Gibraltar and Minorca. During the war, Spanish forces fought in Italy, achieving some notable success. In 1746 Philip V died. During his reign the energy and efficiency of Orry, of Alberoni, and of the marqués de Patiño had done something to restore Spain's prosperity and in particular to improve the royal revenue and develop efficient dockyards.

Ferdinand VI.—Ferdinand VI (*q.v.;* 1746–59), Philip V's fourth son by his first wife, was a pacific, unambitious man. The treaty of Aix-la-Chapelle (1748) ended the War of the Austrian Succession and assigned Parma, Piacenza, and Guastalla to Don Philip. In 1750 Spain concluded an agreement with England ending the *asiento* in return for £100,000. Don José Carvajal y Lancáster, who directed Spanish policy until 1754, was determined to restore the greatness of Spain and decided that the best policy was to tolerate one expensive but powerful friend, England. Under his guidance Anglo-Spanish relations entered upon an era which has been called "the Seven Years' Peace." The other great minister of the reign Cenón de Somodevilla, marqués de la Ensenada (*q.v.*), though he inclined to a pro-French policy and had been responsible for the second Family Compact of 1743, accepted the pacific policy of Ferdinand VI and occupied himself with reviving agriculture, stimulating mining and trade, and developing the navy. By the middle of the century the population of Spain had increased from about 5,700,000 in 1700 to 7,500,000, though estimates of the exact figures vary considerably. In 1746 the first of the economic societies was founded to revive an interest in agriculture. Academies of history and of arts had been set up in the last years of Philip's reign, and a royal college was begun to educate sons of the nobility. The Inquisition, which under Philip V had held 728 autos-da-fé and imposed 14,000 sentences, gradually declined in importance under Ferdinand VI. Unfortunately after the death of Carvajal in 1754 it was impossible for the new minister Richard Wall to prevent increasing friction with England, which arose mainly out of colonial conflict in the Indies. (J. O. Li.)

Charles III.—The reign of Charles III (*q.v.;* 1759–88), eldest son of Philip V and Isabella, and half brother of Ferdinand VI, can be characterized as one of enlightened despotism, for in that period reforms were carried out by a monarch who was much more absolute than his predecessors, using ministers who were influenced by reforming ideas, then current throughout Europe. He had two problems: first, the economic condition of Spain, an amalgam of chronic unemployment, huge untilled estates, a chaotic currency, and the absence of industry; and second, the political and administrative problem, the lack of effective agencies both for the formulation of policy and for its application. Charles III and his ministers set themselves to redress this situation, but their achievements should not be exaggerated. Moreover, Charles III's absolutism was less effective in practice than in theory: as late as 1787 there were in Spain over 10,000 towns and villages subject to seignorial jurisdiction and thereby removed from direct royal control. Nevertheless Charles III did give Spain a more modern political appearance and some measure of economic recovery. His experience as king of Naples (1735–59), his conviction of his reforming mission, and his devotion to work and duty adequately compensated for his rather limited intelligence. Above all he was successful in his choice of advisers and ministers, selecting them neither from the traditionalists nor from the aristocracy but from a small and enlightened group of men who supplied the ideas and set the pace of reform.

Yet Charles III's early ministerial appointments were not popular in Spain and his preference for foreigners was resented. The most prominent of his ministers were the Italians, the marqués de Squillace, secretary of state for war and finance, and the marqués de Grimaldi, secretary of state for foreign affairs. Spain's disastrous participation in the Seven Years' War (*q.v.;* 1756–63), the rise in the price of food caused by general inflation and a series of bad harvests between 1763 and 1765, and the higher taxes demanded by Squillace to pay for public works increased the unpopularity of the foreign advisers. Finally, Squillace's attempt to enforce an old law forbidding men to wear in Madrid their broad-brimmed slouch hats and long capes, on the ground that they hindered the detection of criminals, provoked violent riots in the capital in March 1766 which spread to other towns in central Spain. The king was forced to accept the terms of the insurgents: the exile of Squillace, the revocation of the dress order, and the lowering of food prices. The conde de Aranda (*q.v.*) was called in to restore order (holding office to 1773), and was given the presidency of the Council of Castile. Within a year he was able to revoke most of Charles's concessions.

Suppression of the Jesuits.—The riots of 1766 were now used as an excuse to attack the Jesuits who were regarded as a threat to the supreme authority of the crown and to its interests in Spanish America. It was rumoured that the riots had been planned by groups of nobles and clergy who wished to evict Squillace and discourage Charles from further reforms. The truth of these reports is still an open question, but they were useful to the government and a royal commission found the Jesuits guilty of using their influence to incite riots. On this evidence Charles III ordered the expulsion of the Jesuits, and the decree was enforced in Spain and the colonies in April 1767. The subordination of the church to the state in Spain was completed by the promotion of churchmen of regalist and antipapal views, by the decree forbidding the publication of papal bulls in Spain without royal permission, and by the curtailment of the jurisdiction of the Inquisition. In 1776 Charles III removed his last prominent foreign adviser when the conde de Floridablanca (*q.v.*) replaced Grimaldi.

Administrative and Economic Reforms.—Charles III continued and perfected the administrative reforms introduced by Philip V, whereby individual ministries gradually replaced the Habsburgs' conciliar system of government, and the concentration of power in the hands of a small number of men continually in contact with the king gave policy a new vigour. These ministers promoted policy, extended the central power throughout Spain, stimulated the intendants, and inaugurated administrative reforms in the collection of revenue, in national defense, and in local government. Finally, the various ministries were coordinated by Floridablanca into a *Junta de Estado*, a type of cabinet, which met frequently to discuss policy collectively and which in 1787 was given a definite constitution. Charles III's most urgent problem was, however, the economic condition of the country. The rise in population (from about 6,000,000 to more than 10,000,000 in the course of the 18th century), combined with inflation, led after 1750 to a rush for land in which the already powerful landowners increased their holdings, while the condition of tenants and landless labourers in Castile became even more desperate. In spite of the limitation of the pasture privileges of the Mesta (sheepfarmers' guild) in the interests of farmers of arable land and the creation of agricultural colonies in the Sierra Morena, nothing effective was done to remedy the greatest defect in the Spanish land system: huge and largely untilled estates held by the church and nobility. The planners' main concern, however, was not agriculture but colonial commerce and domestic manufacture. Support was given to the quasi-official economic societies, formed to promote the prosperity of the country by applied science. To improve communications the state itself invested money in public works, such as road and canal building. Official encouragement was given to industry. In 1770 a royal decree announced that there was no incompatibility between aristocratic or gentle status and engaging in a craft. A number of state-supported factories employing foreign craftsmen were set up to produce luxury goods, hitherto imported. Domestic manufacture was protected: by 1771 the import of all foreign cotton cloths was prohibited, a policy particularly favourable to Catalan industry.

Charles III's most effective economic measure, however, was the opening of colonial commerce to freer trade. In 1765 Spain ended the policy of restricted ports of clearance from Spain and of entry to the Indies. Trade with the colonies was now open to all the main Spanish ports, and trade with Spain was open to many more Caribbean ports. Gradually the free-trade area was extended, and by 1789 Spain had abandoned the commercial regulations held sacred for more than 200 years. These reforms injected new life into the domestic and imperial economy. The industries of Catalonia and Biscay (Vizcaya), especially cotton and metallurgy, whose growth had been one of the influences determining the measures of freer trade, further benefited from the resulting expansion. Shipping between Spain and the colonies multiplied rapidly, and the whole trade with Spanish America is said to have expanded by as much as 700% in the period from 1778 to 1788.

Colonial Wars.—Commercial objectives were closely connected with imperial defense. Spain's most powerful rival was Great Britain, and Charles III's anxiety to maintain a colonial balance of power was one of the main objectives of his foreign policy. For this reason he made with France (August 1761) the third Family Compact, which created an offensive and defensive alliance between the two Bourbon powers and took Spain into the Seven Years' War when England, already at war with France, declared war on Spain in January 1762. The war proved a disaster for Spain as well as for France. England captured Havana and Manila, and although both were restored at the Peace of Paris in 1763, Spain was forced to cede to Britain Florida and all Spanish territory in North America east of the Mississippi River. From France, Spain gained Louisiana—and a new frontier to be defended against Britain. Despite this humiliation, however, Charles hesitated to intervene against Great Britain in the American War of Independence (1775–81), for Spain could not profit from the precedent of colonial independence. When Spain finally followed France into the war in 1779, it pursued its own interests without giving much direct support to the British colonies and without recognizing their independence. Spanish forces reconquered Florida, occupied the British outposts in the Bahamas, and expelled the enemy from their logging bases in Honduras. A prolonged siege of Gibraltar (1779–83) was unsuccessful, but an expedition to Minorca recovered the island for Spain in February 1782. By the Peace of Versailles (Sept. 3, 1783) Spain regained Florida and retained Minorca but restored the Bahamas to England and gave the English certain trading rights in Honduras. Despite its defeats Spain came out of the war with moderate success.

Charles IV.—The reign of Charles IV (*q.v.;* 1788–1808), the second son of Charles III, was marked by the rule of the favourite, Manuel de Godoy (*q.v.*), and the impact of the French Revolution. The king himself, intellectually feeble and lacking the willpower of his father, was powerless to direct events. At first he tried to continue the reforming policy of Charles III and he retained Floridablanca as his secretary of state. The moribund *Cortes* of Castile, convoked in 1789 to recognize Charles IV's son Ferdinand (later Ferdinand VII) as heir to the throne, received with apathy a royal proposal to limit the privilege of entail of the nobility. But events in France ruined the chance of any reforming program. The flood of revolutionary propaganda crossing the border, while it found little response in Spain, startled Floridablanca into a series of panic-stricken repressive measures at home and a policy of hostility toward the revolutionary cause in France, a policy which was regarded in Spain as endangering the life of Charles's cousin, Louis XVI. Floridablanca was therefore dismissed in February 1792, but his successor, the conde de Aranda, fared little better and was replaced in November 1792 by Godoy, a former guards officer who, under the protection of the queen María Luisa, had risen rapidly in military rank and government position and whose appointment scandalized the Spanish public.

Relations with France.—Godoy's attempt in 1793 to save the life of Louis XVI by bribing the vote in the Convention failed. Pressed by England and bound by the Family Compact, Charles IV declared war on the French republicans in 1793. A French army invaded Spain (*see* FRENCH REVOLUTIONARY WARS) and was only bought off by the Treaty of Basel in July 1795, whereby Spain ceded to France its portion of the island of Hispaniola, known as Santo Domingo. Spain, weak in material resources and faced with a hostile Britain, was now tied ever more closely to French interests. In 1796, by the Treaty of San Ildefonso, Charles joined France in the war against England, during the course of which the Spanish suffered a naval defeat off Cape St. Vincent (February 1797; *see* ST. VINCENT, BATTLE OF) and lost Trinidad and Minorca. The French, suspecting Godoy's loyalty to the alliance, procured his dismissal in 1798. The advent of Napoleon Bonaparte to power from 1799 meant further subordination of Spain to France. In a second Treaty of San Ildefonso (1800) France promised, in exchange for Louisiana, to aid Spain to enlarge the duchy of Parma. Pursuing his hostility against England, Napoleon forced Spain to attack Portugal in 1801, and Godoy himself, now back in power, led an invading army which forced Portugal, in return for a Spanish guarantee of Portugal's territorial integrity, to undertake to close its ports to England. In 1802 a general peace was secured by the Treaty of Amiens: Spain recovered Minorca but not Trinidad.

In spite of disastrous conditions at home, heavy taxation, and an empty treasury, Spain was forced to pay heavy indemnities to France and was dragged into yet another war with England (December 1804; *see* NAPOLEONIC WARS). In October 1805 a Franco-Spanish fleet was destroyed off Cape Trafalgar (*see* TRAFALGAR, BATTLE OF), a crushing blow to Spanish morale and to Spain's ability to protect its overseas possessions. This was not all. Charles IV and Godoy, subservient to Napoleon in spite of themselves and subject at home to the growing opposition of the Fernandista Party (the supporters of Charles IV's heir Ferdinand), subscribed to the Treaty of Fontainebleau (October 1807) by which Napoleon and the Spanish government agreed upon the conquest and partition of Portugal. The French army which conquered Portugal (1807) also occupied parts of northern Spain, and Napoleon, whose intentions were now becoming clear, claimed all

of Portugal and certain provinces of northern Spain. Unable to organize government resistance, Godoy persuaded Charles IV to imitate the Portuguese royal family and escape to South America. The journey from Madrid was halted at Aranjuez, where a revolt organized by the Fernandista faction (March 17, 1808) procured the dismissal of Godoy and the abdication of Charles IV in favour of Ferdinand. Napoleon, taking advantage of the situation, ordered Gen. Joachim Murat to occupy Madrid and by a mixture of threats and promises induced both Charles and Ferdinand to proceed to Bayonne for conferences. There, on May 5, 1808, Napoleon forced Ferdinand to abdicate in favour of Charles, and Charles in favour of himself. In exchange Napoleon promised that Spain should remain Catholic and independent, under a ruler whom he would name. He chose his brother Joseph Bonaparte. However on May 2 the people of Madrid had already risen against the invader and the war for Spanish independence (1808–14) had begun. For details of the war *see* the article PENINSULAR WAR. (Jo. L.)

The War of Independence.—The rising in Madrid began the movement which ultimately proved fatal to Napoleon's power. Although the Madrid revolt was ruthlessly suppressed by the French, provincial insurrections took place throughout Spain, and the Spaniards showed great capacity for guerrilla warfare. The French were repulsed from Valencia and Gen. Pierre Dupont, who had advanced into Andalusia, was compelled to retreat and ultimately to capitulate with all his army at Bailén (July 23). The Spaniards now advanced upon the capital and expelled Joseph Bonaparte (August).

To some extent this was a civil war, for a small but not unimportant section of Spaniards, who had been much influenced by the doctrines of the French Revolution, continued to see in France the bulwark of liberal principles and looked to Joseph and the constitution he had drawn up at Bayonne as marking the dawn of responsible government in Spain. The struggle for independence, though in itself inevitably productive of new currents of thought in a progressive minority, was to seem in retrospect to have been a struggle no less against the spirit of the age. But the combined defection of the crown, of its senior statesmen, of a large part of the higher nobility, and of the intellectual *Afrancesados* made the reaction of the Spanish people, thus left leaderless, the more remarkable. The provincial committees which had been set up now created a single central junta which sat at Aranjuez. The military superiority of the French ruled out hopes of an immediate victory, but Napoleon himself was compelled to come to Spain to direct operations. The French campaign, leading to the capture of Madrid (December 1808), forced the junta to retreat southward to Seville. In January 1810 Gen. Nicolas de Dieu Soult began the conquest of Andalusia and with the fall of Seville in the same month the central junta fled to Cádiz. Only the obstinate resistance of the duke of Wellington in Portugal, the continuous activity of the *guerrilleros,* and dissensions among the French saved the peninsula from final submission.

The Spanish *Cortes* had become in the course of the 18th century little more than a rubber stamp. Under Philip V the *Cortes* of the other regions had been merged in the *Cortes* of Castile. Now, under the influence of the ideas of revolutionary France, the reformist patriots of Cádiz planned their reestablishment on a more democratic though still national basis, and in September 1810 a new *Cortes* was convened in that city. It included for the first time delegates of the colonies, chosen like those of Spain itself, from among those citizens who were present in Cádiz. The third estate (bourgeois), with 184 deputies, was sufficient to swamp the nobles and the clergy, and the debate on the new constitution soon produced a deep cleavage between *Liberales*—it was then that the term took on its party political connotation—and *Serviles* (supporters of absolute monarchy). The former won the day, and the constitution of 1812 was duly promulgated. Its 384 articles gave to the Spanish nation the exclusive right to determine its fundamental laws. The king would legislate with the *Cortes,* now a single chamber; but royal decrees, to be valid, required the countersignature of a minister, and ministers were responsible to the *Cortes.* Freedom of the press, though not of religious belief, was assured, and legal codes were made uniform for the whole country. The constitution was unworkable in many respects and never achieved more than a symbolic importance; but by vesting in the people for the first time the responsibility for determining under what laws they should live, it closed an epoch in their history and established constitutionalism as the primary characteristic of the new age. Its influence was also profound in two other directions. Its radical idealism, offending the mass of the people, was partly responsible for the support Ferdinand received at his restoration in 1814 (*see* below) and prepared the way for one of the most reactionary regimes in Spanish history; the tendency persisted thereafter for Spanish politics to swing from extreme to extreme. It also gave later warrant for the identification of liberalism with both anticlericalism and centralization.

From his base in Portugal, which he had successfully defended, Wellington in 1812 began his gradual advance into Spain. His defeat of Marshal Jean Baptiste Jourdan at Vitoria on June 21, 1813, finally decided the issue in the peninsula. Joseph Bonaparte withdrew from Spain, and Wellington fought his way across the Pyrenees into France. Napoleon, after his crushing defeat at Leipzig (October), recognized the impossibility of retaining his hold on Spain by releasing Ferdinand, who had been detained by the French at Valençay since his abdication in 1808.

Ferdinand VII.—Before entering Spain in March 1814 Ferdinand VII (*q.v.;* 1808 and 1814–33) had undertaken to maintain the constitution of 1812, and on reaching Figueras he was met by a formal demand in this sense from the *Cortes.* But he was quick to appraise the true temper of the nation, and in rejecting the demand now he was supported alike by the army, the church, and the people. Had he been capable of governing well, he might have ruled despotically for a considerable time. He was, however, cruel, cowardly, and without real ability. Dominated by a *camarilla* of favourites, he changed his ministers incessantly and was consistent only in his ferocious persecution of the *Liberales.*

Several minor revolts between 1814 and 1817 were easily crushed, but that of January 1820, beginning as a mutiny at Cádiz in protest against the sending of an expeditionary force to crush rebellion in the South American colonies, became a revolution. Its leader and hero, Gen. Rafael del Riego, was one of a number of officers, many of them former prisoners in France, who now subscribed to liberal doctrines. Six years of tyrannical excesses had turned the army largely against the king and from 1820 to 1823 Ferdinand was a prisoner in the hands of a section of his subjects, who proclaimed once more the constitution of 1812. Ferdinand appealed for help to the Holy Alliance (*q.v.*). As early as 1820 the emperor Alexander I of Russia had vainly proposed joint intervention. In 1822 the question arose again at the Congress of Verona, France proposing to march into Spain to restore Ferdinand.

Great Britain saw in the demand only a pretext for reviving traditional Bourbon ambitions in the peninsula and protested vigorously; but the mandate was granted by a majority of the powers and on April 7, 1823, Louis Antoine de Bourbon, duc d'Angoulême, crossed the Bidassoa (Bidasoa) River at the head of a powerful army. The result was a startling proof of the flimsy hold of liberalism on the Spanish people. What Napoleon had failed to accomplish in years, Angoulême seemed to have achieved in a few weeks. Napoleon, however, had sought to impose upon Spain an alien dynasty; Angoulême came to restore the Spanish king "to his own" and found some Spaniards forthcoming with active support and the majority at least tacitly cooperative. The *Cortes,* taking the king as prisoner, fled again to Cádiz, where after a brief siege it surrendered (September). Ferdinand, alleging compulsion, refused to honour the amnesty he had sworn, and the French troops witnessed with helpless indignation an orgy of cruel reaction begun under the protection of their bayonets.

The Spanish Colonies.—The emancipation of Spain's American colonies was an indirect result of Napoleon's detention of Ferdinand and invasion of Spain in 1808. Juntas in the various administrative capitals overseas assumed power, first in the name of Ferdinand against Joseph Bonaparte, then in the name of the

colonies themselves against Spain. Some of these had already become independent *de facto* by 1810, and many more became so in the ten years following; and the recognition of their independence *de jure* was, for Great Britain at least, merely a question of time. When the French invasion of Spain was seen to be inevitable in 1823, George Canning, the British foreign secretary, informed the French government that Great Britain would not tolerate any subjugation of the Spanish colonies by foreign force. The United States, though declining to act in concert, followed with the famous message of Pres. James Monroe (Dec. 2, 1823; *see* MONROE DOCTRINE) intimating his country's opposition to any interference by Europe in the Americas. The Republic of Colombia had already been recognized by the United States in 1822. Great Britain recognized Mexico and Colombia in 1824 and only delayed recognition of the other new states until they should have given proof of stability. There now remained to Spain, of its far-flung empire, only the islands of Cuba, Puerto Rico, and the Philippines.

The Succession Question.—Ferdinand, still childless on the death of his third wife María Josefa Amalia in 1829, married María Cristina I (*q.v.*) of the Two Sicilies in that year and in so doing threatened the mounting hopes of his brother Don Carlos (*q.v.*) regarding the succession. Don Carlos was also an extreme reactionary, and the clash of ambitions threw María Cristina on the support of the Liberals. The birth of a daughter Isabella in 1830 greatly complicated the issue. By the ancient law of Castile and León women could rule in their own right. This right had, however, been abrogated by an act of 1713 designed to prevent any union of the crowns of Spain and France; and, although Charles IV had restored the former position in 1789, his enactment had never been published and its validity was now hotly disputed. Hence the birth of Carlism (*q.v.*), the name of the movement by which the supporters of Don Carlos and his heirs were known, which was for more than half a century to be a disrupting factor in the history of Spain. When, on Sept. 29, 1833, Ferdinand died, his daughter was proclaimed queen as Isabella II (*q.v.;* 1833–68), with María Cristina as regent. The First Carlist War (1833–39) broke out almost immediately.

The Regency for Isabella II (1833–43).—María Cristina would have ruled despotically if she could, but the support of the army and control of the machinery of government were not in themselves sufficient to resist the Carlists, in whose ranks were to be found not merely the clergy and other ultraconservative elements but also large numbers of Basques, Catalans, and other peoples of the periphery who had been persuaded that their regional *fueros* were in danger and who supplied the real fighting strength of the movement in defense of them. The necessities of her position thus drove the regent to accept parliamentary institutions under a new constitution in 1837 which, though much strengthening the royal prerogative, was liberal and bore traces of the English Reform Bill of 1832. The constitution established a *Cortes* of two chambers, the upper house being elective, but gave the crown an absolute veto. The return from England and France of many exiles of the preceding decade also brought support to the liberal cause. But neither María Cristina nor her daughter ever genuinely accepted the principle of representative government and there ensued, within the Carlist War, a bitter struggle between Parliament and the parliamentary cause on the one hand and the crown on the other.

The First Carlist War, which was for the crown a dynastic issue and for the nation a conflict between irreconcilable political principles, involved Great Britain in military commitments on the government's side and threatened more than once to involve Europe in war. It ended with the Convention of Vergara (Aug. 31, 1839), a recognition of Carlist defeat. Baldomero Espartero (*q.v.*), the Liberal general, undertook to recommend to the *Cortes* the substantial confirmation of the Basque *fueros*, and Don Carlos left Spain for France. But the Carlist cause remained alive; and the *Cortes,* which insisted, in the name of national unity and uniformity, on whittling down the *fueros* until little more than administrative autonomy remained, ensured continuing Basque support for Carlism. An earlier measure for unity had been the supersession for administrative purposes, in 1833, of the various kingdoms of Spain by 49 provinces on the French model. From that year dates the Catalan *renaixença*, a resurgence of regional feeling at once cultural and political that was to have far-reaching consequences. Another enactment destined to provoke repercussions for many years was the acceptance in the constitution of 1837 of state responsibility for the upkeep of the church and the consequent disposal by the state of ecclesiastical property. The Jesuits, readmitted by Ferdinand VII, had been reexpelled in 1835.

With the removal of the Carlist threat to the throne in 1839, María Cristina was emboldened in her absolutist leanings, and now sought to have the constitution of 1837 overthrown. Failing in this, she directed an attack on the independence of the municipalities in 1840, which provoked a major crisis that culminated in her abdication (October). Espartero was proclaimed as regent in her place, and Spain had its first experience of the army in political control. After three disturbed years a military rising led by Gens. Juan Prim and Ramón Narváez (*qq.v.*), with liberal and moderate backing, drove Espartero likewise into exile (July 1843), and Isabella, then only 13 years old, was declared of age.

Isabella II.—The new reign proved but a continuation of the regency, a confused conflict between the attempts of the crown to rule despotically with a mere pretense of a *Cortes* and the nation's growing demand for honest and efficient parliamentary government. New constitutions (1845, 1852, 1855) alternated between the progressive and the reactionary and caused a spurious importance to be attached to the letter as distinct from the spirit of constitutionalism. Party leaders were almost always generals: Espartero, Narváez (leader of the "Moderates" or extreme conservatives), Leopoldo O'Donnell (*q.v.*), Juan Prim. As such they were unable, even where willing, to further the cause of civic responsibility in the nation at large.

The queen's marriage on Oct. 10, 1846, provoked international repercussions. Louis Philippe of France, seeking to revive the family alliance, constrained Isabella, with the aid of María Cristina, to wed her cousin Francisco de Asís de Borbón, who was believed to be impotent; on the same day Isabella's younger sister María Luisa Fernanda was married to Louis Philippe's son Antoine, duc de Montpensier. The affair broke the entente between England and France—Lord Palmerston having only agreed on condition that the second marriage waited on the birth of an heir to the first—and was a contributory cause of the downfall of Louis Philippe (*see* FRANCE: *History*).

The church's right to hold and acquire property, restored by a concordat of 1851, was again threatened in 1856 when Isabella was compelled to give assent to a bill disposing of church property. Finally, in 1858, an agreement was reached with the pope over the state's compensation for the appropriation of church lands. The major external event of the reign was the Moroccan War of 1859–60, but Spanish successes there led to the acquisition of only a small part of Morocco.

Cumulative discontents against the crown, reflected in the emergence of a republican party in 1854, culminated in a mutiny in September 1868 at Cádiz of a squadron under Adm. Juan Bautista Topete y Carballo, which became the signal for a general revolution. The stand for the queen made at Alcolea by Gen. Manuel Pavía y Lacy (*q.v.*), marqués de Novaliches, was an exception. Even her prime minister, Luis González Bravo, deserted her. Isabella crossed into France an exile, and her reign ended. The revolution of 1868 was the first to be openly directed against the dynasty. The "spurious race of Bourbon," it was boasted, had disappeared; and, as in 1810, sovereignty was again for the moment vested in the people.

J. AMADEO AND THE FIRST REPUBLIC, 1870–75

Most of the nation was, however, not yet ready for a republic, and a constituent *Cortes*, assembled in 1869, voted for a continuance of the monarchy under a different monarch; Gen. Francisco Serrano (*q.v.*), a leader of the revolution, was declared regent until one could be found. In the constitution of 1869 may be found, for the first time since that of 1812, the entire program of the

liberalism of the period. Monarchist opinion on the succession was divided, some favouring Isabella's son Alfonso (later Alfonso XII) under a regency, some the widower Ferdinand of Saxe-Coburg (formerly king-consort and then regent of Portugal), while others preferred the duc de Montpensier, or a Carlist, or even the aged Espartero. The search for a democratic king, said Prim, the prime minister, was like looking for an atheist in heaven. The offer of the crown to Prince Leopold of Hohenzollern-Sigmaringen proved the spark that fired the Franco-German War (*q.v.*).

At length, in August 1870, Amadeo (*q.v.*; 1870–73) of Savoy, second son of Victor Emmanuel II, accepted the offer of nomination. He was proclaimed king on Nov. 16, and landed in Cartagena on Dec. 30. On that day Prim died as a result of an attack made on him by an assassin. Amadeo's reign was a period of turmoil and confusion. The monarch, the tool of intriguing politicians, was looked on as an intruder by men of principle, royalist and republican alike. The Second Carlist War (1872–76) began in the north. There was republican agitation in the towns and when the first honourable excuse presented itself Amadeo resigned (February 1873).

The First Republic.—Disillusionment with monarchy was for the moment complete, and the *Cortes* proclaimed Spain a republic. New elections were held (May) to a constituent *Cortes*, but the republic ran its brief and tempestuous course without a constitution. Beyond initial acceptance of the federal idea there was little agreement on first principles. Four presidents held office within the year: Estanislao Figueras, Francisco Pí y Margall, Nicolás Salmerón, and Emilio Castelar y Ripoll (*q.v.*). Ministries never knew stability, and many regions disavowed the authority of Madrid and proceeded to act independently. In the south, several towns went beyond federalism and pronounced themselves "cantons." The cantonalists in Murcia seized Cartagena and the Spanish navy in its harbour; the fleet was recovered for the government only by the help of British and German squadrons. The republic had further to contend with the Second Carlist War, with monarchist opposition, and with an insurrection in Cuba (the Ten Years' War; 1868–78).

As it grew ever clearer that only the army could restore authority and order, one major plank in the republican program, the taking of the army out of politics, had perforce to be abandoned, and Gen. Manuel Pavía y Rodríguez de Alburquerque quelled the risings in Andalusia. Castelar, reintroducing conscription, provided the means of reducing Cartagena and pursuing the war against the Carlists with vigour. When the *Cortes* reassembled in January 1874 the extreme parties voted against Castelar, whereupon Pavía, then captain general of Madrid, forcibly dissolved the *Cortes*. The federal republic was dead. A unitary regime followed, under General Serrano, until, in December 1874, another army *pronunciamiento* restored the Bourbons. Alfonso XII (*q.v.*; 1874–85), 17 years old and till then at school in England, landed at Barcelona on Jan. 9, 1875. (W. C. An.; R. S. Ll.).

K. The Later Bourbons, 1875–1931

Alfonso XII (1875–85).—National exhaustion and disillusion caused a substantial rallying of the nation around the young king. Pacification, military and political, was his first concern. The Carlists were not finally quelled until the beginning of 1876, when a general amnesty was balanced by the forfeiture by the Basque provinces of their differential system of taxation and military service. Part of the large army thus released was then sent to Cuba, where after a further 18 months of fighting the insurgents accepted the Convention of El Zanjón (Feb. 12, 1878). Arsenio Martínez Campos, Spanish commander in chief, was made governor-general and offered the island the prospect of reforms, including a more liberal tariff and eventual self-government.

Alfonso's first premier, the conservative Antonio Cánovas del Castillo (*q.v.*), ruled at first dictatorially without the assistance of Parliament. Royal decrees set aside most of the legislation and reforms of the previous regime, respecting only—for a while—universal suffrage, on which basis a new *Cortes* was convoked in 1876. The constitution of that year was a compromise between that of 1845 and the more democratic principles of the charter of 1869; it was to remain in force for 47 years. By it the franchise was limited to taxpayers; the Senate became in part elective; religious dissidence was tolerated though not recognized; and a considerable freedom of association was granted. Constitutional guarantees might, however, be suspended when the security of the state so demanded, a provision that left a tempting reserve of power in the hands of the crown, while the parallelism of civil and military districts facilitated its invocation.

The Liberals under Práxedes Mateo Sagasta formed the Fusionista Party, whose name was indicative of its heterogeneity. But the real opposition came from the Socialist-Republican left. From the founding in 1869 of a Spanish section of the First International, labour organization developed rapidly. The split between Karl Marx and Mikhail Bakunin at a congress of the International in 1872 was reflected in Spain on a regional basis, labour being predominantly Marxist in Castile and Anarchist in Catalonia. Regionalism, likewise at its most active in Catalonia, provided another outlet for political dissatisfaction with Madrid.

Cánovas' first administration lasted till February 1881, when, under suspicion of heading the country once again toward absolutism, he gave way to Sagasta and the Liberals. Thereafter, to the end of the century, the two parties alternated in power on an agreed rotatory system based on "made" elections that soon became one factor in the gradual discrediting of the restored monarchy. In August 1883 an attempted republican rising, though soon quelled, led to the discovery of widespread revolutionary activities throughout the country. Fifty Republicans were returned to the *Cortes* at the general election of March 1893. In the country at large the last two decades of the 19th century were scarred by an increasing resort to direct action and crimes of violence, with corresponding waves of harsh official repression.

Alfonso's first wife, his cousin Mercedes, daughter of the duc de Montpensier, had died in May 1878, five months after the wedding. His second marriage (1879) was with María Cristina II (*q.v.*), a second cousin of the Austrian emperor Francis Joseph. On Nov. 25, 1885, Alfonso died, leaving two daughters. Six months later on May 17, 1886, a posthumous son, Alfonso XIII (*q.v.*; reigned 1886–1931), was born. The queen was proclaimed regent.

Regency for Alfonso XIII.—The regency began with five years of Liberal rule, marked by the civil code of 1889, another major step in the consolidation of the country's still heterogeneous legislation, by the restoration of universal suffrage and of trial by jury, and by a liberalizing of the laws affecting the right of association and the liberty of the press. Attempts to curb the political proclivities of the Army were less successful and added military to Conservative opposition. The elections of 1890, which returned the Conservatives to office, showed that very few of those lately enfranchised had voted; they showed also that the home office's control of the voting continued as before.

Cánovas now based his economic policy on protection, completely reversing the moderate free-trade policy which had been so beneficial to Spanish foreign trade since 1868. Heavy increases in duties on agricultural imports, particularly on breadstuffs, and the denunciation of all treaties of commerce that contained most-favoured-nation clauses soon made their adverse effects felt and evoked loud protest. Industrial interests alone benefited, and imports of raw materials, notably chemicals, coal, and coke, increased. In 1892 Cánovas admitted that never since the restoration in 1875 had the budget been balanced and that the exchequer was bankrupt. Political bankruptcy could be seen further in the failure of both parties to tackle the problem of the *latifundios* and a land-starved peasantry. Signs of labour unrest multiplied, chiefly in Catalonia.

Morocco and Cuba.—The Rifi tribes bordering the Spanish zone in Morocco had been a continual source of trouble. In 1894 an incursion into Spanish territory resulted in defeat for the Spanish forces and a severe shock to public opinion at home. An army of 25,000 men, with 30 generals, was sent to Melilla, but hostilities were averted for the time being by a treaty signed in Fès and by the sultan's agreement to pay an indemnity of 20,000,000 pesetas. Earlier, between 1884 and 1886, Spain had established dominion

over the vast territory of Río de Oro, south of Morocco, in furtherance of a new African policy of pacific penetration advocated by the influential jurist and economist Joaquín Costa.

The position in Cuba had again grown threatening, and press censorship left Spanish opinion unprepared for its outcome. Cuban autonomist deputies in the *Cortes* had formally tabled the demand for home rule in 1878, but this was rejected, as was a Conservative measure of 1893 for political and administrative reforms tending in that direction. In 1895 a much weaker Cuban Reform Bill was passed by all parties, but before it could be promulgated a separatist rising had broken out near Santiago de Cuba, backed by a strong body of sympathy in the United States. Sagasta dispatched a preliminary force of 12,000 men to reinforce the 15,000 Spanish troops already in the island, and Cánovas, succeeding once more to the premiership (March 1895), insisted that submission must precede any concessions. Gen. Valeriano Weyler y Nicolau (*q.v.*) applied himself to the stamping out of disaffection with a ruthlessness which brought repeated diplomatic representations from the United States, which at the same time was supplying the rebels with arms.

Realizing the danger of more direct U.S. intervention, Cánovas presented to the *Cortes* a bill granting Cuba a substantially enlarged autonomy. His assassination by an anarchist on Aug. 8, 1897, brought Sagasta back into office. Recalling Weyler from Cuba, Sagasta hastily enacted conciliatory legislation and sought to propitiate the United States, even offering a treaty of commerce which would have allowed the U.S. every advantage in trade with the West Indies possessed by Spain itself. But things had already gone too far, both in Cuba and in the United States, where public opinion pressed irresistibly on Congress and president alike.

Spanish-American War.—On Feb. 15, 1898, in circumstances never clearly established, the U.S. battleship "Maine" blew up in the port of Havana, and events moved quickly to a crisis. Spain appealed to courts and governments and to the pope for mediation, but none was disposed to go beyond mild representations to Washington. In April Pres. William McKinley approved a resolution demanding the cessation of hostilities in Cuba with a view to evacuation by the Spanish forces. Sagasta gave the U.S. minister in Madrid his passports and severed relations with the United States; the known inadequacy of Spain's resources for a struggle could not prevent it from yielding to the urge to defend the last remnants of its colonial empire.

The sequel is soon told. The Spanish fleet in the Far East was defeated in Manila Bay (May), the Caribbean fleet was destroyed off Santiago de Cuba (July), and all communication between Spain and the colonies was severed. Within a fortnight of the landing of a U.S. force near Santiago the Spanish garrison surrendered (July 17; *see* SPANISH-AMERICAN WAR). At the end of July Spain sued for peace. The agreement of Aug. 9 stipulated clearly that Spanish rule in the new world must be considered at an end; and by the Treaty of Paris (Dec. 10, 1898) Spain renounced unconditionally all sovereign rights over Cuba and Puerto Rico and ceded to the United States, in consideration of a payment of $20,000,000, the Philippine Islands, the Sulus, and the largest of the Marianas. The Carolines and a few other Pacific islands were sold to Germany in 1899, and two islands omitted in the Treaty of Paris were purchased by the United States in 1900. Apart from holdings in North and West Africa, the Spanish Empire was at an end.

The Reaction in Spain.—The Liberal government was held responsible for the debacle; but the seeds had been sown long before, and the change to a Conservative cabinet was the least significant among its consequences. From the mood of sober analysis and heart-searching that ensued there emerged the "generation of '98," a group of thinkers who through their chosen medium—the essay—effected a renovation not merely of literature but of the country's intellectual life (*see* SPANISH LITERATURE). The debate on the country's role in history led to a cleavage between *europeizantes* and *africanizantes;* it led also to a new interest in Africa as the one remaining field of overseas activity. With the radical reduction of its commitments the Army became top-heavy, at once an intolerable drain on the nation's economy and more prone than ever to meddling in politics. The wave of solidarity with Spain that swept through Latin America went far to neutralize remaining antagonisms in this sphere.

Long-due reforms in taxation and the servicing of the nation's debts were carried through by a vigorous Conservative minister of finance, Raimundo Fernández de Villaverde. Bondholders were called on for even heavier sacrifices than taxpayers, and the interest on colonial debts, like that on internal debts and on Spanish holdings of external debts, was made payable in pesetas. But in the main the political scene remained untouched by the new spirit. Sagasta, premier once more in March 1901, was confronted by two major problems. One was the growth of the religious orders and congregations, in defiance of the Law of Associations of 1887 by which all congregations were required to register their members and, when not specifically recognized under the concordat of 1851, to apply for authorization. A proposal now to enforce the law greatly incensed the church. The papal nuncio announced that Rome would only consent to discuss the question on condition that all requests for authorization would be granted. To avoid a crisis just when the young king was about to become of age, Sagasta accepted a *modus vivendi.* A final settlement was not reached with the Vatican until 1912.

Industrial unrest constituted a more pressing peril. Fomented by socialist agitation and by anarchist direct action, it led in January 1902 to serious riots in Barcelona and Saragossa and on Feb. 16 to a general strike in the former city. General Weyler's methods in restoring order all but brought the government down. They provided no solution to anarchism, which under the influence of Georges Sorel was now moving toward syndicalism as a basis for organized opposition to all government as such.

Personal Government of Alfonso XIII.—Alfonso XIII was enthroned on his 16th birthday, May 17, 1902. He was still then under clerical and reactionary influences, and his contemptuous treatment of ministers showed from the beginning an intention of intervening actively in politics. His first official act was to countermand a decree closing the military colleges, aimed at reducing the disproportionate officer strength of the Army. Sagasta died in 1903. Conservatives and Liberals alike were now suffering from weakened leadership, and there ensued a period of instability and frequent changes of government.

A tendency to disintegration in the political parties was symptomatic of the growing subordination of national to sectional interests. It became an important factor in the eventual collapse of representative government. This political instability contributed to the general unrest; and the old antagonisms, born of the conflict between the industrial north and the agricultural south and of the now traditional cleavages over regional, social, and religious issues, were increasingly complicated by labour agitation. The notorious Law of Jurisdictions of 1906, under which attacks on officers and military institutions were to be tried before military tribunals, was a serious capitulation of the civil authority, that had far-reaching consequences.

Alfonso's marriage on May 31, 1906, to Victoria Eugénie of Battenberg, princess of Great Britain, marked a stage in his emancipation from clerical-conservative tutelage. Ecclesiastical controversy again loomed large at this period, formal recognition of the validity of civil marriage arousing such agitation among the clergy that the government threatened the bishop of Tuy and the Córdoba chapter with prosecution. A new Law of Associations in 1906 brought the Liberal Party itself to the verge of disruption, and after three changes of premiers the Conservatives returned to power in January 1907 under Antonio Maura y Montaner (*q.v.*).

Maura (1907–09).—Maura, who had already been premier (1903–04), had many qualities of leadership. A sincere Catholic, though no clerical, he pursued first the maintenance of order and the reform of local government. A bill to the latter end was submitted to the *Cortes* in October 1907. Aimed at destroying the power of the *caciques* (local bosses) and at educating the electorate in their privileges and responsibilities, it made voting compulsory, abolished official interference with the polls, and increased considerably the responsibility of local elected bodies. The bill met with strong Liberal opposition because of its alleged re-

gionalistic tendencies and, though finally approved by the lower house in February 1909, did not reach the statute book.

The elections of 1907 were notable for the return of a strong republican minority and the striking victory of the Solidaridad Catalana, a fusion, resulting directly from the opposition to the Law of Jurisdictions, of seven Catalan parties in defense of regionalist interests. Of the 44 seats allotted to Catalonia, 41 were gained by its candidates. Maura, ruling rather as an enlightened despot than as the head of a constitutional government, was concerned also to strengthen the Navy, to reform and to reconstruct the police force, to institute industrial tribunals, and to regulate the conditions of work of women and children.

Trouble long brewing in Morocco broke out in July 1909 with an attack by Rifi tribesmen on Spanish railway workers. The need to reinforce the Spanish troops in Africa had for some time been apparent, but Maura had hesitated to lay the necessary estimates before the *Cortes*. The proposal now of the war minister, Gen. Arsenio Linares, to organize a new field force by calling out the reserves, with its vivid memories of the useless miseries of earlier overseas expeditions, provoked another general strike in Barcelona; and a movement at first directed against "conscription" rapidly developed into a revolutionary attack on the established order in church and state that spread to the whole province. Churches and convents were favourite targets for violence. In Barcelona the rising was suppressed after three days of street fighting (July 27–29).

On July 28 martial law was proclaimed throughout Spain, and a military reign of terror continued until the end of September. In the fortress of Montjuich in Barcelona were imprisoned not only rioters but many others, notably journalists, critical of the government. A sensation was caused by the arrest of Francisco Ferrer, a theoretical anarchist well known in many countries for his anticlerical educational work and especially as the founder of lay schools in Spain. Ferrer was tried by court-martial and shot (October). The execution aroused widespread indignation throughout Europe and produced a ministerial crisis in Madrid. The government, already weakened by a reverse in Morocco (Sept. 30), was furiously attacked by the Liberals under Segismundo Moret y Prendergast at the opening of the *Cortes* in October, and Maura resigned.

Moret and Canalejas (1909–12).—Moret, who now became premier, sent General Weyler back to Barcelona to keep order, released most of the Montjuich prisoners, reduced the forces in Morocco, and reopened negotiations with Rome for a modification of the concordat. In February 1910 Moret was forced to resign through the king's refusal to grant him a dissolution, and José Canalejas y Méndez (*q.v.*), the outstanding Liberal of his day, took office. His government, "inspired by the universal spirit of liberty of conscience," tackled the continuing refusal of the religious houses to comply with the law (in respect of registration and the payment of taxes when engaged in industry and commerce) and announced measures both to restrict the number of religious houses and to implement the religious toleration enshrined in the constitution. A violent Catholic agitation resulted, 72 Spanish bishops and archbishops presenting a formal protest to the government. To attacks from the Vatican Canalejas retorted with the famous "Padlock Law" at the end of 1910 forbidding the settling of further congregations in Spain until a revision of the concordat should have been agreed on with Rome.

On the Vatican's demanding that all the obnoxious measures be rescinded as a preliminary to negotiations, the Spanish ambassador to the Vatican was recalled. Hints were dropped in papal circles of a possible *rapprochement* with the Carlist claimant to the throne, and diplomatic relations were not restored until 1912. The overthrow of the monarchy in Portugal on Oct. 5, 1910, and the expulsion thence by the new republican regime of all religious congregations did not ease the government's difficulties.

French activities in Morocco at this time compelled Spain to adopt a more active policy in its zone. To counterbalance the French entry into Fès (1910) a Spanish force was landed at Larache in June 1911 and took Alcazarquivir and Arcila. The recruiting of native regiments under Spanish officers did something to lessen popular indignation at home. France was prevented from reacting strongly by the crisis resulting from the appearance off Agadir in July 1911 of a German gunboat to "protect German interests," but tension with Spain continued and pervaded the negotiations for a Franco-Spanish agreement. Concluded on Nov. 27, 1912, the treaty reduced the Spanish zone for the third time in 12 years.

A general railway strike in Spain in the middle of 1912 threatened the country with chaos. Canalejas met it by drafting railway workers into military service and compelling them to continue at their posts as soldiers. Such firmness was balanced by a genuine readiness to understand legitimate grievances, which led the Catalanists to make a serious attempt to come to terms. The premier responded by introducing a bill of *mancomunidades* (commonwealths) whereby the four Catalan provincial councils might combine in one joint body (*mancomunidad*) to which the state would grant a substantial measure of devolution. The bill, which was inevitably unacceptable to many Liberals and Conservatives alike, and was regarded in Catalonia as merely a first installment of Catalan demands, had not passed the *Cortes* when, on Nov. 12, 1912, Canalejas too was assassinated by an anarchist.

The Parties in Decline.—The new premier, Alvaro de Figueroa y Torres, conde de Romanones, was natural heir to the Liberal leadership, but he was unable to maintain the firm party discipline of his predecessor. He pursued meantime Canalejas' policy of seeking to attract and absorb the more moderate wing of the republican left. Alfonso himself recognized the need for such an accommodation, and the response of Republican leaders like Melquíades Alvarez led to hopes that the disintegration of the regime might yet be checked.

These hopes were dashed, first by an internal crisis in the Liberal Party over the bill of *mancomunidades*, then by the "implacable hostility" proclaimed by Maura against any policy of collaboration with the antidynastic parties. With the resignation of Romanones in October 1913 the king realized that the Liberal policy which he had consistently followed since 1909 was at an end. Maura's views aroused, however, such intense popular resentment that Alfonso entrusted responsibility instead to his fellow-Conservative Eduardo Dato Iradier (*q.v.*), a choice that initiated the disintegration of the Conservative Party in its turn.

The tangled fate of the bill of *mancomunidades* was resolved, against Dato's judgment, by a royal decree of Dec. 18, 1913, that granted to the four Catalan provinces the right to group together for "exclusively administrative purposes within their competence." This and several similar royal decisions in military matters, which savoured of unconstitutional action, created considerable opposition to the government. But internal difficulties were soon to be merged in the all-absorbing preoccupations and repercussions of World War I.

World War I.—The international situation of Spain at the outbreak of war in 1914 was determined officially by the agreement between Spain, France, and England signed at Cartagena in 1907 and confirmed in conversations held there in 1913 between the French president Raymond Poincaré and Romanones, which stipulated that, "should new circumstances arise tending to alter the territorial *status quo*" in the Mediterranean or on the Atlantic coasts of Europe and Africa, the three powers would "enter into communication" in order to take any necessary measures.

The war was obviously such a circumstance. France and England made no move, but Dato promptly declared for absolute neutrality, Spain having no stake in the conflict. As for the nation, its sympathies were divided. The working classes, most of the intellectuals, and the trading communities—the Liberal left—were pro-Ally. The clergy, most of the Army and the bureaucracy, and the "idle rich"—the Conservative right—were pro-German. But there was agreement on the one essential of avoiding war, and the *Cortes*, resuming its sittings on Oct. 30, 1914, unanimously endorsed the government's decision.

A debate on military reforms brought Dato's government down in December 1915, and Romanones again formed a Liberal ministry. By then the country was profiting from industrial activity on the Allied account, the peseta appreciated, and the foundations

were laid for the repatriation of the foreign debt. This favourable picture had its reverse side, however. In Morocco military operations were again necessary, this time to defend the mines from hostile tribes, while Spanish shipping was victimized by a ruthless German submarine campaign. By September 1916 Spain had lost more than 80,000 tons, 30,000 by torpedo attack. When, on Jan. 31, 1917, Germany intimated an "absolute blockade" of the Allied coasts, Romanones delivered a spirited answer intended partly to test public opinion.

His resignation on April 19, which cost him the leadership of his party, was not solely the result of the nation's failure to rise to his lead. Toward the middle of 1916 the Army had begun to organize secretly its so-called *juntas de defensa* in furtherance first of its professional efficiency, second—and soon chiefly—of its traditional prerogatives, the threat to which came principally from the king's fondness for personal intervention. Before long they were presuming to dictate to the government in a manner which led Romanones to attempt their dissolution. But this new menace to authority from the one bulwark of order loomed larger with each endeavour to exorcize it, and Romanones chose to resign. Under his successor, Manuel García Prieto, an order to arrest the leaders of the juntas provoked an ultimatum (June 1, 1917) that brought the new government down; and Dato, recalled to office, capitulated before them, conceding to their regulations the force of law. This victory of a movement revolutionary in essence encouraged at first, in a public not alive to its implications, hopes of a renovation of political life. Coupled with early reports of the Russian Revolution, it stimulated the ferment at work in the country, and political manifestoes asking for a "renovation" in government and constitution followed one another (Socialists June 12, Catalanists June 16, left coalition June 16).

In the teeth of government opposition, an assembly of senators and deputies convoked in Barcelona formally demanded constitutional reform and set up three commissions to report to a second meeting in Madrid. This movement, which might conceivably have provided the renovation so urgently needed, was defeated by the resort to direct action of extreme labour elements. On Aug. 13, three days before the planned second meeting, a revolutionary general strike broke out, led this time by the Socialist General Union of Labour (U.G.T.) of Madrid and aimed openly at establishing a socialist democratic republic, which soon threatened to paralyze the whole country. A state of war was again declared. The Army now showed its true colours and put down the strike by machine guns, and the juntas remained the only real force left in the country.

Among the victims was the constitution. The new minister for war imposed on the king by the juntas, Juan de la Cierva, dominated a heterogeneous ministry of Maurists, Liberals, and Catalanists. The shortness of its life (November 1917–February 1918), in spite of Army support, showed that a severe blow had been struck not merely at party government but at cabinet government. Brought down by a telegraph strike, it was followed by a hiatus that only ended when Alfonso, at a midnight meeting of former premiers at the palace on March 21, constrained them under threat of abdication to form a ministry of all the talents, under Maura. This lasted till Nov. 6. The end of World War I in the same month had a natural accompaniment in the return of the Liberals to power, first under García Prieto, then (Dec. 3) under Romanones, whose first act as premier was to visit Pres. Woodrow Wilson in Paris.

The Postwar Situation.—Spared the war, Spain had passed nonetheless through a very disturbed and anxious four years. The submarine war cost it 65 vessels, in all 140,000 tons, which Germany had agreed to indemnify on a ton-for-ton basis. Against this, the influx of capital went far to revolutionize its economy. Gold reserves in the Bank of Spain, £22,680,000 in 1914, had risen to £88,920,000 in 1918. Not merely was Spain able to resorb in large part its foreign industrial debt and almost all of its national debt and to acquire the ownership of the railways, but also new industries had been founded, not all of which disappeared with the return of peace, and in older industries (such as the Basque iron and steel works) plant and conditions were modernized. A mingling of two political currents, one a democratic and constitutional agitation born of the Allied victory, the other a revolutionary agitation akin in the civil sphere to that of the *juntas de defensa* in the military, created an involved situation in Barcelona which resulted in the spring of 1919 in open conflict between the military and the civil authorities. When the former expelled the latter from the town, the failure to take up so flagrant a challenge brought the government down, and again there ensued a period of cabinet instability. Its most notable feature was that "made" elections, the mainspring of the old system, were failing now to produce the desired majorities. On March 8, 1921, Dato, who had been engaged in trying to rebuild Conservative unity, was assassinated.

The murder accelerated the final disintegration of the rotatory system which, ever since the death of Alfonso XII, had served the monarchy as substitute for a real constitutional regime on the English model. To its own internal weaknesses was added a triple assault from outside. There were the antidynastic elements, once grouped together under the common label of Republicans, now moving toward socialism on the one hand and toward different varieties of syndicalist or anarchist revolution on the other hand; there were the Catalan home rulers, gradually evolving from autonomy toward separatism under the stimulus of the respect shown by the conference at Versailles for the principle of self-determination; and there was what might be called the military syndicalism embodied in the defense juntas.

With the last-named, governments lacked the moral courage to deal; toward the violence so frequently associated with the first they opposed more violence, failing to recognize legitimate grievances where they existed and tending always to identify the employers' interests with those of authority. To Catalanism, opposition in Madrid was instinctive, and the failure of a new approach from Barcelona in 1918–19 did much both to strengthen and to unify support for the movement throughout Catalonia. By 1923, when two strong groups, Estat Català and Acció Catalana, of the right and left respectively, were in agreement on the main objective and had concluded a triple alliance with Basque and Galician autonomists, it was clear that this issue was destined to loom larger and larger in the country's politics.

War in Morocco.—Effective occupation and pacification of the Spanish zone in Morocco were still far from complete, and in guerrilla warfare the tribesmen were often more than a match for Spanish troops, with whom service in Morocco had become notoriously unpopular. In 1921 Manuel Fernández Silvestre, in command at Melilla, embarked impulsively, with Alfonso's personal backing, on an operation which ended in crushing disaster at Anual (July 21). To heavy losses in men and matériel was added a forced withdrawal from the whole eastern zone, the fruit of 12 years of slow penetration. General Fernández committed suicide.

The repercussions in Spain were profound. The government fell, and Maura was called on in August to face the emergency. An army of 140,000 men was sent to the rescue of the troops in Morocco. To the growing public demand for an inquiry the new government had at length to give way, and General Picasso was charged with investigating the responsibility for the disaster, his report being submitted in due course to a committee of the *Cortes*. Meanwhile the *juntas de defensa* had reacted strongly with a campaign of threats and accusations against the *Cortes*, which they sought to hold responsible for the general disorganization underlying the Moroccan crisis. The government retorted with a formal disbandment of the juntas, replacing them by *juntas informativas* subject, in theory at least, to the minister for war. But the juntas were the stronger, and in March 1922 Maura was forced from office. His successor, another Conservative, José Sánchez Guerra, was brought down in December by the debate in the *Cortes* on the Picasso report. The insistence in Socialist sectors of opinion that definite responsibilities attached to specified individuals and should be brought home to them charged the issue with high explosive.

A Liberal coalition followed and was confirmed in office at the general election of April 1923. The sensational feature of this was the return of Socialists for five out of the seven seats in Madrid, a success attributed by their party to the strong line that it had

taken on the war in Morocco. The burden which the Moroccan problem imposed on the exchequer (an average of 358,000,000 pesetas annually in 1919–23 as compared with 75,000,000 pesetas for 1909–13 and 146,000,000 pesetas for 1913–19) was aggravated by the gradual wilting of the artificial prosperity of the war years and recurrent industrial troubles at home. For the most part, however, these problems were but facets of the basic inability of government after government to uphold the prestige of authority. In the 21 years since Alfonso's accession to the throne there had been 33 ministries.

Dictatorship of Primo de Rivera (1923–30).—In September 1923 the report of the *Cortes* commission on the Anual disaster was ready. Several politicians were believed to be deeply involved, some as having abetted the king in his fondness for unconstitutional initiative. The reaction of the Army, notoriously restive under parliamentary or civilian criticism, was awaited with some concern. Suddenly, on Sept. 13, a dramatic stroke from this quarter brought to an end the system under which the country had lived since 1875. The captain general of Catalonia, Miguel Primo de Rivera (*q.v.*), marqués de Estella, rebelled against the government from his headquarters in Barcelona, rejecting its authority and threatening its members with jail. His manifesto declared it imperative to liberate the country from "the professional politicians, the men who for one reason or another are responsible for the period of misfortune and corruption which began in 1898 and threatens to bring Spain to a tragic and dishonourable end." "The wide net of greedy politics," he added, "has caught and imprisoned in its meshes the royal will itself." He announced the setting-up of a military directorate which would find for the Moroccan problem a "quick, dignified, and sensible" solution and which would bring home responsibility promptly and justly by the formation of tribunals "of recognized moral authority."

The premier proposed to the king strong action against the rebels, who included now the military authorities of Bilbao and Saragossa and a council of high officers in Madrid. The king asked for time for reflection, and the government resigned. Two hours later a deputation of generals was received at the palace and the king requested Primo de Rivera, who was still in Barcelona, to take office. Administratively the new directorate struck an original note in that the only minister was Primo de Rivera himself, ministerial departments being left in charge of permanent officials while the "directors" concerned themselves with specific problems as they arose. The one significant exception was the home office, which was entrusted to Gen. Severiano Martínez Anido with the rank of undersecretary of state. Public opinion received the *pronunciamiento* without demonstration, save for a protest from the Socialist Party. The directorate declared the country to be in a state of war. A strict press censorship and a ban on political manifestations were imposed. The *Cortes* was dismissed, and the seizure of the files of the Commission on Responsibilities, shelving the further inquiry into the Moroccan disaster which was due to begin on Sept. 15, made clear the reason for the timing of the coup. When, two months later, the king refused to acknowledge his obligation under the constitution to convoke a new *Cortes* within three months of a dissolution, he made equally clear his personal responsibility in the collapse of the parliamentary regime and rendered it inevitable that the eventual fall of the dictatorship should involve that of the throne. In November king and dictator paid an official visit to Italy and expressed open admiration of Fascism.

The compelling problem was still Morocco, and it was there that the dictator won his greatest success and the nation's gratitude. The first phase, initiated in the summer of 1924, was one of cutting losses and retreating to a base line near the coast sufficient to cover Ceuta and Melilla. Primo de Rivera claimed later that in this he was interpreting the popular will, "which was weary and sceptical of any other possible solution." He himself led the operation in the field. It evoked opposition from the Army and was not effected without heavy loss, but by the end of the year it was completed.

The sudden attack of Abd-el-Krim (*q.v.*) on the French protectorate in 1925, which led to combined action by France and Spain against the rebellious tribes, made possible a complete reversal of policy. The brilliant landing (Sept. 8) in Alhucemas Bay, again led by Primo de Rivera in person—the other generals having refused—was a thrust at the heart of the enemy's power, and within a month Abd-el-Krim was in flight. The pacification of the entire Spanish zone was not completed until 1927, but already Spain's position and prestige in Africa had been retrieved.

Primo de Rivera was now tempted to consider his mission at an end and withdraw, but there was substantial popular support for the view that the dictatorship could contribute no less to the solving of the nation's internal difficulties. He decided, however, that the military directorate had by now served its purpose and, replacing this in December 1925 by a predominantly civilian cabinet, sought to base his regime on more normal political foundations. The Unión Patriótica (1926) was an attempt to form a nonparty organization of all such good citizens as would accept his own motto of "Country, Religion, Monarchy"; the "supreme national assembly" (September 1927) represented a pseudo *Cortes* largely nominated by the dictator and given a merely advisory function.

Primo de Rivera remained the fount of power and, in his zeal for reform, was driven inevitably to abuse that power, both by attempting to force the pace and by riding roughshod over dissident susceptibilities. His successes lay accordingly in such spheres as public works, finance, and public order. In the conciliation of opposition elements he failed to make headway. The Confederación del Ebro, associating the many interests concerned in the utilization of that river, was a notable innovation and the forerunner of others of the same nature. Ambitious schemes for the electrification of the railways and for road development were set in motion. Income tax was introduced in December 1926 and gradually worked its way into the national finances. Morocco had ceased to be an intolerable drain on the exchequer, and in 1927, for the first time in 20 years, no loan was raised. Economic policy was directed to the correction of the heavy adverse balance of trade by stimulating home industries, canalizing investment, and creating state monopolies. In the labour field an elaborate system of corporations was initiated; deriving from the Italian model, it won the approval of many Socialist leaders. An agrarian credit bank gave promise of new life to the small farmer; the other great problem in agriculture, the breaking-up of the *latifundios* (large estates), the regime left untouched.

In 1925 the government undertook the reorganization of provincial and municipal administration. The right to levy certain rates was delegated to municipalities, with a corresponding widening of their range of action; and female and proportional suffrage were instituted in local elections. One important provision in the decree of provincial reform was the abolition of the *mancomunidad*, the central body in Barcelona responsible for coordinating the work of Catalan local councils. In its Catalan policy the directorate had soon reversed the sympathetic attitude to autonomy voiced in Primo de Rivera's first utterances, which were evidently directed to winning support in Catalanist circles for his *coup d'état*. Army opposition to any measures of Catalan devolution combined with the inherent logic of dictatorship to create an attitude and a policy described in Catalonia as one of systematic persecution.

In external relations interest centred chiefly on the League of Nations. In 1926 Spain claimed a permanent seat on the council of the League in fulfillment of repeated promises by the French and British governments. Similar claims by Brazil, Poland, China, and Persia and opposition from Germany, Sweden, and others to any increase in the number of permanent seats led to a confused situation and to much resentment in Spain. At the June meeting of the council Spain's representative was ominously absent, and on Aug. 9 Spain signed a treaty with Italy stipulating mutual neutrality in the event of unprovoked aggression. Further to stress this independent mood, Primo de Rivera announced in August his desire for such a drastic change in the statute of Tangier (1923), regulating the status of Tangier, as would have amounted to the setting up in Tangier of a Spanish protectorate. The assembly and the council of the League decided in December to institute a special semipermanent seat on the council for Spain. Spain refused this and gave notice of withdrawal from the League. Nego-

tiations with France over Tangier led finally to a number of French concessions (1928) which improved the security position in the Spanish zone of Morocco. This was a factor when, on the invitation of the president of the League council in March 1928 to reconsider its attitude, Spain announced its readiness to remain a member of the League and to accept such status on the council as the League might determine.

The year 1926 witnessed three attempts to overthrow the dictatorship by force. A military rebellion at Valencia in June, led by old General Weyler and Gen. Joaquín Aguilera, was nipped in the bud. Trouble with the artillery corps arose from a measure in July abolishing its cherished tradition of accepting promotion by seniority only as a guarantee against favouritism. A compromise settlement was, the corps claimed, broken by Primo de Rivera, and in September the artillery officers too prepared a revolt. Its failure was followed by the abolition of the corps. An amnesty later restored all but a few to their posts, but much ill-feeling remained. In November Col. Francesc Macià, exiled leader of the separatist Estat Català party, launched a futile attempt to invade Catalonia from France and set up a provisional government there for the restoration of "the national sovereignty."

By the end of 1928 civil discontent with the dictatorship had combined with Army unrest to render the dictator's position more difficult. Continual dissensions between dictator and king, riots in the universities, and protests from intellectuals were symptoms of the general unease. The world economic crisis of 1929 was soon reflected in a serious fall in the peseta. In January 1930 the finance minister, José Calvo Sotelo, resigned; and when, shortly after, the captains general of the Army intimated that they no longer stood behind the dictator his fall became inevitable. On Jan. 28, 1930, Primo de Rivera resigned and retired to France. He was already broken in health and died in Paris on March 16. (W. C. An.; X.)

Fall of the Monarchy.—A temporary government under Gen. Dámaso Berenguer now sought to prepare the way for a return to constitutional rule. Unrest, however, continued. Revolutionary strikes multiplied all over the country. The ranks of republicanism were swollen by all those who refused to absolve the king from his responsibility in the abrogating of the constitution, and a revolutionary committee, pledged to bring about a republic, concluded with Catalan autonomists the Pact of San Sebastián (Aug. 17) whereby it undertook, in return for their support, to introduce a bill for regional autonomy as soon as it was in power. The peseta, which had dropped from 29 to the pound sterling to 36.50 during 1929, fell to 51 by the autumn of 1930. An abortive military rising took place at Jaca (Dec. 15). On Feb. 14, 1931, the government resigned. A new administration under Adm. Juan Bautista Aznar was charged with convoking, first, municipal elections, and then parliamentary elections for a constituent *Cortes*, but it was understood that the former would be taken as a test on the one great issue and that, if the outcome were held to warrant it, the revolutionary committee would attempt to seize power.

Full results of the municipal elections of April 12, 1931, which went off quietly enough, were never published. The only figures issued showed that 22,150 seats had gone to Monarchists and 5,875 to Republicans. Although large numbers of Monarchists were elected in the country districts, all the provincial capitals showed Republican majorities. In Barcelona the Republicans had a three-to-one majority of votes, in Madrid, Niceto Alcalá Zamora (*q.v.*), leader of the revolutionary committee, headed the poll. Alcalá Zamora took the bold course of demanding Alfonso's abdication and the immediate transfer of powers to his committee. The king, after some parleying, agreed to leave Spain, but without formal abdication, and on the night of April 14 he embarked at Cartagena for France. The revolutionary committee became the provisional government of the Second Republic.

L. The Second Republic, 1931–36

King Alfonso's action, and the fact that the army was no longer behind the monarch, allowed the establishment of the new regime without bloodshed. But the initial optimism did not allow for long the illusion that the nation as a whole had turned to fervent belief in responsible democracy or achieved the competence to exercise it. Right-wing elements—Army, church, landed aristocracy—were prepared to tolerate the republic so long as it respected what they believed to be their prescriptive rights. Left-wing elements—Anarcho-Syndicalists and the more extreme Socialists (Communism was as yet an insignificant force in Spain)—regarded its advent as but a stage to the overthrow of the established order of society. The regional autonomists, fortified by the Pact of San Sebastián, appeared to Madrid to be intent on disrupting the national unity. Moreover, the new government, attempting to steer a middle course through all these dangers, was forced by the quantity of social problems facing it to invade sphere after sphere of vested interests. The dictatorship had been merely the culmination of a long denial to the mass of the people of education in political responsibility. The Republic found itself heir to the consequences, and the challenge to the political stature of the nation was as searching as that to the statesmanship and administrative ability of those now come to authority.

During the Republic there were a series of revolutionary strikes, anticlerical outrages, and incipient revolts against the regime, now of the right, now of the left, now of the regions of the periphery. In May 1931 churches were burned by the mobs in Madrid, Valencia, and in parts of southern Spain. The monarchist newspaper *ABC* was suspended and the primate, Pedro Cardinal Segura y Sáenz, whose anti-Republican activities were notorious, was expelled from Spain. A Communist rebellion was crushed in January 1932, a right-wing rising under Gen. José Sanjurjo the following August. The economic life of the country was seriously jeopardized by its constant subordination to political agitation.

Parliamentary elections held in June 1931 gave the left-wing parties 315 out of the 466 seats in the new constituent *Cortes*. The Monarchists had 1 deputy. The *Cortes* met in July, and on Dec. 9, 1931, the new constitution was promulgated. This defined Spain as "a democratic republic of workers of all classes," with authority "emanating from the people," and no official religion. It bound Spain to respect for the rules of international law and to the renunciation of war as an instrument of national policy. Provision was made for the granting of autonomy to regions which could justify the demand for it, but any federation of autonomous regions was forbidden. Titles of nobility were abolished. Free expression of opinion was guaranteed, and both sexes were given the vote from the age of 23. All property and wealth were declared subordinated to the interests of the national economy; forcible expropriation of property, against adequate indemnity, might be authorized by the *Cortes*. Marriage was to be founded on sex equality, and unions were made dissoluble "as a result of mutual disagreement or on the petition of either party." Primary education would be, not merely compulsory and free, as nominally under the monarchy, but exclusively secular.

Legislative authority was assigned to a unicameral *Cortes*, with four years as the maximum length of each Parliament. The president, to be chosen by an electoral college comprising all deputies and an equal number of *ad hoc* members, was to hold office for six years. His powers included the nomination of the premier and considerable authority over the *Cortes*. Amendment of the constitution required an absolute majority in the *Cortes* and the lapse of at least four years from the date of promulgation.

The most hotly debated clauses related to religion. New measures in this field included the disestablishment of the church, the cessation from December 1933 of state payment of the clergy (a form of compensation for past expropriations agreed upon in the Concordat of 1851), and the submission to a special law of all religious confessions and congregations. Freedom of conscience and of worship was recognized, subject to government sanction for all public manifestations of religion. Religious orders requiring a vow of obedience to an authority foreign to the state were to be dissolved and their property nationalized for educational and charitable purposes. Other orders would likewise be dissolved if considered a peril to the safety of the state. None might hold more property than was necessary to its sustenance, and none could engage in industry, commerce, or education. Finally, it would be "permissible" for the property of all the orders to be nationalized.

These provisions occasioned a first serious political crisis. Manuel Azaña (*q.v.*), who as war minister had been active in reorganizing the Army and reducing its officer strength, violently attacked the religious orders in the *Cortes*, announcing as the government's intention the dissolution in Spain, under the "vow of obedience" clause, of the Society of Jesus (this took effect on Jan. 23, 1932) and the application of the other clauses with the utmost rigour. When, after an all-night sitting, the clauses were approved by a half-empty house (178 votes to 59), the premier, Alcalá Zamora, and the home secretary, Miguel Maura, resigned (Oct. 14, 1931). Azaña thereupon became premier, and in December Alcalá Zamora was elected first president of the republic.

A week after Azaña's accession the surging passions constrained the government to introduce the Law for the Defense of the Republic, under which certain constitutional guarantees might be suspended at the discretion of any minister. By it the home secretary was empowered to cancel meetings, dissolve associations, close clubs, and suppress newspapers. Eleven "acts of aggression" against the republic were named, some in the vaguest terms, and made punishable with fines or exile.

Left-Wing Government (1931–33).—Once the constitution was promulgated, Azaña recast his government, which became predominantly Socialist, because of the withdrawal from it of the Radical element under the leadership of Alejandro Lerroux (*q.v.*). The chief measures of 1932 were the Agrarian Law and the Statute of Catalan Autonomy. The agrarian problem, one of the most pressing inherited by the republic, hinged on the prevalence throughout central and southern Spain of huge estates held by absentee landlords and worked by peasants at starvation wages. The new law (Sept. 15) provided for expropriation at discretion, against payment in government bonds to the value of the estate as assessed for taxation. Its administration was entrusted to a state-subsidized Institute of Agrarian Reform, representative of both owners and landworkers.

The movement for regional autonomy was strong in the Basque provinces (although not successful there until Oct. 1, 1936, after the outbreak of the Civil War); it was weaker among Galicians and Valencians, who were far less united in their claims. In Catalonia its urgency was extreme. The fall of the monarchy had been followed there by the immediate proclamation of a "Catalan state," in a manifesto envisaging a "confederation of Iberian peoples." Appeals from Madrid secured the substitution of the historic and less provocative term *generalidad* of Catalonia and resignation to the status of an autonomous region within a unitary Spain. The necessary plebiscite of the Catalan people (Aug. 2, 1931) gave a vote of 592,961 for the draft Autonomy statute and of 3,276 against. Despite considerable opposition within the *Cortes* and from various interests without, the statute, much revised, was approved on Sept. 9, 1932, by 314 votes to 24. By it the *generalidad* was empowered, among much else, to organize its own police services and be responsible for public order, to execute much state legislation, and to found and maintain its own educational institutions. Catalan and Castilian were to be "co-official" languages. Though bitterly disillusioned by the mutilations suffered by the statute, Catalonia, under its presidents Francesc Macià (*q.v.*) and —on his death in 1933—Lluis Companys, became on the whole the most contented part of Spain and the least affected by strikes and revolts. The Law of Religious Confessions and Congregations, implementing the clauses of the constitution dealing with religious activities, also had a stormy passage, lasting seven months (October 1932–May 1933). The proposal that teaching by the orders was to cease after 1933 made necessary alternative provision for the education of at least 350,000 children—an impossible undertaking. Already in June 1931, in Madrid alone, 50,000 children were without schools. The *Cortes* had authorized the creation of 27,000 new schools, 7,000 to be finished within a year; but efficient teachers could not be thus improvised. Misgivings were widespread and were shared by the president himself, who deferred signing the law until the expiry of the period allowed by the constitution.

During these two years, despite the government's efforts to cope with its many major problems, opinion had been running strongly against the left. Opposition was much exacerbated in January 1933 by the ruthless handling of disturbances in the village of Casas Viejas, near Medina Sidonia. In April partial municipal elections, held in 2,500 rural districts, gave impressive Conservative majorities. In September Azaña resigned; and, after the Radical leader Alejandro Lerroux had failed to form a government, the *Cortes* was dissolved and elections fixed for Nov. 19. In them the new female vote, totalling 6,500,000, could be foreseen as likely to strengthen any swing to the right.

Centre-Right Government (1933–35).—The election results showed how deep was the disillusion with the republican experiment to date. In the new *Cortes* the right held 207 seats, the centre 167, and the left 99. The centre thus held the balance of power, and—with occasional infiltrations from the right—governed for the next two years, generally under Lerroux. Changes of government were numerous (seven in 1935 alone). The most important, in October 1934, admitted three members of the right and precipitated revolutions in Catalonia and Asturias. In Barcelona, on Oct. 6, the *generalidad* again proclaimed the Catalan state of the federal Spanish Republic. The central government reacted strongly. After some street fighting, Companys and his fellow-ministers were arrested, the Autonomy statute was annulled in fact if not formally, and Catalonia was again ruled from Madrid. The extremist revolt in Asturias (Oct. 6–13), which resulted in 1,335 killed and 2,951 wounded, was liquidated with great severity, reprisals being taken all over Spain. It was during this period that the Falange (*q.v.;* Spanish Fascist Party) was founded by José Antonio Primo de Rivera, son of the former dictator. The right-wing leader in the *Cortes* was José María Gil Robles (*q.v.*).

Government policy during this biennium was directed broadly to obstructing previous legislation of the left until, in December 1935, revision of the constitution should become legally possible. Thus the substitution of secular for religious schools was indefinitely postponed, the clergy had part of their state stipends restored to them, the transfer of powers to Catalonia was delayed, and a new Agrarian law treated the large landowners with respectful benevolence. But as the date in question drew near it became evident that this negative policy and the severity of the government's repressive measures were in their turn producing a widespread reaction. Azaña, arrested in 1934 on suspicion after the Catalan revolt, tried, set free, and later rearrested, became a popular hero and the centre of progressive hopes, the "strong man" who would soon return to power and carry to completion the reforms already begun. Some looked on him as a possible dictator of the left. By the end of 1935 the centre-right could hold on to power no longer. The president acceded to a dissolution, and new elections were announced for Feb. 16, 1936.

The Popular Front (February–July 1936).—From the political maneuvering preparatory to the elections Azaña emerged as the leader not only of his own party but also of the group of parties and trade unions (comprising Republicans, Left Republican Union, Socialists, and Trotskyists) that now banded together to ensure electoral victory for the left. He received widespread support from the Syndicalists, Anarchists, and Communists. This Popular Front, so-called after a similar French combination, won by a substantial majority, 256 to the left (including 16 Communists), centre 52, and right 165. Soon, however, disturbances of all sorts began to occur. Officials dispossessed during the preceding two years clamoured for reinstatement; prisons were burst open; bands of peasants seized land and defied the law to eject them; churches, seminaries, monasteries, and the offices and houses of capitalists were set on fire; there was a wave of murders by gunmen. During the first four months of Popular Front rule 113 general and 218 partial strikes took place, while 170 churches, 69 clubs, and the offices of 10 newspapers were fired by incendiaries, who attempted to burn a further 284 buildings, 251 of them churches.

Politically, the event of this period was the deposition, on the charge of unconstitutional dissolution of the *Cortes*, of President Alcalá Zamora. Having tried for so long to steer a middle course, the president found himself under attack by both right and left. Azaña had gone so far as to break off relations with him. Only 5 deputies voted for the president, 238 against. On May 10 Azaña

was elected, almost unanimously, as his successor.

The new premier, Santiago Casares Quiroga, proved wholly unable to cope with the indiscipline of the country. While his government worked out fresh measures of reform, Spanish fascism grew, much as republicanism had grown in 1930–31, as a protest against the virtual usurpation of power by extremist violence. The premier flung down a challenge: "In its relations with fascism, the government is a belligerent." The challenge was taken up, and soon violence and counterviolence had relegated the activities of the *Cortes* to insignificance.

On July 13, in revenge for the murder of a Communist lieutenant in the shock troops, José Calvo Sotelo, a former finance minister of Primo de Rivera, was arrested and murdered by men in officers' uniform. Four days later an Army mutiny broke out in Spanish Morocco, led by Gen. Francisco Franco (*q.v.*), which spread next day to the mainland and within 48 hours had involved the whole country in civil war.

M. The Civil War, 1936–39

Everywhere garrisons revolted. In the south, Seville (under Gen. Gonzalo Queipo de Llano), Córdoba, Granada, and Cádiz supported the rebels; in the north, all Galicia, most of León, and part of Asturias, with Burgos (for three years the Nationalist headquarters), Salamanca, Valladolid, Segovia, Pamplona, and Saragossa. Both in Madrid and in Barcelona the mutinies failed (otherwise there might have been no war), and the government began the struggle with the whole of eastern and most of southern Spain under its control. The Basque provinces illustrated the tug of conflicting loyalties. Biscay (Vizcaya) and Guipúzcoa, setting autonomy above all else, threw in their lot with the regime pledged to secure this for them; Álava sided, as also did Navarre, with the insurgents, who claimed to have risen in defense of religion and traditional values.

The Nationalists, as the insurgents called themselves, representing that theirs was no mere *pronunciamiento* but a *movimiento nacional*, lost several outstanding men early, notably General Sanjurjo and Gen. Emilio Mola, killed in airplane crashes, and Gen. Manuel Goded Llopis, shot after the unsuccessful rising in Barcelona. They had with them, however, the bulk of the Army, the armed police, and the arsenals. The Popular Front had to rely chiefly on untrained volunteers and on such arms, outside the arsenals of Madrid and Barcelona, as the unions had been storing for their own purposes. Much time was necessary to turn improvised militias into an efficient army, and by then the Nationalists had greatly improved their positions. Casares Quiroga had resigned the premiership on the outbreak of war, unwilling to surrender to the clamour of the working class organizations for arms to be distributed to them. Azaña's choice of a moderate, Diego Martínez Barrio, was vetoed by the leaders of the extremist groups, because of the not unfounded suspicion that Barrio was disposed to treat with the rebels. He then turned to a personal friend, José Giral, the minister of the Navy in the previous government, who hastily formed another government of Republicans, with the Socialists still unrepresented. Giral proved unable to resist the pressure from the workers and on July 19 orders were given for arms to be distributed to them. It was by then too late; the rebels had already won many provinces.

By the end of July 1936 foreign intervention had begun. The Nationalists received combatants, technicians, and large supplies of war matériel from Italy and Germany, the fruit of prior consultation with Rome and Berlin. From the U.S.S.R., the All-Union Central Council of Trade Unions made the Republicans a first payment of 12,145,000 rubles; both the U.S.S.R. and Mexico began to send war matériel; extremist volunteers (mainly Communist) from many countries entered Spain and formed the International Brigade. Realizing the danger of international complications, France approached the leading powers, and as a result formal embargoes were placed on the export of war matériel and a nonintervention committee, representing 27 nations, was set up in London. Intervention continued nonetheless throughout the war.

With the capture of Badajoz in the middle of August the Nationalists were enabled to unite their northern with their southern forces; and while the former took Irún and San Sebastián early in September, the latter advanced rapidly upon Madrid. The capital was saved by the detour made to relieve the Nationalist garrison imprisoned in the Alcázar of Toledo (Sept. 27). Thanks to this delay the International Brigade was able to take up a position in strength before the arrival of the Nationalists on Nov. 6; and instead of submitting to a triumphal entry the Madrileños held out through a 28-months' siege.

ROBERT CAPA—MAGNUM

SPANISH CIVIL WAR

A Loyalist infantryman falls, fatally wounded during the 1936 attack on Córdoba

As the threat to the capital developed, the government, now under Francisco Largo Caballero (*q.v.*) and enlarged by the inclusion of both Communists and Anarcho-Syndicalists, moved to Valencia (Nov. 7), leaving Madrid in charge of a defense council under Gen. José Miaja. President Azaña had already gone to Barcelona. There the *generalidad* was allowed to retain nominal power and even to take over services reserved by the Autonomy statute to the state; but real power had passed to the unions, which collectivized industry and commerce (Oct. 28) and were responsible for the ban on public worship that only ended with the final defeat of the Republicans.

Apart from the taking of Málaga (February 1937), little was achieved by the insurgents in the winter of 1936–37, two attempts to cut the roads to Madrid conspicuously failing. In April 1937 a new and successful offensive was begun in the north, with Bilbao as its first objective. The city's "iron ring" of defenses held out for nearly two months, but on June 19 the Basque capital fell and one of Spain's two autonomous regions was again centralized. Santander was taken on Aug. 25, and Gijón, the last town to hold out in the north, on Oct. 21.

On both sides the war was marked, especially in its early stages, by a ruthlessness which astounded the civilized world. Churches were burned or desecrated, and public religious observances forbidden throughout republican Spain; ten bishops and many thousands of priests, religious, and devout members of the laity were murdered in cold blood, for no political activity or crime. The Nationalists for their part organized mass executions (*e.g.*, in Badajoz), and the Basque town of Guernica was subjected to terroristic air bombardment.

On Oct. 1, 1936, General Franco assumed the leadership of Nationalist Spain and outlined the features of his proposed "broadly totalitarian" rule. These included abolition of popular suffrage and of regional autonomy; equitable and graduated taxation; complete religious tolerance and a new concordat with Rome; in foreign affairs, suppression of all "Sovietic contacts" and preferential treatment for "nations of related race, language, or ideology." A technical council of state was set up, to be succeeded in January 1938 by a cabinet on the orthodox pattern. On April 19, 1937, the two main Nationalist groups—Falangists and Traditionalists—were fused into a single party, the Falange Española Tradiciona-

lista (F.E.T.), and all other parties were dissolved. The organization of the F.E.T. was developed four months later by the creation of a national council of about 50 members and a *junta política* or party cabinet invested with considerable powers.

Much constructive legislation was promulgated even during the war, such as the Labour Charter of March 9, 1938, which, under captions like "Work for All," "Remuneration and Security," "Protection and Production," gave the worker an undertaking, somewhat vaguely worded, that the "new state" had not come simply to restore the *ancien régime*. In agrarian policy the objective was to give every labourer a living wage and every peasant family a small holding and to improve sanitary and housing conditions in the villages. Little idea was given as yet of how such aims were to be accomplished. Another feature of the Labour Charter was the creation of vertical syndicates combining into one organism, under state direction, all the elements functioning within a single service or branch of production. This departure earned for the new regime the adjective "national-syndicalist."

The Republican cause suffered from internal differences much more fundamental than were to be found on the other side, and serious clashes occurred. A miniature civil war within the Civil War was caused by an Anarcho-Syndicalist rising that raged for a week in Barcelona (May 3–10, 1937). On this followed a major crisis in the Valencian government, leading to the formation of a new cabinet, under Juan Negrín (*q.v.*), from which the Anarcho-Syndicalists were eliminated; they were not again represented in the government until April 1938. The *generalidad* underwent numerous crises on a smaller scale (six between September 1936 and June 1937 alone), which were due also, principally, to the clash between centralizing and decentralizing groups, though, as the combat became fiercer, these drew more closely together. Large-scale plot and spy trials were frequent, some of them implicating the anomalously placed Trotskyists, whose group, known as the P.O.U.M. (Partido Obrero de Unificación Marxista) was the Cinderella of the Republican family.

By the end of the first year of war, General Franco held 35 of Spain's 50 provincial capitals and 119,690 sq.mi. of territory. But the Republicans had as yet scarcely organized their "new army," and the struggle was far from over. The loss of the north and the consequent unification of front simplified their practical problems; and when, on Dec. 4, 1937, they launched an unexpected offensive in Aragon, it won them the city of Teruel (Jan. 9, 1938). This was to be their last, as it was almost their first, important gain and was lost once more on Feb. 22. With the spring the Nationalists regained the initiative, relieved Huesca, penetrated into Catalonia, captured Lérida (April 3), and pursued the enemy up the Segre Valley to Tremp. An almost simultaneous push farther south carried them to the Mediterranean (April 15), and soon they had carved themselves a corridor 40 mi. wide, threatening Castellón de la Plana to the south and Tortosa to the north. Republican Spain was thus cut in two, and its government, which had removed from Valencia to Barcelona on Oct. 31, 1937, deputed General Miaja, the defender of Madrid, to act as civil and military governor of its central and southern territory.

International Activity.—The efforts made by the London Committee to prevent, or at least to lessen, foreign intervention were many and complicated. Though the entry of foreign combatants was made illegal (Feb. 21, 1937), they continued to enter, despite a coastal control system inaugurated on March 7. All the powers save Great Britain and France soon withdrew from this, and Great Britain proposed its supersession by port control and by the granting to both parties of belligerent rights as soon as the withdrawal of foreign combatants already in Spain should have made substantial progress. In fact General Franco never withdrew more than 10,000 of these during the war, and the Republicans, though more amenable, had not completed withdrawal at its close. Belligerent rights were, therefore, never granted.

In the summer of 1937 intensified pro-Nationalist submarine warfare in the Mediterranean led to a nine-power conference being held at Nyon, Switz. This organized (Sept. 14) a new system of maritime zone patrols and agreed that any submarine attacking a non-Spanish merchant ship should itself be attacked. As the fighting moved more and more toward eastern Spain, attention became increasingly centred on the bombing of Mediterranean ports, which often resulted in great damage to neutral shipping and much loss of life among civilians. Protests were made by several powers and by the Vatican against the repeated bombing of open towns (Granollers suffered 700 casualties in one air raid, and Falset, near Gandesa, was almost completely destroyed).

The idea of peace negotiations, which the powers would have supported, was rejected throughout by General Franco, who insisted on the unconditional surrender which he eventually obtained. Signs that the Republicans were envisaging defeat first came in April 1938, when Negrín formulated 13 points as his irreducible minimum. At the time of the Munich agreement on the Sudetenland (September 1938) these were given great publicity, and there was talk of a more moderate Republican government which might canvass for an armistice. That this was not formed may have been because of a checking of the Nationalists on the Ebro (July 26–Nov. 18, 1938), which gave the Republicans new hopes. These proved vain. Their Army was exhausted, their people half-starved, their territory cumbered by about 3,000,000 refugees from the west. When at length the Nationalists were able to recross the Ebro, even the severe winter could not hold back their final sweep to victory.

The End of the War.—The war had been one of lightning offensives and wearisome delays, and it conformed to type to the end. The last great offensive began on Dec. 23. A general advance in Catalonia soon developed into an attack on Barcelona which, menaced from both northwest and southwest, fell on Jan. 26, 1939, after a campaign of only 34 days. The government fled north to Figueras, where, on Feb. 1, 62 members of the Republican *Cortes* held a meeting in the castle vault and Negrín hurled defiance at the enemy, while announcing that his 13 points were now reduced to 3. Pursued northward along the three main roads to France, the Republican forces soon lost the rest of Catalonia. Hundreds of thousands of refugees poured into France, able-bodied men of military age being at first sent back, though eventually admitted and interned. When Gerona fell on Feb. 4, the government moved to a village near the French frontier. Across this, a day later, passed a succession of presidents: Azaña, Companys, and the Basque president José Antonio de Aguirre.

On Feb. 28 Azaña resigned, Negrín and his cabinet returning to Madrid to continue the struggle, while the Nationalists prepared to move south to attack the capital. With no heating in the city, only two ounces of food daily, and hundreds dying of starvation, attack proved unnecessary. On March 5 all parties except the Communists formed a new council of defense, again under General Miaja, ousting the Negrín government. A revolt against the council by Communist and other Negrín supporters broke out two days later but was suppressed after a week of heavy fighting, and the council addressed itself to preparing for surrender. On March 28, without resistance, 200,000 Nationalist troops marched into the capital. By the following day the rest of Republican Spain had surrendered, and Burgos radio announced that the Civil War was over. It had lasted two years and 254 days and was estimated to have cost 1,000,000 lives ("un millon de muertos"), including the tens of thousands who succumbed to the effects of chaos and privations. The figure for war dead is now thought to be about 600,000.

N. General Franco's Government

Great Britain and France had recognized General Franco's government on Feb. 27; the United States followed suit on April 1. A victory parade in Madrid on May 19 was followed within a few days by the departure from Spain of the German and Italian combatants; and during the summer the various ministries returned to the capital. These few months saw the first stages of reconstruction, to speed which General Franco "adopted" the most severely damaged towns and granted extraordinary credits for house building, the repair of public buildings, and the rehabilitation of the transport system. The Falange, reorganized, outnumbered the Traditionalists in the government, and Ramón Serrano Súñer, appointed minister of the interior and later foreign minister,

emerged as the second strongest figure in the administration. He was brother-in-law to the *caudillo* (leader), as General Franco was now called in his capacity as head of the Falangist Party.

World War II.—With the outbreak of hostilities between Germany and the Allies in September 1939, General Franco was quick to reaffirm Spain's neutrality. Six months earlier, in March, Spain had joined the anti-Comintern bloc; but a much more significant action, as it proved, was the conclusion only four days after this of a treaty of friendship and nonaggression with Portugal. This agreement, reinforced by a protocol of July 1940, seemed calculated to rule out any step capable of involving Spain in war with Great Britain's oldest ally. The Soviet-German Pact (August 1939) came as a great shock to Spanish idealism, and the overrunning of Catholic Poland brought consternation to "the most Catholic country in Europe." Germany's attack on the U.S.S.R. in June 1941 helped to restore the perspective, and Spain dispatched its Blue Division of volunteers to fight by Germany's side in the Ukraine. In a speech of July 18, 1942, General Franco reiterated his stand: "Yesterday, today, and tomorrow, for the countries of Europe there exists only one danger—Communism." Earlier he had promised 1,000,000 volunteers for the defense of Berlin should this ever be threatened by the Russians.

Meanwhile, Spain had used the occasion of Italy's entry into the war to alter the status of Tangier, and on Nov. 24, 1940, the administration of the International Zone was formally combined with that of Spanish Morocco. England refused to recognize this unilateral action and protested against attacks on British property in Tangier; but a *modus vivendi* was reached on Feb. 26, 1941, when Great Britain, while reserving all rights, accepted the new situation meantime. The redefinition, likewise in June 1940, of Spain's own status as one of nonbelligerency and the emphasis laid on its new-found "will to empire" reflected confidence in an impending Axis victory and the desire to share in any redistribution of colonial territories. The presence of German troops at the Pyrenees after the French collapse weighed heavily thereafter on Spanish policy; but although General Franco journeyed to Hendaye, Fr., on Oct. 23, 1940, to meet Hitler and to Bordighera, It., to meet Mussolini four months later (Feb. 12, 1941), he succeeded in avoiding irrevocable commitments. The creation in November 1940 of the Council of Hispanity, its concern being "all activities tending to the unification of culture, economic interests, and power throughout the Hispanic world," was aimed at the recovery of prestige and influence in Spanish America. The atmosphere across the Atlantic, as country after country identified itself with the Allied cause, was, however, unsympathetic in the main toward Spain and its totalitarian leanings, Argentina being the notable exception.

By late 1942 the course of the war was giving General Franco cause for second thoughts. The dismissal of the violently germanophile Serrano Súñer from the foreign ministry in September was one sign that Spain was cautiously reverting from nonbelligerency to neutrality. The Allied landings in North Africa in November, followed by the German occupation of the whole of France, threatened Spain for a moment with involuntary involvement in hostilities, and partial mobilization was decreed. But the danger passed, and in December an Iberian bloc was formally cemented with Portugal. This, described by Gen. Francisco Gómez Jordana y Souza, now foreign minister (it was he who had signed the original nonaggression pact with Portugal in 1939), as "an instrument of peace capable of playing its part in international policy," confirmed the new direction; and with the collapse of Italy in July 1943 Spain began to urge the desirability of a compromise peace, in which it aspired to play the role of mediator. The Allies, however, were still not convinced that Spanish neutrality was genuine and in February 1944 the United States, to encourage its stricter observance, imposed an embargo on oil shipments to Spain. This bore fruit in an agreement of May 2, Spain giving satisfaction on various outstanding Allied grievances concerning exports of wolframite to Germany, German agents in Tangier, Italian ships in Spanish ports, and the Blue Division.

Ostracism by the United Nations.—The ending of the war in Europe in May 1945 and the political landslide to the left that swept the continent, from Great Britain to Greece, as its first consequence confronted the Spanish regime with a position of acute difficulty. Overtures from Spain for a voice in the peace settlement and urging the need for a Western entente against Soviet expansion were disregarded. The Potsdam declaration of July branded Franco's regime as unfit to associate with the United Nations, one-third of whom were Spain's daughter nations in Latin America. In August, at an international conference in Paris on Tangier, the U.S.S.R. but not Spain being represented, Spain was given a month in which to remove its troops from the city. In that same month Mexico, which had never recognized the new Spanish regime, granted hospitality and recognition to a Spanish Republican government in exile. The culmination of this general ostracism was the decision of the United Nations in December 1946, by 34 votes to 6, to demand the replacement of the regime by one "deriving its authority from the governed," coupled with the recommendation that all member nations should recall their ambassadors or ministers from Madrid. Apart from Portugal and Argentina, Spain had few friends left. This danger had already been foreseen. Once the tide of war had clearly turned, the "totalitarian state" had given way to a new term, "organic democracy." A pseudo *Cortes* had been inaugurated (March 1943), and Franco had begun to drop hints of a restoration of the monarchy. In October 1944 a first approach to elections took place, for the renewal of the syndical organization. This last was the residual sphere of influence of the Falangist Party, whose program of national syndicalism had earlier supplied the doctrinal basis for the regime. But the party's entanglement in government had come to be felt as an embarrassment. Party control of press, radio, and propaganda had been transferred in August to the ministry of education. The Falangist salute had been abolished by decree in the following September, and the title of *caudillo* was supposed to recede with the party into the background.

Relations with the Church.—The appointment as foreign minister, in the far-reaching cabinet reshuffle (July 1945), of Alberto Martín Artajo, former president of Catholic Action and not a party member, was likewise aimed at foreign opinion and particularly at closer relations with the Vatican. Repeal of the divorce law, restoration of religious education and of budget provision for the church, and the return to the Jesuits of their property and prerogatives had followed close on victory in the Civil War, but there had been no progress toward the new concordat promised in the program of the "new state." A *modus vivendi* reached in June 1941, to allow the filling of vacant sees, then numbering about 20 out of 55, merely revalidated the first four articles of the concordat of 1851. The church distrusted certain of the totalitarian implications of the regime and by 1945 was demanding (as in a pastoral letter issued on Aug. 28 by the archbishop of Toledo and primate of Spain, Enrique Pla y Deniel, to mark the end of the world war) an end to the "constituent period" and a return to a constitution "in conformity with Spain's historic tradition and the degree of political education of the Spanish people."

The new concordat between the Holy See and Spain, confirming Spanish loyalty to the Church of Rome, was signed only on Aug. 27, 1953. Concessions were made to ecclesiastical objections to some reforms in Spanish schools before the Secondary Education Act passed the *Cortes* at the end of February 1953.

Franco and the Monarchy.—As during the war, however, a restoration of the kingship was still to Franco only the last resort. In March 1945 the pretender Don Juan (1913–), conde de Barcelona, third son of Alfonso XIII (who had designated him as his successor before his death in Rome on Feb. 28, 1941), had called on Franco to resign, offering Spain a new constitution to be determined by popular vote, a legislative assembly, guarantees of political liberty and individual rights, recognition of regional characteristics and an amnesty. Only the mention of an amnesty provoked any response. On April 29 it was announced that all political charges against Republicans were being dropped and exiles were invited to return. In October came a proclamation of full pardon for all prisoners convicted of political offenses committed during the Civil War. Coming six and a half years after the end of hostilities, the detail was eloquent of the animosity with which the

regime had persecuted its political opponents. The execution on Oct. 17, 1940, of Lluis Companys, former president of Catalonia, handed over by the Vichy government, was perhaps the blackest and most impolitic stain on the record.

Political evolution did not signify any abandonment of "the fundamental principles of the state," which were declared to be sacrosanct in a charter of rights, the *Fuero de los Españoles*, submitted to the *Cortes* on July 19, 1945; and it gradually became clear that by a restoration Franco understood not a return to the traditional monarchy but only such a solution for the problem of the succession as would guarantee the continuity of those principles under a guise more acceptable to public opinion. This was made explicit in a Law of Succession published on April 1, 1947, by which Spain constituted itself a kingdom and the successor to Franco was declared to be that person of royal blood with the best right who fulfilled certain conditions, swore to observe the fundamental laws of the regime and was approved by a two-thirds vote of the *Cortes*. Failing such a king, a council of the kingdom, acting jointly with the government, would nominate a regent.

In December 1954 Franco and Don Juan agreed on the 16-year-old Infante Don Juan Carlos (1938–), eldest son of the latter, as presumptive heir to Franco (then aged 62). The step provoked renewed Falangist opposition to the idea of an eventual restoration.

In December 1966 a new Organic Law of the State was approved in a national referendum. By this law Spain was to have a prime minister, chosen by the head of the state from a list of three names proposed by the Council of the Realm. His term of office was to be five years, but he would be removable by the head of the state. The law also provided for the election to the *Cortes* of two deputies elected from each province by heads of families and their wives. The membership of the Council of the Realm was to be increased from 13 to 17, the additional 4 members being elected by the *Cortes*. The new law also proclaimed religious liberty for non-Catholics as an established legal right.

International Relations.—The exclusion of Spain from aid under the Marshall Plan (*q.v.*) and the failure of repeated approaches to the United States for loans were aggravations of the acute economic difficulties that had weighed on the regime ever since the end of the Civil War. A succession of droughts and bad harvests, together with the refusal of foreign credits, gravely impeded the nation's recovery but did not suffice seriously to threaten the regime. Apathy, born of the devastation of the Civil War and the desire at all costs to avoid another, had become a major political factor. The Army was still behind the regime, and the psychological effect of the ostracizing of Spain by the United Nations had been to bind Spaniards together in defiance of outside interference. By late 1950 the withdrawal of ambassadors was seen to have had the opposite effect to that intended, and on Nov. 4, by 38 votes to 10, with 12 abstentions, the United Nations rescinded its resolution of December 1946.

A week later Spain was admitted to full membership of FAO (to UNESCO not until November 1952, and to the United Nations on Dec. 14, 1955). Diplomatic relations were resumed with the United States (December 1950), Great Britain (January 1951), and other powers; and in July 1951 the United States initiated discussions with Spain on the latter's contribution to Western defense. These fructified in the bilateral agreements of Sept. 26, 1953, providing for the construction and joint use of naval bases at Cádiz and Cartagena and air bases near Madrid, Saragossa, and Seville; economic aid to Spain; and aid in the organization and equipping of Spain's defense forces. On its expiry in 1963, the pact was renewed for a further five years, and a four-year economic development plan for 1964–67 between Spain and the United States was signed in the same year. In April 1956 Spain concluded an agreement with Mohammed V of Morocco, by which the Spanish protectorate of Morocco, as laid down by the treaty of November 1912, was incorporated into the new independent state of Morocco, except the enclaves (plazas) of Melilla, Ceuta, Alhucemas, Chafarinas, and Peñón de Vélez. The remaining colonies of Spain now consisted of five small islands off Morocco, Ifni and Spanish Sahara in northwest Africa, and the Fernando Po Islands and Río Muni in West Africa. (W. C. An.; W. H. Ca.)

VI. POPULATION

The population of Roman Spain in the 1st century A.D. was estimated at between 6,000,000 and 7,000,000. By the end of the empire, this had risen by about a third. Modern demographic data began with the 1787 census which gave a total population for Spain and the "adjacent" islands of more than 10,000,000. By the end of the 19th century this had nearly doubled (though it was a relatively smaller increase than that for many of the newly industrialized countries of Western Europe). From 1900 to 1930 the population increased by more than 25%, and despite the ravages of the Civil War (1936–39), the rate of increase in the following decade did not decline. The 1950 census registered a population of 27,976,755 (an increase of 8.1% over 1940), and the 1960 census a total of 30,430,698 (an increase of 8.8% over 1950), with 1,061 women to 1,000 men. The death rate in 1960 was 8.8 per 1,000, compared to 28.9 in 1900, while the corresponding figures for the birth rate were 21.8 and 34.8.

The majority of the rural population is distributed around the pluviose coasts and in the valleys stretching inland from the coast, in a predominantly dispersed pattern. The remainder is concentrated in villages in the more arid central plains. Urban increased considerably after the Civil War, though the proportion of population living in towns of more than 150,000 is still lower in Spain than in most other Western European countries. In 1950 Madrid, Barcelona, Valencia, Seville, Málaga, Saragossa, Bilbao, Murcia, Las Palmas, Córdoba, La Coruña, and Granada (*qq.v.*) had populations exceeding 150,000; by the mid-1960s the number of such towns included Palma (reflecting the rapid rise of Majorca as a tourist centre) and Valladolid. (M. C. S.)

VII. ADMINISTRATION AND SOCIAL CONDITIONS

1. Constitution and Government.—The "new state" promulgated no formal constitution but five fundamental laws of the nation: the Labour Charter (March 1938), the Law of the *Cortes* (July 1942), the *Fuero de los Españoles* (a charter of rights; July 1945), the National Referendum Law (October 1945), and the Law of Succession (approved by the *Cortes* June 1947). To these was added the Organic Law of the State (November 1966; approved by national referendum, December 1966).

The Law of Succession defined Spain politically as "a Catholic social and representative state hereby constituted, in accordance with tradition, a kingdom." General Franco, head of state, would be succeeded eventually by a king, being a person of the blood royal acceptable to the Council of the Realm, the government, and the *Cortes* or, failing such, by a regent equally acceptable.

The Council of the Realm, instituted by the Law of Succession, is the supreme consultative body, advising the head of state on all matters of his exclusive competence, such as declaring war or making peace, and the choice of a successor. Seven of its 17 members are ex officio, 7 elected by the *Cortes*, and 3 nominated by the head of state. The *Cortes*, established by a law of July 1942 as the "'superior organism for the participation of the Spanish people in the work of the state," discusses and may propose laws; the head of the state, who with his ministers normally initiates legislation, has the power of veto. By the Organic Law of the State, 100 of the *Cortes'* members, previously either directly appointed by the head of state or elected by official organizations such as the syndicates, municipalities, and professional bodies, are elected by heads of families and their wives, two members being elected to represent each province. Elected members of the *Cortes* serve for three years. The *Cortes* functions both as a whole, to consider all major legislation, and by commissions. It has no control over the government, though members may ask questions of ministers. Laws of outstanding importance are submitted further to national referendum. The head of the state presides over the cabinet, on which the secretary general of the Falangist party sits without portfolio. The Organic Law of the State provided for the division of executive powers between the head of state and a prime minister. The former selects the latter from a list of three names submitted by the Council of the Realm. The prime minister serves for five years but can be removed by the head of state although not by the *Cortes*.

Since April 1937 there has been only one legally constituted political organization, the Falange Española Tradicionalista y de las Juntas de Ofensiva Nacional Sindicalista, with the *caudillo* as president of the Junta Política. This single party (commonly called the Falange) is controlled by a National Council of about 100 members.

Local Government.—The 50 provinces (except for Navarra, the three Basque provinces, and the Balearic Islands) are named after their capitals; they are intermediary administrative units between state and municipality (*see* Table). The provincial councils

Population of the Provinces of Spain

Provinces* by historic regions	Area in sq. mi.†	Population: 1960 census	Population: Density per sq mi	Capital city	Municipal population (1960)
Andalusia					
Almería	3,388	360,777	106.5	Almería	86,808
Cádiz	2,851	818,847	287.2	Cádiz	117,871
Córdoba	5,296	798,437	150.8	Córdoba	198,148
Granada	4,838	769,408	159.0	Granada	157,178
Huelva	3,894	399,934	102.7	Huelva	74,384
Jaén	5,211	736,391	141.3	Jaén	64,917
Málaga	2,809	775,167	276.0	Málaga	301,048
Sevilla	5,406	1,234,435	228.4	Seville	442,300
Aragon					
Huesca	6,050	233,543	38.6	Huesca	24,377
Teruel	5,715	215,183	37.7	Teruel	19,726
Zaragoza	6,639	656,772	98.9	Saragossa	326,316
Asturias					
Oviedo	4,079	989,344	242.5	Oviedo	127,058
Basque provinces					
Álava	1,177	138,934	118.0	Vitoria	73,701
Guipúzcoa	771	478,337	620.4	San Sebastián	135,149
Vizcaya	856	754,383	881.3	Bilbao	297,942
Catalonia					
Barcelona	2,985	2,877,966	964.1	Barcelona	1,557,863
Gerona	2,273	351,369	154.6	Gerona	32,784
Lérida	4,644	333,765	71.9	Lérida	63,850
Tarragona	2,426	362,679	149.5	Tarragona	43,519
Extremadura					
Badajoz	8,362	834,370	99.8	Badajoz	96,317
Cáceres	7,701	544,407	70.7	Cáceres	48,005
Galicia					
La Coruña	3,041	991,729	326.1	La Coruña	177,502
Lugo	3,785	479,530	126.7	Lugo	58,264
Orense	2,810	451,474	160.7	Orense	64,153
Pontevedra	1,729	680,229	393.4	Pontevedra	50,483
León					
León	5,972	584,594	97.9	León	73,483
Salamanca	4,763	405,729	85.2	Salamanca	90,498
Zamora	4,077	301,129	73.9	Zamora	42,060
Murcia					
Albacete	5,737	370,976	64.7	Albacete	74,417
Murcia	4,369	800,463	183.2	Murcia	249,738
Navarre					
Navarra	4,023	402,042	99.9	Pamplona	97,880
New Castile					
Ciudad Real	7,625	583,948	76.6	Ciudad Real	37,081
Cuenca	6,587	315,433	47.9	Cuenca	27,007
Guadalajara	4,707	183,545	39 0	Guadalajara	21,230
Madrid	3,807	2,606,254	684.6	Madrid	2,259,931
Toledo	5,934	521,637	87.9	Toledo	40,651
Old Castile					
Ávila	3,107	238,372	76.7	Ávila	26,807
Burgos	5,509	380,791	69.1	Burgos	82,177
Logroño	1,944	229,852	118.2	Logroño	61,292
Palencia	3,100	231,977	74.8	Palencia	48,216
Santander	2,042	432,132	211.6	Santander	118,435
Segovia	2,683	195,602	72.9	Segovia	33,360
Soria	3,972	147,052	37.0	Soria	19,301
Valladolid	3,167	363,106	114.7	Valladolid	151,807
Valencia					
Alicante	2,264	711,942	314.5	Alicante	121,527
Castellón	2,579	339,229	131.5	Castellón de la Plana	62,493
Valencia	4,155	1,429,708	344.1	Valencia	505,066
Balearic Islands	1,936	443,327	229.0	Palma	159,084
Canary Islands					
Las Palmas	1,569	453,793	289.2	Las Palmas	193,862
Santa Cruz de Tenerife	1,239	490,655	396.0	Santa Cruz de Tenerife	133,100
Total	194,883	30,430,698	156.1		

*The North African plazas are not included. †1 sq.mi. = 2.59 sq.km.

(*diputaciones provinciales*) are elected partly from representatives of municipalities, and partly from economic, cultural, and professional bodies. Half of the members change every three years. Civil governors are appointed by the government and are responsible for provincial administration to the Minister for the Interior. The *mancomunidad*, or interprovincial association for the integration of common interests, continues in the Canary Islands (two provinces since 1927). A municipal law of 1945 renewed the bases of local administration.

In towns of more than 10,000 inhabitants, the mayor (*alcalde*) is appointed by the government. In smaller communities, he is selected by the civil governor of the province. One-third of the town councilors (*regidores*) are elected from representatives of heads of families; two-thirds in equal numbers from those of syndical organizations, and of economic, cultural, and professional bodies. The council (*ayuntamiento*) is mainly advisory in function since the mayor is regarded as a delegate of the central government. All Spaniards over 21 years of age have an electoral vote.

2. Welfare Services and Living Conditions.—The Labour Charter (1938) provided for a better distribution and remuneration of the working classes. Under the Law of Family Subsidy (1939) all working people are compelled to contribute 1% of their earnings (6% being added by employers) in a system of social insurance which entitles all families with 2 to 12 children under 14 years of age to a proportional monthly allowance, with an extra allowance to families over 12. A bonus is also given on marriage and motherhood. Since 1942 health insurance has been obligatory; various occupational insurance plans have been started, and since 1949 there have been old age and disablement pensions.

After the Civil War, a commission for devastated regions was set up and many new villages were planned and constructed. A National Housing Institute was founded which since 1957 has been a ministry. By the mid-1960s more than 250,000 dwellings were being completed annually.

Despite the economic improvement of the nation as a whole, many Spaniards in the mid-1960s were living just above the poverty line. Average annual income was below $400 (£142). However, through collective bargaining, wages were on the rise and demands far exceeded the minimum daily rate of 60 pesetas ($1.05 or 7*s.* 6*d.*) which was decreed in 1963. This in turn has released a rise in prices, particularly in foodstuffs, and the cost of living was soaring faster than anywhere else in Western Europe.

The income tax structure is based mainly on the narrow (rich) section of the population where evasion is comparatively easy.

3. Justice.—"Pending the definitive organization of the administration of justice in the New State," the Supreme Court (*Tribunal Supremo*) was reconstituted in 1938. The president and 15 of its 20 members are direct government appointees and 5 are indirect. There are 15 Territorial High Courts (*Audiencias Territoriales*), and each province has its own court of first instance (*Audiencia Provincial*) for criminal offenses. Below these are more than 550 district courts which hear civil cases in the first instance, and more than 9,000 municipal and related courts which deal with petty offenses and minor civil issues. The jury system exists except for military trials. Spanish law is founded on Roman law, Gothic common law, and surviving elements of medieval customary and local law. (M. C. S.)

Police.—The maintenance of public order is the responsibility of the Ministerio de la Gobernación (Ministry of Home Affairs) which exercises executive power through the police, both armed and unarmed, and the Guardia Civil. The police forces operate in the more important urban areas, while the Guardia Civil is confined to rural sections of the country.

4. Education.—The country is divided into 12 educational districts, and each has a university as a centre. In the mid-1960s there were more than 100,000 primary schools with about 4,000,000 pupils, and more than 2,600 secondary, technical, and vocational schools with 1,070,000. Primary education is free and compulsory, in theory, up to the age of 14.

All instruction is given in Castilian. Boys and girls are usually segregated except in small schools in rural areas. Secondary education caters to students between 10 and 17 years of age. The course (*bachillerato*) is divided into two parts: the elementary, lasting four years; and the advanced, lasting two years. At the end of the advanced stage, state examinations (prerequisite to university entrance) are conducted by the universities. Choice of textbooks in secondary schools as in primary schools is limited, being subject to official approval. Sexes are again segregated. Both state institutions and private ones (run mainly by religious orders) exist.

There are 12 state universities: at Salamanca (founded in 1218); Valladolid (1346); Barcelona (1450); Valencia (1500); Santiago de Compostela (1501); Seville (1502); Madrid (1508); Granada (1526); Saragossa (1542); Oviedo (1608); La Laguña in the Canary Islands (1701); and Murcia (1915). The Pontifical University of Commillas was founded by Pope Leo XIII in 1892 and the Ecclesiastical University of Salamanca dates from the 12th century. There are seven basic faculties in state universities: philosophy and letters; science; law; medicine; pharmacy; veterinary science; and political, economic, and commercial science. Some of the universities do not have all of these, while Cádiz has a medical and science faculty affiliated with the University of Seville. The teaching syllabus comprises two stages: graduation and doctorate, but Madrid alone can confer the doctorate. The halls of residence (*colegios mayores*) retain a medieval tradition but have also developed a modern character. At first they were centres of further studies, but are now mainly for undergraduates. Under the Law of University Ordination (1943) religious instruction is compulsory; there is no university autonomy and teaching staff are state-appointed. However, by the mid-1960s students, notably in Madrid and Barcelona, were actively opposing the Sindicato Español Universitario (the government-controlled organization to which they were all compelled to belong) and were agitating for academic freedom and for a new university order.

By law each factory is required to maintain a technical school or to send apprentices to one. Pupils attend these directly after primary school. In 1957 the Technical Teaching Law marked the beginning of the union of technical institutes with the universities, and by the 1960s there were four technical universities and more than 90 technological colleges and institutes in Spain.

Seminars take place in Santander, Seville, and Barcelona universities. There are also many courses for foreigners, especially organized in summer schools. Outside Spain there are important Spanish colleges and institutes. Other educational centres include primary teacher training colleges, naval and military schools, and schools of art, architecture, and music. Apart from research carried out in the universities and technical institutes, a Higher Council for Scientific Research was started in 1949, and an Atomic Energy Board was founded in 1951.

5. Defense.—The standing army consists of eight corps of two divisions each, corresponding to the eight military regions of Madrid, Seville, Valencia, Barcelona, Saragossa, Burgos, Valladolid, and La Coruña. A ninth military region (Granada) was constituted in 1944, with one division. There are two general commands in the Balearic and Canary islands, and an independent cavalry division and a general artillery reserve. Some military units are stationed overseas, mainly in the places of Spanish sovereignty in Morocco (Ceuta and Melilla) as well as at Villa Cisneros, in Spanish Sahara. The Navy is rather poorly equipped by Western European standards; the main naval docks are at El Ferrol del Caudillo, Cartagena, and Cádiz. The Air Force was created in 1939, and Spain is divided into five air regions, with two more in the Balearics and the Canaries.

On Sept. 26, 1953, the United States and Spain signed three 10-year agreements covering the construction and use of military facilities in Spain by the U.S. in return for military and economic aid to Spain. These agreements were renewed in 1963 for a further five years. (M. C. S.)

VIII. THE ECONOMY

The Spanish economy suffered three severe economic shocks caused by the general world depression in the early 1930s, the Civil War of 1936–39, and World War II which was followed by a United Nations "boycott" (1946–50). Agricultural production did not regain its 1929 level until 1951, by which time industrial production was only about one-third above the 1929 level. A remarkable economic upsurge began after 1953 when the United States supplied aid for the improvement of communications and the building of major dams for irrigation and hydroelectric power. This was quickly followed by U.S. private investment and in the 1960s by a flood of foreign capital; the financial position of the country vastly improved, supplemented by the booming tourist trade which brought in higher receipts than ever before. But rapid economic expansion was accompanied by a rising cost of living, the threat of inflation, and labour unrest. The utilization of labour in agriculture remained a problem, especially in the typically feudal regions of Andalusia and Extremadura.

A. Production

1. Agriculture.—In the mid-1960s about 40% of the nation's labour force scraped a bare living from the land, contributing little more than 25% to the gross national product which amounted to less than $280 (£100) a head. Agricultural methods lacked mechanization and were often primitive. Much potential farm labour was wasted in seasonal unemployment and there were about 2,000,000 redundant workers. By the 1960s there was a migratory trend of rural labour toward industry, both inside Spain and abroad. The areas under wheat, vines, and olives are proportionately extensive but soil erosion is widespread because of many years of overtilling and overgrazing. In addition, much of the cropped area is dependent upon the vagaries of an uncertain rainfall whose variability has at times greatly affected the national income, for wheat is the chief crop in 27 provinces and olives in 5. Frost is frequently a formidable threat to the Valencian orange crop.

The problem of the latifundia, large landed estates where the rural population can find employment only during the crop seasons, is a continuing one. In 1930 there were about 7,000 proprietors (mostly absentees) who owned about 14,826,800 ac. (6,000,000 ha.) in Extremadura, Andalusia, and La Mancha. In 1939 the National Institute of Land Settlement was established, but by 1956 only 494,230 ac. (200,000 ha.) had been divided. After that a new policy was adopted, involving the concentration of small holdings into larger units, and by the mid-1960s more than 1,500,000 ac. (607,000 ha.) had been redistributed with 4,500,000 ac. (1,821,000 ha.) under negotiation. The supply of fertilizers, formerly an acute problem, has greatly improved and wheat yields have increased. But there is evidence of inefficient marketing of farm products, including lack of storage facilities.

Crops.—Wheat is the most important cereal, accounting for about one-quarter of the value of all agricultural produce. It is grown chiefly in the districts of Tierra de Campos, Tierra del Pan, Coca, Cuéllar, and Medina del Campo (all in Old Castile); Saragossa (in Aragon), La Sagra and La Mancha (in New Castile); and La Serena and Tierra de Barros (in Extremadura). Barley, formerly more widespread, is still the important grain of the southeast; it is now chiefly a fodder crop. The cereal lands are still farmed according to traditional practices, especially the biennial rotation of crops (*año y vez*). Only one-eighth of the cultivated land is cropped continuously with legumes or sugar beet. Pulses are a staple crop of the Meseta. Rice is grown in Valencia and the Ebro Delta and some is exported to the Federal Republic of Germany. Potatoes (an important crop in Galicia), beans, and onions are also significant.

The olive dominates in Andalusia, as does wheat in the Castiles. It covers nearly half the cultivated lands of Jaén, one-third of those of Córdoba, and one-quarter of those of Seville. Almost the whole olive crop is utilized for the production of oil, but yields are quite variable. The vine is especially important in La Mancha, in La Rioja, and in the hills of Catalonia, Valencia, and Málaga. As third largest world producer, Spain suffers from over-production of inferior wines, because imports into France have declined and Spaniards drink less than formerly. But brandy and sherry from Cádiz and Jerez de la Frontera, and dessert wines from Valencia and Málaga find a steady demand. Fruit trees are widespread but production tends to be specialized in certain areas: apples and pears in the north and northwest; figs and almonds in Majorca (Mallorca) and Alicante; peaches and apricots in Murcia; and citrus fruits especially in Valencia and Castellón de la Plana. Esparto or alfa grass covers a large area in the southeast. Hemp and cotton are produced in appreciable quantities. Tobacco is grown on the southeast coast, but sugar beet is much more important.

Irrigation.—With periodic drought affecting about 40% of all

(LEFT) KIT ROBBINS—RAPHO GUILLUMETTE; (ABOVE) STIG T. KARLSSON FROM BLACK STAR

THE FISHING AND WINE INDUSTRIES ARE IMPORTANT

(Left) A wine taster in southwestern Spain; (right) fishermen hauling in tunny in Galicia Province

Spanish land, irrigation has become a necessity. The first large canals and dams were built at the end of the 18th century, such as the Canal Imperial de Aragón on the Ebro and the Pantano de Puentes near Lorca. The first national project dates from 1898, and was followed by those of Pantano de Gassel (1911), the Río Guadalhorce (1926), Pardo (1933), and Río Peña (1939). It is estimated that nearly one-third of the total river flow is controlled by about 140 major reservoirs. A system of more than 200 mi. (320 km.) of modern trunk canals irrigates about 1,200,000 ac. (485,600 ha.), and older canal systems a further 750,000 ac. (303,500 ha.). An area as large again is irrigated by other methods. In the late 1950s the rate of development was accelerated, water storage capacity being increased by about two-thirds; and by the mid-1960s (under the 1964–67 development plan) more than 30 dams were planned or under construction, involving irrigation improvement of almost 6,000,000 ac. (2,428,000 ha.) and the supply of hydroelectric power (see *Economic Planning*, below).

Livestock.—Pastoralism has been fundamental in the history of Spain and is still important. More than half of the cattle in Spain are concentrated in the humid north, especially dairy cows in the Cordillera Cantábrica. Sheep are characteristic of the Meseta, with three basic breeds: the long-wooled churro, the short-wooled merino, and the manchegan. Transhumance is still practised from the mountains of León and Old Castile to Extremadura. Goats are best suited to the arid conditions of the southeast, and pig rearing is common in the holm oak woodlands of Extremadura.

2. Fisheries.—Despite the lengthy Spanish coast line, there are few natural harbours except in Galicia, which has the richest fishing grounds and where Vigo is the main fishing port. About three-quarters of the total annual catch (averaging more than 1,000,000 metric tons) normally comes into the ports along the northern and western Atlantic coasts. The chief fish landed are sardines, anchovies, hake, cod, tunny, bream, mollusks, and crustaceans. About 800 fish-packing plants preserve sardines and tunny.

3. Forestry.—To counteract erosion, between 1939 and the early 1960s more than 3,700,000 ac. (1,500,000 ha.) were reafforested. The total area of forest and potential forest was estimated at 58,000,000 ac. (23,471,000 ha.), more than half of which was practically denuded of trees. The commonest trees are the evergreen oak, the maritime pine, and the Aleppo pine, and there is much cork oak, mainly in Andalusia and Catalonia. (*See* also *Vegetation*, above.)

4. Mining and Minerals.—Until the early 1930s more than half the mining investments were in foreign hands and mining was geared to the needs of Western European markets. By the mid-1960s the mining industry was undergoing radical reorganization. Marginal workings were closed and the number of miners diminished, while investment in pits offering the greatest economic potential was increased. The rich veins of hematite (rubio iron ore) of Bilbao, which once accounted for two-thirds of the national output, now represent less than 3% of the national reserves. In the mid-1960s Spain produced 3,000,000–4,000,000 metric tons of crude steel and about the same quantity of pig iron annually, at costs above those of other Western European countries. Production was more than double that of the mid-1950s, largely because of the opening (1957–58) of ENSIDESA (Empresa Nacional Siderurgica, SA) open-hearth steel mills at Avilés by the Instituto Nacional de Industria, established by the government in 1941.

Coal produced, chiefly from Asturias but also from the lower hills of Sierra Morena, is inadequate (about 14,000,000 metric tons annually in the mid-1960s) and of inferior quality, and considerable quantities of coking coal have to be imported.

Spain, however, is rich in other minerals, most of which are exported: mercury (from the ancient mines of Almadén in Ciudad Real), wolframite, tin, zinc, lead, and manganese. The development of electricity production has stimulated the expansion of electrometallurgical industries; these produce electrolytic copper, aluminum at Alicante and Sabiñánigo, and ferrotungsten and ferromanganese in the Basque provinces.

Oil refining is a relatively new development. Until 1949 there was only one refinery at Tenerife (1930); since then the plant at Escombreras, near Cartagena, has nearly trebled its refining capacity and that at La Coruña can deal with 2,000,000 bbl. of crude oil a year. A hydrogenation plant at Puertollano in Ciudad Real, to produce lubricants from shale oil, was also opened in the 1950s.

5. Power.—Power shortages have seriously handicapped industrial production, especially during seasons of drought. Most of the registered output (more than 24,300,000,000 kw-hr. in the mid-1960s) was provided by hydroelectric stations of which there were more than 1,000. About one-fifth of the water power potential was being utilized, although the proportion varied considerably from region to region. One of the most ambitious schemes was the chain of hydraulic plants on the Miño-Sil river system. In the mid-1960s major dams were being built on the Duero (Douro) at the Portuguese frontier and in the Aragonese Pyrenees. An atomic pile was constructed in Madrid and an atomic power station at Sobrón on the upper Ebro was planned.

6. Manufacturing Industries.—Despite significant changes in the growth of manufacturing in the 20th century, the location of industry is still strongly marked by historical inertia. Bilbao has been a steel centre since the Middle Ages; Éibar in Guipúzcoa has been noted for its production of armaments since the 14th century; the ceramic wares of Manises in Valencia Province, and the paper manufacture of Játiva go back to Moorish times; and Béjar, Segovia, Soria, and Burgos have long been woolen centres. Another feature is the marked concentration of the chief industries in a few districts, namely Barcelona, the province of Valencia, the north coast, and Madrid, in that order of importance. The Barcelona region has about one-quarter of all the national enterprises, with three-quarters of the cotton spindles and looms, and one-third of the entire manufacturing labour force; in addition to textiles, chemicals and engineering are also important. There are considerable manufactures at Valencia and its neighbouring towns, and at Murcia, Alicante, and Palma; apart from food processing, liqueurs, and wines, there are textiles, timber, footwear, light engineering, and chemical industries. The northern heavy industrial centres include the coal-mining towns of Oviedo and the metallurgical centres of Santander and the Basque provinces, of which Guipúzcoa and Vizcaya provinces also have en-

gineering, paper, chemicals, and ancillary industries. By the 1960s Madrid had rapidly developed a wide range of industries, notably food and drink, textiles, and engineering, the latter including the Pegaso motor vehicle factory and the Hispano aircraft works. Out of nearly 250 towns with more than 10,000 inhabitants each, only about 45 may be classified as manufacturing centres while a further 30 are largely concerned with food and drink processing.

7. Economic Planning.—In the backward areas of Spain problems of transport and unemployment have high priorities. Previously, many of the national projects for the extension of electric power, irrigation, communications, and particularly industries, were treated in isolation, but in 1952 the government began to set up regional development plans. Thus under the Badajoz plan the Río Guadiana was harnessed to produce by the mid-1960s about 250,000,000 kw-hr. annually, to irrigate 105,000 ha., to facilitate the settlement of 10,000 colonists, and to employ permanently nearly 60,000 agricultural workers. The pivot of the plan was the construction of the Cijara and Orellán dams. Similarly, in the Jaén plan ten dams on the upper Guadalquivir, Guadiana Menor, and their tributaries were to be built, perennial irrigation was to be extended, and new settlements and industries established. In 1954 a similar 14-year plan for the right bank of the Ebro was launched, designed to benefit the steppe lands of Cinca, Las Bárdenas, and Monegro. Plans on similar lines were announced for Cáceres, Ciudad Real, Almería, and Galicia. By the 1960s the first plants of the Miño-Sil hydropower projects had begun to function.

In 1959, because autarky often proved unworkable, Spain asked the World Bank for advice on economic reforms. The conspicuous result was a four-year development plan (1964–67) which envisaged a total investment of 334,997,000,000 pesetas (more than $5,500,000,000) in both the public and private sectors, with transport, housing, irrigation, and education taking precedence. The plan coincided with the improvements due to foreign investment; its goal was an annual 6% increase in the gross national product.

B. Trade and Finance

1. Foreign Trade.—Food commodities (notably olives and olive oil, oranges, fruits, and vegetables) average more than half of the total value of exports, while manufactures and raw material (chiefly iron ore and pyrites) represent each from one-fifth to one-quarter. Most of the imports consist of raw materials (mainly cotton, petroleum and products, and fertilizers) and manufactured goods, the remainder being foodstuffs. By the mid-1960s more than one-third of both imports and exports came from the European Economic Community; the European Free Trade Association (especially the United Kingdom) and the United States were the other chief sources of reciprocal trade. A secondary trade was maintained with Latin America. After restrictions on foreign trade were lifted in 1959 most imports entered the country freely.

With government sponsorship, the tourist trade increased at a rapid rate in the 1950s, and by the mid-1960s, together with emigrants' remittances, brought to Spain foreign exchange almost equal to the total value of all visible exports. The number of foreign tourists (excluding day-trippers) exceeded 14,000,000 annually. This source of foreign capital inflow helped to fill a chronic trade gap caused by expanding imports (mainly machinery).

2. Currency, Banking, and Finance.—The unit of currency is the peseta, divided into 100 céntimos. In 1959 the peseta was stabilized at a devalued rate (U.S. $1 = 60 pesetas; £1 = 167–168 pesetas), and by the mid-1960s it had become one of the stronger European currencies.

The central bank is the Banco de España, in whose profits the government has participated since 1922. The commercial banks are regulated by ordinance and have a strongly entrenched control over industry. Largely because the public prefers to keep its savings on deposit rather than invest in securities, the banks control more than half of the issued capital of the Spanish limited-liability companies; thus seven leading banks control about two-thirds of industry. After 1941 all foreign banks were voluntarily nationalized.

By the mid-1960s government annual revenue and expenditure both averaged more than 120,000,000,000 pesetas. About one-quarter of the revenue came from direct taxation and more than one-half from indirect, with a small proportion from state monopolies and properties. Chief expenditures were on defense, public works (particularly dams), and the servicing of the public debt.

C. Transport and Communications

1. Roads.—Most of the large centres were originally sited on the well-developed Roman road system. But this was allowed to fall into decay, and only in the 18th century were the royal routes improved. There was little development or repair after the Peninsular War, and not until the 1920s was much done. The greatest improvements were made after the start of U.S. aid in 1953. The Madrid-Cádiz highway is now in excellent condition, and there have been considerable improvements on the Madrid-Irún and Madrid-Saragossa roads. By the mid-1960s the total length of the roads exceeded 83,000 mi. (134,000 km.) but only about three-fifths were macadamized. However, prospects for improvement are good, for vehicles are heavily taxed and tourists also pay a tax destined for the development of communications. Regular motor coach and bus services carry more passengers annually than the railways.

2. Railways.—The first railway (Madrid-Aranjuez) was opened in 1848, and the active period of trunk-line construction was between 1855 and 1900. Of the total length (11,226 mi.; 18,066 km.) one-quarter is narrow-gauge, only one-seventh has double track, and one-sixth is electrified. In 1941 the broad-gauge lines were nationalized and taken over by the organization known as RENFE (Red Nacional de los Ferrocarriles Españoles). After World War II traffic greatly increased, trains were faster, and delays were less common. However, Spain is one of the few European countries that remains deficient in rail transport and there is an urgent need for the railways that were planned long ago but not built: Madrid-Galicia, Madrid-Burgos-Santander, Asturias-Galicia, Vitoria-Bilbao, Zafra-Portugal, Jerez de la Frontera-Almadén, and especially Baeza-Albacete-Utiel-Teruel-Caspe-Lérida linked to a Lérida or Saragossa line to Luchon, France. In 1964 the World Bank loaned $65,000,000 to RENFE to finance the first stage of a 10-year railway modernization program.

3. Shipping.—As river and canal transport virtually do not exist, except for a short stretch on the Guadalquivir to Seville, coastwise traffic is important. It amounts to nearly twice the tonnage of the oversea shipping services. During and after World War II the Spanish fleet increased, but only about one-third of the merchandise entering and leaving Spanish ports is carried by Spanish shipping. The Empresa Nacional "Elcano" de la Marina Mercante, a state-organization, operates about 10% of a national gross tonnage which totaled about 2,000,000 in the early 1960s.

4. Air Transport.—The principal Spanish airlines, Iberia and Aviaco, have considerable traffic, especially on the internal routes. Iberia has regular services to North Africa, the Balearic and Canary Islands, Europe, South America, and North America.

5. Telecommunications.—The Compañía Nacional de Telegrafía sin Hilos holds a government concession for a public telegraph service with shipping. Transradio Española, SA operates radio-telegraph circuits; Bilbao is linked with the United Kingdom by cable, and Barcelona is linked with Marseilles, France. Radio broadcasting and television services (both government and commercial) are controlled by an agency of the Ministry of Information and Tourism and relay the official nation-wide news services.

(J. M. Ho.)

Bibliography.—*Geography:* Maps published by Instituto Geográfico y Catastral, Depósito de la Guerra, Instituto Geológico de España; publications of Instituto Geográfico y Catastral, Oficina Central Meteorológica, Servicio Meteorológico Nacional; *Resumen* (annual) and *Memorias; O clima de Portugal*, 7 nos. 1942–53; V. Masachs Alavedra, *El régimen de los rios peninsulares* (1948); M. Willkomm and J. Lange, *Prodromus Florae Hispanicae*, 3 vol. (1861–80, with supp., 1893); 10th International Pflanzengeographischen Exkursion, *Die Pflanzenwelt Spaniens* (1956); E. H. del Villar, *Los suelos de la peninsula luso-ibérica*, Eng. trans. by G. W. Robinson (1937); L. Martín Echeverría, *España, El País y los Habitantes* (1940); J. Sermet,

L'Espagne du sud (1953); M. Sorre, *Espagne-Portugal*, vol. 7 of *Géographie universelle*, ed. by L. Gallois and P. Vidal de la Blache, pp. 69–258 (1934); M. de Terán (ed.), *Geografía de España y Portugal*, vol. 1, 2, 4, 5 (1952–58); W. B. Fisher and H. Bowen-Jones, *Spain, a Geographical Background* (1958); J. M. Houston, *The Western Mediterranean World* (1964). (J. M. Ho.)

The People: J. Caro Baroja, *Los pueblos del norte de la península ibérica* (historico-cultural analysis; 1943), *Los pueblos de España: Ensayo de etnología* (1946), *Los vascos: Etnología* (1949); A. Panyella, "Etnología española," *Enciclopedia Espasa*; R. Violant y Simorra, *El arte popular español a través del Museo de Industrias y Artes Populares* (1953); *Revista de dialectología y tradiciones populares* (1944). (Ao. P.)

History: General: B. Sánchez Alonso, *Fuentes de la historia española e hispanoamericana*, 3rd rev. ed., 3 vol. (1952), *Historia de la historiografía española*, 3 vol. (1941–50), 2nd rev. ed. (1947–); A. Ballesteros y Beretta, *Historia de España y su influencia en la historia universal*, 9 vol., 2nd ed. (1943–); R. Altamira y Crevea, *Historia de España y de la civilización española*, 5 vol. (vol. v by P. Zabala y Lera, 1900–30), *A History of Spanish Civilization*, Eng. trans. by P. Volkov (1930), *A History of Spain*, Eng. trans. by M. Lee (1949); R. Menéndez Pidal (ed.), *Historia de España* (1935 in progress), *The Spaniards in Their History*, Eng. trans. by W. Starkie (1950); J. P. Oliveira Martins, *A History of Iberian Civilization*, Eng. trans. by A. F. G. Bell (1930); W. C. Atkinson, *Spain: a Brief History* (1934); A. Mousset, *Histoire d'Espagne* (1947); H. V. Livermore, *A History of Spain* (1958).

Archaeology and Prehistory: M. Almagro Basch *et al.*, *Ars Hispaniae*, vol. i–ii (1947); H. Obermaier, *Fossil Man in Spain*, Eng. trans. by C. D. Matthew (1924); H. Breuil and G. Zbyszewski, *Contribution à l'étude des industries paléolithiques du Portugal*, 2 vol. (1942–45); G. Zbyszewski, *La Classification du Paléolithique ancien . . . de Portugal* (1943); L. Pericot García, *La cueva del Parpalló (Gandía)* (1942), *La España primitiva* (1950), *Los sepulcros megalíticos catalanes y la cultura pirenaica*, 2nd ed. (1950), *Historia de España*, 2nd ed., vol. i (1958); P. Bosch-Gimpera, *Two Celtic Waves in Spain* (1941); E. MacWhite, *Estudios sobre las relaciones atlánticas de la península hispánica en la edad de Bronce* (1951); F. López Cuevillas, *La civilización céltica en Galicia* (1953); A. García y Bellido, *La dama de Elche* (1943), *La península ibérica en los comienzos de su historia* (1953); A. Schulten, *Tartessos*, 2nd ed. (1945), *Numantia*, 4 vol. (1914–31).

Phoenician and Greek: A. García y Bellido, *Fenicios y carthagineses en occidente* (1942), *Hispania graeca*, 3 vol. (1948).

Much material is also contained in the periodicals *Archivo español de arqueología* and *Ampurias*.

Roman: C. H. V. Sutherland, *The Romans in Spain, 217 B.C.–A.D. 117* (1939).

Medieval Spain: A. Castro, *España en su historia: cristianos, moros, y judíos* (1948), Eng. trans. by E. L. King, *The Structure of Spanish History* (1954); R. Menéndez Pidal, *The Cid and His Spain*, Eng. trans. by H. Sunderland (1934); H. J. Chaytor, *A History of Aragon and Catalonia* (1933); É. Lévi-Provençal, *Histoire de l'Espagne musulmane*, 2nd ed., 3 vol. (1950–), *l'Espagne musulmane au Xe siècle* (1932); H. Mones, *La Chute du califat umayyad de Cordoue* (1948).

The 16th, 17th, and 18th centuries: R. B. Merriman, *The Rise of the Spanish Empire*, 4 vol. (1918–34); H. C. Lea, *A History of the Inquisition of Spain*, 4 vol. (1906–07); J. M. Batista i Roca, "The Hispanic Kingdoms and the Catholic Kings," *The New Cambridge Modern History*, vol. i, ch. xi (1957); R. Carande Thobar, *Carlos V y sus banqueros*, 2 vol. (1943–49), further volumes in progress; E. J. Hamilton, *American Treasure and the Price Revolution in Spain, 1501–1650* (1934); F. Braudel, *La Méditerranée et le monde méditerranéen à l'époque de Philippe II* (1949); H. G. Koenigsberger, "The Empire of Charles V in Europe," *The New Cambridge Modern History*, vol. ii, ch. x (1958); G. Mattingly, *The Defeat of the Spanish Armada* (1959); J. H. Parry, *The Spanish Theory of Empire in the Sixteenth Century* (1940); S. de Madariaga, *The Rise of the Spanish American Empire* (1947), *The Fall of the Spanish American Empire*, rev. ed. (1963); C. H. Haring, *The Spanish Empire in America*, 2nd ed. (1952); L. U. Hanke, *Bartolomé de las Casas* (1951); H. and P. Chaunu, *Séville et l'Atlantique, 1504–1650*, 8 vol. (1955–59); J. H. Elliott, *Imperial Spain, 1469–1716* (1963), *The Revolt of the Catalans* (1963); J. Lynch, *Spain under the Habsburgs*, vol. 1 (1964), further volume in progress; J. Sanabre Sanromà, *La acción de Francia en Cataluña en la pugna por la hegemonía de Europa, 1640–1659* (1956); A. Domínguez Ortiz, *Política y hacienda de Felipe IV* (1960); V. Palacio Atard, *Derrota, agotamiento, decadencia, en la España del siglo XVII* (1949); J. Vicens Vives, *Historia social y económica de España y América* (1957–59); E. J. Hamilton, *El florecimiento del capitalismo y otros ensayos de historia económica* (1948); P. Chaunu, "La société espagnole du siècle d'or," *Annales*, vol. vii (1952); P. Vilar, "Le temps du Quichotte," *Europe*, vol. xxiv (1956), *La Catalogne dans l'Espagne moderne*, 3 vol. (1962); J. Reglà Campistol, *Els virreis de Catalunya els segles XVI i XVII* (1956), "La expulsión de los moriscos y sus consecuencias," *Hispania*, vol. xiii (1953); H. Lapeyre, *Géographie de l'Espagne morisque* (1959); J. O. McLachlan, *Trade and Peace with Old Spain, 1667–1750* (1940); G. N. Desdevises du Dézert, *L'Espagne de l'ancien régime*, 3 vol. (1897–1904); F. Soldevila Zubiburu, *Historia de España*, vol. vi (1957); R. Herr, *The Eighteenth-Century Revolution in Spain* (1958); J. Sarrailh, *L'Espagne éclairée de la seconde moitié du XVIIIe siècle* (1954); A. Domínguez Ortiz, *La sociedad española en el siglo XVIII* (1955).

The 19th and 20th centuries: M. A. S. Hume, *Modern Spain, 1788–1898*, 3rd ed. (1923); S. de Madariaga, *Spain* (1942); A. Ramos Oliveira, *Politics, Economics and Men of Modern Spain, 1808–1946*, Eng. trans. by T. Hall (1946); G. Brenan, *The Spanish Labyrinth*, 2nd ed. (1950); G. Maura Gamazo, *Bosquejo histórico de la dictadura*, 2 vol. (1930); L. Jiménez de Asúa, *Proceso histórico de la constitución de la república española* (1932); H. Buckley, *Life and Death of the Spanish Republic* (1940); E. A. Peers, *The Spanish Tragedy, 1930–1936* (1936), *Spain in Eclipse, 1937–1943* (1943); F. Borkenau, *The Spanish Cockpit* (1937); C. A. Gómez, *La guerra de España, 1936–1939*, 2 vol. (1939); H. S. Thomas, *The Spanish Civil War* (1961); J. Beneyto Pérez, *El nuevo estado español*, 2nd ed. (1939); J. Pemartín, *Que es "lo nuevo"?*, 3rd ed., (1940); J. M. de Areilza and F. M. Castiella, *Reivindicaciones de España*, 2nd ed. (1941); J. Pla, *Gibraltar*, Eng. trans. by D. Round (1955); G. Jackson, *The Spanish Republic and the Civil War, 1931–1939* (1965). (X.)

The Economy: J. M. Houston, "Irrigation as a Solution to Agrarian Problems in Modern Spain," *Geog. J.*, vol. cxvi (1950); J. M. Ruiz Morales, *Apuntes de estructura económica mundial y de España*, 2 vol. (1945–50); I. Olagüe, *La decadencia española*, 4 vol. (1950–51); T. E. Rogers, *Spain: Economic and Commercial Conditions in Spain* (1957); International Bank for Reconstruction and Development, *The Economic Development of Spain* (1963); J. Klein, *The Mesta; a Study in Spanish Economic History, 1273–1836* (1964). (J. M. Ho.)

SPALATIN, GEORG (1484–1545), German humanist and friend of Martin Luther to whose diplomatic influence may be largely attributed much of the early progress of the Reformation in Germany, was born Georg Burkhardt on Jan. 17, 1484, at Spalt (whence he assumed the name Spalatinus), near Nürnberg. He studied at the University of Erfurt, where he took his bachelor's degree in 1499. There Nikolaus Marschalk, the most influential professor, made Spalatin his amanuensis and took him to the new University of Wittenberg in 1502. In 1505 Spalatin returned to Erfurt to study jurisprudence and joined a group of humanists led by Mutianus Rufus. He became a teacher at the Georgenthal monastery, was ordained priest in 1508, and in the same year became tutor to the heir of Frederick III the Wise, elector of Saxony. In 1511 he was given a canon's stall at Altenburg and made preceptor to the elector's nephews, Otto and Ernst of Brunswick-Lüneburg, who were studying at Wittenberg University, where Spalatin met Luther. But the elector valued his advice so highly that in 1516 he summoned Spalatin back to the court to be librarian and chaplain.

Spalatin managed all the elector's correspondence, and it was he who in 1518 persuaded Frederick to give Luther his protection during the indulgence controversy. He also won for Luther a more favourable reception at the Diet of Worms (1521) than the emperor Charles V was at first prepared to give. It is not certain whether Spalatin fully appreciated the difference between humanism and Lutheranism. He several times tried to dissuade Luther from publishing his opinions, but he nonetheless translated Luther's Latin writings for the elector, as well as those of Philipp Melanchthon and Erasmus.

Spalatin was also the valued adviser of the two succeeding electors of Saxony, John and John Frederick. He became superintendent of the church in Saxony, where he organized the Reformation. From 1530 he worked with Melanchthon on the preparation of the Augsburg Confession, and he took part in the formation of the Schmalkaldic League. He died at Altenburg on Jan. 16, 1545.

Spalatin's writings include *Annales reformationis*, edited by E. S. Cyprian (1718), and "A Life of Frederick III" published in *Georg Spalatins Historischer Nachlass und Briefe*, edited by C. G. Neudecker and L. Preller (1851).

See I. Höss, *Georg Spalatin* (1956); H. Volz, "Bibliographie der im 16. Jahrhundert erschienenen Schriften Georg Spalatins," *Zeitschrift für Bibliothekswesen* (1958).

SPALDING, ALBERT (1888–1953), U.S. violinist and composer of international repute, was born at Chicago, Ill., on Aug. 15, 1888, the son of a well-to-do manufacturer of sporting goods. At age seven he commenced his musical studies under Ulpiano Chiti in Florence, continuing them with Juan Buitrago in New York City and Narcisse Lefort in Paris. Spalding's debut was made in 1905 in Paris, where he appeared with Adelina Patti

a year later. His first American appearance was with the New York Symphony Orchestra in 1908, and he toured with that orchestra in Europe, as violin soloist, in 1920. During World War I he served with the U.S. air service in Italy. He played at La Scala, Milan, in 1919, and in 1922 became the first American violinist to play regularly at concerts of the Société des Concerts du Conservatoire, Paris.

His compositions include many works for the violin, as well as a suite for orchestra, a string quartet (in E minor), and works for violin and piano. He wrote an autobiography, *Rise to Follow* (1943), and a novel, *A Fiddle, a Sword and a Lady* (1953). He retired in 1950 and died May 26, 1953 in New York City.

SPALDING, a market town and urban district in the Parts of Holland, Lincolnshire, Eng. Pop. (1961) 14,824. Area, 12.2 sq.mi. It is the centre of an extremely fertile fenland area, visited in spring for its flowering bulb fields; it cans locally grown fruit and vegetables and has a large beet-sugar factory and a livestock market. The Welland River passes through the town by a deep channel. The parish church (once the priory church) of SS. Mary and Nicholas dates from 1284. Ayscoughfee Hall (15th century, now belonging to the town) was the house of Maurice Johnson, a founder of the Spalding Gentlemen's Society (1710), probably the earliest antiquarian society in England; its museum includes birds and antiquities.

SPALLANZANI, LAZZARO (1729–1799), Italian priest and biologist, a skilled experimenter in physiology, was born at Scandiano in the province of Modena on Jan. 12, 1729. At the age of 15 he was sent to the Jesuit college at Reggio di Modena, and soon went to study law at the University of Bologna, where his kinswoman, Laura Bassi, was professor of physics; she, it is said, first interested him in science. After taking orders, he was professor at Reggio, Modena (1756), and later at Pavia, where he enriched the museum by his collections from journeys along the shores of the Mediterranean. In 1785 he visited Turkey, where he made many observations. He visited Vesuvius and Sicily in 1788 and embodied his researches in *Viaggi alle due Sicilie* (1792–97).

Spallanzani studied the circulation of the blood, respiration, gastric digestion, the senses of bats, the electricity of the torpedo, the breeding of eels, and the regeneration of different appendages of Amphibia, examining the muscles, nerves, and bones with a microscope. By filtering semen he proved that spermatozoa were necessary for fertilization in different animals, and succeeded in artificially inseminating a bitch. In a controversy with J. T. Needham, he disproved (1765) the spontaneous generation of Infusoria and other microscopic organisms, which did not develop in vegetable infusions that had been boiled long enough and kept in properly closed vessels. Some of his experiments on spontaneous generation were similar to those later performed by Pasteur.

Spallanzani's great works are the *Prodromo di un'opera da imprimersi sopra le reproduzioni animali* (1768) and *Dissertazioni di fisica animale e vegetabile*, 2 vol. (1780). He died at Pavia on Feb. 11, 1799. (A. C. Cr.)

SPANDAU, a district of West Berlin in the *Land* (state) of B[illegible], Federal Republic of Germany. Pop. [illegible]. The citadel (1560–94) is the oldest well-preserved Italian-style fortification of Renaissance time in north Germany. St. Nicholas' Church (mid-14th century) is the oldest Gothic church in Berlin. Since 1946 the prison on the Wilhelmstrasse has housed the war criminals condemned by the Allies. Spandau is now the chief industrial district of Berlin, with the electrotechnical firm of Siemens in the suburb of Siemensstadt, but its parks and forests make it a weekend resort for West Berliners. Civic rights were granted in 1232; it was incorporated into Berlin in 1920. (Gü. B.)

SPANGENBERG, AUGUST GOTTLIEB (1704–1792), German bishop of the *Unitas Fratrum* (*see* MORAVIAN CHURCH), successor to Count Nikolaus Zinzendorf (*q.v.*), and founder of the Moravian Church in North America, was born on July 15, 1704, at Klettenberg in the southern Harz Mountains. He studied law at Jena but changed to theology after his conversion to Pietism in 1722. In 1728, during a visit by Zinzendorf to Jena, Spangenberg was drawn into his circle, and in 1730 he visited the Moravian community at Herrnhut, on Zinzendorf's estate of Berthelsdorf. In 1732 he was appointed to the theological faculty at Halle, but differences with the Halle Pietists, chiefly over his connection with Zinzendorf, led to his expulsion in April 1733.

Spangenberg then went to work as Zinzendorf's assistant and until 1762 devoted himself to superintending and organizing Moravian mission work. In Pennsylvania (1736–39) he founded the North American branch of the *Unitas Fratrum*. In 1741–42 he was in England organizing a branch of the movement, and received licence to preach. In 1744 Spangenberg was made a bishop of the Moravian Church and returned to supervise work in North America, where he made perhaps his most important contribution to the movement's development. After an interval in Europe again (1749–51) he extended the mission work to North Carolina. John Wesley in his *Journal* mentions several conversations with Spangenberg, which left a profound impression on Wesley and throw an interesting light on Spangenberg's theological thought.

In 1762 Spangenberg finally returned to Germany to become a member of the governing body founded to direct the Brethren after Zinzendorf's death (1760). In 1777 he was commissioned to draw up the *Idea fidei fratrum* which became the accepted statement of the Moravian belief. As compared with Zinzendorf's own writings this book exhibits the finer balance and greater moderation of Spangenberg's nature. It was chiefly through his efforts that the movement was prevented from splitting over internal differences and that it maintained its good relationship with the main body of the Lutheran Church. In his last years Spangenberg devoted special attention to education, a field in which the Brethren made valuable contributions. He died at Berthelsdorf on Sept. 18, 1792.

In addition to the *Idea fidei fratrum*, Spangenberg wrote, besides other apologetic works, a life of Zinzendorf (1772–75) and several hymns which found their way into the hymn books of many other denominations.

See G. Reichel, *August Gottlieb Spangenberg, Bishop der Brüderkirche* (1906); T. Bechler, *A. G. Spangenberg und die Mission* (1933).

SPANISH AMERICAN ARCHITECTURE: *see* IBERO-AMERICAN ARCHITECTURE.

SPANISH AMERICAN LITERATURE: *see* IBERO-AMERICAN LITERATURE.

SPANISH-AMERICAN WAR, a brief and decisive conflict between the United States and Spain in the year 1898. It is often cited as marking the emergence of the United States as a world power.

Origin.—The war originated in the Cuban struggle for independence from Spain, which began in February 1895. (*See* CUBA: *History*.) The destructive Cuban conflict was injurious to U.S. investments in the island, which were estimated at $50,000,000, and almost ended U.S. trade with Cuban ports, normally valued at $100,000,000 annually. On the insurgent side the war was waged largely against property and led to the destruction of sugarcane, sugar mills, and other forms of property. Of more importance than its effect on U.S. property interests was the appeal to American humanitarian sentiment. Under the Spanish commander, Capt. Gen. Valeriano Weyler y Nicolau (nicknamed the "Butcher"), supposedly loyal Cubans were herded into so-called reconcentration areas in and around the larger cities; those who remained at large were treated as enemies. In the reconcentration areas the Spanish authorities made no adequate provision for shelter, food, sanitation, or medical care for the *reconcentrados*, thousands of whom died from exposure, hunger, and disease. These conditions were graphically portrayed for the U.S. public by the sensational newspapers, notably Joseph Pulitzer's *New York World* and William Randolph Hearst's recently founded *New York Journal*. Humanitarian concern for the suffering Cubans was added to the traditional American sympathy for a colonial people struggling for independence. While these aspects of the war created a widespread popular demand for action to halt it, the U.S. was continually inconvenienced by the necessity of patrolling the coastal waters to prevent gunrunning to the

insurgents, and by demands for aid from Cubans who had acquired U.S. citizenship and then had been arrested by the Spanish authorities for participating in the rebellion.

The popular demand for intervention to stop the war and assure Cuban independence gained support in Congress. In the spring of 1896 both the Senate and the House of Representatives declared by concurrent resolution that belligerent rights should be accorded the insurgents. This expression of congressional opinion was ignored by Pres. Grover Cleveland, who firmly opposed intervention, as did his successor, William McKinley, during his first year in office. In the fall of 1897, indeed, prospects for a peaceful settlement in Cuba brightened. In Spain the Sagasta ministry, taking office upon the assassination of Prime Minister Cánovas del Castillo, announced a new policy for Cuba: Captain General Weyler was to be recalled; his reconcentration camps were to be abandoned; and a considerable degree of autonomy was to be granted the Cubans.

These concessions came too late. The insurgent leaders would now settle for nothing short of complete independence. The war went on in Cuba, and a series of incidents brought the United States to the brink of intervention. Riots in Havana in December led to the sending of the battleship "Maine" to that port, as a precaution for the safety of U.S. citizens and property. On Feb. 9, 1898, the *New York Journal* printed a private letter from the Spanish minister in Washington, Dupuy de Lôme (filched from his correspondent in Cuba), describing President McKinley as "weak and a popularity-hunter." De Lôme immediately resigned and the Spanish government tendered an apology. The sensation caused by this incident paled beside that of six days later. On the night of Feb. 15 a mighty explosion sank the "Maine" at her Havana anchorage, carrying more than 260 seamen to their deaths. Responsibility for the disaster was never determined. A U.S. naval board found convincing evidence that an initial explosion outside the hull (presumably from a mine or torpedo) had in turn touched off the battleship's forward magazine. The Spanish government offered to submit the question of its responsibility to arbitration, but the American public, prompted by the *New York Journal* and other sensational papers, held Spain unquestionably responsible. "Remember the 'Maine,' to hell with Spain!" became a popular watchword.

President McKinley, still desirous of peace, now found himself under intense pressure for intervention; the pressure came from members of his own party, the Republican majority in Congress, as well as from the Democratic minority. U.S. business interests, in general, opposed intervention and war. But such opposition diminished after a speech in the Senate, March 17, by Sen. Redfield Proctor of Vermont, who had just returned from a tour of Cuba. In matter-of-fact and unsensational language, Proctor described his observations in the war-torn island, the suffering and death in the reconcentration areas, the devastation elsewhere, the evident inability of the Spanish to crush the rebellion. His speech, as the *Wall Street Journal* remarked on March 19, "converted a great many people in Wall Street."

Yielding to the pressure, the president on March 27, 1898, sent what amounted to an ultimatum to Madrid. Let Spain, he asked, abandon reconcentration, declare an armistice, and accept U.S. mediation in peace negotiations with the insurgents.

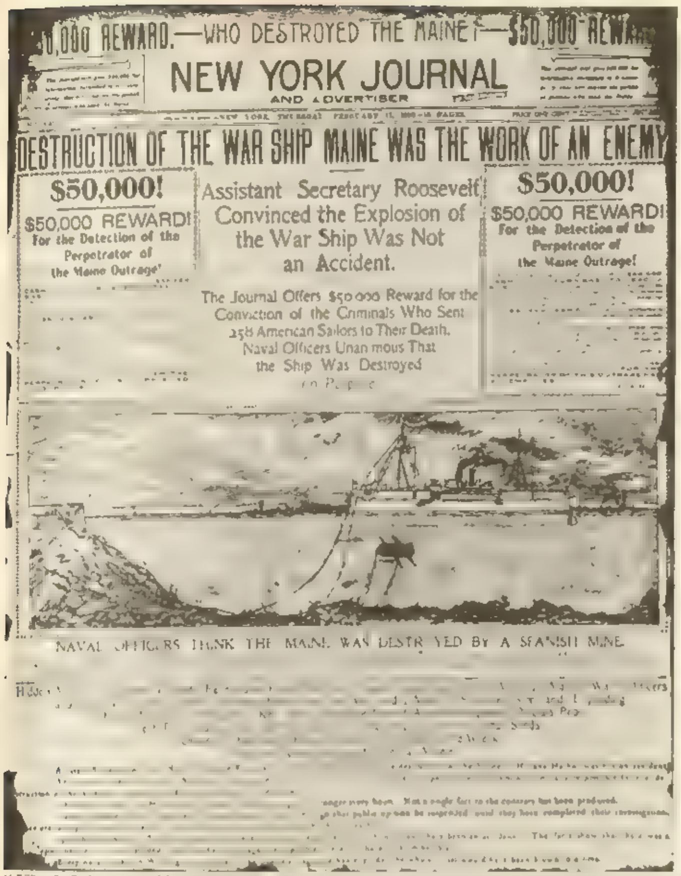
The Spanish government was caught upon the horns of a cruel dilemma. It had not readied its army or navy for war with the United States, nor had it warned the Spanish public of the necessity of relinquishing Cuba. War meant certain disaster. The surrender of Cuba might mean the overthrow of the government, perhaps of the dynasty. The ministry clutched at the only straws in sight. On the one hand, it sought support from the chief European governments. Aside from the British, these governments were sympathetic to Spain but were unwilling to give it more than weak verbal support. Their six ambassadors in Washington (including the British ambassador) called upon McKinley on April 6 and begged him in the name of humanity to refrain from armed intervention in Cuba. McKinley replied urbanely that if armed intervention should come, it would be in the interest of humanity. Meanwhile, Spain was going far in the acceptance of McKinley's terms of March 27—so far that the U.S. minister in Madrid, Stewart L. Woodford, assured his government that, granted a little time and patience, Spain could work out a solution acceptable to both the United States and the Cuban insurgents. Spain would end the reconcentration policy in fact as well as in name. Instead of accepting U.S. mediation, Spain would seek the pacification of the island through the Cuban Cortes, about to be elected under the autonomy program. Spain at first stated that an armistice would be granted only on application from the insurgents, but on April 9 announced one on its own initiative.

By this time, however, President McKinley had yielded to the war party in Congress. Brushing aside the Spanish concessions as unsatisfactory, he declared, in a special message to Congress on April 11, that "the war in Cuba must stop." From Congress he asked authority to use the armed forces of the United States "to secure a full and final termination of hostilities between the Government of Spain and the people of Cuba." Congress responded emphatically. In resolutions of April 20 it declared that "the people of Cuba are, and of right ought to be, free and independent," demanded that Spain at once relinquish authority over Cuba and withdraw its armed forces from the island, and authorized the president to use the army and navy of the United States to enforce that demand. A fourth resolution renounced for the United States any idea of annexing Cuba. As it was proposed by Sen. Henry M. Teller of Colorado, it came to be known as the Teller amendment. Upon being informed of the signing of the resolutions, the Spanish government at once severed diplomatic relations and on April 24 declared war upon the United States. Congress declared war April 25 and made the declaration retroactive to April 21.

The War.—The war thus begun was pathetically one-sided. Spain, as noted above, was in no sense prepared for war with a formidable power. Neither was the U.S. Army but the outcome of the war was dependent on sea power, and in this element the United States completely outclassed its opponent. Spain had nothing to match the four new battleships, "Indiana," "Iowa," "Massachusetts," and "Oregon," which formed the backbone of the North Atlantic squadron (the "Oregon" after her notable voyage from the Pacific coast). Even more superior to their antiquated antagonists at Manila were the protected cruisers of Commodore George Dewey's Asiatic squadron. Thanks largely to the energy and enthusiasm of the young assistant secretary of the Navy, Theodore Roosevelt, the U.S. ships had engaged in battle maneuvers and target practice and were well supplied with fuel and ammunition. Officers and men were confident and aggressive, whereas their Spanish opponents knew that they were doomed to defeat.

The first blow fell in Manila Bay, May 1, 1898. Commodore Dewey, picked by Roosevelt for the command, led his squadron into the bay before dawn and in a leisurely morning engagement destroyed the anchored Spanish ships by gunfire. American casualties amounted only to seven men slightly wounded. Dewey remained in control of the bay while a military force was sent out to assist him in taking possession of the city of Manila. By the end of July some 11,000 U.S. troops under Maj. Gen. Wesley Merritt had arrived in the Philippines, and on Aug. 13 they occupied Manila.

Meanwhile, attention had centred on Cuba. Upon the declaration of war a Spanish fleet of four armoured cruisers and three destroyers, commanded by Adm. Pascual Cervera, had steamed westward from the Cape Verde Islands. Its whereabouts remained unknown until late in May, when it was located in Santiago Harbour on the south coast of Cuba. The North Atlantic squadron under Rear Adm. William T. Sampson and the so-called Flying squadron under Commodore Winfield Scott Schley thereupon blockaded the harbour entrance. An army of regulars and volunteers (including Theodore Roosevelt's regiment of "Rough Riders," minus their horses) embarked at Tampa and landed on the Cuban coast east of Santiago. The American objective now was to catch Cervera between Army and Navy, thus forcing him either to surrender or to come out and fight. On July 1, in the hard fought Battles of El Caney and San Juan Hill, U.S. troops pene-



$50,000 REWARD.—WHO DESTROYED THE MAINE?—$50,000 REWARD.

NEW YORK JOURNAL

AND ADVERTISER

DESTRUCTION OF THE WAR SHIP MAINE WAS THE WORK OF AN ENEMY

$50,000!

$50,000 REWARD! For the Detection of the Perpetrator of the Maine Outrage!

Assistant Secretary Roosevelt Convinced the Explosion of the War Ship Was Not an Accident.

The Journal Offers $50,000 Reward for the Conviction of the Criminals Who Sent 258 American Sailors to Their Death. Naval Officers Unanimous That the Ship Was Destroyed on Purpose.

$50,000!

$50,000 REWARD! For the Detection of the Perpetrator of the Maine Outrage!

NAVAL OFFICERS THINK THE MAINE WAS DESTROYED BY A SPANISH MINE.

(Left) Front page of *New York Journal* Feb. 17, 1898, announcing the sinking of the USS "Maine"; (above) mast of the sunken "Maine" in Havana Harbour; (below) U.S. Rough Riders crouch under fire; (top right) Maj. Gen. William R. Shafter leading forces into Santiago, July 17, 1898

(LEFT) THE GRANGER COLLECTION, (TOP CENTRE, TOP RIGHT) BROWN BROTHERS (BOTTOM) CULVER PICTURES

trated the outer defenses of Santiago. Their hold here was so precarious, and the incidence of malaria and other diseases was so widespread, that their commander, Maj. Gen. William R. Shafter, considered withdrawing to await reinforcements. This idea was abandoned when, on July 3, Admiral Cervera, under orders from Havana, led his squadron out of Santiago Harbour and tried to escape westward along the coast. In the ensuing battle all of Cervera's ships, under heavy fire from the U.S. fleet, were beached in a burning or sinking condition. American losses were insignificant. Two weeks later the city of Santiago surrendered to General Shafter.

The Peace.—The war was now over for all practical purposes, and on July 18 the Spanish government requested the good offices of France in arranging a termination of hostilities. Before the fighting ended, however, another American expeditionary force, commanded by Gen. Nelson A. Miles, occupied Puerto Rico. Armistice negotiations conducted in Washington ended with the signing of a protocol, Aug. 12, 1898. Besides ending hostilities, this agreement pledged Spain to surrender all authority over Cuba and to cede Puerto Rico and an unnamed island in the Ladrones (Marianas) to the United States. In the Philippines, Spain consented that the United States should occupy the city and harbour of Manila until the conclusion of a peace treaty that would determine the final disposition of the islands. Peace commissioners were to meet in Paris not later than Oct. 1.

The great question now confronting McKinley and his advisers was what, if anything, to demand of Spain in the Philippines. It seems certain that McKinley, in proposing intervention in Cuba, had had no thought of acquiring an empire on the other side of the globe; nor is there any reason to suppose that many members of Congress or of the public at large contemplated such an outcome of the war. Dewey's dramatic victory at Manila, however, had called sudden attention to a spot of great potential strategic importance. Theodore Roosevelt and his friend, Sen. Henry Cabot Lodge, devotees of the sea power doctrines of Capt. A. T. Mahan, saw in Manila Bay a base that might greatly enhance the sea power of the United States. Recent European aggressions in China seemed to many businessmen to threaten the American market. Manila appealed to them as a base from which U.S. interests in China could be defended. The Protestant churches saw the easy victory at Manila as a divine summons to missionary work in the Philippines. The British and Japanese governments, furthermore, let it be known that they would be pleased to see the United States keep the islands. There were difficulties about any other course. Restoration of Spanish rule promised only chaos similar to that from which Cuba had just been rescued: Nor did the Filipino people have the education, training, or experience needed for successful self-government, though an active group headed by Emilio Aguinaldo had actually proclaimed independence.

Swayed by these varied considerations and by his appraisal of popular sentiment, McKinley decided, after long deliberation, that the United States must take possession of the entire archipelago of about 7,000 islands and 7,000,000 inhabitants. This demand was reluctantly agreed to by Spain, with the stipulation that the United States should pay Spain $20,000,000, nominally for public buildings and public works in the Philippines. The treaty, signed Dec. 10, 1898, conformed to these terms. Spain relinquished Cuba and ceded to the United States the Philippines, Puerto Rico, and Guam in the Ladrones. The treaty was vigorously opposed in the U.S. Senate as inaugurating a policy of "imperialism" in the Philippines and was approved on Feb. 6, 1899, with a margin of a single vote. Two days earlier hostilities had begun at Manila between U.S. troops and Aguinaldo's insurgents. For over three years the Filipinos carried on guerrilla warfare against U.S. rule. (*See* PHILIPPINES, REPUBLIC OF THE: *History*.)

The Consequences.—The Spanish-American War, short as it was, and relatively inexpensive in both resources and human life, was an important turning point in the history of both antagonists. Though disastrous for Spain in immediate results, it was followed by a remarkable renaissance in Spanish life, both intellectual and material. As Salvador de Madariaga has written: "Spain felt then that the era of overseas adventures had gone, and that henceforth her future was at home. Her eyes, which for centuries had wandered to the ends of the world, were at last turned on her

own home estate." (Salvador de Madariaga, *Spain: a Modern History*, Frederick A. Praeger, Inc., New York, 1958.) There ensued two decades of significant progress in agriculture, development of mineral resources, industry, and transportation. At the same time there came to the front a brilliant group of thinkers and writers, the "Generation of 1898," who gave Spain a prominence in the intellectual and literary life of Europe that it had not enjoyed for centuries.

For the victorious United States the consequences were entirely different. It is a truism that the United States became a world power as a result of the war. It now had insular possessions in the Caribbean and stretching across the Pacific, including Hawaii, whose peaceful annexation had been hastened by the war. The war made certain that an American canal would be cut through the Isthmus of Panama. It stimulated enthusiasm for the Navy, which soon grew from fifth or sixth to second place among the world's war fleets. It prompted drastic reform in the U.S. Army, which had been poorly prepared for war and had lost far more men to disease than to enemy weapons. It also promoted the career of the nation's first world-minded president, Theodore Roosevelt. Within a few years the U.S. had made the Caribbean an American lake, was taking a leading part in the politics of the Far East, and was preparing in spite of itself to play a determining role in the affairs of Europe.

See also references under "Spanish-American War" in the Index.

BIBLIOGRAPHY.—Orestes Ferrara, *The Last Spanish War: Revelations in Diplomacy* (1937); F. B. Freidel, *The Splendid Little War* (1958); Margaret Leech, *In the Days of McKinley* (1959); Salvador de Madariaga, *Spain: a Modern History* (1958); Walter Millis, *The Martial Spirit* (1931); J. W. Pratt, *Expansionists of 1898* (1936); H. F. Pringle, *Theodore Roosevelt: a Biography* (1931). (J. W. PR.)

SPANISH LANGUAGE. Among the modern languages descended from Latin (*see* ROMANCE LANGUAGES), Spanish shows the greatest vitality and promise of future importance over a vast territory. In the Christian north of the Iberian Peninsula, the earliest Hispanic vernacular texts (glosses) go back to the 10th century, the first works of fine literature, displaying strong dialectal colouring, to about 1150. By 1270 Alfonso X the Wise, king of Castile and León, created standard Old Spanish. The late 15th century witnessed the introduction of preclassical Spanish. Classical Spanish, the vehicle of the Golden Age literature (*c.* 1530–1680), was followed by Modern Spanish.

Spanish is also known (paradoxically, more so in Latin America than in Spain proper) under the name of Castilian. Historically, this is a reminder of the fact that at the outset it represented merely one northern variety of Hispano-Latin folk speech. It emerged after A.D. 800 in Old Castile (around Burgos), then by the 11th century spread fanwise southward to New Castile (Toledo, Madrid) in the wake of the Christian "reconquest" and, especially after the merger of Castile with León and in the late 15th century with Aragon, became the official language of all Spain. In the peninsula, Castilian has gradually crowded out, except in secluded rural districts, the Hispano-Roman dialects of León, Asturias, Santander, Navarre, and Aragon.

In the south, where the Moorish occupation lasted generally until the early 13th century (much longer around Málaga and Granada), Western Arabic dialects were used among Muslims, while a strikingly archaic Hispano-Roman dialect ("Mozarabic") survived until about 1100 amid the dwindling Christian minority. Modern southern dialects (Extremeño, Andalusian, Murcian) represent prongs of Castilian overlying dimly recognizable earlier Romance speech forms.

The language of the 500,000 Spanish Jews (found chiefly in the Balkan Peninsula, North Africa, Israel, New York City, Buenos Aires), called Ladino or Judeo-Spaniolic, is based squarely on 15th-century Spanish, containing heterogeneous admixtures. Portuguese, in the extreme west, and Galician, confined to the northwest, jointly form a language apart, with close grammatical and lexical, but not phonic, ties to Spanish. In the northeast, Catalan, including Balearic and diluted Valencian, constitutes a separate Romance language, influenced by Provençal. Basque, a relic of a pre-Indo-European family, straddles the western Pyrenees.

Outside the peninsula, Spanish is spoken or understood in virtually all South America except Brazil, in most of continental and insular Central America, in the Canary Islands, in the Tangier zone and adjoining sections of Morocco, along stretches of the Atlantic coast of equatorial Africa, and, on a decreasing scale, in the Philippines. Its hold on the New World is secure: if it has yielded ground to Portuguese in southern Brazil and is retreating before English in California, Arizona, New Mexico, Colorado, and Texas, it is steadily gaining ground, by compensation, in areas long dominated by indigenous Indian languages (Paraguay and the Yucatán Peninsula of Mexico are characteristic bilingual zones) and is eagerly accepted by immigrants' families in Buenos Aires and Montevideo. The influx of Cubans and, later, Puerto Ricans has made Spanish the second-ranking language in the New York metropolitan area. The latest political events have brought another wave of Cuban refugees to southern Florida.

Phonology and Orthography.—The sound system is characterized by five sharply delineated, short to medium-length vowels, lacking relevant variants (*i, e, a, o, u*). Two important rising diphthongs (*ie, ue;* also, only after *c, ua*), typically confined to the stressed syllable, outbalance several infrequent falling diphthongs (*ai, au, ei, eu, oi, ui*). Consonants, except for *n* in special cases (*ennoblecer, perenne*), are not doubled: *abad, Meca, ilustrar; rr* intervocalically contrasts with *r* (*perro* ~ *pero*), forming with *r* word-initially (*rey*) and after *n, s* (*Enrique, Israel*) a separate sound normally produced by five vibrations of the tongue. Voiceless stops (*p, t, k;* the last spelled mostly *c* or *qu*, according to environment) are uttered virtually alike in all positions. The corresponding voiced stops *b, d, g* are reduced to spirants between vowels, as in *abogado*. The typical syllable consists of consonant plus vowel; hence few consonants occur word-finally (*-d, -l, -n, -r, -s, -z*, anciently also *-ch, -ll, -ñ*, etc.); of these, *-d, -n, -r*, and especially *-s* are inflectionally important. Dialect speech goes further, allowing *-d, -r, -s* to disappear, or *-l* and *-r* to merge.

Until *c.* 1600, Spanish contained a greater variety of consonants: *x* in *exe* (mod. *eje*) equaled English *sh*, *j* in *ojo* was like *z* in English "azure" (these two merged, and their product became [χ], like *ch* in German *Bach*). The affricates *ç* [=ts] and *z* [=dz] collided; their amalgam was pronounced either [θ], like *th* in "thin" (most of Spain), or like voiceless *s* in "mass" (part of Andalusia, Spanish America); nowadays, *c* before *e, i*, and *z* otherwise are merely variant spellings of the same sound, as are also *g* before front vowels and *j*. The lone remaining affricate *ch*, pronounced as in English, represents a unit phoneme and ranks as a separate letter. Voiced *s*, as in English "as," became voiceless between vowels (*casa*), surviving where adjacent to another voiced consonant (*desde*). The [v] sound, as in Old Spanish *bever, Córdova*, has disappeared; today *v*, tolerated for historical reasons, duplicates the orthographic role of *b:* both letters represent the same sound, released as a stop or as a spirant according to its position. Colloquially, palatal *l* (conventional spelling, *ll*, rated as a single letter) tends to become the semiconsonant [*j*], more rapidly so in the new world than in Spain.

Noteworthy are the traditional frequency of *a* and *s* and the infrequency of stops and of *f* (compare *hijo* "son" with its counterparts *filho, fils, figlio*, etc., in cognate languages). The sound [h], absent from standard Spanish, occurs dialectally, in local equivalents of words like *holgar* and *ojo*. Hence the letter *h* is predominantly ornamental, being used in deference to Latin tradition (*hora*) or as a holdover from earlier native practice (*hierro*, from older *fierro*); it is functional only in the segment *hue-*, where its role is comparable with that of mute *u* in *gue-, gui-*, and as an ingredient of *ch*.

Grammar.—Inflection, derivation, composition, and syntax show the typical features of a conservative Romance language, pointing up its resemblance to Portuguese, Old (but not Modern) French, and Italian as regards, for instance, the sparing use of definite and indefinite articles and of subject personal pronouns, the scope of imperfect and preterit, the formation of future and conditional, the partial dependence of the gender on the final vowel. Pronouns alone show traces of declension (*yo, me, mí*) and, originally of the "neuter" (*lo mío, lo otro;* secondarily, *lo bueno*). Plural formation depends on interplay of termination and accen-

tual pattern (*hombre-s, mujer-es, cortés ~ corteses, lunes* invariable). Nouns and pronouns in *-o* produce feminines in *-a;* otherwise the feminine is more sharply characterized in substantives (*vendedor-a*) and in adjective-substantives (*trabajador-a*) than in pure adjectives (*cortés* invariable).

The conjugation is based on a closely integrated system of tenses (simple or compounded with *haber*) and moods (indicative, subjunctive; the latter suggesting uncertainty or purpose); the reflexive construction has overlaid the passive voice. Peculiarly modern are the almost interchangeable use of two past subjunctives (*-se, -ra*) and the obsolescence of the future subjunctive. The only productive conjugation class is *-ar;* for verbs derived from substantives: *-ear* (*telefonear*); the *-er* and *-ir* verbs form closed groups. Stray relics of the Latin paradigm are found among "strong" (=radical-stressed) preterits (*dije, hice*) and past participles (*dicho, hecho*).

Spanish strikes the speaker of English through its extraordinary wealth of derivational models, a feature in which it equals Italian and surpasses French by a wide margin. Abstracts, for instance, can be derived from adjectives in a wide variety of seemingly unpredictable ways: *pobre* → *pobr-eza* "poverty," *mudo* → *mud-ez* "dumbness," *sordo* → *sord-era* "deafness," *calvo* → *calv-icie* "baldness," *avaro* → *avar-icia* "avarice," *pleno* → *plen-itud* "fullness," *pesado* → *pesad-umbre* "heaviness," *bizarro* → *bizarr-ía* "gallantry," *glotón* → *gloton-ería* "gluttony," *loco* → *loc-ura* "madness," *verde* → *verd-or* (poetical) "greenness"; and there exist comparably elaborate devices to transform a verb into an abstract, to extract a female agential from its male counterpart, to designate a fruit tree or an orchard using as the basis the name of the specific fruit or vegetable, etc. The typical diminutive suffix is *-ito, -ita* (*cas-ita, señor-ita*) in preference to older *-illo*, but some dialect speakers favour *-ino, -ico*, or *-uco*, while for names of young animals, especially whelps, one finds also *-ezno* and *-ato*, hence *lob-ezno* "wolf cub" beside *ballen-ato* "whale calf." Prefixes, all Latin in origin, are few, colourless, and by and large confined to verbal stems: *a-, des-, en-, entre-, es-* (*ex-*), *re-, so-, sobre-, tras-;* rarely *ante-, per-, pos-*. Compounds are on the wane, except for the one pattern *matafuego* "fire extinguisher" (literally "killfire"; *cf.* English *killjoy, lack-luster*), *rascacielos* "skyscraper."

Vocabulary.—Viewed in the historical (stratigraphic) perspective, only a handful of important Spanish words, including some animal names, may be safely ascribed to the chronologically oldest layer, having been absorbed by Hispano-Romans from the language of ancestral Iberians and Celtiberians or of neighbouring Basques; *e.g., ard-illa* "squirrel," *bruja* "witch," *izquierdo* "left," *perro* "dog," *pizarra* "slate," *zorro* "fox." The core of the lexicon consists of a relatively archaic layer of Latin words, reflecting (frequently in contrast to French and to Italian) the preference of colloquial Latin throughout the last two centuries of the Republic; characteristic are *hermoso* "beautiful" as against French *beau*, Italian *bello; arena* "sand" as against French *sable*, Italian *sabbia; día* "day" as against French *jour*, Italian *giorno;* and *comer* "to eat" as against French *manger*, Italian *mangiare*. Moreover, Latin, as the medieval language of church, administration, and scholarship, contributed numerous learned and semilearned words, which show "retarded" sound development, having skipped certain evolutionary stages; *e.g., iglesia*, Old Spanish *eglesia* "church." Since the early days of humanism, there has been an incessant flow of Latinisms and Hellenisms into Spanish, less numerous, however, by a considerable margin than those absorbed into English (thus, Spanish lacks counterparts of "obstreperous" and "preposterous" and infrequently uses those of "acrimonious" and "exacerbate").

Additional lexical strains that have contributed to the present wealth of the Spanish vocabulary are traceable to Gothic (*hato* "herd," *ropa* "clothing," *brotar* "to sprout," and numerous proper names, including *Alfonso, Fernando*, and *Elvira*); to Byzantine Greek (*botica* "drugstore"); to Arabic (frequently recognizable by the initial segments *al-, a-* or the termination *-í: al-calde* "mayor," *a-zucena* "lily," *balad-í* "frivolous," "paltry"); and to Old French and Old Provençal, especially words relating to the monastic and the chivalrous or courtly spheres (*monje* "monk," *fraile, Fray, Frey* "friar," *ruiseñor* "nightingale," *vergel* "flower and fruit garden" and others ending in *-el*).

Italian was very influential in the period 1450–1625, supplying hundreds of specific terms which, in many instances, continue to be used in numerous arts, crafts, and trades, such as painting, navigation, and banking. French reached the peak of its influence in the 18th century (when the controversial Afrancesados introduced short-lived *remarcable, retornar*, etc.). English, in particular U.S., civilization began to colour the lexicon of sports, science, commerce, and politics after 1900. Over the centuries, the indigenous New World languages have supplied only terms in restricted domains (mostly fauna and flora).

The Royal Spanish Academy has exercised a restraining influence on the spontaneous development of the language through the 18 successive editions of its dictionary, much less so through its normative grammar, at no time nearly so influential as A. Bello's and V. Salvá's. Of late, the national academies of the Spanish American countries have established a salutary rapport with the Madrid Academy.

BIBLIOGRAPHY.—*Frequency Data:* (sounds) T. Navarro, *Estudios de fonología española* (1946); (phraseology) H. Keniston, *Spanish Idiom List* (1929), (constructions) *Spanish Syntax List* (1937) and *The Syntax of Castilian Prose: the Sixteenth Century* (1937); (vocabulary) V. García Hoz, *Vocabulario usual, común y fundamental* (1953). *Phonetics:* T. Navarro Tomás, *Manual de pronunciación española*, 4th ed. (1950). *Historical Phonology:* A. Alonso, *De la pronunciación medieval a la moderna en español*, vol. i (1955 *et seq.*). *Phonemics:* E. Alarcos Llorach, *Fonología española*, 3rd ed. (1961). *Grammar:* (inventory of difficulties) M. M. Ramsey, *A Textbook of Modern Spanish*, rev. by R. K. Spaulding (1956); (historical) R. Menéndez Pidal, *Manual de gramática histórica española*, 9th ed. (1952). *Language History:* (introductory) R. Lapesa, *Historia de la lengua española*, 4th ed. (1959); (advanced) R. Menéndez Pidal, *Orígenes del español*, 4th ed. (1956). *Syntax:* (descriptive) S. Fernández Ramírez, *Gramática española: los sonidos, el nombre y el pronombre* (1951). *Lexicon:* (standard dictionary) Real Academia Española, *Diccionario de la lengua española*, 18th ed. (1956); (bilingual) E. B. Williams, *Holt's Spanish and English Dictionary* (1955); (overseas) A. Malaret, *Diccionario de americanismos*, 3rd ed. (1946); F. J. Santamaría, *Diccionario general de americanismos*, 3 vol. (1942); (etymological) J. Corominas, *Diccionario crítico etimológico de la lengua castellana*, 4 vol. (1954–57). *Historical and Regional Varieties:* (substratum) R. Menéndez Pidal, *Toponimia prerrománica hispana* (1952). *Foundations:* (Romance) A. Kuhn, *Die romanischen Sprachen* (1951); C. Tagliavini, *Le origini delle lingue neolatine*, 3rd ed. (1959); (Hispano-Arabic) E. K. Neuvonen, *Los arabismos del español en el siglo XIII* (1941); (medieval) R. Menéndez Pidal, *Cantar de Mio Cid*, rev. ed. (1944–46); (new world) R. J. Cuervo, *El castellano en América* (1946) and *Apuntaciones críticas sobre el lenguaje bogotano*, 9th ed. (1955); A. Alonso (ed.), *Biblioteca de dialectología hispanoamericana*, with emphasis on Argentina, Santo Domingo, old and new Mexico, 7 vol. (1930–49); C. E. Kany, *American-Spanish Syntax*, 2nd ed. (1951): draws on written sources; (Gypsy-Spanish) C. Clavería, *Estudios sobre los gitanismos del español* (1951); (Judeo-Spaniolic) C. M. Crews, *Recherches sur le judéo-espagnol dans les pays balkaniques* (1935). *Bibliography:* (general) H. C. Woodbridge and P. R. Olson, *A Tentative Bibliography of Hispanic Linguistics* (1952); H. Serís, *Bibliografía de la lingüística española* (1964); (specialized) M. W. Nichols, *A Bibliographical Guide to Materials on American Spanish* (1941); (current) in the two leading journals, *Revista de Filología Española* (Madrid) and *Nueva Revista de Filología Hispánica* (Mexico City).

(Y. M.)

SPANISH LITERATURE. The literature of Spain falls into three divisions according to language: Castilian, Catalan, and Galician (*see* SPANISH LANGUAGE; PORTUGUESE LANGUAGE; CATALAN; GALICIA). This article provides a brief historical account of the development of each of these three literatures. For the historical background, *see* SPAIN: *History;* and articles on the main kingdoms and provinces of Spain (*e.g.*, ARAGON; CASTILE; CATALONIA; LEÓN). For the literature of the former Spanish colonies in the Americas, *see* IBERO-AMERICAN LITERATURE. *See* also biographies of many writers mentioned; the general articles DRAMA: *Modern Drama* and NOVEL: *Early Modes of Narrative and the Novel;* and articles on other forms, such as SATIRE and SONNET.

Although literature in the vernacular did not begin to be written until the medieval period, Spain had already made considerable contributions to literature. The two Senecas, Lucan, Martial, Quintilian, and Prudentius are among writers in Latin who lived in,

or were born in, Spain before the separation of the Romance languages (*q.v.*). For their writings, *see* LATIN LITERATURE. Later, the writings of Spanish Muslims and Jews form an important branch of Arabic and Hebrew literature (*qq.v.; see* also JEWISH PHILOSOPHY).

This article is divided into the following main sections:

I. CASTILIAN LITERATURE

A. MEDIEVAL PERIOD

1. The Beginnings of Vernacular Literature.—By the time of the Muslim invasion (beginning in 711), the Latin spoken in the Iberian Peninsula was in process of transformation into Romance. The 10th-century *Glosas emilianenses* and *Glosas silenses* —marginal notes in a vernacular in manuscripts belonging to the monasteries of San Millán de la Cogolla and Silos, in La Rioja, and preserved in the Academia de la Historia, Madrid, and the British Museum, London, respectively—reveal traces of a vernacular already substantially developed. Mocádem, a blind poet living near Córdoba at the beginning of the century, was recorded 200 years later as having taken "expressions in vulgar Arabic or Romance, called them *markaz,* and built on them his *zéjels* or *muwashshahs.*" As a result of research since 1948, numerous examples of the *markaz* (theme stanza), written in a Romance idiom transliterated into Hebrew or Arabic, have been recovered from Hebrew and from Arabic *muwashshahs* (*q.v.*) from the 11th century onward. These provide not only the earliest texts in Mozarabic (the Romance dialect of Spaniards living under the Muslims) but also evidence of a popular poetry that may have begun in the time of Mocádem or earlier: they can now be seen to explain much in the traditional Spanish lyric types (*e.g.,* the *villancico,* "carol") of the later Middle Ages and the Renaissance. The *markaz* as a whole was generally a love song sung by the woman (comparable to the later Galician *cantiga de amigo; see* PORTUGUESE LITERATURE), and the motif, in Romance, was a cry of passion on which the whole poem was an elaboration or up to which it led.

2. The Rise of Heroic Poetry.—The earliest surviving monument of Spanish literature, and one of its most distinctive masterpieces, is the *Poema* (or *Cantar*) *de mío Cid,* an epic poem (*cantar de gesta*) of the mid-12th century (the existing manuscript is an imperfect copy of 1307). This work, of 3,735 irregular lines (of 11–18 syllables, with marked caesura and traces of a four-beat accentual rhythm), grouped in assonating *laisses,* tells of the fall from and restoration to royal favour of the Castilian noble Rodrigo Díaz de Vivar (*c.* 1043–99) known by the Arabic title of *sidi* ("lord"). The historical setting and personages, the topographical detail, the realism of tone and treatment, and the proximity in time of the anonymous poet to his hero have made possible the advancing of "historicity" as a first characteristic of this poem, and of the Castilian epic in general. The latter two of its three *cantares* (sections), however, while as convincing in detail as the first, are untrue to recorded history; the conception of the whole work is essentially poetic and handled with much imaginative artistry; and the according of a mere seven lines to the Cid's greatest achievement, the taking of Valencia from the Muslims after a nine-months' siege in 1094, makes it clear that the scale of values is subjective. The poem caught the popular imagination and lived on in epic, chronicle, ballad, and drama, helping to fix the popular conception of the Castilian character. The only other surviving epic text (apart from a 100-line fragment of uncertain filiation on *Roncesvalles,* of the early 13th century; *see* ROLAND, CHANSON DE) is the wholly fanciful and decadent 14th-century *Cantar de Rodrigo* (also called the *Crónica rimada*), telling of the Cid's early manhood; it was from this that the later legend of the Cid took shape. (*See* also CID, THE.)

Frequent allusions in vernacular chronicles to the heroic narratives of minstrels (*juglares*) make it clear that the *mester de juglaría* ("minstrel's craft") was much richer than the scarcity of manuscripts would suggest. Thanks to the vernacular chronicles, which, from their beginning late in the 13th century, accepted these narratives as being of historical validity and "prosified" them at length, it has been possible to reconstruct the themes and even some fragments of the text of epics on *Los siete Infantes de Lara* ("The Seven Princes of Lara") and *Fernán González* (both set in the 10th century; the first about seven brothers who were betrayed by their uncle, killed in battle, and later avenged; the second about a Castilian hero; *see* below); *El Cerco de Zamora* ("The Siege of Zamora"; again involving the Cid, and perhaps artistically the best); and others. *Bernardo del Carpio,* about a fictitious Spanish warrior who helped to defeat Roland at Roncesvalles, provides a hero in defense of Spanish prestige against the French Carolingian epic; the other themes are an integral part of the feudal history of 10th- and 11th-century Castile, close to the time of telling and less probably related in their origins to the very different French epics (*chansons de geste; q.v.*) than to a remote Visigothic past.

3. Learned Narrative Poetry.—A new school of erudite poetry, much indebted to France and linked not with the itinerant *juglar* reciting to popular audiences but with the monastery and a literate public, became known as the *mester de clerecía* ("clerkly craft"). Regularly adopting the French Alexandrine in the set mold of the *cuaderna vía* ("fourfold way," *i.e.,* four-line single-rhyme stanza with a 14-syllable line), and dealing with religious, didactic, or pseudohistorical matter, it was a product of the study. It is best and probably first exemplified in Gonzalo de Berceo (*c.* 1195–*c.* 1268), the earliest Spanish poet known by name. Apologizing for inadequate Latinity, he versified unwearyingly in the vernacular the lives of Spanish saints, the miracles of the Virgin, and other devotional themes with an ingenuous candour and an accumulation of picturesque and affectionately observed detail that largely atone for his prolixity (about 13,000 lines) and fidelity to his sources. Also of the 13th century and in the same form are the *Libro de Alexandre* (*see* ALEXANDER ROMANCES), much more erudite and pretentious; the *Libro de Apolonio,* a version of the Latin romance on Apollonius of Tyre (*q.v.*); and the *Poema de Fernán González,* a clerical reworking of a lost epic on the 10th-century count of Castile who won its independence from León.

4. The Beginnings of Prose.—A major influence on prose was exercised by Arabic. The gates of Oriental learning and story were opened to Spain by the capture (1085) from the Muslims of Toledo, which became a centre of translation from Oriental languages. In 1120 Pedro Alfonso (Petrus Alfonsi, Rabbi Moisés

Sefardí), an Aragonese Jew converted to Christianity, introduced Eastern fable to the non-Arabic reader with his *Disciplina clericalis*, a collection of tales, with ethical instruction, translated into a barbarous Latin. The anonymous translation from the Arabic (1251) of the "beast fable" *Kalila e Dimna* is the first essay in storytelling in the Spanish language. (*See* also FABLE; BIDPAI, FABLES OF.) The Oriental romance of the Seven Sages, or Seven Wise Masters (*q.v.*), was translated from a Latin text as the *Libro de los engannos e assayamientos de las mujeres* (1253; also known as the *Sendebar;* Eng. trans., *The Book of the Wiles of Women,* 1956). Other collections of Eastern stories followed.

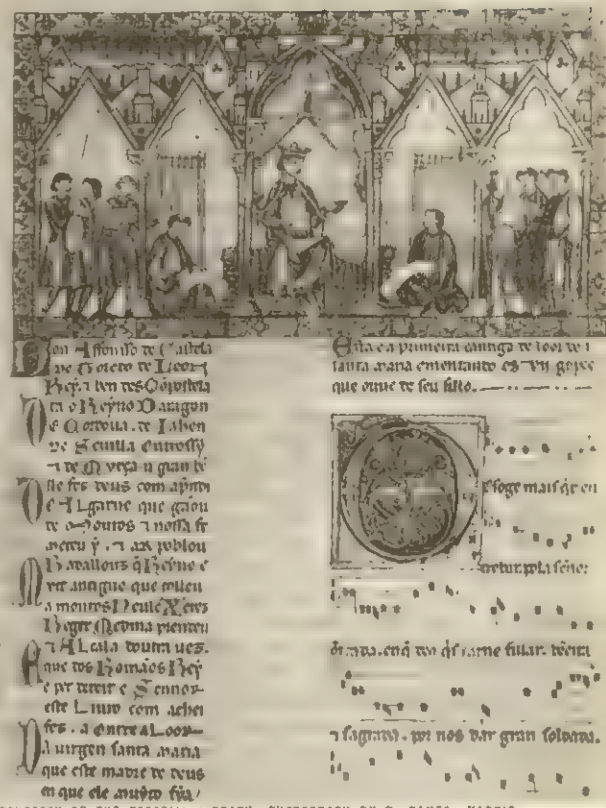
5. Alfonso the Wise.—The middle of the 12th century saw the recovery of Córdoba, Valencia, and Seville by the Christians, the kingdom of Granada alone remaining in Muslim hands. An indication of the more propitious intellectual atmosphere resulting was seen in the founding of universities (*e.g.,* Salamanca *c.* 1218; charter, 1243); and under the stimulus of Alfonso X of Castile and León (Alfonso the Wise [*El Sabio*]; 1252–84) literature achieved official prestige. Alfonso, in whose chancery Castilian replaced Latin as the language for public documents, may be described as the father of Castilian prose. He organized a vast enterprise of translation and compilation aimed at fusing in the vernacular all available knowledge, classical, Oriental, Hebrew, and Christian. The works undertaken, often under his personal editorship, included the great legal code *Las siete partidas,* published in sections, a mine of information on the life and customs of the time, and compilations from Arabic sources such as *Los libros del saber de astronomia, Los tablas alfonsíes* (on astronomy), *El lapidario* (a treatise on the magical properties of precious stones), and *El libro de los juegos* (a famous medieval work on games, especially chess). With *La primera crónica general* (or *Crónica general*), covering the history of Spain from its first settlers, and the *General Estoria,* an attempt at a universal history from the Creation, Alfonso was the founder of Spanish historiography.

The *Crónica general,* carried by Alfonso to A.D. 711 and completed by his son, Sancho IV, gave rise to a series of chronicles over the two centuries following and was the most influential single work of the Spanish Middle Ages; the *General Estoria* did not reach later than the parents of the Virgin Mary. Himself a poet, Alfonso was responsible too for one of the greatest collections of medieval poetry and music, the *Cantigas de Santa María,* in Galician (then the accepted language for lyric; *see* below).

6. The 14th Century.—The period of translation and compilation was succeeded by one of brilliant original creation. It is best represented in the prose of Alfonso's nephew, the infante Don Juan Manuel (1282–1348?), and in the poetry of Juan Ruiz (*c.* 1283–*c.* 1350), archpriest of Hita. Juan Manuel's *Libro de los exemplos del Conde Lucanor et de Patronio* (1323–35), a collection of 50 moral tales, still drew heavily on Arabic sources; but by its individuality of style and treatment it ranks high in the beginnings of Spanish fiction.

Earlier, under the influence in part of the Arthurian, or Breton, cycle (*see* ARTHURIAN LEGEND), which from the turn of the 13th century had been circulating in Spain in translation, there had appeared (*c.* 1305) the first Spanish romance of chivalry (*libro de cabellaría*), and the first full-length Spanish novel, *El caballero Cifar,* on the Eustace theme (*see* ROMANCE). Cifar's squire, Ribaldo, is a distant precursor of Don Quixote's Sancho Panza, and thus an early token of the constant impinging of the real on the ideal so characteristic of Spanish literature. Probably at about the same time there was circulating the famous *Amadís de Gaula* (*q.v.*), a chivalric romance also related to the Arthurian cycle, and destined to father a progeny that held the imagination of kings, courtiers, and scholars throughout the 16th century. Sentimental idealism, lyrical atmosphere, and supernatural adventure notwithstanding, the traditional ascription of this to Portugal (*see* PORTUGUESE LITERATURE) is, at best, inconclusive. The author may have been a Galician or Portuguese troubadour in exile at the Castilian court.

Juan Ruiz, a man of the people, was the most intensely alert



BY COURTESY OF THE ESCORIAL LIBRARY; PHOTOGRAPH BY D. RAMOS, MADRID

THE LITERARY COURT OF ALFONSO THE WISE

The first illumination from a 13th-century manuscript of the *Cantigas de Santa María,* a collection of medieval poetry and music compiled by Alfonso X

and individualized of early writers. His *Libro de buen amor* (1330; expanded, 1343) represents a collection of disparate elements—Ovid, Aesop, the 12th-century Latin *Pamphilus de amore,* the 13th-century *Carmina Burana,* and the liturgy—with a strong suggestion of Muslim ways of thinking in the mingling of eroticism with devotion, and the invitation to the reader to interpret for himself the author's equivocal teachings. But this diversity of adventure was presented as a satiric autobiography and stamped with the author's exuberant personality; and his Trotaconventos, ancestress of La Celestina (*see* below), is the first great character creation in Spanish literature. The dominant metre was still the *cuaderna vía,* handled with a new narrative vigour and plasticity; but the interspersing of the text with lyrics, religious, pastoral-farcical, amorous, or satirical, of great metrical variety, pointed to its approaching supersession.

Continuing evidence of the exotic strands that were being woven into the texture of Spanish literature may be found in the anonymous *Poema de Yuçuf* (also assigned to the 14th century), and in the *Proverbios morales* (*c.* 1355) of Šem Tob ibn Ardutiel ben Isaac (Santob de Carrión de los Condes; *c.* 1290–*c.* 1369). The former, an Aragonese version in *cuaderna vía* of the story of Joseph, is based not on the Bible but on the Koran, and is written in Arabic characters; it is the chief representative of the so-called *aljamiado* ("barbarian") literature, Spanish in language, Arabic (or Hebrew) in tone and script. The *Proverbios* of Šem Tob, comprising 725 Alexandrine couplets with internal as well as end rhyme, introduced to Spanish the grave sententiousness of Hebrew poetry with its extreme aphoristic concision; his chief sources were the Old Testament, the Talmud, and the Hebrew poet and Arabic philosopher Ibn Gabirol.

The chancellor of Castile, Pedro López de Ayala (1332–1407), dominates the later part of the 14th century both in poetry—with his *Rimado de palacio,* the last major relic of the *cuaderna vía,* the 8,200 lines of which distilled the essence of a lifetime of experience and disillusion in high places—and in prose as a his-

torian of things seen. His chronicles (the last incomplete) of the reigns of Peter I (Pedro the Cruel), Henry II (of Trastámara), John I, and Henry III of Castile infused new life into the tradition of the Latin *cronicón*, which dates back in Spain to the 5th century, and greatly stimulated the writing of personal, contemporary history. An early Humanist, he translated and imitated Livy and Boccaccio, Boëthius, St. Gregory, and St. Isidore.

7. The 15th Century.—The early 15th century witnessed a tentative renewal of poetry under Italian influence. The contrast is strong, during the reign of John II (1406–54), between the social anarchy of a feudalism in its death throes and the cultivation of polite letters, which was becoming an expected mark of birth and breeding. Collections such as the *Cancionero de Baena*, made for the king in 1445 by Juan Alfonso de Baena (a converted Jew, and himself a poet), containing 583 poems by 55 poets ranging from the highest nobles to the humblest versifiers on the fringes of their service, show not merely the decadent tradition of the Galician-Portuguese troubadour school but also the new stirrings of a much more intellectual poetry using symbol, allegory, and classical allusion in the treatment of themes of high moral, philosophical, or political intent. Francisco Imperial, son of a Genoese settled in Seville, and a leader of the new poetry, drew on Dante to present a challenge taken up by the marqués de Santillana (1398–1458), whose work was included in the *Cancionero;* Juan de Mena (1411–56); and others. Santillana, poet and scholar, soldier and statesman, was a collector of the masterpieces of foreign literatures and a stimulator of translation. His *Proemio e carta al condestable de Portugal* ("Preface and Letter to the Constable of Portugal"; 1449), the earliest work of literary history and criticism in Spanish, draws on his reading in contemporary foreign languages, and in the translated classics. Imperial had already sought to acclimatize the Italian hendecasyllable. Santillana's 42 sonnets *"fechos al itálico modo"* (*i.e.*, in the Italian style) marked the beginnings of the formal enrichment of Spanish poetry at the dawn of a new age. His role as precursor is still outstanding, though his sonnets, like his long poems *La comedieta de Ponza*, *El diálogo de Bias contra fortuna*, and *La defunción de don Enrique de Villena*, in thought and structure likewise showing the poet at school in Italy, are neglected in favour of his charming *canciones*, *decires*, and *serranillas* (rustic songs describing an encounter between a nobleman and a mountain girl) of native inspiration.

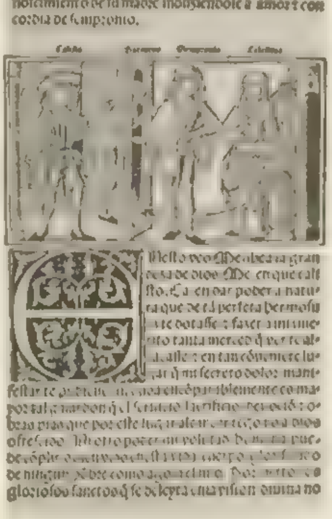
Juan de Mena's vast allegorical poem of the drama of history past, present, and to come (*El laberinto de fortuna*, 1444) was a more conscious attempt to rival Dante; felicitous often in isolated narrative passages on Spain's heroic past, it is weighed down by pedantry, and by over-Latinization of syntax and vocabulary. His metre, the common vehicle of learned poetry throughout the 15th century, is the *verso de arte mayor*, characterized less by its variable length (9–14 syllables, with 12 as the norm, divided into two parts of 6 each) than by its two hemistichs and strong ternary rhythm; accentual verse is not found again in Spanish until the 19th century. The other outstanding 15th-century poet is Jorge Manrique (1440?–79), who achieved immortality with one elegiac poem, *Coplas por la muerte de su padre* (1476).

An outstanding anonymous poem of the early part of the 15th century, the *Danza de la muerte*, is the finest extant example of a kind popular in the Middle Ages with poets, painters, and composers—the Dance of Death (*q.v.*). It is related to the earlier *Danse macabre* at Paris, but is written with greater satiric force, and introduces characters (*e.g.*, a rabbi) not included in the French cycle. Consisting of 79 *coplas de arte mayor* (*i.e.*, stanzas of 8 lines, each of 12 syllables), it is in the form of a dialogue between inexorable death and his protesting victims—who range from pope and emperor, king and cardinal, to lawyer's clerk and humble cleric, and so present a cross-section of society. Although not intended for dramatic presentation, it formed the basis for later drama.

In prose the age saw the first Spanish books of travel, notably an account of a journey to Samarkand in 1403–06 (*Vida del gran Tamerlán*) by Ruy González de Clavijo (d. 1412), who had gone there as Henry III's ambassador; and an account of the travels in Europe and the Middle East (1435–39) of Pedro Tafur (*Las andanças e viajes por diversas partes del mundo* ["Adventures and Journeys in Different Parts of the World"]). Contemporary history took on a biographical dress in two admirable collections of portraits of eminent Castilians: *Generaciones e semblanzas* ("Generations and Biographical Sketches"; the third, and only original, part of a larger *Mar de historias* ["Sea of Histories"], written *c.* 1450; published 1517), by Fernán Pérez de Guzmán (*c.* 1376–1460), and the *Libro de los claros varones de Castilla* ("Book About Distinguished Castilian Men"; 1486) of Hernando del Pulgar (1436–93).

B. The Renaissance and the Siglo de Oro

1. The Beginning of the Golden Age.—The unification of Spain in 1479 under the "Catholic kings," and Columbus' discovery of the New World (1492), following the introduction of printing (1474), and concurrent with the full play of cultural traffic with Italy (where Naples had been since 1443 a dependency of Aragon), may be taken as opening the era of the Renaissance in Spain. With the exception of the more popular veins associated with oral tradition, the medieval poetic achievement lapsed into oblivion: the *Poema de mio Cid* and the corpus of the *mester de clerecía* were not printed until 1779; the mid-12th-century *Auto de los reyes magos* (*see* below) not until 1863. National heroic themes survived through ballad and chronicle, early beneficiaries of the printing press.



BY COURTESY OF THE HISPANIC SOCIETY OF AMERICA

PAGE FROM FIRST EDITION OF "TRAGICOMEDIA DE CALIXTO Y MELIBEA" ("LA CELESTINA"), 1499

What gave Spanish literature its distinctive mark in this period (the *siglo de oro*—Golden Age) was, however, the wealth of new experience born of overseas adventure, reflected in historical and scientific writing and in the detached and questioning attitude, so strikingly exemplified in the theatre, to the rediscovered norms of classical authority. The first Spanish Humanists were also the first grammarians and lexicographers of any Romance tongue: Antonio de Nebrija's grammar of the Castilian language (1492) was written under the sense of impending imperial responsibilities. Juan Luis Vives (1492–1540), a figure of European eminence, the brothers Juan (*c.* 1498–1541) and Alfonso (d. 1532) de Valdés, and others among the best brains and noblest minds of contemporary Spain were friends or followers of Erasmus (*q.v.*), whose writings circulated in translation from 1520 until their suppression after his death in 1536, and whose influence has been seen in St. Ignatius of Loyola, and, later, in Fray Luis de León (*see* below).

The masterpiece of the early Renaissance is the *Tragicomedia de Calixto y Melibea* (1499), a novel in dialogue form published anonymously but generally attributed to a converted Jew, Fernando de Rojas (d. 1541). The dominant character, a go-between elaborated from Juan Ruiz's Trotaconventos and depicted with a realism unsurpassed in Spanish letters, soon caused the work to be rebaptized, after her, *La Celestina* (*q.v.*), thereby obscuring its significance as a tragedy of idealism in the new world of the senses liberated by the Renaissance. The analysis of passion and the dramatic conflict with nemesis that its pursuit involves were worked out with such psychological intensity on the intersecting planes of the noble and the ignoble, of medieval and Renaissance, as to make this the first masterpiece of Spanish prose.

2. Romancero.—Spanish ballads (*romances*) form the strongest link in a chain of tradition from medieval heroic epic to 20th-century poetry and drama; they lie at the heart of the national

consciousness; and their expansion and capacity for survival, from Salonika to Chile, from the Low Countries to North Africa, reflect the far-flung boundaries of Spain's prestige in its age of greatness. The ballad form is clearly defined: the 16-syllable line divided into two equal hemistichs (or two octosyllables), with a single assonance, suggests a regularizing of the old epic versification, though it is doubtful whether (as some scholars have claimed) it may be inferred from this that the *romancero* originated in the fragmentation of the decadent epic. The earliest datable *romances* treat of frontier incidents (late 14th century) or lyrical themes (early 15th century), and it has been shown that only 5 out of about 35 "old" *romances* of the Cid bear relation to known epic sources, while two-thirds of the whole vast corpus are lyrical and betray French or other foreign influence. It is on the ballads on medieval heroic themes, nevertheless, that the peculiar importance of the kind in Spain hinges: they formed everyman's source book on national history and the national character. Circulating first as *pliegos sueltos* (*i.e.*, "broadsides") the traditional ballads were collected in the *Cancionero sin año* (printed at Antwerp, *c.* 1550) and in the *Silva de varios romances* (3 parts, 1550–51), and thereafter repeatedly. Soon the form was exploited for lyrical purposes in the *romances artísticos* of the most famous poets of the age, to remain the chosen medium for popular narrative verse.

3. Early Drama.—The origins of drama in Spain may be presumed, in default of evidence, to have been like those of drama in France and England, originating in the liturgy (*see* DRAMA: *Medieval Drama*). The *Auto de los reyes magos*, an incomplete play belonging to an Epiphany cycle from the mid-12th century, is the only surviving text of medieval Spanish drama, and the second surviving liturgical dramatic text in any vernacular (the first being the French *Jeu d'Adam* of the early 12th century). It is extant in a single 13th-century manuscript, and was apparently based on an earlier liturgical play written in Latin in Orléans, since it resembles it, and differs from other Epiphany plays, in some important respects (*e.g.*, that the three magi are introduced separately). The characterization of the magi, of Herod, and of his advisers is realistic; and the way in which the verse is manipulated foreshadows, although imperfectly, the later development of the drama in Spain.

That there was a tradition of liturgical drama in the vernacular may be inferred from references in legal and other documents, and a reference in Alfonso X's legal code (*Las siete partidas*, *c.* 1255; *see* above) encouraging the clergy to take part in liturgical plays, but forbidding their participation in *juegos de escarnios* (*i.e.*, burlesque pantomimes of sacred subjects) suggests also the existence of some form of secular drama, but no texts have survived. However, these *juegos*—short, satiric entertainments given by traveling players in marketplaces, or outside churches—were the forerunners of the later *farsas*, and of the numerous short plays and interludes—*paso, entremes, sainete*, etc.—that form one of the main Spanish contributions to dramatic genres.

Juan del Encina (1468–1529?) marked the first stage in the emancipation of the drama from ecclesiastical leading strings by way of performances for a noble patron. The *representaciones* and *églogas* in his *Cancionero* (1496) are pastoral-religious dramatic dialogues, in a rustic dialect, *sayagüés*, later accepted as a convention in comic pastoral; but he soon turned to the debating of purely secular themes or to vividly depicted farce. Three of his plays dating from a long stay in Italy (the *Églogas de Zambardo*, *de Cristino y Febea y Cardonio*, and *de Plácida y Vitoriano*) show Encina's conception of drama transformed by his Italian experience from ingenuous medievalism to bold Renaissance experimenting in the realm of sensuous indulgence, fortified by classical myth and allegory.

Encina's Portuguese disciple Gil Vicente (*c.* 1465–*c.* 1537), a court poet at Lisbon who wrote 11 of his plays entirely, and 17 partly, in Castilian, showed a great advance in naturalness of dialogue, acuteness of observation, and humour of situation. His weakness lay in plot, his distinction in a lyricism that expressed itself in the finest Spanish poetry of the period, and in a vivid, plastic representation of allegory that looks forward to Calderón. Bartolomé de Torres Naharro (1484?–1520?) went to Naples, and published there in his *Propalladia* (1517) six of his eight comedies, all performed before a noble audience. These show a sense of plot and a first awareness of dramatic theory, with a division in form into five acts (soon to be reduced to three) and in kind into *comedias á noticia* and *á fantasía* (the former realistic, as the *Comedia soldadesca* and the *Tinellaria*, satirizing respectively an army of occupation and the servants' hall in a cardinal's palace; the latter romantic). The *Serafina* and the *Ymenea* are of particular interest as foreshadowing the "cloak-and-sword" comedy, with the *pundonor* ("point of honour") as theme.

The emergence of the drama from court and nobleman's house to the marketplace and the creation of a public were largely the work of Lope de Rueda (*c.* 1510–65), who toured Spain with his modest troupe, his still more modest properties, and a repertoire of his own composing. His four prose comedies all owned Italian models and were clumsy in construction; it is in the ten *pasos* (comic interludes performed between the acts of longer plays) that his dramatic merit resides, and with it the distinction of fathering the one-act play which, constantly renewing itself, has some claim to be regarded as the most living and popular dramatic form in Spain.

The first dramatist to realize what might be made of the *romancero* in the theatre, in the sense of stirring the public's most responsive chord, was Juan de la Cueva (1543?–1610). Most of his 14 comedies and tragedies were taken from classical antiquity and written in the manner of Seneca, but in three (*Los siete infantes de Lara* ["The Seven Princes of Lara"], *El reto de Zamora* ["The Challenge of Zamora"], *La libertad de España por Bernardo del Carpio* ["The Liberation of Spain by Bernardo del Carpio"]) he turned to themes of Spain's early heroic story, already familiar in ballad, and thereby set an example that accounts in large part for the immense fertility of the *comedia*, and for the close accord thenceforth established between dramatist and public. He was not an accomplished dramatist, but he helped to found a drama "national" in subject matter and free from pseudoclassical influence.

4. Poetry.—The attempt to Italianize Spanish poetry had failed through an overambitious intellectualizing of theme, on a basis of pedantic imitation, when language and metrics were still incapable of sustaining the burden. The Catalan Juan Boscán Almogáver (*c.* 1490–1542; see *Catalan Literature*, below), by reintroducing Italian metres (sonnet, *ottava rima, canzone*, tercet, blank verse), prepared the way for a much greater poet, his friend Garcilaso de la Vega (*c.* 1501–36), with whom the lyric was reborn. The importance of the new metres went much beyond form: in Garcilaso the most significant novelty is the verse paragraph of the *silva* (a free combination of 7- and 11-syllable lines), allowing a new concern with the analytical expression of thought and emotion. To a mastery of the poetic process derived from a study of Petrarch, Boccaccio, and Sannazzaro (and also of Theocritus, Virgil, Tibullus, and Horace), he added the gift of infusing an intense personal note into characteristic Renaissance themes. His meagre output of eclogues, elegies, and sonnets won him recognition for centuries as "the prince of Castilian poets" and largely determined the course of lyric poetry throughout the *siglo de oro*.

Fray Luis de León (1527–91), adopting chiefly Garcilaso's *lira* stanza (7, 11, 7, 7, 11) for poems charged with emotional sincerity, typified a so-called "Salamanca school," with its emphasis on content rather than form. Fernando de Herrera (1534–97) headed a contrasting school of Seville which, deriving equally from Garcilaso, was concerned rather with Petrarchan subtleties of refined sentiment; in a quartet of remarkable odes he gave vibrant expression to topical heroic themes. A movement of protest in defense of the short native metres, associated chiefly with Cristóbal de Castillejo (1490?–1550) was greatly reinforced by the printing of the *romanceros* after *c.* 1550, and by the evolving drama, which, while employing much metrical variety for specific effects, found in the octosyllable the accepted vehicle for narrative and description. Throughout the 17th century poets were to write in both the old and the new styles and metres, their practice still al-

lowing a notable degree of equation between the old styles and metres and "verse," and between the new and "poetry."

For epic poetry the metre was the *octava real* (*ottava rima*, *q.v.*), and the models Ariosto and Tasso, but the themes and heroes were those of overseas conquest and expansion, or defense of the empire and of the faith at home. Alonso de Ercilla y Zúñiga (1533–94) came nearest to achievement with his *La Araucana* (1569–90), telling at firsthand of native resistance to the conquest in Chile; its defects spring from a wavering concept of epic theory, now martial-heroic unrelieved, now amorous-fantastical with admixtures of prophecy and pedantry. Bernardo de Balbuena's *Bernardo* (1624), on a familiar medieval theme, and Lope de Vega's *Dragontea* (1598; a verse history of Sir Francis Drake's last voyage and death), and *Corona trágica* ("The Tragic Crown"; 1627; on Mary, Queen of Scots), may stand for the multitudinous activity in a literary kind that laid its compulsion on every poet of ambition.

5. Prose.—Prose before the Counter-Reformation produced some notable Lucianesque dialogues, especially from Alfonso de Valdés (the *Diálogos de Lactancio y un arcediano* and *de Mercurio y Carón* ["Dialogues Between Lactancio and an Archdeacon" and "Between Mercury and Charon"], 1528). His brother Juan wrote in Naples a *Diálogo de la lengua* (1535; published 1737) of high critical value. In style these stand at the other extreme from two of the works of Antonio de Guevara (1481?–1545), which enjoyed European notoriety as the Spanish counterpart of euphuism (*q.v.*): the *Reloj de príncipes* (1529) and the *Epístolas familiares* (1539–42). Both works were rendered into English by Elizabethan translators—the *Reloj* by Sir Thomas North and also by Lord Berners (*qq.v.*), the letters by Sir Geoffrey Fenton (*q.v.*). History in the tradition of Alfonso the Wise continued to be cultivated, patriotism waxing higher as Spain's imperial greatness casts its shadow over Europe; its last flowering was seen in the Latin *Historiae de rebus Hispaniae* (1592–1621) of Juan de Mariana (1536–1624), the translation of which into Spanish (*Historia general de España*, 1601) by the author himself marked the triumph of the vernacular for all literary purposes. Diego Hurtado de Mendoza (1503 or 1506–75) gave, in his *Guerra de Granada* (a history of the suppression of the Morisco rising of 1568–71; publ. 1627), an extreme example of the concern to write history upon the classical model; the *Anales de la corona de Aragón* ("Annals of the Crown of Aragon"; 1562–80) of Jerónimo de Zurita y Castro (1512–80) by contrast exalted serious documentation and impartiality over style.

But the landmarks in historical writing came from the New World, and showed the transmuting of vital experience into literature with a vividness unknown in Spain. The letters of Columbus and his accounts of his voyages, the *Cartas y relaciones al . . . Carlos V* ("Letters and Accounts to . . . Charles V") of Hernán Cortés, and many another narrative by humble *conquistadores* without literary artifice or pretension opened up new horizons to the reader and, as in the attempt to capture exotic landscapes in words, enlarged the resources of the language. Most engaging was the *Verdadera historia de la conquista de la Nueva España* ("True History of the Conquest of New Spain"), written in old age by Bernal Díaz del Castillo (*c.* 1492–*c.* 1581), who had a phenomenal memory, an infallible eye for the telling detail, and a passion for distributive justice. Bartolomé de las Casas (*q.v.*; 1474–1566), the "Apostle of the Indies," wrote history for noble yet partisan purposes: on his *Brevísima relación de la destrucción de las Indias* ("A Very Brief Account of the Destruction of the Indies"; criticizing Spanish colonial policy and the ill-treatment of the native population) rests a large measure of responsibility for the *leyenda negra* ("black legend") which Spain's enemies were for so long to use against it.

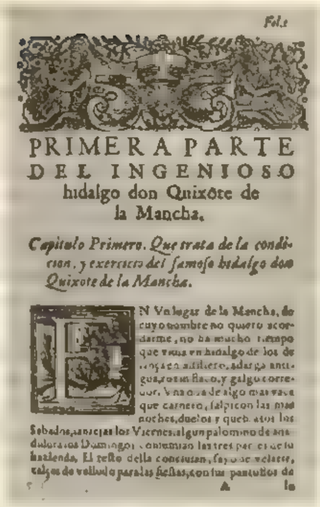
6. Mysticism.—The great period of Spanish mysticism coincided closely with the Counter-Reformation, though it had its antecedents. Ramon Llull had been a notable precursor in 13th-century Catalonia; and the *Dialoghi di amore* (published 1535; Eng. trans. *The Philosophy of Love*, 1937) of the expatriate Spanish Jew León Hebreo (Judah Abravenel; *c.* 1460–*c.* 1521), which, both in Italian and in Spanish translation, exercised a profound influence on 16th-century and later Spanish thought, allowed the infiltration of many Platonic images and doctrines. The literary importance of the mystics derives not only from their subject matter but also from the consideration that in the constant striving to transcend the limitations of language there were liberated previously untapped resources of expression, especially in the sphere of psychological probing. In the writings of St. Teresa of Ávila (1515–82), for example, *El libro de su vida* (her autobiography), *Las moradas, El castillo interior* (1588), and her letters, many of the gifts of a great novelist can be appreciated in embryo. Fray Luis de León, humanist and theologian, did not attain to the highest flights of the mystic ascent, but in his prose treatise *De los nombres de Cristo* (1583–85), as in his poems, he moves by his passionate devotion, his sincerity, and his profound feeling for nature, the whole finding expression in a style of singular purity. St. John of the Cross (1542–91) achieves preeminence by virtue of three poems expressing in exalted style and language the experience of Mystic Union.

7. The Novel.—Popular taste in the novel was dominated for a century by the monstrous progeny of the *Amadís de Gaula* (published 1508). These interminable romances kept alive certain ideals of medieval chivalry, but, having lost touch with life at every point, represented pure escapism, and in due course evoked, along with moral censure, various literary reactions. The pastoral novel was one (*see* PASTORAL). Akin to the eclogue naturalized from Italy by Garcilaso, it was modeled on the *Arcadia* (1504) of Jacopo Sannazzaro (1456–1530), and inspired by nostalgic echoes of an Arcadian Golden Age. Its shepherds were courtiers and poets who, no less than the knights-errant of chivalric romance, turned their backs on reality, with the purpose, however, of exploring more freely, against an idealized and carefree setting, the realm of the emotions. In Spain the first and best example was the *Diana* (1559?) of Jorge Montemayor (1520?–1561?); to the vogue it started, Cervantes with his *Galatea* (1584) and Lope de Vega with his *Arcadia* (1598) alike subscribed. In Gaspar Gil Polo's *La Diana enamorada* (1564), which was outstanding among the many sequels and imitations, the story became little more than a framework for a lyrical anthology.

A more positive reaction was seen in the picaresque novel (*q.v.*), initiated in 1554 with the brief, anonymous *Lazarillo de Tormes*. The word *pícaro*, of doubtful origin, does not appear in this text; the kind, native to Spain, was to exert influence in many foreign literatures. The *pícaro* was essentially the antihero, innocent of ideals and concerned only to keep alive, who lived by his wits, and, as he passed from master to master, saw and depicted life from underneath. Important for its part in guiding fiction back to direct observation of life, the picaresque formula had little to contribute to the development of the novel as an art form. The seed sown by *Lazarillo* did not bear fruit for nearly 50 years; and with the *Guzmán de Alfarache* (1599–1604) of Mateo Alemán (1547–1614?), whose interminable moralizings showed at its heaviest the Counter-Reformation's influence on literature; and Quevedo's *La vida del buscón* ("A Scoundrel's Life"; 1626), in which realism has become a cynically grotesque caricature of human values, the picaresque kind, though it continued abundant throughout the 17th century, may be held to have given what it had in it to give.

Cervantes (Miguel de Cervantes Saavedra, *q.v.*; 1547–1616), the culminating figure in his country's literature, produced in *Don Quixote* (part 1, 1605; part 2, 1615) what may be taken to be the prototype of the modern Spanish novel. Nominally a satire on the already moribund romances of chivalry (*see* ROMANCE), the conception allowed of the presentation of reality on two levels: the "poetic truth" of Don Quixote and the "historic truth" of his squire, Sancho Panza. In pursuing the constant interaction of these equally valid but rarely compatible attitudes to life, Cervantes revealed the novel's scope as a philosophical commentary on existence; and in tracing the constant interplay of character between master and squire—in contrast with the static characterization of previous fiction—he established psychology as one of the touchstones of its excellence. In the *Novelas ejemplares* ("Exemplary Novels"; 1613) Cervantes particularized


EL INGENIOSO
HIDALGO DON QVIXOTE DE LA MANCHA,
Compuesto por Miguel de Ceruantes Saauedra.
DIRIGIDO AL DVQVE DE BEIAR, Marques de Gibraleon, Conde de Benalcaçar, y Bañares, Vizconde de la Puebla de Alcozer, Señor de las villas de Capilla, Curiel, y Burguillos.
Año, 1605.
CON PRIVILEGIO,
EN MADRID, Por Iuan de la Cuesta.


Fol. 1
PRIMERA PARTE
DEL INGENIOSO
hidalgo don Quixote de la Mancha.
Capitulo Primero. Que trata de la condicion, y exercicio del famoso hidalgo don Quixote de la Mancha.

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM; PHOTOGRAPH BY J. R. FREEMAN AND CO

TITLE PAGE (LEFT) AND FIRST PAGE OF TEXT FROM THE FIRST EDITION (1605) OF "DON QUIXOTE" BY CERVANTES

his claim to be the first to write *novelas* (short stories in the Italian manner) in Spanish by differentiating between those which interest by the action and those of which the merit lies in the mode of telling, "making something out of nothing."

8. Later Drama.—The *comedia* (*i.e.*, the drama) achieved its formula and its splendour in the genius of Lope de Vega (1562–1635), to fill thereafter with an incredible fertility of invention the 100 years before the death of Calderón in 1681. Its manual was Lope's own poetic treatise, *El arte nuevo de hacer comedias en este tiempo* ("The New Art of Making Dramas in This Time"; 1609), which firmly rejected the classical and neoclassical "rules," opted for a blend of comedy and tragedy (*comedia* is in Spanish an all-embracing term for drama) and for metrical variety, and made of the *vulgo* (*i.e.*, public opinion) the ultimate arbiter of taste.

The *comedia* was essentially, therefore, a "social" drama, ringing a thousand changes on the accepted foundations of society: respect for crown, for church, and for the human personality, the latter being symbolized in the *pundonor* ("point of honour") that Lope commended as the best theme of all "since there are none but are strongly moved thereby." This *pundonor* was a matter largely of convention, "honour" being equivalent, in a very limited and brittle sense, to reputation. Here too the drama was social; characters, to merit esteem, had perforce to abide by the prescriptions of their fellows. It followed that this was a drama less of character than of action and intrigue, which rarely, if ever, grasped the true essence of tragedy (*q.v.*) or knew the cathartic function. What the great Spanish playwrights did possess was a remarkable sense of stagecraft and the ability to make the most intricate plot grip to the end, however arbitrary the final tying or untying of the knots.

Lope, the "prodigy of nature," as Cervantes called him, who claimed authorship of 1,800 *comedias*, and more than 400 *autos sacramentales* (of which about 500 of plays and *autos* together are extant), swept public and fellow-dramatists alike into vassalage by his powers of creation. Little that he wrote was perfect, but he had an unerring sense for the theme and detail that could move an audience conscious of being on the crest of its country's greatness to respond to a mirroring on the stage of some of the basic ingredients of this greatness. Because of him the *comedia* became a vast sounding board for every chord in the Spaniard's consciousness, a "national" drama in the truest sense. His plays have been classified into religious, mythological, pastoral, and novelesque, etc. In reality only two of these groups really matter: the native historical, and the *capa y espada* ("cloak-and-sword") plays of contemporary manners. Lope ransacked medieval chronicle, the *romancero*, and popular legend and song for heroic themes, chosen for the most part as throwing into relief some aspect of the national character, or of that social solidarity on which contemporary Spain rested. Among the best are *Fuente Ovejuna*, *Peribáñez y el comendador de Ocaña*, *El Caballero de Olmedo*, and *La estrella de Sevilla*. (There is some doubt about the authorship of the last-named play.) (For translations, see *Five Plays of Lope de Vega*, 1961.)

The "cloak-and-sword" play, a type which weighed more and more in the *comedia* after Lope, was pure diversion, written to a number of hollow conventions, with much disguising, mistaking of identities, duelling, falling in and out of love and false alarms about "honour," and a duplicating of the plot below stairs between *gracioso* (the gallant's humorous or would-be humorous lackey) and lady's maid. Though contemporary, it could in no sense be held to give a true picture of society, but it gave pleasure by its infinite resource and dexterity, its sparkling dialogue, its tripping convolutions in the relations between the sexes, and its constant implied paean to the enjoyment of the pleasures of the moment in an irresponsible world. The essentially lyrical character of much of the *comedia* further explains its charm.

The greatest of Lope's immediate successors was Tirso de Molina (*c.* 1584–1648), whose *El burlador de Sevilla* (1630; Eng. trans., *The Trickster of Seville*, in *The Classic Theatre*, volume three [1958] edited by Eric Bentley) presented the Don Juan legend on the stage for the first time (*see* also DON JUAN). His *Prudencia en la mujer* ("Prudence in a Woman"; 1634) figures among the greatest of Spanish historical dramas, as does *El condenado por desconfiado* (1635; Eng. trans., *The Doubted Damned*, 1956) among theological, while his "cloak-and-sword" comedies include some of the liveliest of the kind. Juan Ruiz de Alarcón y Mendoza (*c.* 1581–1639) struck a distinctive note which some have sought to relate to his Mexican birth and upbringing, others to his spinal deformity. His 20 plays were sober, studied, and imbued with serious moral purpose; *La verdad sospechosa* ("Truth Suspected"; 1634) inspired Corneille's *Le menteur*. The dramatic conflict between love and honour posed in *Las mocedades del Cid* (1599?), by Guillén de Castro y Bellvís (1569–1631), was similarly the source of Corneille's *Le Cid*. But moral purpose and dramatic conflict are infrequently found among the plays of Antonio Mira de Amescua (1574?–1644), Luis Vélez de Guevara (1579–1644), Juan Pérez de Montalbán (1602–38), and the many others who, with more or less distinction, if with dwindling originality, bridged the transition from Lope to Calderón.

9. Culteranismo and Conceptismo.—In poetry and prose the early 17th century was marked by the rise and spread of two interrelated movements. That known as *culteranismo*, of which Luis de Góngora y Argote (1561–1627) was archpriest, resumed, as one of its facets, earlier attempts to ennoble the language by re-latinizing it. In Góngora's *Polifemo y Galatea* (1612) and *Soledades* (1613–17; Eng. trans., 1931), this process was not only carried to extremes in vocabulary, syntax, and word order, so that the reader must needs be something of a latinist, but expression was garbed in the full panoply of classical myth and allusion, and further complicated by every subtlety of which metaphor is susceptible. Yet this poetry was something other than a deliberate cult of obscurity. Theme and treatment make clear the attempt to achieve a poetry not merely "pure" but timeless, to write poems which, rediscovered after the lapse of hundreds or thousands of years, could be acclaimed still fresh and immortal. Indeed, after 300 years Góngora was so rediscovered; but in his own age his example invited many who lacked his genius to slavishly imitate his procedures.

The twin affectation, *conceptismo*, played on ideas as *culteranismo* did on language. The *conceptista* purveyed thought, and sought to shock the reader into attention by both the subtlety and the violence of his conceits. Aiming always at the semblance of profundity, the quintessential, his style was concise and epigrammatic, and demanded for comprehension not the well-stocked mind but one infinitely nimble and alert. *Conceptismo* thus belonged properly to prose, inclined to satire, since it was much concerned with stripping off appearances, and found its most adequate outlet in the essay. Francisco Gómez de Quevedo y Villegas, (1580–1645), the master satirist of his age in the literary, the political, and the social sphere, and one of Spain's greatest wielders of language, was, in Los Sueños (1627), its outstanding exponent.

Baltasar Gracián y Morales (1601–58), its preceptist, reduced all its refinements to an exact code in his *Agudeza y arte de ingenio* (1648), as he also sought to codify in a series of treatises the art of worldly living (*El héroe*, 1637; *El discreto*, 1646; *El oráculo manual y arte de prudencia*, 1647). Gracián, if only doubtfully a philosopher, was certainly a thinker, and his allegorical novel *El criticón* (1651–57), a pessimistic interpretation of life as a "daily dying," was to influence La Rochefoucauld, Schopenhauer, and Nietzsche.

10. Calderón.—Pedro Calderón de la Barca (1600–81) spanned the century of Spain's decline, and his drama constitutes a studied expounding of patterns of behaviour (the *pundonor*) or belief. He wrote 80 *autos sacramentales* and these, with *La vida es sueño* (*Life Is a Dream;* several English translations), *La devoción de la cruz* (Eng. trans. by F. Honig, *Devotion to the Cross*, in *Four Plays*, 1961), and *El mágico prodigioso* ("The Marvelous Magician"), which give what dramatic tension is possible to a highly intellectualized exposition of Catholic dogma, make him the perfect literary representative of the Counter-Reformation. In the so-called honour plays—*El médico de su honra* (Eng. trans. by Roy Campbell, *The Surgeon of His Honour*, 1960), *A secreto agravio secreta venganza* (*Secret Vengeance for Secret Insult*, in Honig's translation mentioned above), *El pintor de su deshonra* ("The Painter of His Dishonour"), and *El mayor monstruo los celos* ("The Greatest Monster, Jealousy")—the cold logic with which an unreal code is pursued to the extreme becomes repugnant. *El alcalde de Zalamea* (*The Mayor of Zalamea*, Honig's translation), a reworking of a play by Lope, is perhaps his most successful play on the human level. *La vida es sueño*, to the extent that it raises profound philosophical issues on a plane of universality, has given him great fame abroad.

Francisco de Rojas Zorrilla (1607–48; *Del rey abajo, ninguno*, 1650), and Agustín de Moreto y Cabaña (1618–69; *El desdén con el desdén*, 1654) stand out among those who revolved in Calderón's orbit. After Calderón's death Spanish drama lay dead for a hundred years. *Culteranismo* and *conceptismo*, though symptoms rather than causes of decline, had played their part in the stifling of imaginative literature, and by the close of the century nothing but the shell remained of the greatness that had been the *siglo de oro*.

C. The 18th Century

With the advent of the Bourbon dynasty in 1700, and its establishment after the War of the Spanish Succession (1701–13), a profound renewal of the country's intellectual life took place. Numerous academies were created, the most influential being the Real Academia de la Lengua Española, founded in 1713 to maintain the purity of the language (*Diccionario de autoridades*, 1726–39; *Gramática*, 1771). Men of letters again went to school abroad, and discovered how far Spain had deviated from the intellectual courses of Western Europe.

Moreover, a new spirit of inquiry into the national heritage was to carry back to Luis de León, to Herrera, and to Garcilaso an awareness that in the early decades of the century had seemed limited to Góngora, Quevedo, and Calderón (worst of models for an imitative generation), and was indeed to go back earlier still, to unearth the forgotten medieval literature. Gregorio Mayáns y Siscar (1699–1781), librarian at the royal library in the 1730s, edited Juan Luis Vives (in Latin), Juan de Valdés, and Luis de León, and, in his *Vida de Cervantes* (1737), initiated Cervantine criticism. The Augustinian church historian Enrique Flórez (1702–73), embarking on the vast *Teatro geográfico-histórico de la iglesia de España* (better known as *España sagrada*, 29 volumes, 1747–55; continued by others), helped resurrect the whole cultural background of medieval Christian Spain. A landmark of even greater importance was the publication, in Tomás Antonio Sánchez's *Poesías castellanas anteriores al siglo XV* (1776–90), of the *Poema de mío Cid*, the works of Gonzalo de Berceo, the *Libro de Alexandre*, and Juan Ruiz's *Libro de buen amor*. From all this critical activity there resulted a debate between old and new that, waged throughout the middle decades of the century, compelled both sides to reason their positions and marked the birth of a new critical approach to literature.

Two names stand out: Ignacio de Luzán Claramunt (1702–54), educated in Italy and well-grounded in foreign literatures, whose work on poetics (1737) introduced to Spanish readers the aesthetic of Muratori, Boileau, and Le Bossu, and launched the great neoclassical polemic in Spain; and Benito Jerónimo Feijoo y Montenegro (1676–1764), a Benedictine whose *Teatro crítico universal* (8 volumes, 1726–39) and *Cartas eruditas y curiosas* (15 volumes, 1742–60) comprised more than 280 essays ranging over the whole of human knowledge. Assailing error, prejudice, and superstition wherever he found them, and so bitterly assailed in his turn that the king intervened to protect him, Feijoo made a monumental contribution to the intellectual emancipation of Spain. Imaginative prose produced the *Noches lúgubres* (publ. 1789–90) of José de Cadalso y Vázquez (1741–82), looking forward to Romanticism; and the *Fray Gerundio . . .* (1758) of José Francisco de Isla (1703–81), a satire, in the guise of biographical fiction, on affectation in the pulpit, which looked back to the picaresque novel.

Poetry, moribund for nearly 100 years, raised a timid head in a small group at Salamanca, led by the Augustinian poet and orator Diego González (1731–94), which toward 1775 turned for inspiration to Luis de León, just as two decades later a group at Seville sought to revive the glories of Herrera; and again (as in the 16th century), though the lines were now more blurred, Salamanca stood for content, Seville for form. Juan Meléndez Valdés (1754–1817), who learned to think from John Locke and to feel from Edward Young, best exemplifies the combination of new influences at work. A conscious artificer rather than a great poet—though the nearest to one among the Salamancans—he helped poetry through the painful apprenticeship necessary to its rehabilitation.

For drama, the second half of the century saw the great battle of the neoclassical "rules" (meaning chiefly the unities of place, time, and action) that the *siglo de oro* had spurned. *La Raquel* (1778) of Vicente Antonio García de la Huerta y Muñoz (1734–87), combining a traditional Spanish theme with neoclassic construction, showed the capabilities of the reformist school. His vast compilation of earlier drama, the *Theatro Hespañol* (16 volumes, 1785–86), made apparent by its omission of Lope, Tirso, and Alarcón how unsubstantial were the grounds on which condemnation of the early drama rested.

It fell to Ramón de La Cruz (1731–94) to bridge the gap, unwittingly but effectively, by his resurrection on a somewhat ampler scale of the earlier *paso* (one-act prose skit) or longer *entremes* of Lope de Rueda, Cervantes, and Luis Quiñones de Benavente (*c.* 1589–1651). Based on acute satirical, though in the main good-humoured, observation of the Madrid scene, his one-act *sainetes* ("sketches") could not, because of their brevity, transgress the unities or offend the purist; at the same time they delighted the *vulgo* and brought the drama back from pseudoantiquity to the function of commenting on life and society. Leandro Fernández de Moratín (1760–1828) applied the lesson to the full-length play, and produced comedies imbued with deep social seriousness which, written to a meticulous observance of the unities, were yet good theatre (*El viejo y la niña* ["The Old Man and the Girl"], *La comedia nueva* ["The New Comedy"], *El sí de las niñas* [Eng. trans., *When a Girl Says Yes*, in *Spanish Drama*, edited by Angel Flores, 1962]. *La comedia nueva*, a remorseless flaying of such writers as Luciano Francisco Comella y Villamitjana (1751–1812), who had reduced to absurdity what they fondly believed to be the Calderón tradition, absolves one from any need to read their work. Moratín's dialogue in this play ranks among the best prose of the 18th century.

D. The 19th Century

1. Romanticism.—The first third of the 19th century proved but a prolongation of the 18th, the Napoleonic Wars and their long aftermath casting a blight on letters. The inspiriting odes of a Juan Nicasio Gallego (1777–1853), or a Manuel José Quintana (1772–1857), or the impassioned prose of a Gaspar Melchor de Jovellanos (1744–1811), show neoclassic objectivity and proportion set aflame with the emotion of outraged patriotism. War

against France did not mean repudiation of French literary influences. Most of the thousands of liberals driven into exile by Ferdinand VII after 1823 sought refuge in France. Having drunk there the heady wine of Romanticism, they flocked back to Spain on Ferdinand's death in 1833, a date often taken as marking the beginning of the Romantic movement in Spain.

Such fortuitous contacts were far from being the sole source of Spanish Romanticism. The ground for it had been prepared in Cádiz from 1814 onward in a notable debate on literary values initiated by the German Johann Niklaus Böhl von Faber; in Barcelona with the founding in 1823 of the review *El Europeo* ("The European"), edited by a Catalan, an Englishman, and an Italian; and in Madrid with Agustín Durán's critical essay on the drama of the *siglo de oro* (1828), and with his vast *Colección de romances antiguos* ("Collection of Ancient Ballads"; 1828–32). Romanticism in France, moreover, was largely a revolt against native literary tradition and a turning abroad, notably to Spain, for new sources of inspiration; in Spain it implied the rejection of an alien neoclassicism and to that extent a return to the spirit of its own earlier classics. All the formal characteristics of Spanish Romantic drama—the mingling of kinds, the rejection of the unities, metrical variety (now extended to include prose)—had characterized the drama of Lope and his contemporaries, who had, moreover, treated many of its themes (taken, by choice, from the Spanish Middle Ages).

It is from a study of these themes that the novelty and significance of the new school becomes apparent. Whereas the drama of the *siglo de oro* had rested on social solidarity, Romantic drama was in revolt against society: its protagonist staked everything on his right to be himself, *especie única,* and, deliberately flouting convention, defied society to stand in his way. In France, tragedy was by Romanticism deflected into drama; in Spain, drama into tragedy.

The Romantic movement in Spain, coming to fruition a generation later than elsewhere, knew a correspondingly short and checkered life. The six plays that epitomize its dramatic achievement fall between 1834–44: *La conjuración de Venecia* ("The Venetian Conspiracy"), by Francisco de Paula Martínez de la Rosa (1787–1862); *Macías,* by Mariano José de Larra (1809–37); *Don Álvaro o la fuerza del sino* ("Don Alvaro, or the Force of Destiny"), by Ángel de Saavedra, duque de Rivas (1791–1865); *El trovador,* by Antonio Garcia Gutíerrez (1813–84); *Los amantes de Teruel,* by Juan Eugenio Hartzenbusch (1806–80); and *Don Juan Tenorio,* by José Zorilla y Moral (1817–93). (The poetry of José de Espronceda [1808–42] also belongs to this period). Of these six plays, *Don Álvaro* came nearest, in the theatre, to expressing a "philosophy" of Romanticism; and the preface, by the critic and orator Antonio Alcalá Galiano (1789–1865), to Rivas' narrative poem, *El Moro expósito* ("The Abandoned Moor"; 1834) came nearest to supplying a foundation of precept. But the movement never possessed the consistency of a school, a leader, or a corpus of doctrine. Espronceda was the one Romantic who lived his Romanticism, if with an element of Byronic posturing; his *Estudiante de Salamanca, Canciones,* and the unfinished *El diablo mundo* ("The Devilish World") were the only subjective lyricism of value that the period produced, and marked, too, a milestone in the enlargement of poetic form. In comparison, the *Romances históricos* (1841) of Rivas, and the *Cantos del trovador* ("Troubadour Songs"; 1840–41) of Zorrilla were little more than exercises, brilliant within the limitations of their kind, in the retelling of picturesque legend.

2. Costumbrismo.—Side by side with, and beginning a little earlier than, Romanticism flourished a movement called *costumbrismo*—a kind of realistic prose writing, often within a narrative framework, having affinities with the De Coverley papers of Joseph Addison and Richard Steele, and particularly suited to the Spanish genius. Its chief vehicles are the *cuadro de costumbres* and the *artículo de costumbres*—short literary sketches, describing customs, manners, or character, the *cuadro* inclining to description for its own sake, the *artículo* being more critical and satirical, and both generally concentrating on a particular locality or region.

The forerunners of the movement were Santos López Pelegrín (1801–46) and Sebastian de Miñano (1779–1845) whose *Cartas de un pobrecito holgazán* ("Letters from a Poor Idler"; 1820), attacking the monarchy and satirizing Spanish customs and politics, is sometimes regarded as the first *costumbrista* work. The three most important *costumbristas* were Larra, the outstanding prose writer and critical mind of his age, who dissected society pitilessly in his *artículos,* collected in 1835–37; Ramón de Mesonero Romanos (1803–82), whose *Escenas matritenses* ("Scenes of Madrid"; 1836–42) give a vivid picture of the social life of the period, especially in Madrid; and Serafín Estébanez Calderón (1799–1867), whose *Escenas andaluzas* ("Andalusian Scenes"; 1847) portray the manners, folklore, and history of Andalusia. These, and other *costumbrista* writings, helped to prepare for the revival of the novel.

3. Post-Romantic Poetry.—Three poets bestride the third quarter of the century and, in their diversity, reveal how one of Romanticism's major services was its liberation of the individual personality. Gustavo Adolfo Bécquer (1836–70), accounted a Romantic, was concerned only to give expression to his own tortured emotions. His 76 brief *Rimas,* which owe something to Heine, bear the stamp of a deeply moving sincerity. In his *Leyendas* ("Legends") Bécquer proved himself also a prose stylist in a richly poetic and imaginative vein. Ramón de Campoamor y Campoosorio (1817–1901) affected the epigrammatic in his *Doloras, Pequeños poemas,* and *Humoradas* (new poetic forms of his own invention), making a sustained attempt at originality and at bringing poetry back into the realm of ideas. Gaspar Núñez de Arce (1832–1903) likewise endowed poetry with a purpose, the social purpose of stimulating the civic virtues and defending the body politic against dangerous heresies. His *Gritos del combate* ("Combat Cries"; 1875) represented a noble if despairing attempt, at times too declamatory, to stem the tide of the abuse of democracy. Among somewhat later poets were Salvador Rueda (1857–1933), an exuberant local colourist who has been regarded as a precursor of Modernism; José María Gabriel y Galán (1870–1905), whose works extoll the traditional virtues of the Castilian countryside; and Rosalía de Castro (1837–85), a true poet of Galicia (*see* also below), whose one volume of poems in Castilian, *En las orillas del Sar* (1884), invites comparison with Bécquer.

4. The Novel.—The last-named poets, each identified with his *patria chica* (native region), shared with the novelists a quality that was to characterize Spanish fiction on its rebirth. For more than two centuries, the kind in which Spain had made its most striking contribution to world literature—the novel—had been virtually extinct. Its place had been taken during the Romantic period by the *cuadro de costumbres,* and, although attempts had been made by Larra, Rivas, Espronceda, and others to write historical novels in the manner of Sir Walter Scott, this kind of fiction never took root in Spain. Indeed, the novel, on its revival, was both an extension of the realistic and regional elements of *costumbrismo,* and a continuation of the tendencies of earlier Spanish novelists. The first novels of the revival were more interesting for their powers of observation and description than for imaginative or narrative quality.

It was Fernán Caballero (1796–1877), who, in 1849, with *La gaviota* (*The Sea Gull*), determined the particular technique of observation that was to rehabilitate the novel. This simple tale of romantic disillusion captivated less by its plot than by its Andalusian setting and vividness of colour and detail. In 1874, coinciding with the restoration of the Bourbon monarchy, the great period of the regional novel began, for, in this year Pedro Antonio de Alarcón (1833–91), a waverer between the esoteric-romantic and the pedagogic, achieved his masterpiece in *El sombrero de tres picos* (*The Three-Cornered Hat*), a sparkling tale of Andalusian peasant malice. A fellow-Andalusian, Juan Valera y Alcalá Galiano (1824–1905), brought psychological insight and a classical note of cultured detachment to four novels which depended less on local realism, acutely observed though this was, than on knowledge of men and women: *Pepita Jiménez* (1874), *Las ilusiones del doctor Faustino* (1875), *El comendador Mendoza* (1877), and *Doña Luz* (1879). José María de Pereda (1833–

1906), graduating through a long series of sketches on life in the mountains of his native Santander, emerged as the greatest of the regional writers; his realism was no longer merely picturesque but a powerful re-creation of nature as the abiding reality, more important than the individual and therefore than the plot. *Sotileza* ("Subtlety"; 1884) and *Peñas arriba* ("Up the Mountains"; 1893), perhaps his best novels, revealed him as a champion too of traditional values—religion, the family, country life—in a dissolving society typified by the big city.

Emilia Pardo Bazán (1851–1921) attempted, in two novels of Galicia, *Los Pazos de Ulloa* (1886) and *La madre naturaleza* ("Mother Nature"; 1887), to combine the Spanish Christian outlook with French Naturalism, which she had already examined curiously and with many reserves in the essays in *La cuestión palpitante* ("The Crucial Issue"; 1882–83). The attempt made clear their incompatibility, and in later novels her return to a genuinely Spanish realism, in which spiritual values also found their place, was complete. Armando Palacio Valdés (1853–1938) was the novelist of Asturias even when the stage was transferred to Andalusia, as in the perennially popular *La hermana San Sulpicio* (1889; Eng. trans., *Sister Saint Sulpice,* 1890), or to Valencia, as in *La algería del capitán Ribot* (1899; Eng. trans., *The Joy of Captain Ribot,* 1900). He also experimented, somewhat halfheartedly, with Naturalism, in *La espuma* (1890; Eng. trans., *The Froth,* 1891) and *La Fe* (1892; Eng. trans., *The Faith,* 1892), but was saved by his most engaging quality, humor.

Besides this cult of regionalism, the fecundity in the novel of Benito Pérez Galdós (1843–1920) created a world in itself. Going to Madrid from the Canary Islands in 1862, Pérez Galdós was better able than the regional writers to see Spain in the round; and, after living through one of the most turbulent chapters in Spanish history, he addressed himself to imaginatively re-creating its antecedents, beginning with the war against Napoleon. The five series, containing 46 novels of his *Episodios nacionales* ("National Episodes"; 1873–79 and 1898–1912), embraced 70 years of the country's history, from Trafalgar to the early years of Alfonso XII, and taught generations of Spaniards much of what history they knew. His other series (the *Novelas de la primera época* and the *Novelas españolas contemporáneas,* in which he treats contemporary society) completed the revelation of Spain to Spaniards; embodying the author's liberal and challenging views, they constituted a concerted campaign to regenerate and Europeanize Spanish society and thought.

5. Drama.—The drama, disorientated after the brief Romantic efflorescence, produced few outstanding works and no clear conception of goal or function. Manuel Tamayo y Baus (1829–98), who began by oscillating between and even attempting to fuse classical tragedy and Romantic drama, gradually evolved from verse to prose and from the historical-heroic to the contemporary-social; he achieved fame by the isolated *Un drama nuevo* (1867; Eng. trans., *A New Drama,* 1915), a powerful portrayal, in which the characters are members of Shakespeare's company of actors (including Shakespeare himself). The hero is the clown, Yorick, who is compelled to act on the stage his personal tragedy of betrayal by his wife, and retributive murder. Adelardo López de Ayala (1829–79) used considerable technical gifts to pillory bourgeois vices, in, for example, *El tejado de vidrio* ("The Glass Roof"; 1856), and *Consuelo* (1870). The Nobel Prize for Literature (1904) set a seal on José Echegaray y Eizaguirre (1832–1916), whose 63 plays fall into two main groups. In the first, imbued with a violent neo-Romanticism, drama becomes melodrama with a constant striving after effect and a basic falsity of character, passion, and situation alike. The second comprises a serious, often tragic, drama of social problems in which Ibsen can be seen as one of several foreign influences though here, too, passions were apt to be torn to shreds and denouements to rely on violent dispatch: *El gran Galeoto* (1881; Eng. trans., *The Great Galeoto,* 1922), *O locura o santidad* (1877; Eng. trans., *Madman or Saint,* 1907), *El hijo de don Juan* ("Don Juan's Son"; 1892). Joaquín Dicenta (1863–1917) brought the drama nearer to earth with an unpleasantly realistic study of working-class conditions in *Juan José* (1895). Failure to master the problems involved in the change from the novel to drama robbed Pérez Galdós of full success in the theatre, to which he turned increasingly in his concern to stimulate the social conscience: *Realidad* ("Reality"; 1892); *La loca de la casa* ("The Madwoman in the House"; 1893); *Electra* (1901).

6. The Generation of 1898.—Once again, as in 1700, the turn of the century brought about a radical change in literary values and directions. For two decades and more a mood of scathing analysis had been developing with the growing conviction that behind the peace and order of a nominally representative regime much rottenness lay concealed. Pérez Galdós sensed and gave voice to the disquiet; and it evoked, in Ángel Ganivet's *Idearium español* (1897), one of the most searching analyses of the Spanish character ever written. With 1898 came the Spanish-American War, which left Spain stripped of the last vestiges of its trans-Atlantic and Pacific empire. The imperial cycle begun in 1492 had ended in ignominy, and thinking Spaniards embarked on an analysis of their country's ills, a scrutiny of its possible destiny, and an attempt to shock the national mentality out of its *aboulia* (or lack of will). Among literary consequences were the informing of the novel with a new seriousness of purpose, and the emergence of the essay—critical, psychological, philosophical—to a new importance. Novelists and essayists constitute what Azorín first called the "generation of '98." Rarely in agreement in their diagnosis or on the treatment proposed for the patient, they reinvigorated Spanish letters to a point where these again commanded respect abroad and, as part of the process, they laid Spain open to English, French, German, Russian, and other influences on a scale never known before.

Miguel de Unamuno (1864–1936) dominated the literary scene for a generation. In him the national problem was intertwined with his obsession with the personal problem of immortality. The former he studied acutely in the five essays in *En torno al casticismo* ("On Spanish Purism"; written 1895, publ. 1902), and in the *Vida de Don Quijote y Sancho* (1905; Eng. trans., *Life of Don Quixote and Sancho,* 1927), an analysis in depth of Cervantes' characters. The problem of immortality was examined in his most important work, *Del sentimiento trágico de la vida* (1913; Eng. trans., *The Tragic Sense of Life in Men and in Peoples,* 1921). A deliberately provocative rather than a systematic thinker, he aimed at sowing spiritual disquiet. The novel was to him an extension of the essay, as the theatre was of the novel, a medium for discussion of the fundamentals of personality, hence stripped of externals and as timeless as the *Soledades* of Góngora; representative examples of his fiction are *Niebla* (1914), *Abel Sánchez* (1917), and *Tres novelas ejemplares y un prologo* (1921).

Azorín (José Martínez Ruiz; 1874–1967) concerned himself with the reinterpretation of earlier literary values and of the Spanish countryside, in, for example, *El alma castellana* ("The Castilian Soul"; 1900), *La ruta de Don Quijote* ("Don Quixote's Route"; 1905), and *Clásicos y modernos* ("Classics and Moderns"; 1913). An artist in criticism and a miniaturist of fine sensitivity, he contributed powerfully to the deflation of the hollow rhetoric that had vitiated much 19th-century writing. José Ortega y Gasset (1883–1955) brought to the analysis of the national consciousness a mind formed in Germany in the study of neo-Kantian thought. From criticism and psychology (*Meditaciones del Quijote,* 1914 [Eng. trans., *Meditations on Quixote,* 1961]; *El espectador* ["The Spectator"], 1916–34) he passed to national problems (*España invertebrada,* 1921; Eng. trans., *Invertebrate Spain,* 1937) then to international (*El tema de nuestro tiempo,* 1923 [Eng. trans. *The Modern Theme,* 1933, 1961]; *La rebelión de las masas,* 1929 [Eng. trans. *The Revolt of the Masses,* 1932]). His *Revista de Occidente* ("Magazine of the West"; 1923–36), directed to keeping Spain abreast of contemporary foreign thought, proved a major intellectual force.

Pío Baroja (1872–1956) repudiated tradition, religion, and the cult of the individual, and sought to be an arch-European, fixing his gaze on the future and making a cult of social action. His *Memorias de un hombre de acción* ("Memories of a Man of Action"; a cycle of 22 novels set in the early 19th century) are thus less characteristic than his 11 trilogies on the contemporary scene

—*La raza* ("The Race"; 1908–11); *La lucha por la vida* (1904; Eng. trans., *The Struggle for Life*, 1922–24); *Agonías de nuestro tiempo* ("Agonies of Our Time"; 1926), etc. Insensitive to art for art's sake, Baroja often depresses through his concern to arouse discontent with material conditions; but his novels, full of rude vigour and a fierce independence of spirit, are seedbeds of ideas. There was rude vigour too, with richness of colour and greater narrative skill, in Vicente Blasco Ibáñez (1867–1928), a regional novelist in, for example, *Flor de Mayo* (1895; Eng. trans., *Mayflower*, 1921) and *La barraca* (1898; Eng. trans., *The Cabin*, 1917), on his native Valencia, who wrote also on contemporary social problems from the standpoint less of the generation of '98 than of the anarchist—for example, *La bodega* (1906) and *La horda* (1905). Blasco Ibáñez won international renown with novels of World War I; the most famous was *Los cuatro jinetes del Apocalipsis* (1916; Eng. trans., *The Four Horsemen of the Apocalypse*, 1918). Ricardo León (1877–1943) was a conscious stylist who sought to recapture the traditional Christian values of a society threatened with materialism (*Casta de hidalgos*, 1908; *Alcalá de los Zegríes*, 1909; *El amor de los amores*, 1910). His work, however, belongs less in outlook and style to the generation of '98 than to the 20th century.

E. The 20th Century

Critics and literary historians differ about the dividing line between the generation of '98 and the so-called *Novecentista*, which by some is taken to mean the generation immediately following that of '98, and by others is extended to cover the whole of the period from 1900 to the 1930s. In general, however, the term is applied to the writers of the early 20th century who sought to renew intellectual and aesthetic standards after the passionate involvement of their immediate predecessors.

1. The Novel.—In Ramón Pérez de Ayala (1880–1962) the novel is at once a satisfying art form and a forum for philosophical discussion tinged with Asturian humour. *Troteras y danzaderas* (1913) recaptures something of the vein of Juan Ruiz. *Belarmino y Apolonio* (1921), a projection of the old debate between faith and reason, makes its characters almost symbolic, as do *Tigre Juan* and its continuation *El curandero de su honra* (1926; Eng. trans., *Tiger Juan*, 1933) on the traditional theme of honour. Gabriel Miró (1879–1930) was a lyric poet in prose, with gifts of description and evocation (shown in *Figuras de la pasión del Señor* ["Scenes of the Passion of Our Lord"], 1916–17; *Años y leguas* ["Years and Leagues"], 1928) which retard the action of his novels but have given him the reputation of a supreme artist in words. In these writers can be seen qualities and tendencies to some extent incompatible with the basic art of the novel. This as a literary kind fell under the influence of Ortega y Gasset, who, in *La deshumanización del arte* (1925; Eng. trans., *The Dehumanization of Art*, 1956), propounds the principles of a pure, depersonalized art. With it, he published notes on the novel, predicting its decline. In the following decade Benjamín Jarnés (1888–1950) and others attempted, without complete success, to apply a technique of pure art to the novel; Jarnés' works are the outstanding examples of the Surrealist novel in Spain.

The publication in 1929 of *Imán*, based on experience of military service in Morocco, by Ramón José Sender (1902–), marked a return to realism and social criticism. Like his later novels (the trilogy *Los terminos del presagio* ["The Terms of the Omen"], 1931–34; *El epitalamio del prieto Trinidad*, 1942; etc.), written in Spain, Mexico, and the United States, where he settled in 1942, it is concerned with the human situation in all its horror, and with the search for a satisfying philosophical and ethical system. (An English translation of *Imán* appeared in 1935, entitled in the United States *Pro Patria* and in Great Britain *Earmarked for Hell*. A number of his other novels are available in English translations.)

Sender was followed by Max Aub (1903– ; *El laberinto mágico* ["The Magic Labyrinth"], 1943–45), and Arturo Barea (1897–1957), whose autobiographical trilogy of novels was first published in England (where he settled after the Spanish Civil War) with the title *The Forging of a Rebel* (1947), and had great influence on writers there and elsewhere in Europe. It was published in Spanish in 1951–52 as *La forja de un rebelde*. Novelists of promise to emerge after the Civil War were Camilo José Cela (1916–); Carmen Laforet (1921–); Juan Antonio de Zunzunegui (1901–); and José María Gironella (1917–).

2. Drama.—With the new century the drama achieved new vigour under the stimulus of Jacinto Benavente y Martínez (1866–1954), Spain's greatest playwright since Calderón, and winner of the Nobel Prize for literature in 1922. With him the drama returned to reality by way of social criticism, declamatory verse giving way to prose, melodrama to comedy, formula to experience, impulsive action to dialogue and the play of minds. After a period of effervescence and irony, Benavente, as he mastered his medium, showed a preoccupation with aesthetics and later with ethics. The extent to which he broadened the scope of the theatre is shown by the range of his plays, *e.g.*, *Los intereses creados* (1907; *The Bonds of Interest*), *Los malhechores del bien* (1905; *The Evil Doers of Good*), *La noche del sábado* (1903; *Saturday Night*), and *La malquerida* (1913) (all translated by J. G. Underhill in *The Plays of Benavente y Martínez*, 4 volumes, 1917–24).

By contrast, Gregorio Martínez Sierra (1881–1947 or 1948) remained true to a more traditional Spanish atmosphere, in which idyllic sentiment, though always deep-rooted in Spanish realities, breathed a sensuous joy in nature and life (*Teatro de ensueño* ["Fantastic Theatre"], 1905; *Canción de cuna* ["*Cradle Song*"], 1911; *El Reino de Dios* ["God's Kingdom"], 1916). The brothers Álvarez Quintero (Serafín, 1871–1938; Joaquín, 1873–1944), writing as one, were pure *costumbristas*. If they contributed nothing new to dramatic art, the *joie de vivre* of their long succession of Andalusian comedies never failed to entertain. Among other dramatists of note are Manuel Linares Rivas (1867–1944) and José López Pinillos ("Pármeno," 1875–1922).

The poetic drama was cultivated by Eduardo Marquina (1879–1946), Jacinto Grau Delgado (1877–1958), Francisco Villaespesa (1866–1936), Ramón María del Valle-Inclán (1866–1936), and Luis Fernández Ardavín (1891–).

Federico García Lorca stands far above any of these. His drama was poetic in more than the formal sense, presenting the elemental passions with an intensity that made the characters poetic symbols of man's tragic impotence to arrest his fate (*Bodas de sangre*, 1933 [*Blood Wedding*]; *Yerma*, 1934; *La casa de Bernarda Alba*, 1936 [*The House of Bernarda Alba*]; all in several English translations). The impulse of renovation visible in the novel after the Civil War failed to reach the theatre, which with Edgar Neville (1889 or 1899–), Joaquín Calvo Sotelo (1905–), and Juan Ignacio Luca de Tena (1897–) remained generally faithful to the patterns established by Benavente and the Álvarez Quinteros. Antonio Buero Vallejo (1916–) and Miguel Mihura (1903–) both attempted to revitalize the drama, Buero Vallejo by social criticism, Mihura by introducing the "theatre of the absurd."

One other vein of dramatic activity, generally thought to fall outside the limits of formal criticism, adds persisting vitality to a long and honourable descent. It is the *género chico* (or "little genre"), heir to the *paso*, *entremes*, and *sainete*, which, within the limitations of the short sketch, frequently with music, achieves a great variety of dramatic effect and for almost a century has charmed the *vulgo* whose way of life supplies it with setting and theme. Ricardo de la Vega (1839–1910) and Carlos Arniches (1866–1943) are outstanding in the kind, which the brothers Álvarez Quintero likewise notably enriched, particularly with their libretti for *zarzuelas* (plays between drama and opera, in which recitative alternates with song).

3. Poetry.—*Modernism.*—Rubén Darío (1867–1916), Latin America's greatest poet, took with him to Spain on his first visit in 1892 the movement known as *Modernismo*, of which he was the leading exponent. In general Modernism was a reaction against 19th-century bourgeois materialism and a search for other, and, more specifically, for aesthetic, values in life. In literature it was a reaction against Naturalism, drawing heavily on French Symbolism and Parnassianism alike (*see* Symbolists, The; Parnassians), with elements of a revived Romanticism. Endowed with

rare technical virtuosity, Darío greatly enriched the musical resources of Spanish verse, notably by his daring use of new rhythms and metres; and under his influence poetry in Spain experienced a renewal comparable with that effected by Garcilaso almost four centuries earlier.

Though so nearly coincident in time with the movement represented by the generation of '98, Modernism had scarcely any contact with it. Initially it stood for the cult of the ivory tower, wherein the poet took refuge from his fellows and from life and sought self-realization in the pursuit of esoteric beauty. It produced, however, one outstanding prose writer, Ramón María del Valle-Inclán, whose four *Sonatas* (1902–05) typify the four seasons of his hero's life. Valle-Inclán later wrote historical novels (the trilogy *La guerra carlista,* 1908–09, the two books of the unfinished series entitled *El ruedo ibérico*), but here, too, his vision was purely aesthetic. For the rest, the Modernists were poets, predominantly Andalusian, the contrast stilll holding between form, which interested them, and content, with which the northerners of the generation of '98 were concerned.

Juan Ramón Jiménez (1881–1958), beginning to write under the aegis of Darío, soon discovered poetry to be for him not a withdrawal, but reality itself, and his style, initially ornate, became more and more bare, with a turning to blank verse, as he strove to capture the poetic essence of each passing moment. Using and gradually winnowing all the resources of Modernism, he became the poets' poet, and a powerful influence on the younger generation. The brothers Machado well exemplify the Andalusian-Castilian dichotomy: Manuel (1874–1947), always at heart a loyal son of his native Seville, was the Modernist; Antonio (1875–1939) early lost his heart to Soria, in Old Castile, and dealt with the recesses of the poet's soul or with the unchanging soul of Castile. Unamuno, Valle-Inclán, and Pérez de Ayala all have their place as poets, though their poems were but the prolongation of their prose works. Of the three, Unamuno, the least accomplished technically, is the most rewarding; he seeks to distill into verse the intensity of his spiritual striving.

The poetic stature of García Lorca grew steadily after his death. He is a poet, as he is a dramatist, of fundamentals, and has shown of what starkness of feeling and effect the *copla* and *romance* are still capable. His *Primer romancero gitano* (1928) and *Poema del cante jondo* (1931) reveal the potency of the inspiration that he drew from Andalusian ways and folklore. The other major poets of the mid-20th century were still those of the pre-Civil War generation, contemporaries of García Lorca and heirs to Darío and to Juan Ramón Jiménez: Pedro Salinas (1891–1951), Jorge Guillén (1893–), Gerardo Diego Cendoya (1895–), Dámaso Alonso (1898–), Vicente Aleixandre (1900–), Rafael Alberti (1902–), Luis Cernuda (1904–63), and Manuel Altolaguirre (1906–). The Civil War dealt hardly with them too, driving some into exile, stifling the inspiration of others; in the main, their best work had already been written by that time. The generation immediately after the Civil War was characterized by minority groups, fugitive reviews, and dispersive tendencies.

Despite the appearance of many new poets during the 1940s and 1950s, especially noteworthy being the group who called themselves *la juventud creadora* and who in 1943 founded the review *Garcilaso,* and the poets associated with the review *Espadaña,* no poet of outstanding stature emerged.

II. CATALAN LITERATURE

1. Medieval Period.—*Poetry.*—The Catalan language is a branch of Peninsular rather than of southern Gallo-Romance. It shows nonetheless many traces of kinship with Provençal (*see* PROVENÇAL LANGUAGE), and the literature in its origins used the *langue d'oc* and the poetic forms cultivated by troubadours north of the Pyrenees: the Catalan troubadors Guillem de Bergadà (1140–1203), Hug de Mataplana, Ramon Vidal de Besalù, Guillem de Cervera, and Cerverì de Girona (1250–80) were as genuinely Provençal poets as were those of Limousin, Quercy, and Auvergne (*see* also TROUBADOURS).

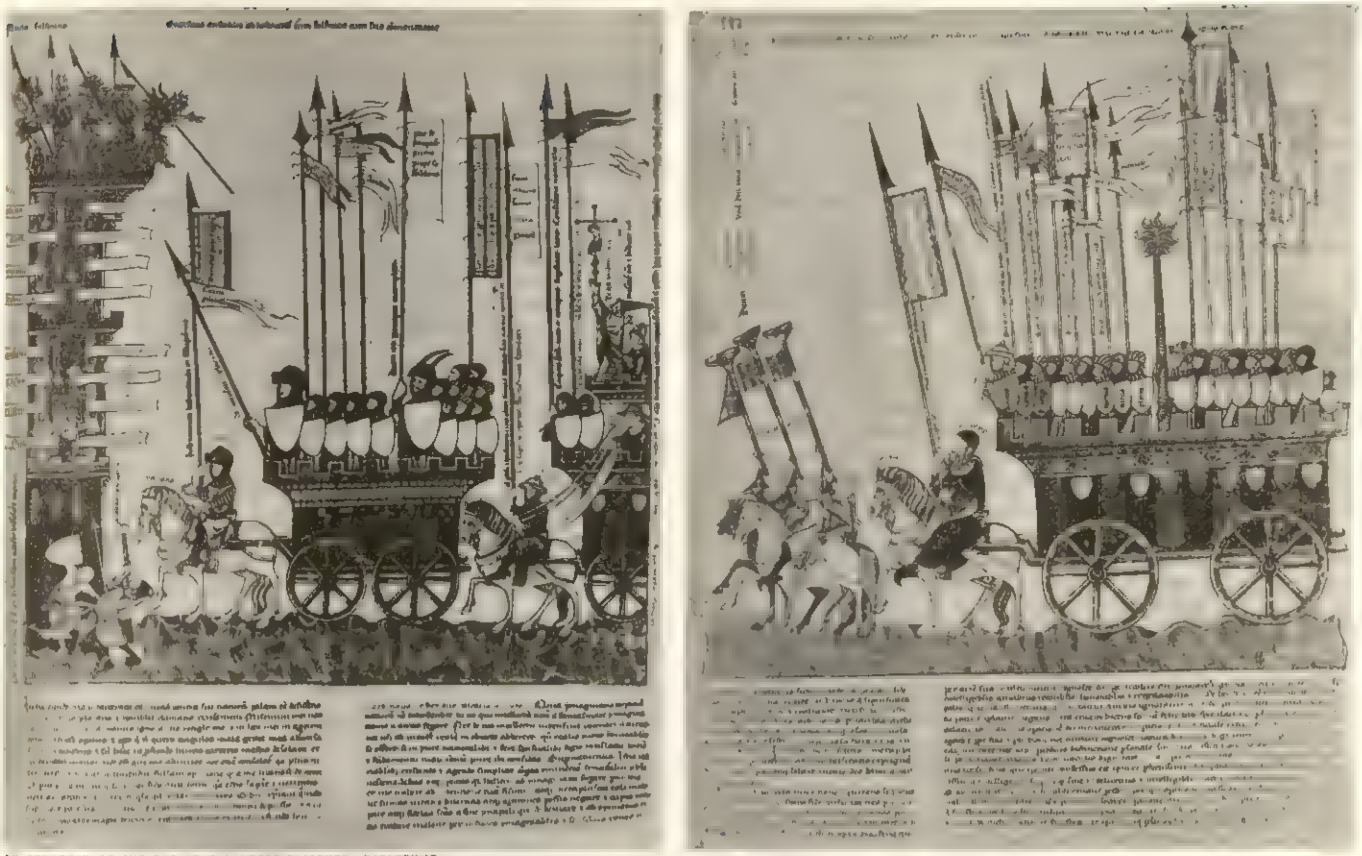
Ramon Vidal was preceptist as well as poet; his *Dreita manera de trobar* became the code for Catalan poets writing in a tongue based supposedly on that of Limoges, and known as *lemosi.* This conventional language of poetry was not proof against the contagion of the spoken tongue, and there resulted a composite idiom well illustrated in the early poetry of Ramon Llull (*c.* 1235–1316). Ramon Muntaner (1265–1336), writing plain Catalan in prose, remains a troubadour in verse. In the later 14th century, there was apparent a gradual waning of Provençal influence and a turning instead for inspiration to northern France, whence derived a cult of long narrative poems on romance themes, including the Arthurian cycle. With the themes came the metre, known as *noves rimades,* a sequence of octosyllabic rhymed couplets. There exist poems of this kind by Pere March (1340–1413), Guillem de Torroella, and Bernat Metge (*c.* 1350–*c.* 1413). Anonymous poems include the *Llibre dels set savis de Roma* ("Book of the Seven Wise Men of Rome"), of more than 3,000 lines, the *Facet o libre de cortesia* ("Book of Courtesy") of about 1,500, and several which carry the new interest in the *langue d'oïl* even to incorporation of passages of French poetry. Another form, the *codolada,* was native in origin; used for narrative but more especially for satire, it combined, again in rhyming couplets, an eight-syllable with a four-syllable line.

The 15th century is the great period of Catalan poetry. It dates from the institution at Barcelona in 1393, under the auspices of John I of Aragon (1387–95), of a poetic academy, the *Consistorio de la Gaya Ciencia,* on the model of that of Toulouse (*see* PROVENÇAL LITERATURE), with *jocs florals* (floral games, including literary competitions). These in themselves produced little poetry of note but typified a new and propitious royal encouragement. This continued under Martin I (1395–1410) and Ferdinand I (1412–16), and contributed to the emancipation of style from alien influences. As the century advanced, Valencia emerged as a new focus of literary activity.

With Johan Berenguer de Masdovelles, Francesc Ferrer, Pere Torroella, Pau de Bellviure, Antoni Vallmanya, and, above all, the Petrarchan Jordi de Sant Jordi (*c.* 1395–*c.* 1440) and Ausiàs March (1397–1459), there developed a school which, as regards versification, is noted for its almost exclusive use of eight-line decasyllabic *cobles* with crossed or "chained" rhymes and a final four-line *tornada* and, in inspiration, shows a turning away from France to Italy. The *Cants d'amor* and *Cants de mort* of Ausiàs March, the most inspired and profound, if also the most difficult, poet of his age in the Iberian Peninsula, contain the finest verses ever written in Catalan; they exerted great influence in 16th-century Castile and continue to influence modern Catalan poets. Very different is *Lo Spill o Llibre de les dones* ("Book of the Natural Gifts"; also called the *Llibre de Consells, c.* 1460) of Jaume Roig (*c.* 1405–78), a bitter and caustic satire upon women, of more than 16,000 four-syllable lines, which derives piquant interest from its vivid portrayal of contemporary Valencian life and manners. Johan Roïç de Corella (1430–1500) is the third great Valencian lyricist, heir to Sant Jordi and March, and perhaps the best representative of the Renaissance spirit, which flowered more fully in Valencia than in Barcelona.

After this bright period, Catalan poetry faded rapidly, the victim largely of political circumstance. From the union of Aragon with Castile there resulted a predominance of Castilian throughout Spain which inflicted a death blow on Catalan literature. The fact that it was a Catalan, Joan Boscà i Almugàver (Castilian, Juan Boscán Almogáver, *c.* 1490–1542), who inaugurated in Castilian a new school of poetry, and that Castilians regard him as a landmark in the history of their Renaissance muse (*see* above), is revealing; by the year of publication of Boscán's works (1543), Catalan poetry had been dead for half a century.

Prose.—The earliest prose works in Catalan are later than the poems of the first troubadours of the Provençal school. Though the oldest document dates from *c.* 1100 (the text of an oath imposed on certain barons by a bishop of Urgel), and the *Homilies d'Organyà* (sermons preserved in an imperfect manuscript, chiefly of linguistic interest) are attributed to the 12th century, literary prose goes back no farther than the close of the 13th. Beginning



BY COURTESY OF THE BADISCHE LANDESBIBLIOTHEK, KARLSRUHE

TWO PAGES FROM A BIOGRAPHY OF RAMON LLULL, "VIDA COETÁNEA," ABOUT 1311

The illumination shows Aristotle, mounted at left, Averroes, mounted at centre, and Ramon Llull, mounted at right, leading an army of personifications of their several philosophies to free Truth from the Tower of Falsehood, where Truth is held captive by an army of demons

as history, often written by prime movers in the events it describes, this prose has the advantage of being original, the language being the everyday speech found in charters from the time of James I's accession to the Aragonese throne in 1213. Its four great 14th-century chronicles are the crowning ornament of medieval Catalan prose. Two—that of James I, apparently compiled shortly after his death in 1276 with the help of memoirs dictated by himself and called the *Llibre dels feyts* ("Book of Facts"); and that of Ramon Muntaner, relating at length the expedition of the Catalan Grand Company to the Morea and the conquest of Sardinia by James II—are distinguished by the artistic skill of their narration and the quality of their language, and invite comparison with those of Villehardouin, Joinville, and Froissart. Bernat Desclot's chronicle deals particularly with the reign of Peter I the Great (1276–85). The account of the reign of Peter IV the Ceremonious (1336–87) is ascribed to Bernat Descoll, but was planned and revised by the king himself. These four works collectively are a perfect expression of the spirit and greatness of medieval Catalonia.

Ramon Llull stands on an eminence apart for his encyclopaedic production, in Catalan, Arabic, and Latin, covering every branch of medieval knowledge and thought. His *Llibre de contemplació en Déu* marks the beginning of Catalonia's golden age of literature; an exhaustive theological treatise, it is also a mine of information on contemporary society. The *Llibre d'Evast e Blanquerna,* his literary masterpiece, laid the foundations of Catalan fiction. *Felix,* or the *Llibre de meravelles* (*c.* 1288) and the *Llibre qui és de l'orde de cavalleria,* a widely translated manual of chivalry, are instructive works with a narrative framework. Inserted in the *Blanquerna* is the *Llibre d'amic e amat* ("Book of the Lover and the Beloved"), a brief masterpiece of mysticism. A century later, Francesc Eximenis (*c.* 1340–1409) represents something of the same intellectual amplitude, if without the originality or driving force, in *Lo Chrestià,* planned as a comprehensive work in 13 books on theology, morals, and politics for the layman, but surviving only in 4, perhaps the only ones written. From its pages emerges a vivid picture of the medieval scene. His *Llibre de les dones,* at once a book of devotion and a manual of domestic economy, is similarly informative on women's way of life.

At the close of the 14th century Bernat Metge inaugurated the Catalan "classical age." Well-versed in Italian literature, he translated in *Valter e Griselda* Boccaccio's story of Griselda from Petrarch's Latin version; in *Lo somni* (1398), which clothes scholastic learning with poetic imagination, he achieved the stylistic masterpiece of Catalan prose. The chivalric romance *Tirant lo Blanc* by Joanot Martorell (d. *c.* 1488) was notable of its kind for the theme, drawn from Muntaner, of the real adventures of the Catalans in the Near East (hence the sparing of it in the famous scrutiny of Don Quixote's library, a satire on the decadent romance). The anonymous late 14th-century *Curial e Güelfa* draws on Desclot, and is the only other Catalan romance in this vein. The beginnings of the drama are represented by the *Misteri d'Elch,* a 15th-century Assumption play based on an earlier *representació;* sung throughout to traditional music, it is still performed at the Feast of the Assumption every year at Elche.

2. Decline: 16th–18th Centuries.—With the loss of political independence, literary and linguistic independence was also lost, and Catalan fell to the level of a patois, kept alive only in the countryside and in the pulpit. The 16th century furnishes a single poet worthy of the name: Pere Serafí, some of whose *Cants d'amor* (1565), written in imitation of Ausiàs March but less obscure, are graceful enough to merit remembrance. In prose, only scholars, chiefly antiquaries and historians, still wrote in Catalan; such were Pere Miguel Carbonell, compiler of the *Chròniques d'Espanya,* Francesc Tarafa (*Crònica de cavallers catalans*), and Anton Beuter, of whose *Crònica general de toda España* the first part (1538; dealing with Valencia) is in Catalan. Forty years of research and abundant documentation give interest to the *Crònica universal del principat de Cathalunya,* a history of the Catalan kingdom, of Jeroni Pujadas, of which also only the first part

(1609) is in Catalan. Thereafter the eclipse was almost complete. Catalan remained only as the language of folksong and ballad; in these—first collected in the *Romancerillo catalán* ("Little Collection of Catalan Ballads"; 1853) by Manuel Milà i Fontanals (1818–84), the historian who played a considerable part in the Catalan revival—it lived on until the reawakening.

3. The Renaixença and After.—In 1814 appeared the *Gramàtica y apologia de la llengua cathalana* of Josep Pau Ballot i Torres, a forerunner of the literary and linguistic renaissance which marked the Romantic period in Catalonia. The pioneers of the rehabilitation of Catalan for literature soon discovered, however, the inadequacies of the ancient language for the expression of their spiritual and intellectual ideas. The very popularity of the satirical poems and farces of a Josep Robreño (1780–1838), for example, was a serious obstacle, because of the corruption of the language in which they were written, as was the later advocacy by Frederic Soler (1839–95) of the use for literature of "Catalan as now spoken" in opposition to literary refinement. The development of modern Catalan literature has been dependent on deliberate enrichment and purifying of the language as a vehicle for contemporary thought; in this the Institut d'Estudis Catalans, founded in Barcelona in 1907, has played a notable part. Bonaventura Carles Aribau's patriotic *Oda a la pàtria* (1832), and the poems of Joaquim Rubió i Ors (1818–99) and Victor Balaguer (1823–1901) prepared the way—as did the reinstitution in 1859 of the *jocs florals*—for the imagination and mysticism of Jacinto Verdaguer Santalo (1845–1902), a great epic poet (*L'Atlàntida*, 1877; *Canigó*, 1886) whose gift is yet essentially lyrical. Miguel Costa i Llobera (1854–1922) cultivated a classical perfection of form; formal beauty is also the mark of his fellow-Majorcan Joan Alcover i Maspons (1854–1926). In Joan Maragall i Gorina (1860–1911) Catalonia found its first great modern poet, one who looks forward rather than back, and whose poetry is suffused with deep spiritual intent. In this spiritual quality, as also, if less happily, in his theory that inspiration absolves the artist from the need for formal perfection, he exerted a powerful influence on later poets. Josep Carner (1884–), Josep-Maria López-Picó (1886–), and Josep-Maria de Sagarra i Castellarnau (1894–1961) are outstanding among many poets of distinction.

The foundations of modern Catalan prose were laid by the critical writings of Rubió i Ors, Francesc Pi i Margall, one of the four presidents of the Spanish Republic of 1873, and Josep Torras i Bages, archbishop of Vich (*La tradició catalana*, 1892). Fiction reached a modest level of achievement late in the century with Josep Pin i Soler (1842–1927), Narcís Oller i Moragues (1846–1930), and Emili Vilanova (1840–1905). One of the best and most influential writers in prose was the essayist Eugeni d'Ors (Eugenio d'Ors y Rovira, pseudonym "Xenius"; 1882–1954), part of whose *Glossari* (1906–17) had already been translated into Castilian before the author abandoned Barcelona for Madrid and himself began to write in that language. His philosophical novel *La ben plantada* (1912) is one of the most notable works in modern Catalan literature.

Catalan dramatists have produced plays of considerable originality. Àngel Guimerà (1847–1924) achieved a European reputation with *Terra baixa* (1896), which inspired a German and a French opera (*Tiefland*, *La Catalane*) and was widely translated. Beginning with historical tragedy (*Gala Placidia*, 1879; *Judit de Welp*, 1883), Guimerà turned to rural drama in *La Boja* (1890), to comedy in *La Baldirona* and *La sala d'espera*, and to modern tragedy in *Maria Rosa* (1894) and *La festa del blat* (1896). The many social dramas of Ignasi Iglésias (1871–1928), inspired originally by the early works of Gerhart Hauptmann, included one near-masterpiece, *Els Vells* (1903). Adrià Gual, author of several works of fantasy, did his best work as director of the *Teatre Intim*, founded in Barcelona in 1898, which familiarized the public with the great drama of all countries and ages. The painter Santiago Rusiñol achieved repute, too, with vivid and imaginative plays.

Hardly had Catalan literature thus won itself once more a place in the sun, when there supervened the dictatorship (1923–29) of Primo de Rivera, who banned the use of any language other than Castilian, and the Civil War of 1936–39. The expatriation of many intellectuals led to the development of a considerable publishing activity in Catalan in Mexico City and Santiago de Chile.

III. GALICIAN LITERATURE

1. Medieval Poetry.—Galician is closely related to Portuguese, and there is no separating the two emergent literatures in the three great repositories of medieval verse, the *Cancioneiros* ["Songbooks"] *da Ajuda, da Vaticana*, and *Colocci-Brancuti* (*see* PORTUGUESE LITERATURE: *Early Period*). Indigenous lyric origins were overlaid by Provençal influence, which had traveled along the pilgrim route to Santiago (*q.v.*) de Compostela, but the dominance of emotion over thought was to remain a constant, so identifying Galician with subjective lyricism that for upward of a century Castilian poets, following Alfonso the Wise (*see* above), made it their medium for works in this vein. Of 116 names in the *Cancioneiro da Vaticana*, 75 have been tentatively identified as Galician; none achieves particular individuality. Macías *El Enamorado* (*fl.* mid-14th century), who died for love and became a legend, was the last Galician troubadour; Galicians thereafter wrote in Castilian, and though echoes of their tradition still abound in the *Cancioneiro de Baena* (*c.* 1445; *see* above), with the Renaissance, and Castilian political hegemony, there descended a sleep even more profound than that affecting Catalan.

2. The Modern Revival.—The Romantic movement, like the Peninsular War, revived local feeling and interest in things Galician, but not in the language. The *Xogos Froraes* (floral games; the equivalent of the Catalan and Provençal *jocs florals*) of 1861, with the first dictionary (1863) and first grammar (1864) of Galician, mark the change. Francisco Añón y Paz (1812–78) was the first notable poet in the resurrected idiom, his most stirring notes being love of country and of freedom. Rosalía de Castro (1836–85; *see* also above), the greatest name in Galician literature, identified herself to a rare degree with the spirit and people of the Galician countryside in *Cantares gallegos* ("Galician Songs"; 1863); her *Follas novas* ("New Leaves"; 1880), introspective to the verge of despair, reflected deep personal sorrows. Eduardo Pondal y Abente (1835–1917) stood somewhat apart: bard of a dimly sensed heroic past, he communed with nature and Celtic mythology (*Queixumes dos pinos*, 1886; *Os Eoas*, an unfinished symbolic epic on the discovery of America). Valentín Lamas Carvajal (1849–1906) is remembered as the voice of the peasant, whose daily life, woes, and simple joys inspired, with growing disillusion, his *Espiñas, follas e frores* ("Thorns, Leaves, and Flowers"; 1871), *Saudades gallegas* ("Galician Sorrows"; 1889), and *A musa das aldeas* ("The Muse of the Villages"; 1890). In Manuel Curros Enríquez (1851–1908) a rebellious temperament, at issue especially with the church, informed the *Aires da miña terra* ("Airs from My Land"; 1880) and *O divino sainete* ("The Divine Farce"; 1888), but he shows too, with genuine feeling for nature, a gentle, pervading melancholy.

Prose showed no comparable achievement. Aurelio Ribalta (1864–1940?), Manuel Lugrís Freire (1863–1935), and Heraclio Pérez Placer expressed in the short story very similar veins of inspiration, but were overshadowed by novelists of stature—Emilia Pardo Bazán, Ramón María del Valle-Inclán—who chose to write for a larger public in Castilian; so too, in her five novels, did Rosalía de Castro.

The 20th century has produced, especially since 1920, a continuing abundance of Galician poets, not yet sufficiently differentiated, who underline the identification of Galician literature with a markedly poetic regional temperament and language.

See also references in PORTUGUESE LITERATURE.

BIBLIOGRAPHY.—*Castilian: Anthologies:* G. Díaz Plaja (ed.), *Antología mayor de la literatura española*, 2 vol. (1958); A. del Río (ed.), *Antología general de la literatura española* (1960); W. T. Pattison (ed.), *Representative Spanish Authors*, 2nd ed. (1958); S. Resnick and J. Pasmantier (eds.), *An Anthology of Spanish Literature in English Translation* (1958); J. M. Cohen (ed.), *The Penguin Book of Spanish Verse* (1956); J. Fitzmaurice-Kelly (ed.), *The Oxford Book of Spanish Verse*, 2nd ed. by J. B. Trend (1940); E. A. Peers (ed.); *A Critical Anthology of Spanish Verse* (1948); E. L. Turnbull (ed.), *Ten Centuries of Spanish Poetry* (1955).

Histories: G. Ticknor, *History of Spanish Literature,* 3 vol., rev. ed. (1910); J. Fitzmaurice-Kelly, *Chapters on Spanish Literature* (1908) and *A New History of Spanish Literature* (1926); E. Mérimée, *Précis d'histoire de la littérature espagnole* (1922; rev. and trans. by S. G. Morley as *A History of Spanish Literature,* 1930); G. T. Northup, *An Introduction to Spanish Literature* (1925; 3rd ed. rev. by N. B. Adams, 1960); A. F. G. Bell, *Castilian Literature* (1938); A. del Río, *Historia de la literatura española,* 2 vol. (1948); G. Díaz Plaja (ed.), *Historia general de las literaturas hispánicas,* 5 vol. (1949–58); G. Brenan, *The Literature of the Spanish People* (1951; 3rd ed., 1957); J. M. Blecua, *Historia y textos de la literatura española,* 2 vol. (1950); R. E. Chandler and K. Schwartz, *A New History of Spanish Literature* (1961).

Middle Ages: W. J. Entwistle, *The Arthurian Legend in the Literature of the Spanish Peninsula* (1925); J. B. Trend, *Alfonso the Sage, and Other Spanish Essays* (1926); F. Whyte, *The Dance of Death in Spain and Catalonia* (1931); M. Menéndez y Pelayo, *Origenes de la novela,* 2nd ed., 4 vol. (1943); A. R. Nykl, *Hispano-Arabic Poetry, and its Relations with the Old Provençal Troubadours* (1946); *Cantar de mio Cid,* ed. by R. Menéndez Pidal, 3 vol., 3rd rev. ed. (1946–57); *The Poem of the Cid,* prose trans. (1957); A. Millares Carlo, *Literatura española hasta fines del siglo XV* (1950); E. S. Procter, *Alfonso X of Castile, Patron of Literature and Learning* (1951); S. M. Stern, *Les Chansons mozarabes* (1953); R. B. Donovan, *The Liturgical Drama in Medieval Spain* (1958).

The Renaissance and the Siglo de Oro: H. A. Rennert, *The Spanish Stage at the Time of Lope de Vega* (1909; new ed. 1964) and *Spanish Pastoral Romances* (1912); H. Thomas, *Spanish and Portuguese Romances of Chivalry* (1920); J. P. W. Crawford, *Spanish Drama before Lope de Vega* (1922); E. A. Peers, *Spanish Mysticism* (1925) and *Studies of the Spanish Mystics* (1927); E. K. Kane, *Gongorism and the Golden Age* (1928); L. von Pfandl, *Historia de la literatura española del siglo de oro* (1933); W. J. Entwistle, *European Balladry* (1939); M. Menéndez y Pelayo, *Historia de les ideas estéticas en España,* new ed., 5 vol. (1940); J. Chuzeville, *Les Mystiques espagnols* (1952); A. A. Parker, *The Approach to the Spanish Drama of the Golden Age* (1957); *El romancero,* ed. by A. Rodriguez Monino, 12 vol. (1957); K. Pond (ed.), *The Spirit of the Spanish Mystics: an Anthology of Spanish Religious Prose from the 15th to the 17th Century* (1958); J. B. Avalle-Arce, *La novela pastoril española* (1959).

The 18th Century: R. E. Pellissier, *The Neo-Classic Movement in Spain During the 18th Century* (1918); A. Gonzalez Palencia, *Eruditos y libreros del siglo XVIII* (1948); R. F. Brown, *La novela española, 1700–1850* (1953); J. Sarrailh, *L'Espagne éclairée de la seconde moitié du XVIII^e siècle* (1954); J. A. Cook, *Neo-Classic Drama in Spain* (1959).

The 19th Century: E. Pineyro, *El romanticismo en España* (1904; Eng. trans. 1934); L. B. Walton, *Pérez Galdós and the Novel of the 19th Century* (1927); I. L. McClelland, *The Origins of the Romantic Movement in Spain,* 2 vol. (1937); F. C. Tarr, *Romanticism in Spain and Spanish Romanticism* (1939); E. A. Peers, *A History of the Romantic Movement in Spain,* 2 vol. (1940; abridged 1949); P. Laín Entralgo, *La generación del noventa y ocho* (1945); J. K. Leslie, *Ventura de la Vega and the Spanish Theatre, 1820–65* (1940); E. Correa Calderón, *Costumbristas españoles,* 2 vol. (1950–51; an anthology and study); M. Aub, *La prosa española del siglo XIX* (1952–); J. Ayuso Rivera, *El concepto de la muerte en la poesía romántica española* (1959).

The 20th Century: A. Valbuena Prat, *La poesía española contemporánea* (1930), and *Teatro moderno español* (1944); E. L. Turnbull, *Contemporary Spanish Poetry* (1945); J. Deleito y Piñuela, *Origen y apogeo del "género chico"* (1949); G. Torrente Ballester, *Panorama de la literatura española contemporánea,* 2nd ed. (1961; studies and selections), and *Teatro español contemporáneo* (1957); R. Castelltort, *La poesía lírica española del siglo XX* (1957); J. L. Alborg, *Hora actual de la novela española* (1958).

Bibliographies, Periodicals: R. U. Pane, *English Translations from the Spanish, 1484–1943* (1944); J. Simón Díaz, *Bibliografía de la literatura hispánica,* 5 vol. (1950–); G. Bleiberg and J. Marías, *Diccionario de la literatura española,* 2nd ed. (1953). *See* also the following periodicals: *Revista de filología española* (1914–); *Boletín de la Real Academia Española* (1914–); *Bulletin of Spanish Studies* (1923–); *Hispanic Review* (1932–); *Nueva Revista de Filología Hispánica* (1947–).

Catalan: J. Massóo i Torrents, *Revista de bibliografia catalana* (1900–07); J. Arús, *Evolucio de la poesía catalana* (1923); M. Garcia Silvestre, *História sumária de la literatura catalana* (1932); O. Saltor, *Les idées litèraries en la Renaixença Catalana* (1934); M. de Riquer, *Resumen de literatura catalana* (1947), *La lírica de los trovadores* (1948; an anthology with notes), and *Los cantares de gesta franceses. Sus problemas, su relacion con España* (1952).

Galician: A. de las Casas (ed.), *Antología de la lírica gallega* (1950); *Poems of Rosalía de Castro,* trans. by C. D. Ley (1964); J. Mouriño, *La Literatura medioeval en Galicia* (1928); B. Varela Jácome, *História de la literatura gallega* (1951). (W. C. An.; X.)

SPANISH MAIN, the name applied during the period of Spain's empire in the Western Hemisphere to the mainland of Spanish America, particularly to northern South America west of the mouth of the Orinoco River and to what is now Central America and Mexico. Early Spanish expeditions first gave the name to these regions. During the development of the Spanish colonial trade the name was also applied to the Caribbean Sea, across which moved the bulk of Spain's colonial commerce in gold, silver, spices, tea, hides, etc. Spain's chief commercial ports on the Caribbean, most of which were strongly fortified against enemy and pirate raids, were Cartagena, Portobelo, Port Royal (Jamaica), Veracruz, Havana, and San Juan (Puerto Rico). (C. F. J.)

SPANISH MOSS, the name given to an epiphytic, herbaceous plant (not a true moss), *Tillandsia usneoides,* of the bromelia family (Bromeliaceae). It hangs in long festoons from the branches of trees in tropical and subtropical American forests, being especially conspicuous on the live oak in the southeastern United States. The whole plant is covered with scaly hairs for the absorption of water, giving it a gray colour. It is used, after processing, as an excellent substitute for horsehair stuffing.

Ramalina reticulata, a lichen called Californian Spanish moss, grows supported on trees in humid valleys of the western coastal range of the United States.

SPANISH SAHARA, an overseas province of Spain on the low, desert Atlantic coast of northwest Africa, comprises the Southern region of Río de Oro, between Cape Blanc (Blanco) and latitude 26° N near Cape Bojador; and the Northern region of Saguia el Hamra, between that line and the 27° 40′ N line of latitude, the boundary with Morocco. Spanish Sahara is bounded by Algeria for a few miles in the northeast and otherwise entirely by Mauritania. The total area is 102,703 sq.mi. (266,000 sq.km.; Río de Oro 71,043 sq.mi. [184,000 sq.km.], Saguia el Hamra 31,660 sq.mi. [82,000 sq.km.]). The capital of the province is El Aiún (Aiun), and it is administered by a governor general.

Along the coasts are occasional cliffs and bold headlands, as at Cape Bojador, but sand dunes are also common. In both regions the coastal plain has rocks of Cretaceous or later age, though they are often masked by sand. These rocks have been prospected for oil, especially in the northwest. The east has Paleozoic and Precambrian rocks, rises to about 1,450 ft. (440 m.), and is rugged with several deep ravines, all permanently dry. Rain occurs only rarely. In the interior the daily range of temperature is commonly 11° to 44° C (52°–112° F). Only along the coast are conditions more temperate, as the result of the influence of the sea, the cool Canary Current, and the northeasterly trade winds.

The total population (1960) was 23,793 mostly in Saguia el Hamra, excluding nomads who spend most of the year in the province. The capital of Río de Oro is Villa Cisneros (pop. [1960] 1,961), a small settlement with white-domed buildings that stands on a 23-mi. (37 km.)-long peninsula. In Saguia el Hamra there is a permanent settlement at Smara, as well as at El Aiún (pop. [1960] 5,251). Most of the people of both regions are Berbers, more or less arabized with some Negro admixture. All, except about 5,300 Europeans, are Muslims.

In neither region is there any agriculture to speak of. Camels, goats, and sheep are raised, and there is some export of their skins, as well as of dried fish. Between Villa Cisneros and the mainland is Río de Oro Bay, which provides a sheltered but shallow anchorage. The town is connected with the rest of the region, with the Canary Islands, and with Spain by air, and there are tracks north to Cape Yubi (Juby) and south to Tichlá and La Agüera, as well as several caravan routes. A shipping service links Villa Cisneros with La Agüera and the Canary Islands.

History.—Río de Oro Bay was discovered by the Portuguese in 1436, mistaken for a river, and named Rio de Ouro ("River of Gold") on account of gold dust obtained from the inhabitants. The Spaniards had already begun the colonization of the Canary Islands, and in 1476 they established the fort of Santa Cruz de Mar Pequeña on the mainland. It was abandoned in 1524, and the territory remained independent until the 19th century. In 1878 a British trading post was established by Donald Mackenzie on Cape Yubi, but it was surrendered to the ruler of Morocco in 1895. Meanwhile French travelers had visited Saguia el Hamra. The Spaniard C. Benitez and others explored the coast in 1882,

and E. Bonelli signed treaties with the inhabitants on behalf of the Sociedad Española de Africanistas. In 1884 a Spanish protectorate was declared over the coastal zone from Cape Bojador to Cape Blanc and nominally subordinated to the captain general of the Canaries. A factory was established at Villa Cisneros in 1885, but it was abandoned in the face of attacks by the Africans. Spanish claims were challenged by the French, who claimed a protectorate over the Sahara, and a Franco-Spanish convention of 1900 established the landward frontier of Río de Oro, awarding the salt deposits of Ijil (Fort Gouraud) to the French. Other agreements (1906 and 1912) resulted in French recognition of Spanish rule in Saguia el Hamra. The strip of desert country bounded on the north by the Wadi Draa and by parallel 27° 40′ on the south was ceded to Morocco in 1958.

(H. V. L.; R. J. H. C.)

SPANISH SUCCESSION, WAR OF THE (1701–1713), the name given to the European war that arose out of the disputed succession to the throne of Spain and to the Spanish Empire following the death (1700) of the childless Charles II of Spain. The war was primarily a struggle to determine whether these vast possessions should pass to the House of Bourbon or to the House of Habsburg, both of which had dynastic claims (*see* genealogical table below), or whether they should be partitioned to preserve the balance of power in Europe.

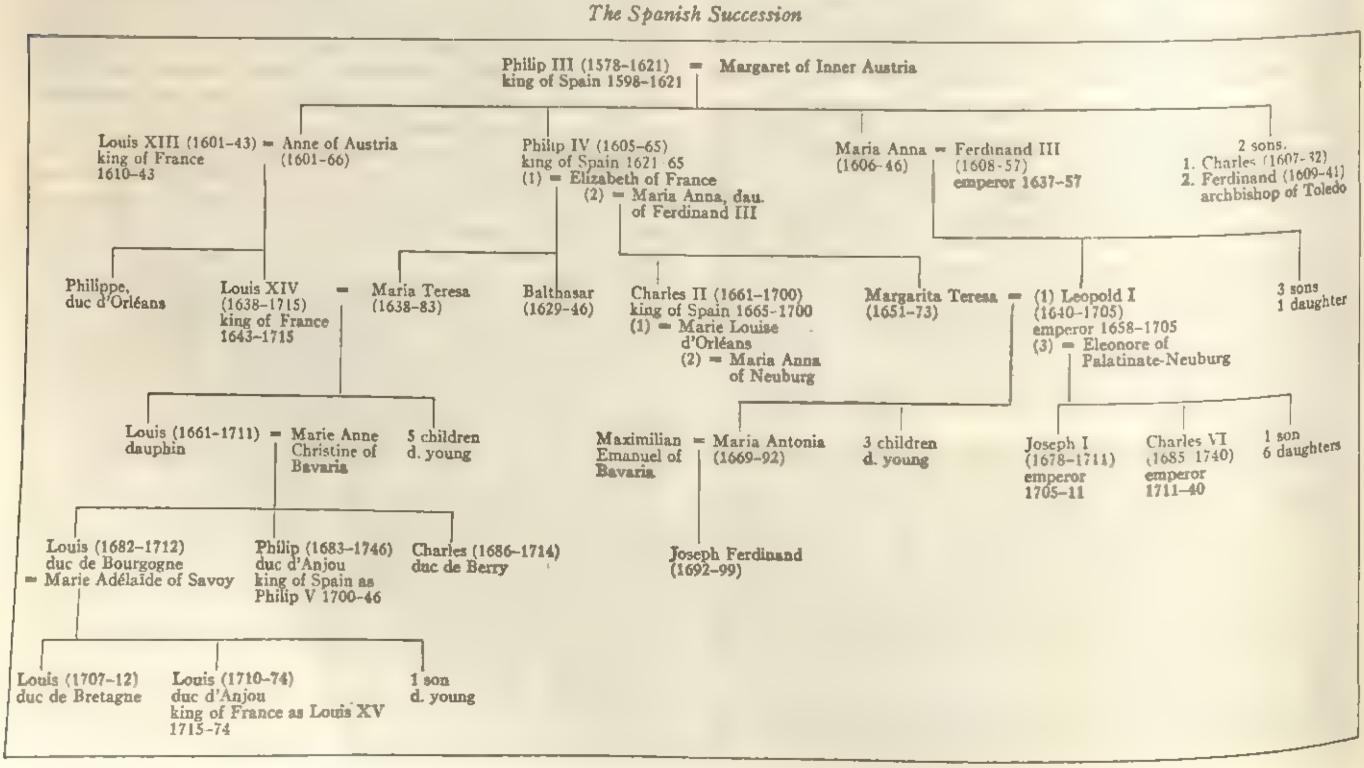
Dynastic Claims.—Of the two daughters of Philip III of Spain (Charles II's grandfather), the elder, Anne, married Louis XIII of France, and the younger, Maria Anna, married the future Habsburg emperor Ferdinand III. Two sons of these marriages, Louis XIV and the emperor Leopold I, respectively, married their Spanish cousins Maria Teresa and Margarita Teresa, the daughters of Philip IV and sisters of Charles II. Maria Teresa explicitly renounced her claim to the Spanish succession on her marriage to the French king, as her aunt Anne had done on the occasion of her marriage to Louis XIII. But Margarita Teresa (d. 1673) made no renunciation and, moreover, was named in Philip IV's will as the next heir after his son Charles (Charles II of Spain). Consequently, when Margarita's daughter Maria Antonia (d. 1692), who in 1685 married the Bavarian elector Maximilian Emanuel, gave birth in 1692 to a son, the electoral prince Joseph Ferdinand, this prince could be regarded as heir presumptive to Charles II. Leopold I, however, had persuaded Maria Antonia to bestow her right to her mother's succession on him, and on the sons of his third marriage, with Eleonore of Palatinate-Neuburg; but the validity of this bestowal, on which the immediate Habsburg claims to the succession were based, was dubious. The Bourbon claim was similarly dubious, being based on a disregard for the French queens' acts of renunciation. The claim of the electoral prince Joseph Ferdinand, on the other hand, appeared superior to both.

In the precarious lifetime of Charles II, French influences at the Spanish court were heavily counterbalanced in the field of international politics by the attitude of the Maritime Powers, England and the United Provinces of the Netherlands. United after 1688 under the personal rule of William III, they were strongly opposed to the Bourbon claim, fearing not only the aggrandizement of French power but also the loss of their well-established trade with Spain and the Indies. Louis XIV, becoming convinced that a Bourbon success would result in a general European coalition against him, and to avoid the Spanish possessions passing to a Habsburg, in October 1698 agreed with William III in a Partition Treaty (signed at The Hague) to recognize Joseph Ferdinand's rights to Spain, the Spanish Netherlands, and the Indies. Milan, however, was to go to the archduke Charles (later the emperor Charles VI), the emperor's younger son (who would presumably be excluded from the imperial throne by his elder brother Joseph); and the rest of Spanish Italy, as well as the Basque province of Guipúzcoa, was to go to the dauphin Louis (1661–1711).

Spain was indignant at this partition treaty, and in November 1698 Charles II made a will naming Joseph Ferdinand heir to the whole inheritance. But in February 1699 Joseph Ferdinand died. A second Partition Treaty (October 1699) between France and England, which was subscribed in March 1700 by the United Provinces, offered Spain, the Spanish Netherlands, and the Indies to the archduke Charles; the dauphin was to get all Spanish Italy except Milan, instead of which he would have Lorraine. The emperor, whose approval of the First Partition Treaty had been sought in vain, now also declined to accede to the second, hoping to win the whole succession for the archduke Charles. To counter this new partition, Charles II in October 1700 made another will, leaving the Spanish dominions, which were to be undivided and to be kept separate from any other crown, to Philip, duc d'Anjou, the second son of the dauphin Louis; failing Philip, to his younger brother Charles, duc de Berry; failing Charles, to the archduke Charles.

When Charles II died on Nov. 1, 1700, Louis XIV was con-

The Spanish Succession



fronted with a dilemma. If he accepted the will, he was faced with the prospect of war with the emperor, who might have the support of the Maritime Powers. If he refused the will and stood by the Second Partition Treaty, the inheritance (by the terms of the will) would pass to the Habsburgs, and France would have to fight for those possessions promised by the treaty. As war appeared inevitable in either case, Louis chose the former alternative, acknowledging Philip of Anjou as Philip V of Spain. The war might have remained a struggle between the king of France and the emperor if Louis had not acted with haughty arrogance and provoked England and the United Provinces to enter the war. In December 1700 Louis solemnly acknowledged the rights of his grandson Philip to remain in the line of succession to the crown of France. On the death (September 1701) of the exiled James II of England, Louis recognized James's son as King James III. During 1701 he also made it clear that French merchants would in future enjoy exclusive rights to share in the slave trade to the Spanish colonies in America.

Strategy of the War.—The War of the Spanish Succession has the complexity of any major war fought between two groups of allies whose interests and ambitions are sometimes in sharp conflict. It was fought on five fronts —the Low Countries, the Rhine, the Danube, Northern Italy, and Spain—as well as at sea. It was a war of some considerable movement, for it was conducted according to the strategic conception of the 18th century, which advocated maneuver and countermaneuver to contain an enemy, rather than the Napoleonic strategy of striking at the enemy's main force regardless of losses. Armies of the 18th century, before the days of national conscription, were too difficult to replace for generals to be eager to risk heavy casualties if these could be avoided. One very interesting fact that emerges in the study of the War of the Spanish Succession is how often John Churchill, 1st duke of Marlborough (*q.v.*), favoured a strategy that was much more forceful and "Napoleonic" than was usual in the early 18th century, and it is clear that his bold plans were sometimes too unorthodox for his imperial ally, Prince Eugene of Savoy (*q.v.*). Other points of interest are the relative importance of the various allies in determining the outcome of the war, the number of times that military operations were interrupted by attempts at peace negotiations, and the extent to which the final outcome of the war was decided not by victories in the field but by political developments in the capitals of the combatants. It is a curious fact that the fate of Spain was decided principally by campaigns in the Low Countries or on the Danube, with very little regard to what was actually happening in the Peninsula itself.

FIG. 1.—SOUTH-CENTRAL EUROPE DURING THE WAR OF THE SPANISH SUCCESSION

THE COURSE OF THE WAR

1701. The Emperor and Louis XIV Begin Hostilities.—The war may be said to have begun in March 1701 when French troops took over from Spanish forces the fortresses in the Spanish Netherlands; but for the rest of that year the war was essentially one between Louis XIV and the emperor Leopold I, and its principal theatre was Northern Italy, where the French had been able to get control of the Spanish possessions in the valley of the Po. Prince Eugene assembled an Imperial Army in Tirol and marched south with great secrecy through the territory of neutral Venice along the east bank of the Adige River. Marshal Nicolas Catinat (*q.v.*), commanding the French, failed to attack a numerically inferior Imperial force that, by a feint at Chiusa, had distracted his attention from Eugene's main move on Verona (May 28), nor did he cross the Adige and force battle on Eugene's main army, though this too was inferior in numbers to his own. Eugene maintained the initiative and forced the French to retreat westward behind the Mincio River. He then succeeded in following them

across this river (July 28), so that, to defend Milan, Catinat had to fall back still farther westward beyond the Oglio River (Aug. 16). These humiliating maneuvers were reported to Louis XIV by Catinat's enemies at the French court, and he was replaced by François de Neufville, duc de Villeroy (*q.v.*); but when the new commander tried to take the offensive, he was sharply defeated at Chiari (Sept. 1). The campaigning season ended with the Imperialists in command of most of the duchy of Mantua, and during the winter of 1701–02 Eugene executed a brilliant raid on Cremona, where he captured Villeroy, a success that led the dukes of Modena and Guastalla to declare for the emperor.

1702. The War Becomes General.—By 1702 the efforts of the diplomats had transformed the struggle into a general European war. France and Spain had the support of two reluctant allies, Portugal and Savoy, who feared to oppose the Bourbons, whose forces controlled Spain and the Spanish possessions in Northern Italy. Two allies who were to be of more value to France were the Wittelsbach brothers, Maximilian Emanuel, elector of Bavaria, and Joseph Clement, elector of Cologne. The emperor had the support of the United Provinces and England by the Treaty of The Hague (September 1701), and of Frederick I of Prussia and a great many minor German princes. But the first year of the general struggle saw little decisive action. In Italy Eugene's communications with Vienna were cut by Louis Joseph, duc de Vendôme (*q.v.*), who had taken over Villeroy's command; nevertheless, Eugene was able to maintain himself and to harass the French. At sea an English expedition under Adm. Sir George Rooke against Cádiz (August–September 1702) was a failure, but Rooke managed in October to destroy the Spanish silver fleet that had taken refuge in Vigo harbour.

In the Low Countries the Maritime Powers were outnumbered by the French so that there was a serious threat to Holland. This had a deep effect on the Dutch deputies, who remained excessively cautious and timid throughout the war. The Franco-Spanish forces had two fortified lines, one running from Antwerp to Huy, on the River Meuse, the other from Antwerp along the Scheldt and Lys rivers to Aire. In July Marlborough assembled an army near Nijmegen and attacked southwest toward Diest, driving back Louis François, duc de Boufflers (*q.v.*), toward his fortified lines. Marlborough outmaneuvered him and reached the lines first, but the Dutch refused to allow him to give battle, being satisfied that by these maneuvers Marlborough had forced Boufflers to abandon his threat to the United Provinces. The Maritime Powers did, however, succeed in capturing Kaiserswerth (June) and Rheinberg on the lower Rhine and thus improved their communications with the emperor.

The 1702 campaign on the Rhine opened uneventfully, but the end of the summer saw a development that was next year to become one of the most important operations of the war. An Imperial Army under the margrave Louis of Baden, which had been gathered round the Neckar River, crossed the Rhine north of Speyer (June 1702) and threatened Alsace. Catinat, now in charge of the German front, was uncertain whether to mass the French forces on the Lauter or the Ill River, and Louis of Baden successfully besieged and captured Landau in September. It was at this point that the development took place that was to grow into the major campaign of 1703–04. The elector of Bavaria now openly declared for France and captured Ulm. This threat in his rear forced Louis of Baden to retreat to the east side of the Rhine. Louis Hector (later duc) de Villars (*q.v.*), who had superseded Catinat, followed the retreating Imperial Army, crossed the Rhine at Hünningen and defeated the Imperialists at Friedlingen (Oct. 14, 1702).

1703. The Tide Turns in Favour of the French.—In 1703 fortune began to favour the French, although in three theatres of war the Imperialists and the Maritime Powers managed to achieve minor successes. In the Mediterranean the English fleet was able to blockade the French fleet in Toulon, and this convinced the Portuguese government that England could provide effective protection against the Bourbons. In May 1703 Portugal, therefore, relinquished its uneasy alliance with France and joined the Maritime Powers. On the lower Rhine Marlborough invaded the electorate of Cologne and in May captured Bonn. By the end of the summer Marlborough, though impeded by the cautious Dutch generals, had managed to push the French out of the country between the Meuse and the Rhine. In Italy Vendôme was justifiably suspicious of the loyalty of his ally Victor Amadeus II, the duke of Savoy, and demanded that he should hand over Turin and Susa to the French and even disband his troops. This demand eventually decided Victor Amadeus to abandon his uneasy alliance with France and join the emperor. This he did in October, but it was too late in the year to have any effect on military developments in Italy that season.

These minor successes of the emperor and his allies, however, were more than outweighed by the French and Bavarian threat to Vienna itself. In March Villars had crossed the Rhine at Kehl and, pushing across the Black Forest, had in May joined the elector of Bavaria near Ulm. The Imperialist forces under Louis of Baden had offered no resistance but had remained to the north in the lines of Stollhofen, on the Rhine, watched by a French force under the comte de Tallart. The situation facing the emperor was very grave, but the elector of Bavaria lacked Marlborough's dash. Instead of pushing on down the Danube and striking direct at Vienna, the elector decided first to establish control of Tirol so as to safeguard his communications with Milan. This more cautious strategy might have been justified, for the elector was able to reach the Brenner Pass at the beginning of July, but Vendôme's force, which he had hoped would be there to join him, did not manage to push north through Italy till some weeks later. By this time the elector had been forced to retreat toward Bavaria. The possibility of seriously threatening Vienna, however, was still not out of the question. Though by now Louis of Baden and the Imperial general Hermann Styrum had brought up their forces to cover any attack on Vienna, they had failed to combine, so that Villars was able to attack Styrum and defeat him at Höchstädt, near Blenheim (Sept. 20). Villars was in favour of making a dash for Vienna, but the elector of Bavaria refused. There was a quarrel, and Villars, one of the most brilliant of the French commanders, was recalled. On the Rhine in September Tallart took Breisach and in November defeated General Thüngen at Speyerbach and recaptured Landau, but these minor French successes could not repair the damage done by the elector of Bavaria's excessively cautious strategy on the Danube.

1704. The French Threat to Vienna Checked.—When the campaigning season opened in 1704, the serious effects of the elector of Bavaria's delays of the previous year were not yet apparent, and it looked as if the French and Bavarian threat to Vienna was as dangerous as ever. In April strong French forces crossed the Black Forest and, in May, joined the elector of Bavaria near Dillingen. Louis of Baden failed to stop the junction of these two forces, and, though Eugene himself had been moved to the Danube front to replace Styrum, it seemed impossible with the forces available that he would be able to stop the French and Bavarian armies from striking at Vienna. Another French force under Tallart was to the north, near Kehl, to protect the communications between France and the French Army on the Danube. A third French force, under Villeroy, was in the Netherlands, holding Marlborough. It was at this point that Marlborough showed his genius for appreciating the strategy of the war as a whole and his consummate skill in moving troops with unexpected speed and complete secrecy. Marlborough knew that he would never be able to gain the consent of the Dutch to a move to relieve Vienna from the French and Bavarian threat if this seemed to leave Holland unprotected. Therefore, he pretended that he was going to turn the flank of Villeroy's force by a move up the Moselle. Marlborough moved across the Meuse to the Rhine and made his way up that river, reaching Mainz at the end of May. Tallart, from his position farther south, at once moved across to the left bank of the Rhine, fearing that Marlborough intended to attack Alsace, but Marlborough only made a feint of attempting to cross the Rhine at Mannheim. Instead, from Mainz he pushed southeast across the Main River, up the valley of the Neckar, across the watershed between the Neckar and the Danube, and by the middle of June joined forces with Louis of Baden just north of Ulm.

The Imperial commanders had been able to discuss plans with Marlborough during his march, and it was decided that Eugene should move westward to contain Tallart at Stollhofen, while Marlborough and Louis of Baden were to attack the French on the Danube. Marlborough was able to take Donauwörth (July 2), thus forcing the French and Bavarian troops, under the elector and the comte de Marsin, Villars' successor, to retreat south up the Lech River, which enabled Marlborough to get between his enemies and Vienna. At this point the French forces were increased, for Villeroy, once he had realized that Marlborough had given him the slip, moved south from the Netherlands to join Tallart on the Rhine in Alsace. Tallart, therefore, was free to cross the Rhine at Kehl (July 6) and join the elector of Bavaria near Augsburg (Aug. 6). On his march Tallart made the serious mistake of wasting five days trying to capture Villingen. As a result, Eugene, who had marched from Stollhofen along the north bank of the Danube, was able to reach Höchstädt and join forces with Marlborough on the same day that Tallart joined the elector. Marlborough's force was then large enough for him to detach Louis of Baden, an uncongenial colleague, to besiege Ingolstadt, the one strong place on the Danube east of Donauwörth held by the Bavarians, and which was a serious threat to communications with Vienna. The French and Bavarians, learning that Louis of Baden had gone, moved north to attack Eugene, crossing the Danube at Dillingen. Marlborough had remained south of the Danube to cover the siege of Ingolstadt but was not out of touch with Eugene and was able to move to his assistance. A forced march brought his army to the north of the Danube and into line with that of Eugene near the village of Blenheim. Marlborough and Eugene attacked the enemy at the Battle of Blenheim (*q.v.*) on Aug. 13, 1704, and achieved a victory that shattered the reputation of the French Army and compelled the French to withdraw to the west of the Rhine. By the end of 1704 the whole situation on the Danube front had been altered. Vienna was no longer in danger. The French invasion had been destroyed, and the elector of Bavaria was now a fugitive in France, the whole of his electorate falling into the hands of his enemies.

Developments in Spain and the Mediterranean during 1704 were also fairly encouraging for the Imperialists and their allies. The accession of Portugal to the side of the Imperialist and Maritime Powers meant that it was now much easier to launch an invasion of Spain itself, and in February the archduke Charles was escorted to Portugal by an English squadron under Rooke to attempt the conquest of Spain. He achieved little on land, and Rooke, who had been ordered to sail from Lisbon into the Mediterranean and join the duke of Savoy in an attack on Toulon, found Savoy unable to spare troops for his enterprise. But on his way back to England, Rooke managed to capture Gibraltar (Aug. 4) and to defeat the French fleet off Vélez-Málaga (Aug. 24), which gave England command of the Mediterranean for the rest of the war.

1705. Stalemate.—Marlborough had intended to launch an attack from the Rhine against France itself by way of Metz, but the Dutch failed to produce enough supplies, and the Rhenish electors did not provide sufficient transport. Moreover, the death of the emperor Leopold I in May 1705 had the effect of deflecting Imperialist forces to cover the election of the new emperor (Joseph I) at Frankfurt. Marlborough, therefore, transferred his attention to the Netherlands, where he succeeded in piercing the French lines at Tirlemont (July 18); but because of the excessive caution of his Dutch allies, Marlborough was not able to make full use of his success. In Italy Eugene pushed the French from the Oglio River as far west as the Adda River, thus relieving some of the pressure on Savoy, where Turin was under siege by the French.

1706. A Year of Disaster for the French.—In spite of French successes in Italy, where Vendôme in April 1706 drove the Imperial Army back into Tirol, Marlborough still had plans for joining forces with Eugene, expelling the French from Northern Italy, and threatening Toulon. In the same month, however, Louis of Baden retreated eastward across the Rhine, and the Dutch were, therefore, too nervous to consider Marlborough's unorthodox scheme. Instead, England provided subsidies that enabled the emperor to buy 24,000 troops from German princes. In July Eugene attacked south down the left bank of the Adige, once more disregarding Venetian neutrality as he had in 1701. This time he struck south across the Po (July 18), then moved with astonishing speed westward up that river to join the duke of Savoy at Villastellone, south of Turin, and in September raised the siege of Turin. This victory decided the issue in North Italy. By the Convention of Milan in March 1707 Louis XIV gave up his attempt to control North Italy and withdrew all his troops.

In the Netherlands Villeroy, receiving information that Marlborough had not yet been reinforced by Prussian and Hanoverian forces, struck toward Liège but was heavily defeated by Marlborough at the Battle of Ramillies (May 23, 1706). Marlborough pursued the retreating enemy so hard that they were unable to reform on the line of the Lys River but had to retreat still farther to Courtrai. The result of this victory was that in less than two weeks Marlborough was in command of all Spain's province of Brabant and most of its possessions in Flanders. Antwerp, Ghent, Bruges, Oudenaarde all capitulated, and in July the fall of Ostend improved Marlborough's communications with England. At the end of August he obtained Menen; in September, Dendermonde; and in October, Ath. The victory in the Spanish Netherlands was overwhelming, and Marlborough's success helped Eugene's 1706 campaign in Italy, because, after the Battle of Ramillies, Vendôme was recalled from Italy to replace the discredited Villeroy. Marlborough's victory also had the effect of paralyzing the French forces on the Rhine, for troops were transferred from that quiet sector to reinforce the shattered French forces in the Netherlands.

In Spain the Imperialists and their allies also achieved successes in 1705–06. A siege of Gibraltar by Bourbon forces was relieved in March 1705. Henri de Massue, earl of Galway (*q.v.*), commander of the allied forces in Portugal, advanced from Portugal into Estremadura and compelled the comte de Tessé to abandon Andalusia. The English admiral Sir Cloudesley Shovell took a squadron into the Mediterranean and enabled Charles Mordaunt, 3rd earl of Peterborough (*q.v.*), to capture the fort of Montjuich near Barcelona in September and to take Barcelona in October 1705, thus inducing the traditionally separatist Catalonia and Valencia to declare for the archduke Charles. In 1706, though Tessé besieged Barcelona, he was compelled to withdraw in May when an English fleet forced the French ships supporting Tessé to retire to Toulon. In June Imperialist and English forces seized Madrid, though Galway was unable to maintain his hold on the capital and had to withdraw to Valencia.

1706. The Opening of Peace Negotiations.—So decisive were the reverses experienced by France and its allies in 1706 that in August Louis XIV made an approach to the Dutch for peace. He was prepared to cede Spain and Spanish America to the archduke Charles if Philip retained Milan, Naples, and Sicily. The Dutch were offered a strong "barrier" of fortresses in the Spanish Netherlands, but the English and Imperial governments would not consider terms that envisaged the partition of the Spanish Empire.

1707. The Imperialists and Maritime Powers Suffer Reverses.—In Germany the emperor's situation became critical because of developments in the Northern War (*q.v.*), which was raging in the Baltic. Charles XII of Sweden had established himself in Saxony in September 1706, and there was a serious danger that where the elector of Bavaria and French troops had failed in their attack on Vienna, Charles might succeed. The emperor was already short of troops because many minor German princes were reluctant to hire troops to him when they might be needed for their own defense. But Charles XII was not prepared to collaborate with France to attack Vienna, and the crisis passed. Its effects were felt, however, on the Netherlands front, where the absence of German troops made it impossible for Marlborough to achieve anything of note against Vendôme. On the Rhine the Imperial forces were weakened by the death (January 1707) of Louis of Baden, whose successor in command, Margrave Charles Ernest of Brandenburg-Bayreuth, proved considerably less able.

FIG. 2.—THE LOW COUNTRIES DURING THE WAR OF THE SPANISH SUCCESSION

Villars managed to take the lines of Stollhofen in May and was able to raid the whole of Swabia until the margrave of Bayreuth was superseded by George, elector of Hanover (later George I, king of Great Britain). In September Villars withdrew in good order to the west of the Rhine after some of his troops had been diverted to the Provence front (*see* below).

In Spain the emperor and his allies experienced considerable reverses. Galway and the archduke Charles quarreled and separated, the archduke retiring to Catalonia. This left Galway with only 15,000 men, who were decisively defeated by James Fitz-James, duke of Berwick (*q.v.*), at the Battle of Almansa, in Murcia (April 25, 1707), partly because most of Galway's Portuguese troops did not engage the enemy. Aragon, Valencia, and Murcia were lost to the allies, and in Catalonia they were compelled to remain on the defensive. In Northern Italy the emperor, having gained control in 1706, decided that Eugene should launch an attack against Toulon. Unfortunately, Savoy delayed sending help, and some Imperial troops under Wierich, Graf von Daun, were also diverted to capture Naples. After the defeat at Almansa the help Eugene expected from Spain was not forthcoming, and instead French troops returning victoriously were able to reinforce Tessé at Toulon. Eugene's siege of Toulon (July–August) had finally to be broken off, and the only success achieved by Eugene's thrust was the scuttling of the French fleet in the harbour.

1707–08. Peace Negotiations Resumed.—In spite of the improvement in French fortunes during 1707, Louis XIV was still eager to end the war, and in the winter of 1707–08 envoys from Louis and Philip V made a fresh overture to the Dutch, baited with offers of commercial advantages at the expense of Spain. But once again the negotiations were checked by the uncompromising attitude of England, where in December in the House of Lords the Whigs passed a resolution refusing to make a peace by which a member of the House of Bourbon should retain any of the Spanish possessions.

1708. The Imperialists and Maritime Powers Achieve Further Successes.—A project for a three-pronged offensive against France planned early in the year proved impracticable, partly because Saxony and Prussia did not provide their proper contingents. Instead, Eugene began his march from Italy to join Marlborough in the Spanish Netherlands, where Vendôme had managed at the beginning of July to recover Bruges and Ghent and so regain control for France of most of the Spanish possessions in West Flanders. Though Eugene's troops had not yet arrived, Marlborough moved to check Vendôme's advance and defeated the French at the Battle of Oudenaarde (*q.v.;* July 11, 1708). Marlborough then wanted to invade France itself and launch a simultaneous invasion of Normandy, but his plan was considered by Eugene to be too unorthodox. The Imperialists and their allies, therefore, concentrated instead on besieging Lille, which finally fell on Oct. 22; this forced the French to relinquish their hold on the Spanish possessions in West Flanders and to retire to France

1709. Third Peace Negotiations.—The weakness of France was clearly shown in the peace terms that Louis XIV was now prepared to consider. When the French envoys, Antoine Louis Rouillé and Jean Baptiste Colbert, marquis de Torcy, again approached the Dutch (March 1709) at The Hague, though at first they once more demanded at least Naples and Sicily for Philip V, they finally made it clear that Louis XIV was prepared to relinquish the whole of the Spanish inheritance. But again the English and Dutch proved implacable. They insisted that Louis XIV should obtain the surrender of the Spanish throne by Philip, and that, if Philip refused, Louis should agree to use force to expel him. It now seems clear that the English secretary of state, Charles Townshend, the emperor, and the Dutch were chiefly responsible for these unrealistic terms and that Marlborough, though he did not exert his influence enough, had a sounder appreciation of what could be reasonably expected from France. As it was, on May 28 Louis XIV rejected the terms offered, and the war was resumed.

1709. Hostilities Resumed.—The lull in hostilities caused by the peace negotiations of 1709 were used by the French to make a prodigious effort to put yet another army into the field under Villars, who took up a strongly defended position in the Netherlands. Marlborough wanted to attack along the French coast, supported by the English fleet, but this plan was again too unorthodox to win the approval of his allies, and instead the Imperialist and Maritime forces besieged Tournai (July–September) and later moved against Mons. Villars advanced against his opponents and met them at Malplaquet, southwest of Mons. Marlborough was in favour of joining battle at once, but this was vetoed, and Villars gained two vital days in which to strengthen his position, with the result that when the Battle of Malplaquet (*q.v.*) was eventually fought on Sept. 11, 1709, although Marlborough achieved a victory, it was at the cost of 20,000 casualties as compared with about 11,000 French. In October Mons fell, but the campaign had been disastrously costly to the Maritime Powers and the Imperialists.

On other fronts the Imperialists achieved very little. The duke of Savoy failed to cooperate with Daun in a move against Dauphiné, and though an Imperial Army under Claudius, Graf von Mercy, crossed the Rhine near Basel, it was defeated at Huningue (Aug. 26). In Spain Galway's Portuguese troops had proved unreliable, and it was decided in 1708 to send them home and replace them with German troops from Italy, released by the

Imperial successes of 1706–07. Early in 1708, therefore, Galway returned to Portugal, but before the Germans under Guido, Graf von Starhemberg, had arrived in Spain, the Bourbons took Tortosa (July 1708), thus separating Catalonia from Valencia. Even after the arrival of the German forces, the Bourbons were able to take Denia (November 1708) and Alicante (April 1709). In September 1708 the English had captured the island of Minorca, but in Spain itself, although some French troops had been withdrawn because of the critical military situation in the Netherlands in the summer of 1709, Starhemberg achieved little during 1709, and Galway, who had invaded Estremadura from Portugal, was sharply defeated on the Caya River (May 17).

1710. Fourth Peace Negotiations.—By this time not only the French were weary of the war. The Dutch, having in October 1709 by the First Barrier Treaty been guaranteed by England the right to garrison nine towns in the Spanish Netherlands and ten more if these were recovered from the French, were less than ever interested in prosecuting the war. Even some of the English statesmen, including Townshend, Sidney Godolphin, earl of Godolphin (*q.v.*), and Marlborough, were prepared to consider allowing Philip V to retain part of the Spanish possessions. But the emperor was vehemently opposed to any partition of the Spanish Empire. The duke of Savoy was bitterly opposed to allowing Sicily to pass to Philip, seeing any increase in Philip's territories in the Mediterranean as a menace to himself. Philip himself, having achieved considerable successes in Spain and being sure of the enthusiastic support of many of the Spaniards, was as obstinately opposed to the idea of dividing the Spanish possessions as was the emperor. Peace negotiations (March–July 1710) were held at Geertruidenberg in the United Provinces between representatives of Louis XIV and the Dutch. Louis XIV went so far as to offer a subsidy to pay troops to drive Philip out of Spain. The Maritime Powers insisted on his using French troops to do this. Marlborough might have been able to make saner councils prevail, but by 1710 his own personal position had been undermined by political events in England (*see* below), and he was not strong enough to overrule his allies. In July 1710 the peace negotiations broke up.

1710–11. The Whole Situation Altered in Favour of France.—By the end of 1710 Philip V was firmly established in Spain. Starhemberg had at first won some successes at Almenara in July and at Saragossa in August, and Imperialist troops had actually been able to occupy Madrid for a second time in September. But, as in 1706, they had been forced to evacuate the capital, and on their retreat eastward they had been defeated at Brihuega (Dec. 8–9) and at Villaviciosa (Dec. 10), so that by the end of 1710 the Imperialists were again confined to Catalonia.

But more decisive for the outcome of the war than victories or defeats in any of the fields of battle were the political developments that were taking place in England. The Whig government's conduct of the war had provoked much criticism, partly because it was thought that not enough use had been made of the victories of 1708, partly because Malplaquet had been regarded as so costly in casualties that it had been looked upon as a defeat rather than a victory, and partly because the First Barrier Treaty with the Dutch in 1709 was criticized as sacrificing English trade with the Netherlands and English friendship with the emperor. There was always a fear that if the Dutch controlled the Spanish Netherlands they might exclude British merchants. The heavy taxation necessary to finance the war had grown increasingly unpopular. The Whigs who made up the ministry were divided among themselves. Marlborough's wife Sarah lost her influence over Queen Anne, and Marlborough found that his personal influence at court had disappeared. During the summer of 1710 the Tory Charles Talbot, duke of Shrewsbury, was appointed lord chamberlain, and the Whigs Godolphin and Charles Spencer, 3rd earl of Sunderland, were removed from office. The Parliament that met in November 1710 was strongly Tory. On April 17, 1711, the emperor Joseph I died without male issue. This event revolutionized the whole international situation, for the heir presumptive was Joseph's brother, the archduke Charles, and no English or Dutch government would continue a war to put the Spanish Empire under the control of the prince who would also be the ruler (as Charles VI) of the Holy Roman Empire and the hereditary possessions of the Habsburgs in Austria, Bohemia, and Hungary.

1711–13. Effective Peace Negotiations.—The death of Joseph I dramatically improved the prospects of the fifth peace negotiations, which had begun at the end of 1710, when communications were established between the governments of France and England. In the Netherlands Marlborough achieved his final success against the French, breaking Villars' *ne plus ultra* defensive lines, which extended from the coast to Valenciennes, by a series of skilful maneuvers (July–August 1711), but on Dec. 31, 1711, Marlborough was dismissed from his command. The peace negotiations continued throughout 1711, and on Jan. 29, 1712, peace discussions were opened formally at Utrecht, in the United Provinces, between representatives of France, England, and the United Provinces. There was at the beginning no representative of Portugal, Savoy, or of the German princes, but it is more remarkable that there was no representative either of the emperor or the king of Spain. The situation became complicated by the deaths of the dauphin Louis (April 1711), his eldest son Louis, duc de Bourgogne (February 1712), and the latter's elder surviving son, Louis, duc de Bretagne (March 1712). This left the French claimant to the Spanish throne, Philip V, the eldest surviving brother of the duc de Bourgogne, next in succession to the French throne if his nephew Louis (later Louis XV), the sickly younger son of the duc de Bourgogne, should die. Through English pressure, Philip renounced his claims to the French throne in May, although the French legal experts considered that no renunciation, however formal, could cancel a hereditary claim based on divine right.

Both Louis XIV and the English Tory government were now determined to end the war. Louis was even prepared to make peace without Philip. Henry St. John, Viscount Bolingbroke (*q.v.*), who was in charge of the English negotiations, adopted an equally ruthless attitude toward his ally, the emperor Charles VI, for early in May 1712 Bolingbroke instructed the duke of Ormonde, who had succeeded the duke of Marlborough as commander in chief, not to take part in any future battle. In November Philip's renunciation of his claim to the French throne was formally made before the Spanish *Cortes*, and later in the same month his younger brother Charles, duc de Berry, renounced any claim to the throne of Spain. In March 1713 Louis XIV's recognition of these renunciations was formally registered by the French *Parlement*. The peace terms on which England and France finally agreed were advantageous to England but were also much more favourable to the Bourbons than the terms that Louis XIV had been prepared to accept in 1710 when he had been ready to pay for troops to expel his grandson from Spain, or in 1709 when he would have been satisfied if Philip could have been allowed to retain Naples and Sicily, or even in 1706 when Louis had been prepared to be satisfied with Naples, Sicily, and Milan for his grandson.

THE TREATY OF UTRECHT

A more accurate name for the series of treaties signed between 1713 and 1714, which finally concluded the war, would be the peace settlement of Utrecht. Most of the documents were signed at Utrecht, but two were signed in the German cities of Rastatt and Baden, and one in Madrid.

On April 11, 1713, France concluded treaties of peace at Utrecht with England, Holland, Prussia, Portugal, and Savoy. By the treaty with England, France recognized the Protestant succession in England and undertook to give no further help to the Stuarts. France ceded Newfoundland, Nova Scotia, the island of St. Kitts, and the Hudson Bay territory to England and promised to demolish the fortifications at Dunkerque (Dunkirk). The Anglo-French treaty was supplemented by a treaty of commerce. In the treaty with the Dutch, France agreed that the United Provinces should annex part of Gelderland and should retain, as a barrier against any future French invasion, certain fortresses in the Spanish Netherlands; these latter territories were to be assigned to

the emperor when he made peace. By a commercial treaty the French conceded to the Dutch privileges similar to those enjoyed by England. In the treaty with Prussia, France acknowledged Frederick I's royal title (king in Prussia from 1701) and recognized his claim to some small territories, including Neuchâtel and upper Gelderland. In return France obtained the principality of Orange. In the treaty with the duke of Savoy, France accepted that he should rule Sicily and Nice. The treaty with Portugal recognized its sovereignty on both banks of the Amazon.

The peace treaties between Spain and its opponents were not signed until a few months later, but an understanding with England had been prepared by the Asiento (*q.v.*) agreement, by which Spain gave to Britain the exclusive right of supplying the Spanish colonies with slaves. The treaty of peace was finally concluded in Utrecht on July 13, 1713. In it Spain ceded Gibraltar and Minorca to England and promised to cede Sicily to Savoy. England and Spain concluded a commercial treaty in December 1713. But though by July 1713 so many peace treaties had been concluded, there were still delays, and Louis XIV was sharply critical of Philip for holding up the negotiations with the Dutch and for quibbling over the terms already conceded to Savoy. On Aug. 13, 1713, the Spanish treaty was concluded with Savoy, Victor Amadeus II overlooking some modifications made by Philip in order to secure Sicily. The peace between Spain and the Dutch was delayed till June 26, 1714, and that between Spain and Portugal until February 1715.

But even then the general pacification was not complete, for the emperor remained at war both with France and with Spain. Eugene was defeated in the Netherlands by Villars at the Battle of Denain (July 1712) and on the Rhine lost Landau (August 1713) and Freiburg (November 1713). On March 7, 1714, the emperor concluded a peace treaty with France at Rastatt. The emperor recovered Breisach, Kehl, and Freiburg and in return ceded Strasbourg and Alsace to France and agreed to allow France's allies, the electors of Bavaria and Cologne, to recover their possessions. In addition, the emperor was recognized by France as ruler of the former Spanish possessions of Milan, the bases in Tuscany, Naples, the Spanish Netherlands, and Sardinia. On Sept. 7, 1714, the princes of the empire accepted these peace terms and concluded with France the Treaty of Baden. On Nov. 15, 1715, a treaty (known as the Third Barrier Treaty) between the United Provinces and the emperor laid down that seven fortresses near the French border were to be garrisoned by the Dutch. This "barrier" was guaranteed by Great Britain.

The Treaty of Baden is usually considered the last of the treaties of the settlement of Utrecht. The emperor remained technically at war with the king of Spain until 1720, but the peace settlement of Utrecht laid the foundation for international relations in southern Europe until about 1733.

Thus 12 years of war and all the brilliance and courage of Marlborough and Eugene had less effect on the development of the international situation than domestic and personal politics in London and the fortuitous death of the emperor without male issue. In spite of Blenheim, Ramillies, Oudenaarde, and Malplaquet, Philip V retained the major part of what he had been bequeathed —Spain itself and all Spain's possessions in America. All he had to cede were the Spanish Netherlands, the Spanish possessions in Italy, Gibraltar, and the islands of Sicily, Sardinia, and Minorca.

See also references under "Spanish Succession, War of the" in the Index.

BIBLIOGRAPHY.—J. W. Fortescue, *A History of the British Army*, vol. 1 (1899); A. von Arneth, *Prinz Eugen von Savoyen*, 3 vol. (1888); *Mémoires militaires relatifs à la succession d'Espagne*, ed. by F. E. de Vault, 11 vol. (1835–62); J. S. Corbett, *England in the Mediterranean*, 2 vol. (1917); O. Troude, *Batailles navales de la France*, 4 vol. (1867–68); P. Roeder von Diersburg (ed.), *Kriegs- und Staatsschriften des Markgrafen Ludwig Wilhelm von Baden über den spanischen Erbfolgekrieg*, 2 vol. (1850). (J. O. Li.)

SPARKS, JARED (1789–1866), U.S. historian and president of Harvard, was also an editor and, to his contemporaries, the "American Plutarch." He was born at Willington, Conn., on May 10, 1789. He attended Phillips Exeter Academy and graduated from Harvard College in 1815. Sparks was installed as minister of the First Independent Church (Unitarian) of Baltimore, Md., in 1819, resigning the post in 1823. From then until 1830, under his ownership and editorship, the *North American Review* became the arbiter of literature in New England. He fostered the growing interest in the American Revolution with his multivolumed "lives and letters" of Franklin, Washington, and Gouverneur Morris, 12 volumes of that war's diplomatic correspondence, and 25 volumes entitled *The Library of American Biography*. Occupying the McLean chair at Harvard, 1839–49, he was the first U.S. professor of secular history. On Feb. 1, 1849, Sparks was chosen president of Harvard. He was harried by administrative difficulties and ill health, and his presidency, lasting four years, closed his most fruitful years of scholarship. He died at Cambridge March 14, 1866.

Sparks believed that patriotism obliged him, when editing source materials, to omit passages likely to cause international ill will and sometimes to "embellish" what "the Fathers" had actually written. Therefore, the exacting scholarly standards of a later age rendered obsolete much of what he had edited. To his admirers, however, he was the patriarch of American history and the peer of his friend George Bancroft (*q.v.*).

Sparks's own manuscripts and papers are at Harvard. His private library and maps are at Cornell University, Ithaca, N.Y.

BIBLIOGRAPHY.—H. B. Adams, *The Life and Writings of Jared Sparks*, 2 vol. (1893); J. S. Bassett, *The Middle Group of American Historians* (1917); Samuel Eliot Morison, "Jared Sparks," in *Dictionary of American Biography*, vol. xvii (1935). (W. T. Hn.)

SPARROW, any of several small, chiefly drab birds having thick bills well adapted for seed eating. They occur in several families of songbirds (suborder Passeres). The true sparrows (*Passer*) are cosmopolitan members of the weaverbird family (Ploceidae). The house, or English, sparrow (*P. domesticus*) is common throughout temperate parts of Europe, Asia, North America (introduced from England in 1851), South America, and Africa. In North America it has replaced many native birds to a considerable extent in cities and towns. This prolific bird (it raises several broods a year) has a pugnacious disposition and is adaptable to urban areas, thriving best in the neighbourhood of man. City birds become sooty, but in the country the male has a gray crown, black throat mark, white cheeks, a chestnut nape, brown upperparts, and a buffy belly; the female and young are dull brown above and dirty white below, with no distinctive head marks. The nest is an untidy tangle of grass, feathers, hair, etc., which may contain three to seven whitish eggs flocked with gray. House sparrows spend much time hopping about on the ground feeding, all the while flicking their tails and uttering repetitive "cheeps." The tree, or mountain, sparrow (*P. montanus*), both sexes of which resemble the male house sparrow except for the brownish crown and doubly barred wings, is a local species in Britain but has been introduced elsewhere.

JOHN H. GERARD

HOUSE SPARROW (PASSER DOMESTICUS)

The numerous native American sparrows belong to the subfamily Emberizinae, of the finch family (Fringillidae). Foremost among them is the song sparrow (*Melospiza melodia*), a variable bird with more than 30 subspecies; its song is sprightly and musical and is among the first heard in spring. The chipping sparrow (*Spizella passerina*), a perky little bird easily told by its chestnut cap and white stripe along the eye, nests in trees in gar-


JOHN H. GERARD
WHITE-THROATED SPARROW (ZONOTRICHIA ALBICOLLIS)

dens and is a common bird in the suburbs. The Savannah sparrow (*Passerculus sandwichensis*), abundant in the Maritime Provinces of Canada, breeds from northern Alaska and most of Canada south to Mexico; about 16 subspecies are recognized in the U.S. and Canada. The grasshopper sparrow (*Ammodramus savannarum*) is common enough, but its skulking habits and insect-like buzzing and ticking calls cause it to be overlooked. One of the handsomest of sparrows is the white-crowned sparrow (*Zonotrichia leucophrys*), with pearl gray breast and bands of black and white on the crown; resembling it closely is the white-throated sparrow (*Z. albicollis*), with an additional white throat patch.

The hedge sparrow, or dunnock (*Prunella modularis*), belonging to the accentor family (Prunellidae), is a small, brown-backed bird with an iron-gray head and neck. It is widely distributed throughout Europe and can be distinguished from true sparrows by its thin, dark bill and shuffling, rather than hopping, gait.

A few species of the weaver finch family (Estrildidae) are called sparrow, including the well-known Java sparrow (*Padda oryzivora*), a handsome, finely feathered gray bird with a black head and large pinkish bill; it is a popular cage bird. (HT. FN.)

SPARROW HAWK, the smallest—and perhaps commonest—of the American falcons, is a little larger than an American robin, rusty-red barred with black on the back and tail, lighter underneath, with a striking face pattern of black and white. It is *Falco sparverius*. Although often swallowlike in flight, it may hover on rapidly beating wings as it surveys fields below for prey. Its ringing call—*killy-killy-killy*—is often heard. Grasshoppers and mice form its chief food, though occasional small birds may be eaten. Sparrow hawks are often found in open suburban areas—even in cities, along railroad or superhighway rights of way. It is not unusual to find one perching on a telephone pole or on a wire. The nest site is a tree hole, ground hole, crevice among rocks, or some secluded nook in a building. There, without benefit of nesting materials, four or five eggs are laid—ivory to light brown, with brown to dull red speckles—which are incubated about 30 days. The hatchlings are down covered but gradually assume adult-type plumage. The young remain in the nest for about 25 days, the females eating the lion's share of the food brought by the parents and rapidly attaining the larger size characteristic of its sex in adult falcons. This species is similar to the European kestrel (*q.v.*). (W. J. BE.)

SPARTA, the capital of Laconia (*q.v.*), was for long the most powerful city of ancient Greece. The name refers in Greek to the city alone, the alternative name LACEDAEMON to both city and territory; "Spartiate" designated a member of the ruling aristocracy, but the people and state were normally "the Lacedaemonians," this term including the subject perioeci (*q.v.*).

Sparta lay at the northern end of the plain of the Eurotas (Evrotas) river on its western bank, where the routes into Laconia from north and northwest join. The central territory was this plain, between the high peaks of Taygetus (Taiyetos) to the west and Parnon to the east, reaching the sea at the port Gythium (Yithion) to the south. Sparta also controlled the mountainous district eastward, with the coast to Cape Malea; some Arcadian border districts to north and northwest; Messenia (*q.v.*) to the west: altogether about two-fifths of the Peloponnese. The Eurotas valley and the Messenian plain of Stenyclarus are, for Greece, conspicuously fertile, and their climate kindly.

Prehistory.—Sparta itself has produced no Mycenaean remains: Amyclae, a few miles south, and Therapne, above the east bank of the Eurotas, were then the important centres, the latter perhaps the capital of the kingdom assigned in the *Iliad* to Menelaus, though no palace comparable to those of Mycenae or Pylos had been found by 1960.

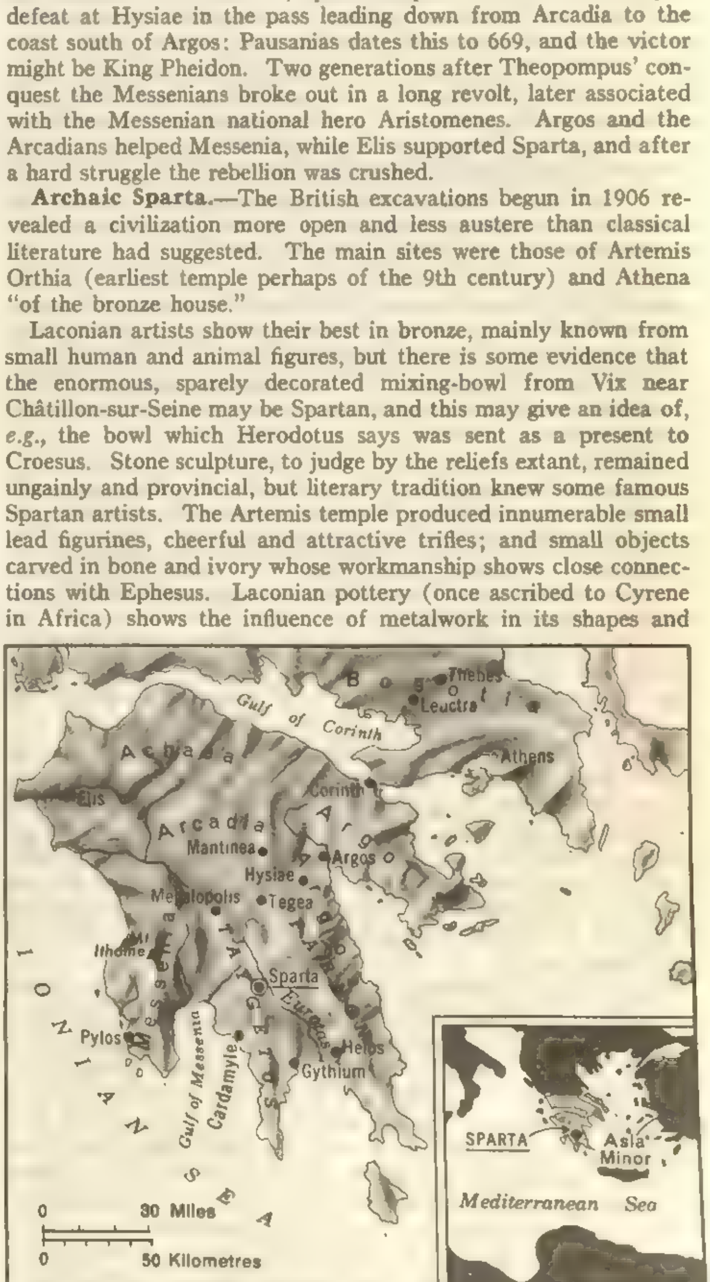
Tradition ascribed the foundation of Sparta to Dorian invaders claiming descent from Heracles (*see* DORIANS). The Dorian dialect, spoken at Sparta, belongs to that division of the Greek-speaking peoples which entered Greece last; and at the Amyclaeum the latest Mycenaean pottery is followed by a variety of Protogeometric which marks a fresh start and does not, like some varieties elsewhere, show stylistic continuity with Mycenaean. The earliest finds from Sparta show some, but not much, contact with the outside world.

Detailed accounts by later Greek writers of Sparta's gradual conquest of Laconia have an improbable ring. Spartan tradition made much of the capture of Amyclae, in the later version the work of the eighth generation of kings, but more likely an incident of the original conquest: importance was also attached to the capture of Helos in the south, in the belief that the serfs called helots (*q.v.*) took their name from it.

Conquest of Messenia.—The First Messenian War (*c.* 735–715) under King Theopompus marks the beginning of authentic Spartan history, though it represents the completion of the conquest rather than Sparta's first venture west of Taygetus. Samos helped Sparta, and the war may be connected with the contemporary Lelantine War (*q.v.*). The best land was divided among the Spartiates, and the Messenians joined the helots in servitude. After the war Tarentum in south Italy, the only Spartan colony, was founded. To the east, Spartan expansion was checked by a defeat at Hysiae in the pass leading down from Arcadia to the coast south of Argos: Pausanias dates this to 669, and the victor might be King Pheidon. Two generations after Theopompus' conquest the Messenians broke out in a long revolt, later associated with the Messenian national hero Aristomenes. Argos and the Arcadians helped Messenia, while Elis supported Sparta, and after a hard struggle the rebellion was crushed.

Archaic Sparta.—The British excavations begun in 1906 revealed a civilization more open and less austere than classical literature had suggested. The main sites were those of Artemis Orthia (earliest temple perhaps of the 9th century) and Athena "of the bronze house."

Laconian artists show their best in bronze, mainly known from small human and animal figures, but there is some evidence that the enormous, sparely decorated mixing-bowl from Vix near Châtillon-sur-Seine may be Spartan, and this may give an idea of, *e.g.*, the bowl which Herodotus says was sent as a present to Croesus. Stone sculpture, to judge by the reliefs extant, remained ungainly and provincial, but literary tradition knew some famous Spartan artists. The Artemis temple produced innumerable small lead figurines, cheerful and attractive trifles; and small objects carved in bone and ivory whose workmanship shows close connections with Ephesus. Laconian pottery (once ascribed to Cyrene in Africa) shows the influence of metalwork in its shapes and



SPARTA AND THE PELOPONNESE

its precise decoration, painted on a white ground. It developed unevenly, influenced by Corinthian art but generally conservative, and seldom achieved satisfactory composition: the well-known cup depicting Arcesilaus of Cyrene watching the weighing and storage of wool is at least fresh and lively. After the middle of the 6th century it declined like all fabrics except Attic.

Of native craftsmen, Gitiadas who made Athena's Bronze House in the middle of the 6th century was best remembered. From abroad, Theodorus of Samos built a famous portico, and Bathycles of Magnesia late in the 6th century the elaborate throne for Apollo at Amyclae. Many foreign poets and musicians came, among the earliest Terpander and Polymnastus from the eastern Greek world, Thaletas from Crete, but these are little more than names. Late in the 7th century Alcman, probably of Lydian origin, is known from the long fragment of a *partheneion* ("maiden song") and many shorter quotations: he had much to say of Spartan girls, horses and food, and was more sensitive to landscape than most Greeks. But Tyrtaeus' military exhortations, composed during the Messenian revolt, are native work, though later he was inappropriately claimed as an Athenian.

Within its fairly narrow limits, this society shows a gay and lively mind, open to the outside world. But the Messenian revolt underlined the dangerous position of a small aristocracy which excluded the free perioeci from government and held down a helot population many times its own number, and this must have hampered Spartan development. The evidence of archaeology and Alcman, however, provides no decisive evidence about the date at which Sparta was bound by its other peculiar institutions.

Political and Military Institutions.—In the 5th century and later, the Spartans believed that their whole system of life was the work of a single early lawgiver, Lycurgus (*q.v.*). Some modern historians substantially accept the Greek theory, as it was eventually fixed, that he carried out a large reorganization in the 9th century; but the readiness of the Greeks to ascribe everything Spartan to Lycurgus, and the character of the institutions themselves, make this doubtful. The institutions may not all be of the same date, and it is possible that later reformers sheltered behind Lycurgus' name, fostering the ascription of everything to him, their own reforms included.

Two kings held equal power, from the separate families of the Agiadae and Eurypontidae. The origin of this unique system is lost in the dark ages. Later the kings held certain priestly and other privileges, and a Spartan army abroad was almost always commanded by one of them: Aristotle characterizes them as "hereditary generals." The two families to the end produced many gifted and influential leaders.

Greek aristocracies regularly curtailed the king's power by setting up other executive magistrates, and there is evidence that the Spartan ephors (*see* EPHOR) once represented "the city" as against the kings.

The *gerousia* (*q.v.*) no doubt descends from a council of nobles summoned in Homeric fashion by the kings. The assembly included all Spartiates over 30 who had undergone the prescribed training and paid their dues to *phiditia* (*see* below). It could pronounce only on issues set before it by the *gerousia*, but in practice it discussed and decided all major questions, such as war and peace. It often tried kings whose conduct had caused dissatisfaction; and through the election of ephors it indirectly controlled the administration.

The relation between kings, *gerousia* and assembly in the process of reaching decisions of state is the main subject of the Great Rhetra which Plutarch quotes as an oracle given to Lycurgus; it also refers to the numbers of the *gerousia* and to tribal organization. It has been accepted as a 9th-century document, but its provisions seem more appropriate to the 7th century, and it is here taken as an enactment of that time, one of the pillars of the classical constitution, designed to secure Sparta from the discontents which in many cities produced tyranny by guaranteeing, under certain limitations, the rights of the assembly.

The army, based originally on the three kinship tribes common to all Dorian communities, was regrouped when the citizen body was reorganized on a local basis, the five villages (*obai*) of Sparta. This system probably dates from the 7th century (*obai* are mentioned in the Rhetra) and was in force at the battle of Plataea (479), where the perioeci fought in separate units but in equal numbers with the Spartiates. Before the end of the 5th century there were six regiments called *morai*, in which Spartans and perioeci served together, and at Leuctra (371) the Spartan element was little more than a cadre.

The navy was commanded by an admiral who held office for one year only.

Social Institutions.—Spartiates lived under state supervision from birth, when weakly children were exposed on Taygetus. From 14 to 20, graded in year-groups, they underwent a severe training in discipline and endurance. Thereafter they might be admitted to the messes called *phiditia* in which male Spartans ate and passed much of their lives, and on admission they became full citizens, called "equals" (*homoioi*): it is not clear whether in the classical period there was any large residue of those who failed for admission or had not land enough to pay their dues, in kind, but a class of "inferiors" is mentioned once in the 4th century. Foreigners were sometimes allowed to undergo this training, even some selected helots (*mothakes*), but these could not become citizens.

BY COURTESY OF THE TRUSTEES OF THE BRITISH MUSEUM

"GIRL RUNNING," SPARTAN BRONZE. ABOUT 500 B.C.; FOUND IN ALBANIA: HEIGHT 4½ IN.

Similar but less strict institutions existed among the Dorians of Crete, and there are many parallels outside Greece, so that the system may be a general inheritance which elsewhere lapsed so early that it was forgotten. But at Sparta it served military needs, and it may have been deliberately preserved and strengthened there. The assertion of equality must also have been a deliberate assertion at some particular date: in practice it proved increasingly difficult to maintain.

The evidence for Lycurgus' division of land into equal lots is suspect, coming mainly from the time of the 3rd-century reformers (*see* below). But land was the basis of a Spartan's citizenship, and for long there were conventional or legal restrictions on its alienation. The accumulation of other forms of wealth was hindered by Sparta's retention of a clumsy currency of iron spits when other Greeks, from the 7th century onward, coined silver; and by the prohibition of private ownership of precious metal. Spartiates might not engage in trade, and there were sumptuary laws, *e.g.*, against elaborate building.

Over-all, this was a hardy society, freed to an unusual extent from the Greek preoccupation with agriculture, using its leisure for hunting and the perfection of its rigorous military training. Plato complained that Lycurgus educated his citizens in the warlike virtues to the neglect of the others.

Peloponnesian League and Persian Wars.—When Croesus of Lydia in the 540s sought Greek help against the rise of Persia, he found that Sparta was the strongest city and made his alliance with it. By the end of the 6th century Sparta headed an organized league covering most of the Peloponnese and reaching out beyond it. The Spartans' rise was due primarily to their military power. They suffered one heavy defeat in the middle of the 6th century when they tried to reduce Arcadian Tegea to a status like that of Messenia, but Herodotus says that their other wars of this period were successful, and they had mastered Tegea before Croesus' appeal. The basis of the settlement with Tegea was alliance, not subjection, an important stage in the development of the Peloponnesian league. They also defeated Argos in the Thyreatis, a border district on the coast which the Spartans now firmly held: this involvement, they said, prevented them from helping Croesus when Cyrus of Persia overthrew him.

A policy of deposing tyrants is ascribed to the powerful Chilon,

who was ephor *c.* 556 (he has also been credited, but by a misunderstanding of the evidence, with increasing the power of the ephorate). Sicyon was liberated soon after 556; an unsuccessful attempt was made, with Corinthian help, to overthrow Polycrates of Samos in 525; Hippias was driven from Athens in 510. It is not a very full record, but Sparta gained the reputation of supporting constitutional government against tyranny, and this must have helped the growth of the league.

King Anaxandridas, who died soon after 520, played some part in all this, but his attitude is not clear and it may not have been consistent. His energetic son Cleomenes I (*q.v.*) quarreled with Athens after its liberation and made several attempts to coerce the Athenians, thwarted mainly by his colleague Demaratus and the Corinthians. These differences were perhaps the occasion for giving a formal constitution to the Peloponnesian league, for the first recorded league conference occurred about 500. If these developments damaged Sparta's prestige, it was restored by Cleomenes' victory over Argos at Sepeia without allied help (*c.* 494), and at the end of his reign he was on good terms with Athens, at whose request he took hostages from Aegina when the Aeginetans gave tokens of submission to King Darius shortly before Marathon (490). But to achieve this he had to get his opponent Demaratus deposed, and overstrained the great personal influence he had in Sparta: the opposition triumphed, and after a brief exile and restoration he committed suicide in prison. The troubles in which this reign ended probably included also a Messenian revolt in 490, but ten years of quiet enabled Sparta to recover, and in 480 the Greeks accepted without question Spartan leadership during Xerxes' invasion.

Since Cyrus' first conquest of Asia Minor frequent appeals had been made to Sparta for help against the Persians. This is the more striking, since after the failure of the expedition against Samos in 525 Sparta consistently refused to send help overseas: this was presumably from military calculation rather than indifference, and several Spartan actions of this period can be interpreted as intended to build up resistance to Persia inside Greece. Above all, Sparta had never shown any sign of compromising with the Persians.

For the operations of 480 and 479 *see* GRECO-PERSIAN WARS. Sparta's performance has been criticized, but under Spartan leadership the war was won: Spartan heroism at Thermopylae was never forgotten, and the last land battle at Plataea was very much Sparta's victory, under the regent Pausanias. But Salamis had demonstrated Athenian naval power. Athens came forward as the leader of the eastern Greeks, and in 478, though Pausanias' expedition to Cyprus and Byzantium was a military success, his conduct so much alienated the non-Peloponnesian allies that they formed the Delian league (*q.v.*) under Athens.

Struggle With Athens.—The Spartans acquiesced in the transfer of the naval hegemony to Athens, and this withdrawal must have weakened the great influence they had acquired. Pausanias (*q.v.*) probably wished to fight Athens instead, but the peace party prevailed, and his eventual disgrace and death did further harm to Sparta's prestige. The Spartans had now to fight to retain their position even inside the Peloponnese.

The Spartan army, trained for a set battle in the Greek heavy infantry style, was ill-adapted for distant and scattered operations, and the Spartans failed to improvise the machinery necessary for the running of an overseas alliance. The peace party saw further expansion as a danger to Sparta's traditions, and trusted the Athens of Cimon (*q.v.*); and the helots were an ever-present menace. Democracy was another danger: in the 6th century the Spartans could appear as the leaders of a crusade for constitutional government, but in the 5th they were left on the defensive as a bulwark of oligarchy, for they could not accept democracy without disrupting their whole system.

Democracy was adopted in Elis (471) and Mantinea: Themistocles, in exile at Argos, had perhaps some hand in this. Tegea also revolted, with some help from Argos, and Arcadia had to be subdued in two hard-fought battles. In 464 a disastrous earthquake caused heavy casualties at Sparta and precipitated a Messenian revolt. Defeated in the field, the rebels were hard to dislodge from their stronghold on Mt. Ithome, and the Spartans called for help from their allies, Athens included. But they then dismissed the Athenian contingent, distrusting its loyalty, and this insult wrecked Cimon's policy of co-operation with Sparta. The Athenians allied themselves with Sparta's enemy Argos (462/461).

The Spartans, having weathered the crisis but still dangerously weakened, took no direct part in the ensuing war except in 457, when they tried to revive the Boeotian league as a counterweight to Athens: but they had a hard fight at Tanagra to secure the army's return, and Athens soon after overran Boeotia. Disaster in Egypt (454) checked Athens' progress, and a truce was made in 451 for five years. In 446 Sparta invaded Attica in support of a Megarian revolt from Athens, but retired without a battle and concluded the Thirty Years' peace, which implicitly recognized the Athenians' right to their empire.

The breakdown of this peace in 431 and the outbreak of the Peloponnesian War (*q.v.*) have been variously explained. Thucydides thought the basic cause was Spartan distrust of Athens, and this is probably correct: incidents just before the war suggested that Athens had not really renounced expansion, while Sparta had recovered and was ready to proclaim a crusade to liberate Greece from Athenian tyranny. But Spartan invasion and devastation were inadequate weapons against Athens, supplied with corn from overseas, and the cities of the empire were not eager enough for liberation. Fighting in the northwest, which could never be decisive, went in Athens' favour. Pylos on the Messenian coast, taken in 425, served as a refuge for fugitive helots, and Spartiate prisoners captured there caused Sparta great anxiety. But in 424 Athens was defeated near Delium by the Boeotians, and the success of the enterprising Spartan general Brasidas in the north gave the Spartans something to bargain with, so in 421 they gave up their crusade and signed the compromise peace of Nicias.

Several Spartan allies refused to sign this peace, which was never fully effective. In 418 Sparta actually fought Athens and allies at Mantinea, where the Spartan victory restored their position inside their own league, and in 414 Athens formally broke the treaty by raiding Laconian territory. The Spartans, resuming open war, fortified a permanent post at Decelea in Attica. The loss of Athens' expedition to Sicily (413) gave a fresh turn: Sparta, hesitantly at first, crossed the Aegean to provoke revolt in Asia Minor, where the remainder of the war was mainly fought to sever or protect Athens' vital access through the Hellespont to the grain of south Russia. The problem of maintaining campaigns at a distance had been partly solved by the use of detachments of allied troops and freed helots under officers called harmosts. In spite of Athenian victories at Cyzicus (410) and Arginusae (406), Lysander finally gained control of the Hellespont at Aegospotami (405) and Athens was starved into surrender (404). But the fleet which won the war was maintained only with Persian subsidies, notably those of Cyrus the younger to Lysander from 407 onward, and Persia's price was the surrender of the Greek cities of Asia Minor.

Spartan Empire.—Lysander (*q.v.*) aimed at a tightly controlled empire, and had secured political control of many cities through committees of ten (decarchies) of his personal adherents; and Athens was governed by "the Thirty" with a Spartan garrison. But jealousy of Lysander's power and distrust of his methods grew at Sparta. The turning point came with Thrasybulus' democratic rebellion at Athens, and King Pausanias was allowed to mediate a settlement by which democracy was restored (403): on his return to Sparta he was put on trial, but Lysander lost this trial of strength, and his decarchies were abolished. Sparta's prestige remained enormous, but Spartan troops and harmosts could not be everywhere.

In 401 Cyrus revolted against his brother Artaxerxes II with some Spartan help. When he was killed and the satrap Tissaphernes began to reconquer the coast of Asia Minor for Persia, the Greek cities there obtained Spartan help in a series of inconclusive campaigns, including those of Agesilaus II (*q.v.*) in 396 and 395. Meanwhile Sparta's continental ambitions in central Greece provoked Thebes to form a coalition with Athens, Argos

and Corinth against Sparta. They were defeated in 394 near Corinth, and again by Agesilaus' returning army at Coronea in Boeotia, but the destruction of the Spartan navy at Cnidus, by the Athenian Conon with a largely Persian fleet, largely compensated this. Sporadic land fighting went on around Corinth, and Thrasybulus in 389 recovered much of Athens' naval power. Thereafter Sparta and Persia rediscovered their common interest: Antalcidas with Persian and Syracusan help gained control of the Hellespont, thus paralyzing Athens, and imposed the king's peace or peace of Antalcidas (spring 386) which assigned all Asia Minor to Persia and proclaimed the autonomy of all other Greek states.

Sparta, led by Agesilaus, made ruthless use of this peace to dissolve the Boeotian league and any other combination thought dangerous. The seizure of the citadel of Thebes in 382 marked the height of Spartan power; but Thebes recovered its independence in winter 379, and an unsuccessful Spartan attempt to surprise the Piraeus alienated Athens, whose Second Confederacy spread widely, defeating the Spartan navy at Naxos. But the war exhausted both Sparta and Athens; peace was signed in 375/374 but immediately broken; at the next conference, in 371, Thebes refused to recognize the independence of the other cities of Boeotia, and when King Cleombrotus tried to impose this by force the Spartans were disastrously defeated at Leuctra (July 371).

Leuctra ended Sparta's domination. Impressive as it looked, it was never fully effective without Persian help, especially against Athens, and its effects were wholly negative. Spartan institutions, however much admired by theorists, were not a model that could be copied in an increasingly democratic Greece; and Sparta's great prestige could not suppress the resentment caused by Spartan interference. Greece as a whole was appreciably weaker since the 5th century.

Decline.—Epaminondas' liberation of Messenia (winter 370/369) cut off nearly half Sparta's resources, but Spartan power was not quite broken and Thebes could not dominate the Peloponnese: the Theban victory at Mantinea (362), marred by Epaminondas' death, only increased the confusion. Sparta held aloof from peace conferences, refusing to recognize the independence of Messenia, while some of the kings capitalized their military skill as mercenary leaders abroad. The Spartans took no part in resistance to Philip of Macedonia: Agis III led an anti-Macedonian movement while Alexander was conquering Persia, but was defeated and killed by Antipater at Megalopolis (331). Toward the end of the 4th century the city, hitherto proudly unwalled, was partly fortified; Areus I, who showed some military energy but was killed before Corinth (*c.* 265) during the Chremonidean War, struck coins in his own name; land had for some time been fully alienable and was falling steadily into fewer hands, but there was less of it since the loss of Messenia, and other territory was transferred to Sparta's neighbours by Philip (338) and later by Antigonus Gonatas. The pattern of Spartan life was changing.

Agis IV by peaceful agitation in 244, and Cleomenes III by force in 227, attempted to restore the old order. For details *see* AGIS; CLEOMENES: the basic plan was to cancel debts, redistribute land on the system ascribed to Lycurgus, bring up the numbers of the citizen population by admitting perioeci and others, and reestablish the ancient discipline. To the Hellenistic world, where the propertied classes had a firm grip on city politics, this was dangerous social revolution. Agis was soon suppressed. Cleomenes, more effective in his seizure of power, so alarmed the Achaean league by his military achievements and ambitions that Aratus (*q.v.*) of Sicyon reversed the policy of years and called in the Macedonian king Antigonus Doson; Cleomenes was defeated at Sellasia in 222, and died soon after in Egypt. The reforms were cancelled, but trouble continued. The Achaean leader Philopoemen defeated and killed a "tyrant," Machanidas, but the still more radical Nabis (*q.v.*) achieved some power and was partly tolerated by Rome till his assassination. Philopoemen then captured Sparta (188) and abolished the remains of the Lycurgan constitution: but many problems remained, especially the numerous exiles left over from this period of trouble, and Spartan discontents served as some excuse for Rome's war on the Achaeans (146).

Roman Sparta.—After Achaea became a Roman province, Sparta, though technically a free city, had no further individual history. Much territory had been lost, many of the perioecic cities being now detached in an independent league, and Sparta was wholly cut off from the sea: but under the wealthy citizen Gaius Julius Eurycles it took Augustus' side in the Roman civil wars and was rewarded with an outlet at Cardamyle (Kardhamili) on the Messenian gulf. Some picturesque features of the Lycurgan training were retained for the young—the flogging ordeal at the temple of Artemis Orthia, suitably brutalized, proved a tourist attraction —but their elders led the ordinary life of Roman provincials.

Medieval Sparta.—Sparta suffered severely during the invasion of Greece by the Visigoth Alaric in A.D. 395-396, and from subsequent invaders. Nevertheless, it remained a prosperous town and bishopric, often known as Lacedaemonia under the Byzantine empire. On the establishment of the Latin empire in Constantinople in 1204, the bulk of the Peloponnese became the Frankish principality of Achaea under the Villehardouin family, who built a new fortress and city named Mistra on a spur of Taygetus near Sparta. Michael Palaeologus recovered Mistra for the Byzantine empire as part of the ransom of William Villehardouin, captured at the battle of Pelagonia in 1259; thereafter it was the capital of the despotate of the Morea (*i.e.*, the Peloponnese), with a brilliant life of about two centuries (*see* MISTRA). From 1460 it was Turkish, with a Venetian interval from 1687 to 1715.

Modern Sparta.—The existing town (modern Gr. SPARTI) was built on the ancient site in 1834, after the establishment of Greek independence, and is the capital of the *nomos* (department) of Lakonia. Pop. (1961) 10,412.

See also references under "Sparta" in the Index.

BIBLIOGRAPHY.—General histories of Greece; H. Michell, *Sparta* (1952); K. M. T. Chrimes, *Ancient Sparta* (1949; very technical and somewhat adventurous). *Archaeology:* principally *Artemis Orthia*, ed. by R. M. Dawkins (1929). *Pottery:* E. A. Lane, "Lakonian Vase-Painting," *Annual of the British School at Athens*, xxxiv, pp. 99 ff. (1933–34). *Sculpture:* E. Langlotz, *Frühgriechische Bildhauerschulen* (1927; some doubtful attributions). *Other Finds From Sparta:* M. N. Tod and A. J. B. Wace, *Catalogue of the Sparta Museum* (1906).
(A. As.)

SPARTACUS, leader in the Gladiatorial War against Rome (73–71 B.C.), was a Thracian by birth. He served in the Roman army but seems to have deserted and been sold as a slave. With about 70 fellow gladiators, he broke out of a training school at Capua and took refuge on Mt. Vesuvius (73 B.C.), where they were joined by other runaway slaves. They defeated two Roman forces in succession, overran most of southern Italy, and ultimately were several thousand strong. Spartacus defeated both consuls of 72 and fought his way successfully toward the Alps, intending his army to disperse to their homes. But the men refused to leave Italy, and he returned to Lucania, hoping to cross to Sicily. He was thwarted by the new Roman commander, M. Licinius Crassus, who had eight legions. Spartacus' army divided; the Gauls and Germans were defeated first, and he himself ultimately fell fighting in a pitched battle (71). Pompey's army met and killed many slaves who escaped northward, and 6,000 prisoners were crucified along the Appian Way by Crassus.

Spartacus was both competent and, apparently, humane, though his name inspired terror throughout Italy. It would be a mistake to call his revolt an attempt at revolution. Nevertheless, he has become the patron saint of revolutionaries, and from Adam Weishaupt in the late 18th century (*see* ILLUMINATI) to the present day the pseudonym "Spartacus" has covered many such; the most important were Karl Liebknecht, Rosa Luxemburg (*qq.v.*), and the other members of the German *Spartakusbund* of 1916–19. The theme of the slave rising has also been attractive to writers of fiction—as Arthur Koestler, *The Gladiators* (1939).

See T. Rice Holmes, *The Roman Republic and the Founder of the Empire*, vol. i, pp. 155 ff., 386 ff. (1923); H. Last in *Cambridge Ancient History*, vol. ix, pp. 329 ff. (1932).

SPARTANBURG, a city in the piedmont section of South Carolina, U.S., the seat of Spartanburg county, is in the foothills of the Blue Ridge mountains, 32 mi. N.E. of Greenville. The name of the city and county commemorate the Spartan regiment

of South Carolina militia which fought in the American Revolution. Established first as a courthouse village in 1785, the town was in an area noted prior to 1860 for iron works, cotton planting and a few cotton mills. Nearby resorts at mineral springs drew a clientele from the entire state. After 1865 growth was accelerated by the intersection of three major rail lines at Spartanburg. By 1900 a number of cotton mills had been established on nearby streams. The trend to manufacturing continued in the 20th century; less than 12% of the county labour force is employed in agriculture and about 45% in manufacturing. Textile mills dominate county industry, which also includes bathroom fixtures, electrical equipment, food products, textile machinery, wearing apparel and chemical products. With the decline of cotton farming, farmers in the Spartanburg area turned to vegetable and peach production. The city was incorporated in 1831 and has a council-manager form of government in effect since 1947.

Spartanburg has a symphony orchestra, a civic band and a public-school symphony. It is the site of Converse college, a private college for women, founded in 1889; Wofford college, Methodist, for men, founded in 1854; Spartanburg Junior college, a coeducational Methodist institution, founded in 1911; and the South Carolina School for the Deaf and Blind (1849). Located near the city is Croft State park. The nearby mountains offer many recreational opportunities.

The population of Spartanburg is about 67% native-born white and 32% Negro. For comparative population figures *see* table in SOUTH CAROLINA: *Population*. (L. P. J.)

SPARTINA, a genus of grasses, most of whose seven or eight species are North American and as a group are commonly called cord, salt, or marsh grass. Salt meadow grass (*S. patens*) and smooth cord grass (*S. alterniflora*) are common on the salt marshes of New England. In Europe *S. stricta* occurs on the salt flats of southeast England, France, Portugal, and the Adriatic. *S. alterniflora* has been known at Southampton since 1829; it was probably introduced by shipping from America.

The genus is remarkable for the sudden appearance of a new distinct form at Southampton in 1870, *Spartina townsendi*. This tall-growing and robust form spread widely on the soft tidal muds of both sides of the English Channel. *Spartina townsendi* is exceptional in colonizing soft mud previously bare, for the vigour of its spread, and for its power of holding silt and raising the level of the mud. It is probably a hybrid between *S. stricta* and *S. alterniflora*. *Spartina spartinae*, gulf cord grass, native to the southern coastal plain of the U.S., produces wiry culms used, in combination with broomcorn, in the manufacture of household brooms and brushes. (F. W. O.)

SPASM: *see* CRAMP.

SPEAKER. Originally the name given to the presiding officer of the English house of commons, "speaker" has been adopted by many legislatures which have based their procedure on that of the British parliament. The office of speaker exists throughout the Commonwealth, but there are usually modifications of the British pattern: for example, in Australia, Canada and New Zealand the office is regarded as the privilege of the party in power and the occupant of the chair changes with the government. The office of speaker in the United States house of representatives is described below.

GREAT BRITAIN

Functions and Powers.—The speaker is the mouthpiece or representative of the house of commons in its relations with the crown, the house of lords and bodies or persons outside parliament. This function is largely confined to formal or ceremonial occasions. The speaker's main tasks are to preside over debates (except when the house is in committee), to enforce the observance of all rules for preserving order and to advise the house in its proceedings (for the presiding officer of the house of lords, *see* LORD HIGH CHANCELLOR). Conscious endeavour over the last 150 years has succeeded in making the office the objective embodiment of the rules and laws of the house, freeing it from any suspicion of partisanship. The speaker rules on points of order as they arise in debate and such rulings constitute precedents by which subsequent speakers are guided. He decides whether a proposal for an emergency debate should have precedence over the business already set down, and whether a prima facie case has been made out in complaints of breaches of privilege. He has the power to select amendments and to accept or reject a motion for the closure of debate. He can check irrelevance or repetition. Disorderly conduct can usually be restrained by a simple call to order; the speaker can direct the withdrawal of offensive expressions, or, in serious cases, "name" a member, which leads to immediate suspension from the house. If general disorder arises the speaker may either suspend the sitting for a period or adjourn the house at once without putting the question. The calling of members to speak in debate is entirely in the hands of the speaker, and his main concern is to ensure that a variety of points of view is heard. In addition to his duties in the chair the speaker controls much of the machinery of the house. His department includes the library of the house, and he is responsible, through an editor whom he appoints, for the production of the official report of debates (*Hansard*). He is also the chairman of the boundary commissions which review parliamentary constituencies, and from time to time he is called on to preside over conferences on electoral reform.

Rules, custom, ceremonial and deliberate policy combine to give the speaker, once elected, great authority. He is elected for a whole parliament (normally of about four or five years duration) and by custom dating back to 1835 is re-elected for as long as he wishes, no matter what changes of government may occur. The usual tenure of office is about 10 or 15 years. The house of commons' mace is carried before him on his daily procession to the house at the beginning of its proceedings. He is robed in a gown and full-bottomed wig; members bow to the chair on entering and leaving the chamber; all speeches are addressed to him; and when he rises to his feet all members must resume their seats and hear him in silence. The speaker's actions or rulings may not be criticized incidentally in the course of debate but only on a substantive motion, and such motions are rare. The speaker receives an annual salary of £5,000, which is charged on the consolidated fund and not voted annually, and in addition the annual allowance for parliamentary expenses paid to all members of parliament; he has an imposing official residence within the palace of Westminster. Formerly the first commoner in the land, the speaker since 1919 takes precedence immediately after the prime minister and lord president of the council. On retirement the speaker is offered a viscountcy and is voted a pension of £4,000 a year for life.

As important as the speaker's authority is his impartiality. On election a speaker severs all his ties with political parties and clubs. He takes no part in debate and votes only in the event of a tie, and then only in accordance with rules which preclude an expression of opinion upon the merits of the question. A. W. Peel (speaker, 1884–95) summed up the position when he said: "I know how necessary it is for any man who aspires to fill that great office to lay aside all that is personal, all that is of party, all that savours of political predilection, and to subordinate everything to the great interests of the house at large." J. W. Lowther (speaker, 1905–21) quoted with approval the remark that the office of speaker "does not demand rare qualities. It demands common qualities in a rare degree." The qualities chiefly required are good temper, patience, tact, judgment, and dignity in voice and bearing. A spice of wit, which can dissolve a strained situation in a burst of laughter, is invaluable.

Election of the Speaker.—At the beginning of a new parliament the sovereign commands the house to elect a speaker. When a new speaker has to be chosen the government of the day makes every effort, by informal discussion with the opposition and all shades of back-bench opinion, to discover a candidate who will command general support. Contested elections for the speakership are rare—the last two contests being held in 1895 and 1951. On the latter occasion W. S. Morrison (speaker, 1951–59), a former minister from the Conservative side of the house, was opposed by members of the Labour party, who put forward Major James Milner, for many years chairman of ways and means and deputy

speaker, as their candidate. The Labour party had not, by the early 1960s, provided the house with a speaker, and the feeling that it should be enabled to do so was widespread within the party in 1951, and again in 1959 when Morrison retired. Although the election of Sir Harry Hylton-Foster in 1959 was not contested by the Labour party, strong dissatisfaction was expressed at the inadequacy of the consultations that preceded his being put forward for the office. Upon his death in 1965, for the first time, a Labour member, Horace King, was elected speaker.

Once elected, the speaker is not normally opposed in his constituency by official party candidates at general elections. Since 1895 such contests have only occurred in 1935 and 1945, though unofficial candidates opposed the speaker in 1950 and 1955. If there is a contest the speaker refrains from any party political campaigning. It has been suggested, in order to spare the speaker any possibility of a contested election and to prevent his constituents being, in effect, disfranchised, that his constituency should be authorized to elect two members, that a fictitious constituency should be created, or that he should be enabled to sit in parliament without a constituency. However, a select committee of the house that reported on the matter in 1939 recommended that no change be made.

History of the Office.—Although the office can be traced back to 1258, when Simon de Montfort presided over the "Mad" parliament at Oxford, the first speaker to be so designated and the first known to have been chosen by the house of commons was Sir Thomas Hungerford, elected to the chair in 1377. Since that date, every parliament has elected its speaker. Although the choice of speaker was strongly influenced by the crown in the medieval period, several holders of the office showed considerable independence. A number of the early speakers met violent deaths, and the position of the speaker, between a strong king and a house concerned to increase its own power, could be an unenviable one. The custom whereby a newly elected speaker shows reluctance to accept office dates from this period. By 1642, after the constitutional struggle between Charles I and parliament, it was established that the first duty of the speaker was to the house, and when the king came to arrest five members, Speaker Lenthall replied to his questions with the memorable words: "May it please your majesty, I have neither eyes to see nor tongue to speak in this place, but as the house is pleased to direct me, whose servant I am here."

Since 1679, when Charles II refused to approve the re-election of Sir Edward Seymour, but failed to secure the election of his own nominee, royal approval of the commons' choice has been a formality, and by the end of the 17th century the speaker's independence of the crown was an established principle. In the 18th century dependence of the speaker on the crown threatened to be replaced by dependence on party. However, Arthur Onslow (speaker, 1728–61) raised the conception of the office to its present level by his resolute independence of all parties and his complete impartiality. His immediate successors failed to maintain his standards and once again the speakership was regarded as in the gift of the party in power, and few speakers took any pains to stand aside from politics. William Grenville, Pitt's nominee, was elected to the chair in 1789 when less than 30 years old, remained speaker for only five months, and later became home secretary and prime minister. Henry Addington, (afterward Viscount Sidmouth, speaker, 1789–1801), stepped straight from the chair to the premiership. However, during the 19th century the firm tradition of an independent and nonpartisan speaker was built up, a process especially associated with Charles Shaw-Lefevre's tenure of the office (1839–57). (H. J. Pa.)

UNITED STATES

The United States constitution (art. 1, sec. 2) provides that "the House of Representatives shall chuse their Speaker and other Officers . . ." In creating this office its authors had in mind the speaker in the colonial assemblies who had been both a presiding officer and a political leader. Early expectations were later realized, for U.S. speakers after 1789 were both parliamentary and political personalities.

The speaker of the house holds an office of dignity and honour, of great power and influence. Usage dictates that he shall always be a member of the house, and as such the constitution requires him to be seven years a citizen and at least 25 years of age. Candidates for the office are nominated by party caucuses prior to the convening of each congress and the speaker is customarily elected by the house on the opening day, barring a protracted contest. Competition for the place has often been keen and the final choice is influenced by knowledge of parliamentary procedure, previous legislative experience, sectional claims, private interests, personal qualities and political considerations.

History of Office.—From the beginning the speakership has generally been held by men with previous experience in state legislatures; many of the speakers of the house of representatives have been men with experience as speakers of their state legislative bodies. Fewer than one-fifth of the speakers who served during the first 170 years of the constitution's history were without any prior state legislative experience, and of these none held the office prior to at least his fourth term in the national legislature. Only three speakers were elected to the office in their first term in the house of representatives: Frederick A. C. Muhlenberg, the first speaker, who had been speaker of the Pennsylvania house and a delegate to the constitutional convention; Henry Clay, who had served several terms in the Kentucky legislature as member and speaker and had been a U.S. senator; and William Pennington, who had once served in the New Jersey assembly. With the exception of these three, every speaker had had prior service in congress, ranging from 1 to 16 terms.

Many famous men have served as speaker of the house—James K. Polk, Schuyler Colfax, James G. Blaine, Thomas B. Reed, Joseph G. Cannon, Champ Clark, Frederick H. Gillett, Nicholas Longworth and Sam Rayburn. Rayburn set a record for the longest service—nine terms, including two half-terms. After Rayburn the next longest term of service as speaker was that of Clay—six terms, five of them consecutive. Andrew Stevenson, Cannon and Clark served four terms each; Nathaniel Macon, Colfax, Blaine, Samuel J. Randall, John G. Carlisle, Reed, Gillett, Longworth and William B. Bankhead each had three terms in the chair.

More than one-third of the speakers have come from the south, about twice as many as from any other part of the country. Southern influence predominated in the organization of the Democratic party, for the "solid south" was overwhelmingly Democratic after the Civil War and repeatedly re-elected the same congressmen. Southern members thus acquired seniority and power in party councils and most Democratic speakers from 1865 to 1961 were southerners. In 1962 John W. McCormack, Democrat of Massachusetts, was elected speaker to succeed Rayburn. (For a list of speakers *see* UNITED STATES [OF AMERICA]: *Table VI.*)

Powers and Duties.—The parliamentary powers and duties of the speaker stem from four sources: the constitution and laws of the United States, the rules of the house, previous decisions of the chair and general parliamentary law. He presides at sessions of the house and opens and closes its sittings, acts as its spokesman and representative, and authenticates documents, announces the order of business, puts questions and announces the vote.

He also decides questions of order, prevents obstruction of house business, refers bills and reports to the proper committees or calendars, chooses speakers pro tempore and appoints the chairmen of committees of the whole. These duties he performs impartially like the speaker of the house of commons, as evidenced by the fact that his decisions are rarely appealed; however, he also enjoys the privileges of an ordinary member of the house and may vote and take part in debate on the floor.

Prior to 1910 the speaker also possessed important political powers. He appointed, deposed and demoted the chairmen and members of committees; he served as chairman of the committee on rules, which enabled him to control the business of the house; and he recognized or refused to recognize members claiming the floor. These powers, gradually evolved by strong speakers like Reed and Cannon, reached a peak of perfection and omnipotence during the latter's regime. So dictatorial and despotic, indeed,

did Speaker Cannon become that the house rebelled, and in March 1910 a coalition of Democrats and insurgent Republicans succeeded in drastically curtailing his powers. They removed him from the rules committee, deprived him of the power to appoint standing committees and restricted his right of recognition.

Although the speaker was removed from the committee on rules in 1910, he continued to sit on the unofficial steering committee that meets in his office and determines party strategy. Though he no longer has the power to appoint standing committees, he retains the power of appointing house conferees, the chairman of the committee of the whole and select committees. His power of recognition was reduced, but he still has a discretionary power of recognition over motions to suspend the rules and on days other than consent calendar days.

Place in Political System.—The Presidential Succession act of 1947 ranked the speaker of the house first in line of succession in case of the death or disability of both the president and vice-president. Relations between speakers and presidents have fluctuated between rivalry and harmony through the years. During periods of congressional supremacy the chair has overshadowed the White House, while popular-leader presidents have dominated congress. Despite the reduction in his powers, the speaker continues to be at once the presiding officer and titular leader of his party in the lower chamber, a private member and the most influential man in the house of representatives. (G. B. Gy.)

Bibliography.—*Great Britain.* J. A. Manning, *Lives of the Speakers* (1850); E. and A. G. Porritt, *The Unreformed House of Commons,* 2 vol. (1903); A. I. Dasent, *The Speakers of the House of Commons* (1911); M. MacDonagh, *The Speaker of the House* (1914); Sir Thomas Erskine May, *Treatise on the Law, Privileges, Proceedings and Usage of Parliament,* 16th ed. (1957); Sir William Ivor Jennings, *Parliament,* 2nd ed. (1957). (H. J. Pa.)

United States. M. P. Follett, *The Speaker of the House of Representatives* (1896); C. W. Chiu, *The Speaker of the House of Representatives since 1896* (1928); F. M. Riddick, *The United States Congress: Organization and Procedure* (1949); G. B. Galloway, *History of the House of Representatives* (1961). (G. B. Gy.)

SPEARMAN, CHARLES EDWARD (1863–1945), British psychologist whose chief contributions dealt with the theory of intelligence and cognition, was born in London on Sept. 10, 1863. As an officer in the British Army, he served in the Burmese war of 1885 and in the South African War, and did not obtain his doctorate in psychology at Leipzig until he was 40. His earliest researches were concerned with the improvement of psychophysical methods, especially as applied to visual perception. On returning to England, he joined a group at Oxford who at William McDougall's suggestion were applying Karl Pearson's correlational techniques to develop tests of intelligence for Francis Galton's projected surveys in British schools. His first results were reported in "General Intelligence Objectively Determined and Measured" (1904). Galton had maintained that intellectual processes depended both on a "general ability" and on a number of "special abilities"; Spearman sought, by simplifying Pearson's methods, to show that "general ability" alone sufficed, but later modified this extreme position; his animated controversies with Pearson and other critics led to a valuable series of factorial researches from both sides. He died in London on Sept. 17, 1945.

Spearman's books include *The Nature of Intelligence and the Principles of Cognition,* 2nd ed. (1927); *The Abilities of Man: Their Nature and Measurement* (1927); and, jointly with L. Wynn Jones, *Human Ability* (1950). (Cy. B.)

SPECIES. The nature of species is a question of considerable importance in general biology. The term is regularly employed as a taxonomic unit, or taxon, in systematic botany, zoology, and mineralogy, but despite its wide use its definition remains elusive. The word "species" is the Latin equivalent of the Greek *eidos, idea* ("class") and originally meant primarily shape or visible form but was extended to cover the sensible character of things generally, and also the intelligible.

PHILOSOPHY

As intelligible character, *eidos* or species was the common essence of individuals or instances of the same kind. As such, it was an object of thought, not of sense; universal, not individual; and, unlike its instances, exempt from generation and decay. Its relation to its instances (which may be compared with that of a law to the events exemplifying it) was a subject of dispute both in Greek and scholastic philosophy. There were also disputes about the relation of species to their common genus, of which they are determinations, and to the properties connected and accidents conjoined with their specific nature in individuals of a species. In particular it was disputed whether species and genera were anything real independently of our minds, or only common names, or else notions of ours, the products of man's classifying activity. A settlement of these questions is not necessary to the work of the systematic biologist, but they must be faced if it is really to be understood what is meant by the evolution of species; for this evolution is a process of which no persisting material thing is the subject. (H. W. B. J.)

BIOLOGY

History.—The concept of species as a kind of animal or plant is an old one in biological literature. Aristotle (384–322 B.C.) used the term for groups of organisms with common characters, and these in turn he subdivided into more species. John Ray (1627–1705) is usually given the credit for using the term for groups with mutual fertility. Carl von Linné (*see* Linnaeus, Carolus; 1707–78) established a system of binomial nomenclature universally used in the 20th-century. Each organism was classified as a member of a species, and each species was included in a higher taxonomic group, the genus. The genera were also classified in still higher taxonomic categories, but the binomial name consisted of the names of the genus and the species together.

International congresses of botany and zoology, respectively, formulate, administer, and interpret the international rules of botanical and zoological nomenclature. The basic principle used for natural classification is homology (*q.v.*). In early modern times homology was a vague concept applied roughly to apparent similarities of structure. During the 20th century, genetics, embryology, ecology, and evolutionary analysis came to place a firm scientific foundation under the concept of homology, with consequent clarification, correction, and verification of systematic biology.

Linnaeus first believed in the fixity of species but later in his life suggested that the genus constituted the ancestral species at the time of creation and that derived species subsequently multiplied by means of hybrid generation or intercrossing between species. During the 19th century, particularly through the work of Lamarck (1744–1829) and Charles Darwin (1809–82), the theory of the evolution of species was accepted by the large majority of biologists (*see* Evolution, Organic).

The importance of isolation in the origin of species was emphasized by M. Wagner (1813–87) and J. T. Gulick (1832–1923). Modern genetic theory, initiated by Mendel (1822–84) by means of his investigations of domestic varieties of plants, was applied to both plants and animals in the early 20th century. Mendelian genetics of individuals led gradually to the establishment of the science of population genetics during the third and fourth decades of the 20th century. Population genetics substantiated the concept of the biological species and greatly improved the understanding of evolutionary processes.

Attempts at Definition.—The species is recognized as a prime biological unit with objective criteria. More than 1,000,000 species of animals and more than 300,000 species of plants had been named by mid-20th century; probably at least twice this number will ultimately be recognized when the knowledge of the world faunas and floras approaches completion.

There are almost as many definitions of "species" as there are biologists who have attempted to condense the major aspects of the concept. A brief definition that includes the most generally recognized criteria of the biological species was given by A. E. Emerson in 1945: a species is an evolved or evolving, genetically distinctive, reproductively isolated natural population. All of these attributes are necessary, and no others would seem to be essential. Each attribute is discussed separately in the following paragraphs.

Since the establishment of the principle of evolutionary change, species have been considered to be dynamic entities, hardly fixed or static, as was once believed. Evolutionary characterization is therefore necessary in the definition. Demarcation of species is sometimes somewhat arbitrary when transitional stages are encountered; however, true evolutionary gradations are not often found, so that ancient continuities are seldom observable between living species, and the paleontological record seldom bridges the gaps between species categories. Although nomenclature is easier when the ancient continuities are unknown, cases in which intergradations between species can be directly observed offer an opportunity for a more exacting study of the processes of speciation, and analytical biologists seek such material for investigation. Also, the early stages in species divergence that have not progressed to the point of complete species division enable the scientists to analyze the dynamic factors more exactly and are consequently sought.

Genetic distinction seems to be the common denominator for all valid qualitative or quantitative species characters, including cytological, physiological, embryological, ethological, ecological, and morphological characters. Taxonomists, of course, seldom can experimentally investigate the genetics of the species they study and must rely upon structural or colour indications of genetic relationship. Anatomy and pigmentation often are sensitive to small genetic influences that are not easily detected by other techniques of observation. Heredity may be inferred through demonstrated germinal continuity in populations and through the experimentally established basis of similar characters in investigated organisms. Hybrids and intergrading populations give some genetic information. If striking structural characters are not correlated with genetic distinctions of reproductively isolated natural populations, the taxonomist usually does not use them for species separation and will correct errors that have been committed in ignorance of the genetic implications. For example, structural changes produced by various conditions at the time of growth, by differences in stages of development, by sexual dimorphism, by castes of social insects, and by various types of polymorphism do not meet the criteria of genetic distinction and reproductive isolation. On the other hand, slight genetic distinction of natural populations that are reproductively isolated may be used for species separation, while gross genetic differences of individuals or of only partially isolated populations are not valid for species distinction.

Reproductive isolation, or lack of gene flow between populations, may result from chronological separation, spatial separation, topographical or geographical barriers to dispersal, habitat or ecological restrictions, cyclic differences in breeding periods, mechanical inability to crossbreed, psychological or ethological barriers to interbreeding, physiological barriers to fertilization, genetic unbalance preventing fusion, inviability of hybrids, hybrid sterility, and selective elimination of hybrids. Regardless of the exact type of isolation or the combination of isolating factors, anything that prevents genes from moving from one population to another has profound evolutionary consequences. Reproductive isolation is the dividing factor in the branching of the phylogenetic tree, and the two populations thus deprived of gene exchange will tend to drift apart through changes in gene incidence and gene or chromosome mutation. Because of its effect upon inbreeding, reproductive isolation also influences gene and character fixation. As most taxonomic characters are the result of complex genetic patterns rather than single genes or chromosomes, the consolidation of these patterns and their transmission as systems becomes a necessity for further adaptive evolution. Fixation of complex genetic patterns is a primary consequence of reproductive isolation. Reproductive isolation may take place gradually and give rise to many subdivisions of the species population before speciation is completed. Theoretically the point of species origin is the moment that genetically distinctive natural populations cease gene exchange. Practically it may not be possible to recognize this point in instances of contemporary speciation. However, the large majority of valid species have been isolated for thousands or millions of years, so that it is a rare case that presents difficulties when an abundance of evidence is accumulated.

An important group of biologists, including Theodosius Dobzhansky (1951) and E. Mayr (1942), excluded extrinsic barriers to gene flow (geographical or ecological) in their definitions of reproductive isolation as applied to species and confined the concept to examples separated by intrinsic barriers (psychological, genetic, and physiological). They did not give full species rank to genetically divergent natural populations completely lacking gene exchange if such populations have not lost their potential ability to interbreed when the extrinsic barriers are removed. Inasmuch as it is often impossible to judge the potentialities of interbreeding under natural conditions, these biologists had to rely upon a rough correlation between interbreeding potentialities and degrees of taxonomic divergence. When two related species occupy the same territory (sympatric species) and no intergrades occur between them, biologists may assume an intrinsic barrier to gene flow. Such an assumption is much less warranted in the cases of closely related but geographically separated distinctive populations (allopatric species), however; and there is a resulting confusion of taxonomic status of numerous groups. One school of taxonomists may refer certain populations to full species rank, while the other school will treat the same populations as subspecies.

If reproductive isolation includes the extrinsic barriers to gene flow, it is possible that species that have not evolved intrinsic barriers may ultimately interbreed when there is an occasional breakdown of temporary extrinsic barriers. The hybrid origin of a species from two preexisting species is thereby a possibility. The school of biologists that confines reproductive isolation between species to intrinsic prevention of gene flow does not recognize full species status for the extrinsically isolated populations and would not allow for the possibility of hybrid origin of species except in special cases in which intrinsic barriers are also broken down (allopolyploidy in plants). In spite of these differences of opinion of the exact definition of reproductive isolation, with consequent confusion over the taxonomic status of some natural populations, the large majority of plant and animal species prove to be intrinsically isolated. Therefore, there is common agreement among biologists concerning most species limitations, together with the criteria for the recognition of the species category.

The final characterization of the species is that it constitutes a natural population. The population is integrated by means of genetic continuity through reproduction; it may also be integrated through sexual reproduction and various other adaptations for attraction and interaction between individuals of the same species. The species population is often subdivided into many subpopulations, but the larger species population is a real entity with a definite boundary. Differences in fundamental biological interactions occur *within* the species population, in contrast with the biological interactions *between* species populations. The species is generally conceived to result from natural evolution. Artificial populations compounded by man through domestication or for scientific investigation are not given species rank even when they are genetically distinctive and reproductively isolated populations. Under natural conditions such man-produced populations would seldom maintain their genetic distinctiveness, if they were able to survive at all! The domestic dog is classified as a full species (*Canis familiaris*), but the separate breeds are not. Corn (*Zea mays*) is a full species; selectively bred types of sweet corn are not.

Related Categories.—Numerous infraspecies categories relate to variations in the qualitative or quantitative attributes of the species. Each term places emphasis upon particular aspects of the species concept. Some of the terms are defined below and are useful in the study of speciation mechanisms.

Sexual Species.—This term is used for species that possess sexual adaptations and interfertility among the component individuals. Sexual recombination of genetic elements greatly increases genetic variability. Many definitions of species confine the concept to the sexually interbreeding population.

Asexual Species.—These are species that have never evolved sexual fusion or have secondarily lost the ability to intercross. All reproduction is confined to division, parthenogenesis, apomixis, or vegetative reproduction. This species category would include

the primary asexual bacteria and many single-celled organisms and would also include the secondarily asexual parthenogenetic animals and apomictic plants. In asexual organisms reproductive isolation between individuals of the same generation is absolute, and genetic continuity is only through descent. In some instances taxonomic distinction is maintained for periods comparable to the maintenance of taxonomic pattern in sexual species (species of flagellate protozoans inhabiting the hind gut of termites). In other instances taxonomic distinction may vary over short periods of time, and the species category is difficult to apply (virulent and nonvirulent strains of viruses and nutritive strains of bacteria).

Subspecies.—A subdivision of a sexual species with all the attributes of the species except that reproductive isolation is partial rather than complete. Subspecies are given trinomial names by taxonomists and are recognized as constituting a valid category in the international rules of nomenclature. A species subdivided into subspecies or races is called a polytypic species.

Race.—Race is not separated by definition from the subspecies. However, it is either used as a synonym for the subspecies category or is given a rank slightly below that of the subspecies, presumably because the degree of reproductive isolation is smaller than that separating the subspecies.

Deme.—A communal interbreeding population within a species either wholly or partially isolated from other demes. The term has little currency.

Supraspecies.—A collection of groups or subspecies that replace each other geographically or physiologically and in which the extremities only have evolved intrinsic isolation. Connecting groups between the extremities interbreed and intergrade. There are several known cases among snakes, birds, and mammals in which intersterility occurs among some of the subspecies within the species.

Coenospecies.—The sum total of possible combinations of a genotype compound as seen particularly through cultivation under artificial conditions. The term has also been used to include groups of species separated only by extrinsic barriers to gene flow.

Ecospecies.—The species as realized in nature under the limitations imposed by the natural environment. It emphasizes the limitations of potential variation produced by natural selection.

Biotype.—A population of individuals with identical genotypical constitution under similar conditions. Biotypes are usually limited to asexual species or to populations within a sexual species that result from asexual reproduction from a single individual. Parallel mutation may possibly account for some biotypes.

Clone.—A group descended from a single individual without sexual combination. The individuals are presumably genotypically identical or isogenic. The term overlaps somewhat with the biotype.

Lineage.—A racial complex of several lines of descent forming a meshed network of evolution within the species.

Variety.—The category for individuals or small groups of individuals that varied in a describable way from the typical individuals in the species. Varieties may be considered somewhat distinctive groups that fall within the range of species variation as a whole but do not necessarily have geographical or ecological correlations.

Cline.—A character gradient within a group—usually within a species. Numerous subdivisions of clines are named when there is a correlation with special factors. Ecoclines are correlated with ecological gradients. Geoclines are correlated with geographical gradients. Taxoclines are correlated with taxonomic gradations associated with hybridization. Chronoclines are correlated with time in paleontological sequences. Ontoclines are gradations in the appearance of characters during the life cycle of individuals.

Physiological Species or Races.—Species or subdivisions of a species that are recognized by physiological characters rather than structural characters. Genetic distinction is usually manifest in structural or anatomical characters, but gene patterns may effect physiological characters without producing sufficient anatomical effects to be recognized. The term "biological race" is essentially synonymous with "physiological race."

Incipient Species.—A diverging, partially isolated population that may become a species when isolation becomes complete.

See also TAXONOMY.

BIBLIOGRAPHY.—G. S. Carter, *Animal Evolution* (1951); Theodosius Dobzhansky, *Genetics and the Origin of Species*, 3rd ed. (1951); G. du Rietz, "The Fundamental Units of Biological Taxonomy," *Svensk Bot. Tidsk.*, 24 (1930); A. E. Emerson, "Taxonomic Categories and Population Genetics," *Ent. News*, 56 (1945); J. S. Huxley, *Evolution, the Modern Synthesis* (1942); E. Mayr, *Systematics and the Origin of Species* (1942), *Animal Species and Evolution* (1963), (ed.) "The Species Problem," *Amer. Assoc. Adv. Sci.*, no. 50 (1957); G. C. Robson and O. W. Richards, *The Variation of Animals in Nature* (1936); G. L. Stebbins, *Variation and Evolution in Plants* (1950). (A. E. E.)

SPECIFIC GRAVITY is a number expressing the relative density (mass per unit volume) of a substance; it is defined as the ratio of the density of the substance to that of a standard. The standard is usually taken to be water at 4° C (39.2° F); accordingly, if the specific gravity of a body at 4° C is less than 1, the body will float in water at that temperature; if it is greater than 1, the body will sink. Specific gravity and density refer to the same physical property (in the metre-kilogram-second system of units they have the same numerical value); it is a matter of convenience which one is used. *See* DENSITY; MECHANICS, FLUID: *Buoyancy;* HYDROMETER; JOLLY BALANCE; and also references under "Specific Gravity" in the Index.

SPECK, FRANK GOULDSMITH (1881–1950), U.S. anthropologist, an authority on Indian tribes of eastern North America, was born on Nov. 8, 1881, in Brooklyn, N.Y. He took his M.A. degree in 1905 at Columbia University under Franz Boas, and his Ph.D. at the University of Pennsylvania in 1908. In the same year he initiated a program of anthropological studies at the University of Pennsylvania, where he became professor in 1925 and continued teaching until his death. He did intensive field research among the Indians of the eastern United States and Canada, particularly the Algonkian and Iroquois. Within the communities he visited, Speck was welcomed as a friend and confidant. Vitally interested in linguistics, he mastered several Indian languages and, in a few instances, recorded some known only to a few remaining speakers. He published such important monographs as *Ethnology of the Yuchi Indians* (1909), *Naskapi, the Savage Hunters of the Labrador Peninsula* (1935) and *Penobscot Man: the Life History of a Forest Tribe in Maine* (1940). He discovered and described the territorial hunting system of the northeastern Algonkian tribes. Speck dealt authoritatively with a wide range of topics—technology and decorative art, social organization, folklore and music. One of his perennial interests was religion and ceremonialism. He also contributed to ethnobotany and ethnozoology. He died on Feb. 6, 1950, in Philadelphia, Pa.

See A. I. Hallowell, "Frank Gouldsmith Speck, 1881–1950," in *Amer. Anthrop.*, 53:67–87 (Jan.–March 1951), with a bibliography by J. Witthoft. (A. I. H.)

SPECTACLES: *see* EYEGLASSES.

SPECTROCHEMICAL ANALYSIS, in its broadest sense, comprises all methods of chemical analysis that depend upon the measurement of the wavelength and the intensity of spectrum lines or bands, and encompasses all frequencies in the electromagnetic spectrum. In its restricted and commoner usage two methods are implied: (1) ultraviolet and visible emission spectroscopy and (2) ultraviolet and visible absorption spectrophotometry. Both methods serve to characterize the qualitative and quantitative composition of matter.

Emission Spectrochemical Analysis.—In emission spectroscopy atoms are excited to higher electronic energy levels above their ground states by means of electrical discharges (arcs or sparks) or flames. Atoms of unknown identity in a specimen so excited emit light of characteristic frequencies when they return to lower energy states. These characteristic frequencies are separated into an ordered sequence (spectrum) by diffraction or refraction in a spectrograph for visual, photographic or photoelectric registration. The process therefore consists of three interdependent steps: (1) the vaporization and electronic excitation of atoms, (2) separation of the emitted light into its component frequencies and (3) measurement of the wavelength and intensity of selected frequencies.

Each chemical element possesses its unique and characteristic

collection of spectral frequencies, whose recognition in the spectrum of an unknown substance therefore provides unambiguous evidence of the presence of that element. Ordinarily emission spectrochemical analysis is applied to the qualitative and quantitative determination of metallic elements, but it is not restricted to them. The method is among the most sensitive of all analytical methods, a few milligrams of a solid sample usually sufficing for the detection of metallic elements present to the extent of only parts per million. Although a given element may provide many thousands of spectrum lines, depending upon its energy level system and its concentration in a sample, it is not necessary to identify each wavelength. Normally the identification of three prominent lines in the spectrum of a sample serves to ensure the presence of an element with high reliability, while the absence of one of its most sensitive lines certifies the absence of the element to the level of spectroscopic detection.

The physical state of the sample suitable for spectrographic analysis may vary widely. Metallic samples are usually fabricated into rods or disks for electrodes in an arc or spark discharge. Powders and nonconducting solids are conveniently placed in a crater formed in a rod-shaped carbon or graphite electrode of high purity for volatilization and excitation. Solutions, biological fluids and tissues are generally converted to dried inorganic residues before vaporization and excitation from graphite or metal electrodes.

Quantitative analysis by emission spectroscopy depends upon the fact that the quantity of light emitted in a given spectrum line is proportional to the number of atoms vaporized and excited. To establish the relationship between the concentration of an element and the quantity of light emitted in one of its spectrum lines, it is necessary to provide a series of standard samples in which the matrix composition is fixed, and the sought element varies in concentration in a known and graded manner from standard to standard. Semiquantitative analyses may then be carried out by visual comparison of the photographic density of a spectrum line in the unknown sample with the densities of the same line in the spectra produced from the standards. More refined quantitative analyses, precise to about 1% of the element concentration, may be made by measurement of the spectrum line densities with a microphotometer. Since it is exceedingly difficult to maintain constant excitation conditions during the excitation of a sample and to reproduce them from sample to sample, absolute measurements of spectrum line intensities are not made. Instead a relative method, called the internal standard method, is employed. Devised by W. Gerlach in 1925, the internal standard method relates the intensity of a spectrum line of the sought element to the intensity of a suitable spectrum line of an element whose concentration remains constant in all standards and unknown samples. Fluctuations in the excitation conditions then affect the intensity of the unknown and standard equally, and their intensity ratio is independent of the unavoidable variations in the excitation and photographic processes. Frequently, as in the analysis of nearly pure metals or alloys of definite composition, the internal standard line is chosen from among the spectrum lines furnished by the base element. When this is not feasible, it is often possible to add an internal standard element in equal concentration to standards and samples alike.

Although the emission spectrographic method has relied greatly upon photographic emulsions for the recording and measurement of spectrum lines, and continues to be useful, increasing use is being made of photoelectric methods for the direct measurement of spectrum line intensities. Particularly in routine metallurgical control analyses, in which speed and a large volume of determinations are of importance, direct-recording spectrometers are available. These instruments receive spectrum lines in electron photomultiplier tubes and record the analytical composition directly on digital counters or print the results on tape.

Absorption Spectrophotometry.—Quantitative analysis by absorption spectroscopy depends upon the fact that many substances characteristically absorb light of a given frequency or range of frequencies. Commonly in analysis the absorbing substance is an ion or molecule in unknown concentration in aqueous solution. Those frequencies of light are absorbed which electronically excite the molecule or ion to a higher energy level. The diminution in intensity of light passing through the solution is related to the length l of the light path and the concentration c of the absorbing species by the Beer-Lambert law:

$$I = I_0 \exp(-\epsilon c l)$$

In the Beer-Lambert law, I is the intensity of light transmitted by the solution; I_0 is the intensity of light transmitted by the solvent alone; and ϵ, the extinction coefficient, is a constant characteristic of the absorbing species. Many ions which alone do not absorb light strongly can be made to absorb strongly by causing them to form complex ions with other ions or molecules. The sensitivity and specificity of detection may often be greatly increased by appropriate selection of complexing agents.

See also SPECTROPHOTOMETRY; SPECTROSCOPY.

BIBLIOGRAPHY.—G. R. Harrison *et al.*, *Practical Spectroscopy* (1948); N. H. Nachtrieb, *Principles and Practice of Spectrochemical Analysis* (1950); J. C. D. Brand and G. Eglinton, *Applications of Spectroscopy to Organic Chemistry* (1965). (N. H. N.)

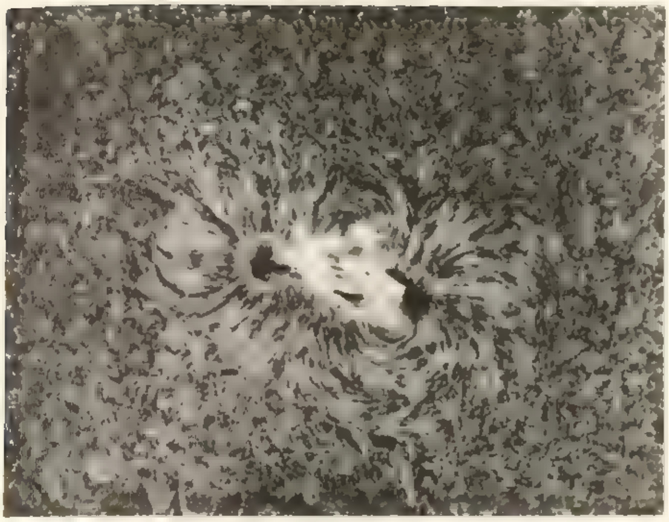
SPECTROHELIOGRAPH, an instrument that utilizes the spectrum for photographing the sun with monochromatic light. In its simplest form it consists of a spectrograph in the focal plane of which is a second slit that transmits monochromatic light to a photographic film. This slit may be set at any wave length in the spectrum, usually on a strong spectral line produced by some particular chemical element; *e.g.*, the K line of calcium or the Hα line of hydrogen. An image of the sun is formed by a telescope objective on the collimator (first) slit of the spectrograph, and the photographic film is mounted almost in contact with the second slit. In operation, the spectroheliograph is moved parallel to itself, admitting to the first slit light from all parts of the sun as the slit moves across the sun's image. Simultaneously the second slit is moving across the photographic film, building up an image of the sun with the monochromatic light transmitted by the slit. Thus the distribution of calcium or hydrogen clouds in the sun's atmosphere may be recorded. Such clouds are invisible to the eye and are not shown on a solar photograph taken in the usual way. Similar solar photographs can be obtained, however, without a spectrograph by using filters that transmit nearly monochromatic light.

The spectroheliograph, originally designed for photographing only the solar prominences at the sun's edge, disclosed in its first application by George Ellery Hale at the Kenwood observatory (Chicago, Ill., 1891) new, unexplored features of the sun's atmosphere. Photographs of the solar disk showed extensive luminous clouds (flocculi or plages) vastly greater in area than the sunspots around and above which such clouds are always found. These clouds may also appear where no spots are visible. Although in the lower, denser vapour the clouds appear bright, in the rarer vapour at high levels the light from below is absorbed, giving rise to dark clouds. The relatively dark clouds are the prominences that appear bright at the sun's edge.

At about the time that Hale was devising his spectroheliograph, H. Deslandres at the Paris observatory introduced a velocity spectrograph which permitted motion in the line of sight to be measured on photographs of the K (or H) line in successive sections of the sun. Such a series of photographs reveals the same features as shown on spectroheliograms.

The first photographs of the sun in hydrogen light were made with the spectroheliograph at the Yerkes observatory (Williams Bay, Wis.) in 1903 using the blue and violet spectral lines of hydrogen. Later, in 1908, at the Mt. Wilson observatory (near Pasadena, Calif.), when methods for making photographic emulsions sensitive to red light had been developed, the red hydrogen line (Hα) was used. The long dark flocculi (filaments) shown on high-level calcium spectroheliograms and identified as prominences projected against the surface of the sun are very conspicuous on spectroheliograms made in the centre of Hα.

In the high region of the sun's atmosphere over sunspots the distribution of the small dark flocculi shown by the hydrogen photographs suggests the operation of definite forces. Surround-



BY COURTESY OF THE MOUNT WILSON OBSERVATORY

HYDROGEN CLOUDS NEAR SUNSPOTS

Hydrogen clouds surrounding a large, magnetically bipolar group of sunspots, photographed at the Mt. Wilson observatory with the 13-ft. spectroheliograph using the red hydrogen line (Hα). The forms of such clouds resemble those of iron filings over a bipolar magnet or flow lines around double vortices rotating in opposite directions

ing large single isolated spots, these flocculi often form great vortices that resemble terrestrial cyclones or tornadoes. About 75% of these appear as huge spirals, clockwise in the southern hemisphere and counterclockwise in the northern. The detection of these vortices led, in 1908, to the discovery of magnetic fields in sunspots.

Spectrohelioscope.—The functions of the spectroheliograph and velocity spectrograph are combined in the spectrohelioscope, an instrument that renders visible to the eye many of the phenomena of the solar atmosphere photographed with the spectroheliograph and also permits their velocity in the line of sight to be measured. The principle of the spectrohelioscope was suggested by C. A. Young in 1870 and was used by him for observing prominences at the sun's limb. Hale, who revived and improved the instrument (1924), succeeded in observing with it the bright and dark hydrogen flocculi and in measuring their velocities.

Imagine a narrow slit, rapidly oscillating, between the eye and a telescopic image of the sun. Through persistence of vision, the image can be seen in white light. A spectroscope becomes a spectrohelioscope when placed in the optical path so that it transmits to the eye through a second slit only the light of the red hydrogen line. Instead of moving the slits, the sun's image may be moved. A method commonly employed is that of placing a rotating prism with a square cross section in the telescope beam in front of the first slit and viewing the second slit through another prism rotating at precisely the same rate as the first.

A "line shifter" consisting of a plane parallel-sided glass plate mounted behind the second slit displaces the spectral line toward the red or violet when rotated. The amount of rotation indicates the wave length of the light being utilized and the velocity of the observed features. In the spectroheliograph the displaced line would be off the slit, and the image of rapidly moving gas would not be recorded.

The spectrohelioscope affords a rapid means of detecting and studying many solar phenomena, since the entire disk and limb can be examined in less than a minute. These instruments have been widely used in the detection of flares that suddenly increase in brightness and then fade more gradually. Such flares are solar clouds and usually occur near sunspots. Flares affect the ionosphere directly, causing interference in radio transmission, and less directly in the production of aurora and geomagnetic storms.

See SPECTROSCOPY, ASTRONOMICAL: *Instruments;* SUN: *Spectroheliographic Investigation;* ECLIPSE: *Information Yielded by Solar Eclipses.* (S. B. N.)

SPECTROPHOTOMETRY is a subdivision of physics and chemistry which deals with the relative measurement of radiant energy (power) as a function of wavelength, frequency, or wave number. Thus, the radiant energy transmitted (reflected) by a body is compared with that transmitted (reflected) by some other system taken as a standard. Absorption spectroscopy and spectrometry are other terms often used synonymously.

Different types of modern instruments cover a wide range of wavelengths of the electromagnetic spectrum. Approximate limits of several spectral regions are: X ray, less than 1 to 100 angstroms (Å); ultraviolet, 10 to 400 millimicrons (mμ); visible, 400 to 700 mμ; infrared, 0.7 to more than 500 microns (μ); and microwave, 0.1 to 1,000 cm. The relationship among these units is given by

$$1\text{Å} = 0.1\text{m}\mu = 10^{-4}\mu = 10^{-8}\text{ cm}.$$

Two laws express the relationship of the absorption of radiant energy to the absorbing medium. According to Bouguer's law, each layer of equal thickness of the absorbing medium absorbs an equal fraction of the radiant energy traversing it. According to Beer's law, the absorptive capacity is directly proportional to the concentration of the solute in a solution. One expression of the combined laws is

$$\log_{10} P_0/P = \log_{10} 1/T = A = abc$$

in which P_0 and P are the radiant power (flux) transmitted by the standard and by the unknown, respectively; T is the transmittance which is equal to P/P_0; A is the absorbance; and a the absorptivity, b the thickness, and c the concentration of solute. Thickness and concentration are often expressed in centimetres and grams per litre, respectively. For a molar solution, the symbols ϵ and a_M are widely used for molar absorptivity. Deviations from Beer's law are common for solutions, but Bouguer's law is said to apply to all homogeneous systems.

Although spectrophotometers may vary widely in their spectrometric, geometric, and photometric designs, they consist essentially of (1) a source of radiant energy, (2) a monochromator, (3) a photometer, and (4) a detector for the unabsorbed radiant energy. An absorption cell holds the liquid or gas samples. With the majority of instruments a photometric value is obtained for each wavelength setting. For extensive scanning work, recording instruments greatly facilitate the accumulation of data. Most rapid of all is a model incorporating an oscillograph with attached camera for recording.

Occasionally monochromatic light sources are used, but generally continuous sources (for given regions) are employed. Examples of the latter are the hydrogen arc (ultraviolet), the incandescent tungsten filament (visible), and the Nernst glower (infrared). The monochromator isolates a narrow spectral band from a continuous source, and the photometer serves to measure P/P_0, or some related value, for this band. The detector varies from region to region, there being no universally usable means. Among those commonly used are photographic plates, photocells, thermopiles, and bolometers.

A curve coordinating photometric values with spectral positions is known as an absorption spectrum. Usually ordinates are T, A, log A, or log ϵ, and the common abscissas are wavelength, log wavelength, wave number, or frequency. Often only a value at a single wavelength is needed.

Ideally, the absorptivity a is a constant, characteristic of a given system. As such, it is comparable to a density or refractive index value. Unfortunately, because of lack of adequate control of instrumental and solution conditions, this ideal is not generally realized. Instrumental variables include spectral band width, cell thickness, stray radiant energy, and general operation of the instrument. Various solution conditions may affect absorption data, among them hydrogen-ion activity, concentration, temperature, interfering ions, irradiation, redox potential, dissolution aids, component interaction, hydrolysis, light fading, fluorescence, and plating.

Absorption data have many applications. One of the most important is the study of solution effects. The data are also widely used to follow the kinetics of chemical reactions and to determine dissociation constants and formulas of complexes. Analysts use

absorption data to identify constituents or to determine their amounts. If analysis of a solution is the purpose of the spectrophotometric measurement, all relevant variables must be under control to prevent misinterpretation of the data.

As each pure substance which absorbs selectively has a characteristic absorption spectrum, the curves are important for qualitative analysis. The contour or form of the bands and their wavelength position are the important items. Ions, molecules, and various groups such as the carbonyl group in organic compounds are involved. The infrared region is particularly useful for study of the last type, for revealing structures and purity. For mixtures, of course, a measured absorbance gives the combined absorbing power of the components. Application of such data in large quantities is made by means of machine-sorted punched cards.

In the visible region, absorption spectra are basic for applying the CIE international system (Commission Internationale de l'Éclairage) of colour specification. A recording instrument, with attached automatic tristimulus integrator, greatly facilitates calculations.

Quantitative analytical uses of absorption data depend upon the fact that the magnitude of the absorbance is a function of the concentration of the absorber. For any solution conforming to Beer's law, the expression $c = A/ab$ is used, a and b being known and A being measured for the unknown solution. If the solution does not conform to Beer's law, a simple calibration curve coordinating A and concentration may be established.

The quantitative application to systems with many components assumes that the values of A and abc for the individual components are additive for the absorption at the wavelengths employed. The concentration of n components may be obtained from absorbances at n wavelengths, where no two curves coincide or intersect, by solving n simultaneous equations. Spectrophotometers passing narrow band widths are needed for such determinations.

See also SPECTROCHEMICAL ANALYSIS; SPECTROSCOPY; and references under "Spectrophotometry" in the Index.

(M. G. ME.)

See I. Sunshine and S. R. Gerber, *Spectrophotometric Analysis of Drugs* . . . (1963); R. Rikmenspoel, "Sensitive Absorption Spectrophotometer . . . ," *Review of Scientific Instruments*, vol. 36 (1965).

SPECTROSCOPY pertains to the investigation of spectra, the phenomena observed when the radiations from a particular source are separated into their constituent colours or wavelengths. Such separation results from refraction (as in a prism) or from diffraction (as in a grating). Instruments designed for this purpose are called spectroscopes if used for direct visual observation, and spectrographs if photography or other methods of recording the spectra are employed. The rainbow, formed by refraction of sunlight in rain droplets, is a beautiful and common spectrum in nature. Through spectroscopy, chemical elements present on distant stars have been identified, and an increased understanding of atomic structure and such phenomena as luminescence (*q.v.*) has been achieved. The discussion to follow deals with electromagnetic wavelengths in the range between ultraviolet and infrared, as does SPECTROSCOPY, ASTRONOMICAL. For a consideration of shorter wavelengths *see* SPECTROSCOPY, X-RAY.

The major divisions of this article are as follows:

I. History
II. Light Sources
III. Spectroscopes and Spectrographs
IV. Methods of Observation
V. Measurements
VI. Applications

I. HISTORY

The composite nature of white light was first demonstrated by Isaac Newton (1664) when he allowed sunlight entering a round hole in a shutter to pass through a glass prism and fall on a screen. This elongated and coloured image of the sun he called a spectrum. In 1800 W. Herschel studied the spectral distribution of heat from the sun with the aid of thermometers and found the maximum temperature beyond the red end, thus discovering the infrared spectrum. In 1801 J. W. Ritter, studying the effect of spectral light upon silver salts, found this action extending beyond the violet, thus discovering the ultraviolet spectrum.

The first connection between spectral colour and wavelength appeared in 1802 when Thomas Young substituted his wave theory of light for Newton's corpuscular theory, explained the colours of thin films, and calculated the approximate wavelengths of the seven colours recognized by Newton. In 1814 Joseph von Fraunhofer modified Newton's solar-spectrum experiment by substituting a narrow slit for a hole and a telescope for a screen. Under these conditions he observed the continuous spectrum of the sun irregularly interrupted by many hundreds of dark lines, still known as Fraunhofer lines. These lines were the first fiducial marks in spectra and were promptly exploited as wavelength standards for the comparison of spectral dispersion of various optical glasses. Fraunhofer constructed the first diffraction gratings by winding a fine silver wire upon two fine parallel screws or by ruling a glass plate with a diamond point, and with these he made surprisingly accurate determinations of wavelengths corresponding to the Fraunhofer lines. Although Fraunhofer and others observed that certain bright lines in the spectra of flames seemed to coincide with dark lines in the solar spectrum, it remained for G. R. Kirchhoff in 1859 to enunciate the general law connecting absorption and emission of light and to emphasize the fact that each species of atom has a uniquely characteristic spectrum.

Kirchhoff and R. W. Bunsen (1861), through systematically comparing the solar spectrum with the flame or spark spectra of the purest elements available, made the first chemical analysis of the sun's atmosphere and thus laid the foundation for spectrochemical analysis and for astrophysics (*q.v.*). In the course of these investigations, they discovered spectroscopically two new chemical elements, cesium and rubidium. These demonstrations greatly stimulated spectroscopic research and increased the need for accurate standards of wavelength. The first useful standards were provided in 1868 by A. J. Ångström, who measured the wavelengths of about 1,000 Fraunhofer lines and expressed them in units of 10^{-10} metre—a unit now known as the angstrom, abbreviated Å, and used for all spectroscopic measurements. Further refinements in standard wavelengths were made in 1887 by H. A. Rowland, who invented and applied the concave diffraction grating.

In 1881 A. A. Michelson invented the interferometer. With this instrument he measured the wavelengths of cadmium lines relative to the standard metre and attained an accuracy far exceeding that of any previous work. The wavelength of the red radiation from cadmium vapour was remeasured (1905) in terms of the standard metre by J. R. Benoît, C. Fabry, and A. Pérot, using the interferometer invented by Fabry and Pérot, and since 1907 their value—6438.4696 Å—has been internationally accepted as the primary standard of wavelength. Several thousand atomic radiations ranging in wavelength from between 2100 and 10,200 Å have been measured relative to this primary standard by interference methods, and many of these have been adopted internationally as secondary standards of wavelengths. The wavelengths corresponding to lines characteristic of any spectrum whatsoever can be obtained by interpolation between secondary standards impressed on the same spectrogram. In this manner spectroscopists have accumulated wavelength data for several million lines observed in atomic and molecular spectra, extending from the extreme ultraviolet to the far infrared and embracing nearly 30 octaves, as compared with the single visible octave that was first recognized in 1802.

These data continued to serve in the 1960s for making chemical identifications; since 1885 they have contributed mightily toward the analysis of spectral structures, interpretation of spectra in terms of quantum theory, and in the development of fruitful hypotheses concerning atomic and nuclear structure. These developments and applications of spectroscopy have unobtrusively but profoundly changed civilization (*see* LIGHT).

Any experiment in spectroscopy involves: (1) a source of light; (2) a spectroscope or spectrograph for forming the spectrum;

(3) detectors for observing or recording details of the spectrum; (4) measurements of wavelengths and intensities; and (5) the interpretation of such measurements either as chemical identifications or as clues to the structure of atoms and molecules.

II. LIGHT SOURCES

The sun and lightning flashes are examples of bright sources of light occurring in nature. Artificial light sources studied spectroscopically may be divided into two groups according as they emit continuous or discontinuous spectra. A continuous spectrum is characterized by an uninterrupted gamut of wavelengths over a considerable range. Discontinuous spectra exhibit bright lines or bands of different colours or wavelengths on a dark background. Continuous spectra are emitted by incandescent solids and liquids (gas mantle, lamp filament, hot molten metal) or by certain electrical discharges (underwater sparks). These spectra are usually dependent only on temperature and are of no use in identifying the emitting substance; their principal use in experimental spectroscopy is to provide continuous backgrounds for the production of absorption spectra of gases, vapours, solutions, or solids. Discontinuous spectra are emitted by atoms, ions, or molecules in a gaseous or vaporous state in which the individual particles are excited to radiate uniquely characteristic spectra. The particles are excited by absorbing energy either from collisions with other atoms or electrical particles or from incident radiation. Light sources producing discontinuous spectra are flames, furnaces, and electrical discharges in arcs and sparks at atmospheric pressure, or in lamps containing gases or metal vapours at reduced pressure. In the order named these sources form a rough sequence in which increasing temperatures or energies of excitation are responsible for producing different spectra.

Flames.—Since the days of Kirchhoff and Bunsen it has been customary in elementary chemistry courses to demonstrate spectrochemical identification by dipping a platinum wire into salt solutions, inserting it in a Bunsen burner, and examining the spectrum of the coloured flame with a Bunsen spectroscope or Amici prism. Because the temperature of this flame hardly exceeds 2,000° C, the kinetic energy of atomic collisions is not sufficient to produce more than the strongest lines of easily excited atoms, like alkalis and alkaline earths. These flames also emit band spectra of carbon compounds and of water vapour present in the gas and air, and other bands due to oxides formed in the flame. Higher temperatures, (about 2,500° C) are attained in an oxyacetylene flame, and atomic spectra are more fully developed. H. Lundegårdh found that when solutions of materials to be investigated were sprayed into such flames 34 elements could be determined spectrographically.

Furnaces.—An evacuated furnace, in the form of a carbon tube heated by forcing large electric currents through it, provides a source of metallic spectra in which the excitation is pure thermal energy. The temperature may be controlled between 1,500° C, where spectra begin to appear, and 3,600° C, where the carbon tube tends to fail. Small samples are placed in a porcelain boat in the tube, and the luminous vapour is imaged on the slit of a spectrograph. If a carbon plug is inserted in one end of the tube, it provides a continuous spectrum for the observation of absorption spectra of metal vapours. Such furnace spectra have been extensively investigated by A. S. King at the Mt. Wilson laboratory, in Pasadena, Calif., and the results have greatly advanced the understanding of atomic radiation processes and aided the structural analyses of atomic and molecular spectra.

Electric Arcs.—Next to the Bunsen burner the simplest source of atomic spectra is the direct-current arc consisting of an insulated holder for electrodes, an applied potential of 100 or more volts, and a resistance to limit the current to 5 or 10 amp. The actual potential drop in such metallic arcs is 20 to 40 v., most of which occurs immediately at the cathode. Atomic spectra are very fully developed in electric arcs because the temperatures range from 3,500° to 8,000° C, where all known substances melt and vaporize. Besides undergoing violent thermal agitation, many atoms collide (especially near the electrodes) with electrons or ions accelerated by the applied potential. For the study of metallic-arc spectra it is best to use solid rods of the pure metals themselves; other forms of material, *e.g.*, powders, salts, precipitates, ashes, filings, foils, refractories, minerals, are readily volatilized by placing some of the substance on the lower (positive) pole of a pure carbon arc. The direct-current arc is characterized by high intensity and effectiveness in generating the first spectra of atoms, but it is subject to erratic wandering and flickering that can impair its usefulness for accurate spectrochemical analysis unless standardized conditions are imposed. A high-voltage (2,200 v.) alternating-current arc, introduced in 1936 for quantitative analysis, reduces some of the irregularities of the direct-current arc at the expense of simplicity and sensitivity of spectral detection.

Electric Sparks.—Compared with the electric arc the high-voltage condensed spark supplies higher excitation and is easier to control. Initially the high voltage was produced by a Rhumkorff induction coil, and a Leyden jar served as a condenser. In its modern forms the spark is animated by discharges from suitable condensers connected (in parallel with the spark gap) to the secondary terminals of a transformer giving from 10,000 to 50,000 v. The condenser is charged on every half cycle to the voltage at which the gap breaks down. An oscillating current then flows in the spark circuit with an initial value of $I = V\sqrt{C/L}$, where V is the condenser voltage, C the capacity in farads, and L the circuit inductance in henrys. This initial current may be hundreds of amperes, and the result is an energetic emission of excited or ionized atoms from the electrodes. Such sparks have higher effective temperatures than arcs. Indeed, in the highest attainable vacuum these sparks produce multiple ionization until atoms are stripped of all outer electrons and the corresponding temperature is estimated at several million degrees. The ordinary spark is somewhat variable, owing to lack of constancy in breakdown potential or in the rate of decay of the oscillating discharges in each half cycle. Several devices have been introduced to control or stabilize sparks used for quantitative spectral analysis. Such devices are the synchronous auxiliary rotating spark gap and the air-blasted auxiliary gap. Any conducting solids may be used directly as spark or arc electrodes.

Other substances will serve as electrodes if they are packed in a hollow carbon rod or if they are pressed into solid form with some binder, such as pure copper powder, but in such cases the spectrum of the substance is contaminated by that of the auxiliary electrode material.

Geissler Tubes.—An important type of light source is that produced by electrical discharges in gases at reduced pressures, first demonstrated and studied by J. Plücker in 1858 (fig. 1). It usually consists of a small-bore or capillary tube of glass (or silica) connecting two bulbs or larger bore tubes containing metal electrodes that can be attached to an electrical potential difference by means of wires leading through the glass. These light sources are prepared by thoroughly exhausting the lamp, then admitting some pure gas (or vapour) through a side tube and sealing it off with a blowtorch when suitable conditions are established. The optimum pressure of gas is usually that equivalent to a few millimetres to a centimetre of mercury. When electrical potential is applied to such tubes, a sheath of ions near the cathode marks the edge of the cathode fall within which electrons released from the cathode by ion bombardment attain sufficient speed to produce new ions by collisions with gas atoms. Outside this negative glow is the Faraday dark space, and beyond it is the positive column extending nearly to the positive electrode. This positive column is responsible for most of the light emission, and being a

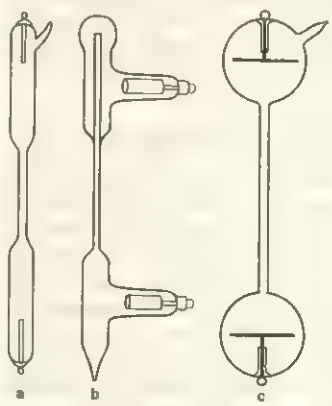


FIG. 1.—GEISSLER TUBES
(a) original, (b) modern, (c) National Bureau of Standards model

region of relatively low potential gradient and electron speeds it favours the first spectrum of the enclosed gas. A glassblower named Heinrich Geissler under Plücker's direction prepared the first practical discharge tubes and Plücker himself referred to them as Geissler tubes, a name that still clings to them. About 1920 a modified form of the Geissler tube became a spectacular type of light source for luminous coloured signs, and shortly thereafter helium, neon, and mercury lamps became commonplace. A more recent modification of the Geissler tube is the fluorescent lamp, consisting of a low-pressure mercury vapour tube whose discontinuous spectrum is partially converted to continuous bands characteristic of the fluorescent materials lining the tubes.

Hollow Cathode Discharge.—Another type of gas discharge of great utility for spectroscopic research is that devised by F. Paschen. It consists of a hollow cylindrical cathode within a cylindrical anode, and a low pressure of pure noble gas. As the pressure is reduced, the discharge retreats within the cathode until, at a pressure such that the mean free path of the electrons equals the cathode diameter, the negative glow fills the hollow cathode. The material of the cathode, or of any other metal within it, is vaporized by bombardment of the rare gas ions and excited by collisions with electrons whose maximum speeds are fixed by the ionizing potential of the sustaining gas. Since this source is operable at low pressures and low temperatures (the cathode may be cooled with liquid air), it yields atomic spectrum lines of extraordinary sharpness.

Electrodeless Discharges.—Gases or metal vapours may also be made luminous at low pressures in closed vessels without internal electrodes by placing the vessel inside a coil carrying a high-frequency current generated by either a Tesla spark generator or a vacuum-tube oscillator of ultrahigh radio frequency. The oscillating high-frequency electric fields accelerate electrons sufficiently in rarefied gases and metal vapours to excite atomic and ionic spectra. These electrodeless discharges are especially convenient for minute quantities of pure materials (such as artificial elements) or for gases and vapours that attack electrodes or that may be contaminated by them.

Linelike Sources.—When viewed through a spectroscope, any of the above-mentioned light sources will be seen as a multitude of images, one for each monochromatic radiation characteristic of the excited atoms or ions in the source. Since all practical light sources are more or less wide, these spectral images will overlap even if the spectroscope has enormous resolving power. In the case of electrical discharges in gases and metal vapours at reduced pressure, the light source is usually constructed with a long constriction to increase the intensity, and this incidentally reduces the confusion of overlapping spectral images. However, in general, it is not possible to make all light sources linelike except by imagining the light source on a narrow slit with straight parallel jaws, in which case the illuminated slit becomes the light source and the spectrum is a series of bright line images of the slit. The same effect may be obtained by substituting for the slit a cylinder of polished metal about one millimetre in diameter and using the virtual image of any light source in this cylindrical mirror as a linear light source. It is emphasized that spectral lines do not occur in nature but are consequences of using an artificial linear light source, because it gives the least confusion of overlapping images.

III. SPECTROSCOPES AND SPECTROGRAPHS

Prism Instruments.—The production of a spectrum by the use of a prism of glass or any transparent material depends upon the fact that light rays of different colours or wave-lengths are refracted differently on passing obliquely from one medium to another of different density. A composite ray of light is accordingly dispersed as well as refracted in passing through a prism, the amount of refraction being usually greater for shorter wave-lengths. The first prism spectroscopes employed by Newton and by Fraunhofer exhibited astigmatism, a type of image defect in which a point source is seen as two perpendicular image lines at different distances from the optical system. This defect was effectively eliminated in Kirchhoff and Bunsen's first spectroscope by using a collimating lens to illuminate the prism with parallel rays. Astigmatism is least and definition best in a prism spectroscope if the prism is traversed by parallel light, the slit is parallel to the prism edge, the light rays are parallel to a section perpendicular to the edge, and the rays pass through the prism symmetrically at minimum deviation. The essential parts of a simple prism spectroscope are sketched in fig. 2. Light from a source under examination enters the spectroscope through a narrow slit S located at the principal focus of the collimator lens C. Parallel rays of complex light thus fall on the prism, and rays of each component colour continue parallel in their passage through the prism and on emerging from it. The colours are refracted differently and, after passing through the telescope lens T, are focused to form a spectrum, which may be observed with an eyepiece or recorded on a photographic film.

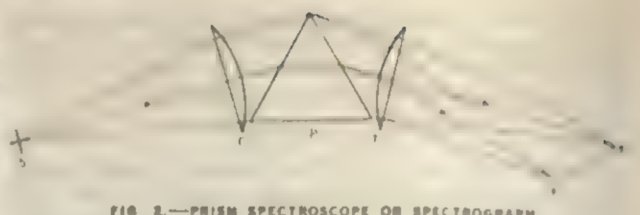


FIG. 2.—PRISM SPECTROSCOPE OR SPECTROGRAPH
S, slit; C, collimator lens; P, 60° prism; T, telescope lens; RV, red to violet spectrum

The brightness of the spectrum depends upon the way in which the slit is illuminated by the light source, upon the focal lengths and apertures of the spectrograph lenses, and upon reflection and transmission losses within the instrument. The maximum illumination is attained if the light source is large enough or so close to the slit that it subtends the same angle as the collimator lens. If the source is too small to fill the collimator with light, the brightness of the spectrum may be increased by using a condensing lens of equal aperture ratio to form an image of the source on the slit. This type of slit illumination of a stigmatic spectrograph gives a point-to-point correspondence between a section of the source, the slit, and the spectral images; it is preferred when the distribution of excitation in a source or of colour in an interference pattern is to be studied. When a perfectly uniform slit illumination is desired, as for the measurement of spectral-line intensity ratios by means of stepped filters, a condensing lens is placed just in front of the slit to form an image of the source on the collimator lens, in which case each point of the source sends light to every point of the slit and to every point of the spectral image. In either case a movable metal plate pierced with a series of openings (known as a Hartmann diaphragm) may be used to illuminate successive portions of a slit and thus record on a stationary photographic plate a number of spectra side by side for comparison and for measurement.

The light efficiency of a spectrograph, like that of a camera or telescope, depends chiefly on the ratio of the focal length F to the diameter D of the lenses. This ratio F/D may be 100/1 or more in spectrographs employed for the study of bright sources like the sun, but extremely faint sources like phosphorescence or the night sky demand unit ratio or less. A ratio of about 15/1 represents a common compromise between desired speed and the necessity of using compound corrected lenses. In such spectrographs single plano-convex or meniscus collimator and camera lenses may be used if the curvatures are chosen so as to minimize spherical aberration and coma, and the plate is tilted and bent to fit the chromatic focus. The inclined plate also usually provides an increased spectral dispersion, and the use of single lenses ensures that light losses resulting from absorption and surface reflections are a minimum.

When all relevant factors are considered, it appears that an isosceles prism with 60° angles is the best form for a spectrograph. As to prism and lens materials, there is a wide choice of optical glasses for the study of visible and near infrared spectra, but for spectral regions absorbed by glass it is necessary to use other materials, such as crystalline quartz or fluorite, in the ultraviolet, and rock salt or other alkali halides in the infrared. Un-

[illegible] natural or artificial crystals are transparent to [illegible] below 1800 A, and [illegible] A, [illegible] opaque [illegible] A and [illegible] A. Furthermore these [illegible] [illegible] homogeneity, and cost. [illegible] material is crystalline quartz [illegible] transmits and produces double [illegible] birefringence and [illegible] property of [illegible] The birefringent effect is [illegible] prisms so that the light travels [illegible] crystal, and the rotation [illegible] parts of nature, left-handed [illegible] quartz spectrograph therefore [illegible] variety of quartz [illegible] prism is made by cementing two [illegible] of the opposite varieties of quartz. The same effect [illegible] material in a Littrow type spectrograph [illegible] quartz lens and a [illegible] quartz prism [illegible] mirror. (*See* fig. 3.) Obviously the rotation doubling

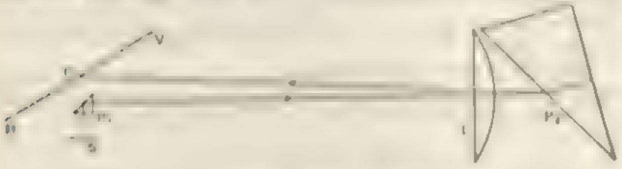


FIG. 3.—LITTROW QUARTZ SPECTROGRAPH

S, slit; P_1, totally reflecting quartz prism; L, autocollimating quartz lens; P_2, Littrow quartz prism; C, camera; [illegible]

[illegible] by light in its forward passage through the lens and prism [illegible] by its [illegible] the same path after reflection from the mirror [illegible] the back of the prism.

[illegible] compound prism, the Amici, and the Pellin-Broca [illegible] A form of direct-vision [illegible] by G. B. Amici, consists of two prisms, ABE and CDF (fig. 4), of crown glass having a [illegible] prism [illegible] of flint glass having a high ratio of dispersion [illegible] The prism angles [illegible] and dispersed. [illegible] power are obtained [illegible] prisms.

A [illegible] constant-deviation spectroscope [illegible] by P. Pellin and A. Broca [illegible] prism ABCD [illegible] BEC [illegible] cancel [illegible] introduced [illegible] axes are fixed [illegible] different wavelengths [illegible] into the field of view by rotating the prism [illegible] a wavelength-calibrated drum is attached.

FIG. 4.—AMICI DIRECT VISION SPECTROSCOPE

ABC and [illegible] crown glass; [illegible] flint glass; [illegible]

Dispersing Power of Prisms.—The power of any spectroscope [illegible] dispersion. It is [illegible] angle of deviation θ with respect to a [illegible] in wavelength λ [illegible] Since θ varies with the refractive index n of the prism and the refractive index varies with the wavelength λ, then the angular dispersion $\frac{d\theta}{d\lambda} = \frac{d\theta}{dn}\frac{dn}{d\lambda}$.

[illegible] depends upon two factors. The second factor $dn/d\lambda$ [illegible] of the prism [illegible] represented by an equation of the (Cauchy) form $n = a + b/\lambda^2$, where a and b are constants obtained from measured values of n for two different wavelengths. $dn/d\lambda = -2b/\lambda^3$. The negative sign merely indicates increase of n with decrease of λ. By this formula the dispersive power of a glass [illegible] varies inversely as the cube of the wavelength. For example, the dispersion at 4000 A is about eight times that at 8000 A. The other factor $d\theta/dn$ depends upon the angle of the prism and on the angle it presents to the incident light. If, as in fig. 6a, the angles of incidence and emergence be represented by i_1 and i_2 respectively, the internal angles of refraction by r_1 and r_2, the angle of the prism by α and the angle of deviation by θ, it is readily deduced from Snell's law of refraction $n = \sin i/\sin r$ that

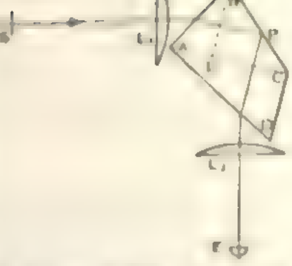


FIG. 5.—PELLIN-BROCA CONSTANT-DEVIATION SPECTROSCOPE

S, slit; L_1, collimator lens; P, Pellin-Broca prism; L_2, telescope lens; E, eyepiece

$$\frac{di_2}{dn} = \frac{\sin(r_2 + r_1)}{\cos r_2 \cos i_2}$$

or, since $d\theta = di_2$, and $r_1 + r_2 = \alpha$

$$\frac{d\theta}{dn} = \frac{\sin \alpha}{\cos r_2 \cos i_2}$$

This expression clearly shows the dependence of the dispersion upon the prism angle and the angle of incidence. A convenient equivalent expression was first deduced by Lord Rayleigh, namely

$$\frac{d\theta}{dn} = \frac{t_2 - t_1}{a}$$

where t_2 and t_1 in fig. 6b are the lengths in the prism traversed by the extreme rays, and a is the breadth of the emergent beam. When the prism is at minimum deviation and completely covered by the beam, $t_2 - t_1$ is simply the base length of the prism. The original expression for angular dispersion now becomes

$$\frac{d\theta}{d\lambda} = \frac{d\theta}{dn}\frac{dn}{d\lambda} = \frac{t_2 - t_1}{a}\frac{dn}{d\lambda}$$

When there is more than one prism, $t_2 - t_1$ is summed for all prisms, and in Amici prisms the algebraic sum is taken, the lower dispersion being regarded as negative. The angular dispersion may be expressed in radians or degrees per angstrom, but it is usually more convenient to measure linear distances x in the focal plane and express linear dispersion $dx/d\lambda$ in millimetres per angstrom. In general, if f is the focal length of the image-forming lens, $dx/d\lambda = f\, d\theta/d\lambda$.

FIG. 6.—DISPERSING POWER OF A PRISM

Resolving Power of Prisms.—It would appear from this equation that by choosing a large focal length it would be possible to separate two juxtaposed images due to light of two wavelengths, however little the wavelengths differed. Such indeed would be the case if an infinitely narrow slit were imaged as an infinitely narrow line. However, owing to the wave nature of light, any optical image is a diffraction pattern of finite size that depends both on the wavelength and on the dimensions of the optical system. An instrument is thus limited in its power to resolve images of slightly different wavelength, or closely adjacent spectrum lines. This resolving power R is expressed numerically as the ratio of a mean wavelength λ to the least difference $d\lambda$ between two wavelengths that can be recognized as a double

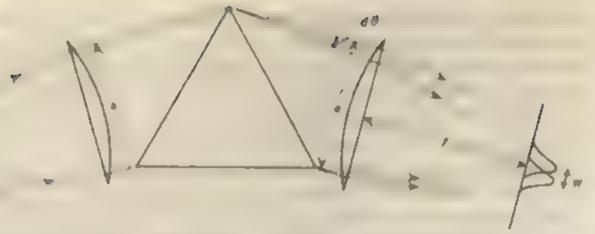


FIG. 7.—RESOLVING POWER OF A PRISM SPECTROSCOPE

image; *i.e.*, $R = \lambda/d\lambda$. Each actual image of the slit is a diffraction pattern having most of the light concentrated in a central band of finite width flanked by secondary bands of negligible intensity. If f be the focal length of the camera lens, and a the effective breadth of the beam passing through a rectangular cross section, the half width w of the central diffraction band is given by the elementary theory of diffraction as $w = f\lambda/a$. This says that the image width increases with focal length and with wavelength but decreases with aperture increase. The condition for resolution is explained in fig. 7, where the spectrum is focused on a screen by a lens of focal length f. The beams of two equal-amplitude closely adjacent wavelengths λ and $\lambda + d\lambda$ are separated by the prism an increment of angle $d\theta$. Each slit image appears as a diffraction band, and (according to Rayleigh's criterion) resolution is just effected when the central maximum of one image coincides with the first minimum of the other; *i.e.*, when the distance between centres of the two bands is w. Under these conditions the integration of the two bands gives a curve with two maxima and, if the intensity of these be rated as unity, that of the minimum between is 0.81. Experience has shown that this intensity difference is sufficient to reveal duplicity of the image, and the condition for resolution is thus $d\theta = w/f = \lambda/a$; or $\lambda = a \cdot d\theta$. By definition, $R = \lambda/d\lambda$; therefore, $R = a\,d\theta/d\lambda$, or in Rayleigh's form

$$R = (t_2 - t_1) \cdot dn/d\lambda$$

When there are several prisms, the algebraic sum is to be taken, as in the case of dispersion. An ordinary flint-glass prism of one centimetre base is found to have a resolving power for sodium-yellow light of about 1,000, but this number increases rapidly with decreasing wavelength, since $dn/d\lambda = -2b/\lambda^3$ approximately.

An infinitely narrow slit was assumed in the derivation of the theoretical resolving power, but such a slit would transmit no appreciable light. In practice it is necessary to use a slit of finite width and sacrifice some of the resolving power for the sake of intensity in the spectral images. The actual separating power with finite slits was called purity by A. Schuster, who defined a normal slit as the product of a quarter wavelength by the aperture ratio $\lambda F/4D$ and calculated that nearly 99% of the theoretical resolving power was conserved by using this normal slit. Thus, if $F/D = 20$, the normal slit for 4000 Å would have a width of 5λ or 2μ. If double the normal slit is used, the resolving power is reduced to 94% of maximum, but the intensity is doubled. Increasing the slit width to about four times normal reduces the resolving power to 80% but trebles the intensity. Still wider slits rapidly reduce the resolving power or spectral purity without appreciably intensifying the spectra of monochromatic images except where they overlap. A wide slit greatly increases the intensity of continuous background relative to that of monochromatic lines. The above definitions of dispersing power, resolving power, and spectral purity are valid also for diffraction gratings discussed below.

Grating Spectrographs.—The use of diffraction gratings instead of refraction prisms for the study of spectra brings advantages of constant and usually greater dispersion and resolution, as well as opportunity to observe extreme ultraviolet and infrared spectra for which no transparent prisms can be found. Modern gratings are made by ruling with a diamond point equidistant parallel lines on a polished plate of glass, of speculum (copper + tin) metal, or of glass on which a film of aluminum has been deposited. Those in most general use for visible and ultraviolet spectra have either 15,000, 20,000, 24,000, or 30,000 lines per inch and effective ruled areas ranging from about two to six inches. A grating with 15,000 lines per inch permits measurements in infrared spectra to 30,000 Å, but for the investigation of longer waves coarser gratings must be used. In 1910 R. W. Wood first ruled 2,000 to 3,600 lines per inch on copper plates and by proper choice of groove form succeeded in concentrating much of the visible light into relatively high orders. Such echelette gratings concentrate the greater part of infrared energy in a first-order spectrum, and they can be ruled to perform in this manner for almost any desired wavelength. Thus, a grating with 900 lines per inch may be made to function well at about 300,000 Å, 360 lines per inch about 900,000 Å, and 80 lines per inch about 1,500,000 Å. These coarse echelette gratings have ruled surfaces up to 10×20 inches in area and are usually cut in a solder surface with a steel tool. A grating ruled on a sheet of transparent glass is called a transmission grating, but it is unfeasible to rule large areas of glass without wearing down the diamond point. For this and other good reasons most gratings are ruled on opaque soft metal surfaces and are called reflection gratings. The fundamental principle is the same in both cases.

Theory of Diffraction Gratings.—The simple theory of the plane diffraction grating gives the wavelength of light as a function of the grating space and of the angles of incidence and diffraction of the light rays. Fig. 8 represents the cross section of a transmission grating in which transparent and opaque lines or spaces alternate and each pair (such as AB) is called the grating space b. If light from an illuminated slit S falls as a parallel beam on the grating at an angle of incidence i with the normal, a portion of the light passes directly through the grating aperture to produce an undeviated image, but a part of it is diffracted in the direction AD, making an angle θ with the normal. After drawing the perpendiculars CB and DB, it is seen that the difference in path traveled by corresponding rays from adjacent apertures (or the retardation as it is called) is given by the sum of the lengths CA and AD. When this retardation is equal to any integral number n of wavelengths λ, bright images of S in colour corresponding to λ are produced where the diffracted rays are brought to a focus F. This condition is simply expressed by the equation

$$CA + AD = n\lambda = b\ (\sin i + \sin \theta_n)$$

in which n represents the order number of the spectrum.

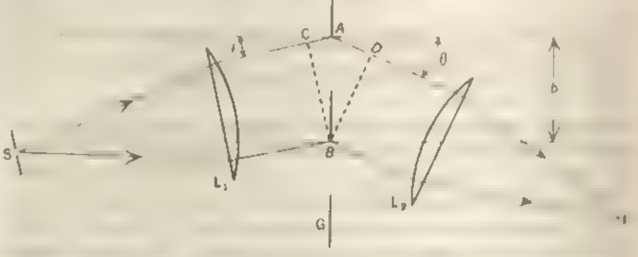


FIG. 8.—TRANSMISSION GRATING

S, slit; L_1, collimator lens; G, grating; L_2, camera lens; b, grating space; i, angle of incidence; θ, angle of diffraction; F, focused spectrum

The law of reflection gratings is similar to that of transmission gratings. Thus in fig. 9 a beam of parallel rays falls on a plane reflection grating with grating space $AB = b$. The incident light makes an angle i with the normal, and part of the light is diffracted at an angle θ. The perpendiculars AC and BD show that BC and AD are the retardations. In this case the total retardation is $AD - BC = b\ (\sin i - \sin\ \theta_n)$. Whenever this

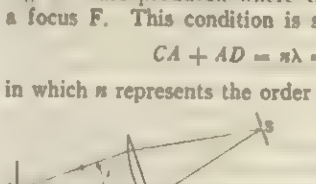


FIG. 9.—REFLECTION GRATING

S, slit; L_1, collimator lens; G, grating; L_2, camera lens; b, grating space; i, angle of incidence; θ, angle of diffraction; F, focused spectrum

equals one or more whole wavelengths $n\lambda$, a bright image is seen in colour corresponding to λ and yields the law of the grating. If the diffracted rays are on the same side of the normal as the incident rays, the two retardations are added; wherefore, the general equation for all spectra must be written $n\lambda = b\ (\sin i \pm \sin \theta_n)$, the positive or negative sign being used when the incident and diffracted rays are on the same or opposite sides of the normal, respectively. If the grating space is not too small, the groove form of the ruled lines can be controlled by shaping the diamond point in such a way that the diffracted light is largely concentrated in one general direction or spectral region. Such gratings may, for example, give an intense visible spectrum and weak ultraviolet, or vice versa. If light is incident normally on any grating, the angle i is zero and the grating formula reduces to $\pm n\lambda = b \sin \theta$, which represents a central zero-order image flanked on either side by first-, second-, third-order spectra, and so on, as $n = \pm 1$, ± 2, ± 3, and so on. It is obvious that the absolute values of wavelengths may be determined from the measurement of the grating space and of one (or two) angles. This was the method employed by Fraunhofer and by Ångström to establish standards or fiducial points in spectra.

Concave Gratings.—All plane gratings, like prisms, require the use of lenses or mirrors to collimate and to focus the incident and emergent light. To eliminate lens aberrations and limitations of light absorption, Rowland in 1881 first ruled a grating on a spherical mirror that collected light from a slit, dispersed the diffracted light into spectra, and focused them on a circle the diameter of which equaled the radius of curvature of the concave grating. All concave gratings mounted on the Rowland-circle principle exhibit strong astigmatism; *i.e.*, a point slit is imaged as a line with a length dependent on wavelength and spectral order. This astigmatism reduces the brightness of the spectra and precludes the use of diaphragms or sectors at the slit, and other observations requiring a one-to-one correspondence in the slit and its spectral images. Fortunately, this defect can be avoided either by placing horizontal lines to be focused on the circle at a greater distance from the grating than the slit or by illuminating the concave grating with parallel light, in which case the astigmatism is practically eliminated and the brightness of spectra greatly increased.

With both plane and concave gratings a great advantage is obtained by forming and observing spectra on the normal to the reflecting surface, for this gives the normal spectrum in which angular or linear distances measured either way from the grating normal are (for a considerable distance) proportional to the change in wavelength observed. Since the concave reflection grating is free from chromatic aberration and focuses all spectra on the same focal surface, Rowland took advantage of the normality of spectra and of the coincidence principle in overlapping orders of spectra to determine the relative values of the Rowland system of standard wavelengths.

Dispersion.—The explanation of dispersion by a grating is easily seen, since the formula shows that the angle θ increases or decreases with the wavelength λ. When composite light falls on the grating, it is obvious that the shorter violet waves will give an image closer to the central image than the longer red waves, and successive orders of spectra will increase in dispersion proportional to the order number.

If θ be the angle of diffraction, dispersion is defined as before by $d\theta/d\lambda$, and since (for a given position of the grating) the angle of incidence i is constant, the grating formula gives $d\theta/d\lambda = n/b \cos \theta$. This equation shows clearly how the dispersion varies directly with the order of the spectrum, and inversely as the grating space; the closer the rulings, the greater the dispersion, irrespective of the total number of grating spaces. Furthermore, the dispersion is smallest when $\cos \theta$ has its maximum value (which is unity for zero angle), *i.e.*, when the spectrum is observed normal to the plane of the grating. In this position also the dispersion is most nearly uniform throughout the spectrum, since it varies with $\cos \theta$, and in the proximity of $\theta = 0$ this trigonometric function changes very slowly. The spectrum given by a grating, unlike that given by a prism, is accordingly normal insofar as $\cos \theta$ can be considered constant. There is a certain amount of overlapping of successive orders, since $n\lambda$ may acquire the same value in different ways. For example, $\lambda = 7000$ Å in the first order will be coincident with $\lambda = 3500$ Å in the second order, and $\lambda = 4000$ Å in the third order will coincide with $\lambda = 6000$ Å in the second. When observation of a particular spectrum is desired, it is usually possible to remove the overlapping spectra by using suitable colour filters in front of the slit or selective detectors at the focus.

Resolving Power.—As in the case of the prism, the resolving power of a grating is a theoretical quantity expressing the image-separating power when the slit is indefinitely narrow; it is represented symbolically by $R = \lambda/d\lambda$ and is given in simple form by $R = nN$, where n is the order number of the spectrum and N is the total number of rulings. Thus, the resolving power of a grating depends only on the total retardation in wavelengths from the first to the last line and not on the grating space. For example, a five-inch grating with 20,000 lines per inch would have a theoretical resolving power of 200,000 in the second order, surpassing an imaginary glass prism of two metres base length. At $\lambda = 8000$ Å, $d\lambda$ would be 0.04 Å, while at $\lambda = 4000$ Å it would be 0.02 Å. This theoretical resolving power is attained only under ideal conditions, since actual resolution depends on various factors, such as the relative intensities of spectral lines, their form or structure, the width and manner of illuminating the slit, the perfection and adjustment of the optical parts, graininess or contrast of photographic emulsions, and so on.

Interferometers.—Since the resolving power of a grating is proportional to the product nN of the order number and the total number of elements, it may be increased in two ways. To increase the number of elements, the number of lines per inch may be increased or a greater number of inches ruled, but the total number of useful lines is usually limited to about 100,000 by the practical difficulties of making all grooves exactly alike and equally spaced. The alternative method of using higher orders is limited in highly dispersive gratings by the great reduction in aperture when the grating is viewed tangentially. These limitations in ruled gratings are circumvented more or less in certain interferometers known as Michelson's echelon interferometer, the Fabry-Pérot interferometer, and the Lummer-Gehrcke interferometer, all of which employ a relatively small number of elements (10 to 30) but very large spectral orders (10,000 to 100,000 or more). Since such high orders are very close together, confusion due to overlapping of many lines must be avoided by crossing these interferometers with any stigmatic spectrograph. The Michelson transmission echelon (fig. 10) consists of a number of glass plates of equal thickness t (*c.* 1 cm.) and refractive index μ, but each shorter than its predecessor. These are wrung into optical contact with each other to form a series of steps; *i.e.*, an echelon. Parallel light from a collimator, with the slit parallel to the edges of the steps, is then passed through the larger end.

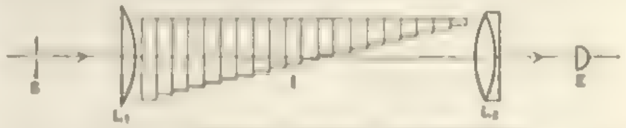


FIG. 10.—MICHELSON ECHELON INTERFEROMETER
S, slit; L_1, collimator lens; I, interferometer; L_2, achromatic lens; E, eyepiece

In each step the light is retarded an equal amount $(\mu - 1)t/\lambda$, and diffraction spectra of very high order are formed when all the transmitted beams are collected and focused by a lens. This interferometer was later improved as a reflection echelon that is not limited by the absorption of glass and has nearly four times the resolving power, since the retardation per step is $2t/\lambda$. An echelon of 30 plates each 1 cm. thick has a theoretical resolving power at 5000 Å of $30 \times 2/.00005 = 1{,}200{,}000$. This instrument has been employed for the examination of spectral line structure and for the precise measurement of relative wavelengths.

When a beam of light passes through a transparent film or plate, multiple reflections occurring at the surfaces produce interference colour effects seen, for example, in soap bubbles or in oil

films on a wet pavement. Such interference bands are relatively broad because the reflecting power is so low that only the first reflection at each surface is effective. When the reflecting surfaces are exactly parallel and the reflection coefficient is increased, the multiple reflections produce very narrow bright fringes on a broad dark background. These conditions are met in the Fabry-Pérot and in the Lummer-Gehrcke interferometers. In the former a large coefficient of reflection is obtained by depositing bright, partly transparent metal films on two glass or quartz planes. These planes are mounted face to face, separated some millimetres or centimetres, and adjusted to exact parallelism. The most common type of Fabry-Pérot interferometer (fig. 11) is the fixed-etalon type, in which the plates are maintained at a fixed distance by being pressed against a separating ring spacer or etalon. When it is used with monochromatic light and viewed axially, circular interference bands known as Haidinger's fringes are seen at infinity. When it is used with complex light the overlapping of fringes belonging to neighbouring spectral lines is avoided by crossing the interferometer with a stigmatic spectrograph. The order of interference at the centre of the pattern for wavelength λ is in general a whole number n plus a fraction ϵ, determinable from the diameters of successive rings. The double distance between the interferometer plates is $2D = \lambda\ (n + \epsilon)$ for wavelength λ, or $2D = \lambda'(n' + \epsilon')$ for another wavelength λ'. Thus the simple proportion $\lambda/\lambda' = (n' + \epsilon')/(n + \epsilon)$ offers an opportunity for the precise comparison of wavelengths, and this is in fact the method employed for the establishment of secondary standards of wavelength measured relative to the primary standard (red radiation from cadmium), which was itself found to be 6438.4696 Å by comparison with the meter. (*See* INTERFEROMETER.) This method of comparing wavelengths has been used extensively also for measuring the ratios of wavelengths in vacuo and in gases and thus determining the refractive indexes and dispersions of gases, notably air. In addition, the Fabry-Pérot etalon interferometer has proved the most effective instrument for the measurement of exceedingly small wavelength differences, which appear, for example, in hyperfine structure, isotope shifts, pressure shifts, and so on. Besides simplicity, stability, and freedom from defects, this interferometer possesses a large advantage in the flexibility of its resolving power, since this is equal to the product of the spectral order n by the effective number of reflections N. With a reflection coefficient of 0.75 the value of N is about 10, and with 0.90 the effective number of reflections is about 30. Thus, a Fabry-Pérot interferometer with silver mirrors reflecting 0.90 is equivalent to a reflection echelon of 30 plates having plate thickness equal to the etalon separator. The echelon plates have a fixed thickness, whereas the Fabry-Pérot plates may be separated as desired to increase resolving power.

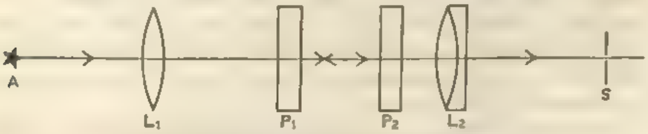


FIG. 11.—FABRY-PÉROT INTERFEROMETER

A, light source; L_1, collimator lens; P_1 and P_2, interferometer plates; L_2, achromatic projection lens; S, spectrograph slit

In fact, the highest resolving powers on record have been attained with Fabry-Pérot interferometers, and the most detailed and reliable results on hyperfine structure must be credited to them.

The Lummer-Gehrcke interferometer consists of a long unsilvered plate with perfectly plane-parallel surfaces in which a high reflection coefficient is obtained by directing the entering light beams so that they are reflected at nearly the critical angle (fig. 12). The greater part of the light is reflected at each internal reflection, but the remaining fractions emerge with equal retardations as parallel refracted beams that may be collected and focused by a lens to form identical sets of interference patterns above and below the optic axis. The resolving power is again nN, where n is the order of interference (retardation) and N the effective number of reflections or emergent beams, which usually ranges from 10 to 20. The practical resolving power of the Lummer-Gehrcke interferometer has been found to be about equal to the length of the plate measured in wavelengths. This instrument has been employed successfully for the observation of hyperfine structure, isotope shifts, and for studying Zeeman patterns.

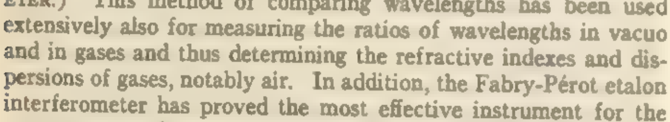


FIG. 12.—LUMMER-GEHRCKE INTERFEROMETER

A, light source; L_1, collimator lens; I, interferometer; L_2, achromatic projection lens; S, spectrograph slit

The most famous of all interferometers is that invented by Michelson in 1881 and fully exploited by him in his classical investigations of "ether drift," sharpness of spectral lines, and evaluation of the international metre in wavelengths of cadmium radiations.

IV. METHODS OF OBSERVATION

There are in general four ways of observing spectra: visually, photoelectrically, radiometrically, and photographically; each has advantages and limitations.

Visual.—Although the average human eye is extraordinarily sensitive for green light (5500 Å), its sensitivity declines rapidly to zero for infrared ($\lambda > 7700$ Å) and ultraviolet ($\lambda < 3800$ Å). Being highly selective, variable, localized, and restricted to one octave, visual methods of observation are relatively unimportant in spectroscopy.

Photoelectric.—As a substitute for the human eye, the electric eye or photoelectric tube may be employed for the investigation of spectra in limited ranges. Briefly, the photoelectric tube consists of two metal electrodes in an evacuated glass tube. The anode is maintained at a positive potential with respect to the cathode. When radiation falls on the cathode, the liberated electrons are drawn across the evacuated space to the anode. This flow of electrons constitutes a photoelectric current that is directly proportional to the incident energy. Since these photoelectric currents are usually very small, they are magnified by the use of thermionic amplifiers, or, in the case of feeble radiation, use is made of a Geiger-Müller tube, in which each electron emitted gives rise to a pulse of current. Since 1940, multiplier phototubes, in which enormous amplification is obtained by a sort of chain reaction, have become popular as detectors and intensity recorders of spectral lines. All photoelectron emitters, *e.g.*, Na, Zn, Cd, Mg, are highly selective and usually most efficient in the ultraviolet. By depositing thin films of heavy metals or compounds, *e.g.*, CsO, PbS, photoelectric sensitivity may be extended to near infrared but not much beyond 40,000 Å.

Radiometric.—A unique and important class of radiation detectors, responding to the heating effect of the radiation, is found in blackened radiometers, radiomicrometers, thermopiles, and bolometers. They are unique because they are nonselective, and important because their sensitivity extends from ultraviolet, through visible, to far infrared spectra. The radiometer is a blackened vane supported some distance from the axis of rotation of a glass or quartz fibre *in vacuo*. When radiation is absorbed by the vane, the supporting fibre is twisted through an angle proportional to the incident radiation. The radiomicrometer is essentially a moving coil galvanometer of the D'Arsonval type. A single thermocouple is suspended by a quartz fibre with the plane of the couple coil in the line joining the poles of a magnet. When radiation falls on the thermocouple junction, the electric current generated acts to turn the couple in the magnetic field. The thermopile consists of one or more junctions of dissimilar metals; *e.g.*, bismuth and antimony or silver, iron, and constantan (alloy of nickel and copper). If radiation falls on alternate junctions, a galvanometer in the circuit will give a deflection proportional to the heating effects. All radiation-detecting and measuring devices that depend on the initial conversion of radiant energy into heat energy and finally into mechanical or electrical energy

are extremely inefficient, as are all heat engines. The feeble final effects of radiation may be magnified up to about 200 times by suitable amplifiers, but thermopile currents that produce galvanometer deflections of the same magnitude as Brownian-motion effects obviously cannot be amplified to advantage.

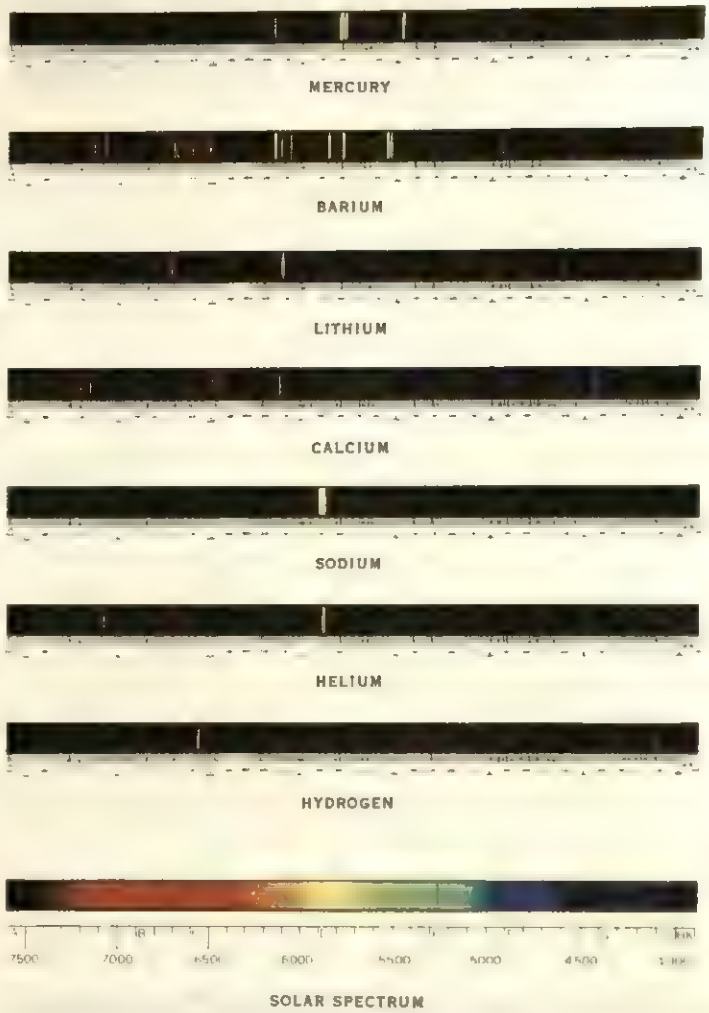
The bolometer, devised by S. P. Langley in 1880, consists of two similar very thin strips of metal, usually platinum, which form two arms of a Wheatstone bridge. The strips are blackened on one side, one strip is shielded, and the other is exposed to radiation. The absorption of radiation by one strip increases its temperature, changing its electrical resistivity and producing a deflection of the galvanometer. With his bolometer Langley in 1888 mapped the solar spectrum to 55000 Å by using a spectrometer with rock-salt lenses and prisms and passing the spectrum over the bolometer strip by rotating the prism. The galvanometer readings were recorded on photographic paper whose motion was coordinated with that of the prism. The same basic principles were used in the 1960s with improved instrumentation in all automatic recording of spectra scanned either by photoelectric or radiometric detectors.

Photographic.—The actinic action of radiation on silver salts is fairly effective throughout many octaves of spectra, from blue to extreme ultraviolet, but the maximum sensitivity is usually in the blue. Photographic image formation rests on a latent and lasting effect produced by light in microscopic silver-halide crystals suspended in gelatin coated thinly on a glass, film, or paper base. The crystals affected by light are readily reduced to metallic silver by a suitable reducing agent or developer, and the unreduced silver halide is dissolved out with a fixer, usually hypo, thus producing images composed of tiny silver grains, the density of which is a function of the intensity or total energy incident upon the emulsion. Although gelatin is highly transparent to waves longer than 3000 Å, it is strongly absorbent for shorter waves and practically opaque below 2000 Å. In order to record spectra below 2000 Å, in 1901 V. Schumann effectively eliminated the shielding effect of gelatin by depositing silver salts on glass plates with only enough gelatin to support them. Another way to circumvent the ultraviolet absorption of gelatin was discovered in 1921; it consists of coating an ordinary photographic plate or film with a thin film of oil or suitable organic substance that fluoresces in ultraviolet light and emits longer wave radiation not absorbed by the gelatin (*see* PHOTOGRAPHY).

In 1873 H. M. Vogel discovered that the light sensitivity of silver salts could be extended to longer waves by bathing ordinary plates in certain dye solutions. By 1900 the incorporation of suitable photosensitizing dyes in emulsions produced panchromatic plates that were sensitive to the entire visible spectrum, and after 1930 the synthesis and application of an important series of polycarbocyanine dyes extended the sensitivity of photographic emulsions into the infrared to about 13000 Å. Since large intervals of spectra and multitudes of spectral details could thus be recorded simultaneously, it follows that more spectrographic observations have been made photographically than by all other methods. In addition to being applicable in an enormous range of wavelengths, the photographic method provides a permanent, detailed record of a spectrum that can be conveniently examined or accurately measured by any number of persons as often as desired without repeating the experiment. Furthermore, photographic exposures are cumulative, and the photographic method thereby derives additional advantages in integrating the light from flickering or intermittent sources and in recording extremely faint images by giving long exposures. Although the absolute sensitivity of a photographic emulsion may vary considerably with wavelength, it can often be regarded as constant over a limited range of spectrum and in any case can be calibrated as a function both of wavelength and intensity, thus providing a convenient method of measuring the relative intensities of spectral lines. This photographic-photometry method was developed about 1925 for the purpose of testing quantum rules for the relative intensities of certain lines, and since 1930 it has become standard practice in quantitative spectrochemical analysis.

It should be emphasized that any selective receiver (eye, phototube, or photographic plate) employed for measuring the relative intensities of spectral lines requires calibration by comparison with nonselective receivers; *e.g.*, thermopiles.

BRIGHT-LINE SPECTRA



WELCH SCIENTIFIC CO

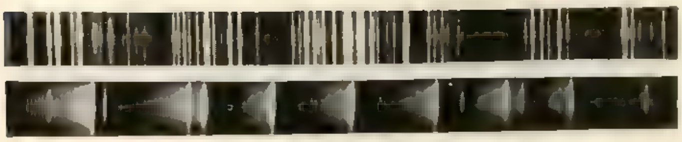
Reproductions of the bright-line spectra of selected elements found on the sun, and of the solar spectrum with principal Fraunhofer lines A, B, C, D, E, F, G, H, K. A and B are caused by terrestrial oxygen (O_2); C and F belong to hydrogen; D represents sodium; E, G, H, K indicate calcium mainly. Also shown are spectra for mercury and barium

V. MEASUREMENTS

Spectroscopic measurements are concerned essentially with energy distribution as a function of wavelength. This function is characteristic primarily of the light source and secondarily of its operating conditions, assuming that the spectrograph is nonselective. If the energy changes slowly from point to point in large intervals of spectra, the spectra are called continuous. They may be either emission spectra or absorption spectra and consist usually of one broad band.

If the spectral intensity changes rapidly from point to point and undergoes sudden inflections, the spectra are called discontinuous. Such spectra also may be either emission spectra or absorption spectra; they usually consist of hundreds or thousands of lines. The division of observable spectra as continuous and discontinuous, either in emission or in absorption, is a broad but useful classification. Each class contains an enormous number of examples, some of which may be grouped according to certain criteria into smaller classes. The following examples will illustrate the principal classes of spectra that are subjected to observation and measurement.

Emission Spectra.—All self-luminous sources yield emission



LINE AND BAND SPECTRA OF NITROGEN

spectra, which are of different types according to the nature of the light source. Spectrally dispersed light from glowing gas mantles, electrically heated metallic filaments, or hot molten metals are continuous over great ranges of wavelengths, and for blackbody radiation the wavelength of maximum intensity is an inverse function of absolute temperature. It follows that such substances cannot be distinguished by means of their spectra as long as they remain in the solid or liquid state.

The effects are quite different when the substances examined are in a state of luminous gas or vapour. They then emit characteristic discontinuous spectra by which they can be uniquely identified, whether they are excited in the laboratory or in a celestial body far away. These spectra consist of bright lines or bands variously spaced on a dark background.

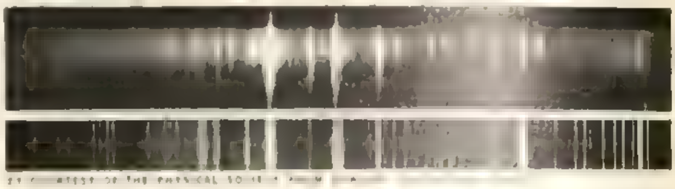
Line spectra are characteristic of atoms or ions, while band spectra originate in molecules. The former consist of a certain distribution of slit images of different intensity and usually of various types, such as partially self-reversed lines, sharp lines, diffuse lines, unsymmetrical lines, and complex lines with two or more components. Partially reversed or easily absorbed lines always indicate that the normal state or some low energy state of the atom or ion is involved. The relative sharpness or diffuseness of lines is determined by such factors as pressure, temperature, mass, and excitation. Complexity of spectral lines is referred either to different isotopes of the same element (isotope shift) or to interaction of valence electrons with a spinning isotopic nucleus (hyperfine structure).

In a band spectrum a great number of fine lines are closely and regularly placed in each band, becoming more and more crowded near the band head. The band heads are on the short-wave side of the line groups in some band systems, and on the

"RED" SPECTRUM OF HYDROGEN SHOWS ITS CHARACTERISTIC BALMER SERIES OF LINES IN THE VISIBLE AND NEAR ULTRA-VIOLET REGIONS

long-wave side in others. In pure gases the commonly observed band spectra characterize molecules consisting of two similar atoms, such as H_2, He_2, N_2, O_2. The line spectra of metals excited in flames, arcs, or sparks operated in the free atmosphere are usually accompanied by band spectra characteristic of monoxides (such as CaO, BaO, ScO, LaO) formed by the oxidation of the metals. Arc spectra frequently show band spectra identified with NO and CN, as well as line spectra of N, O, Ar, and C, the principal constituents of the atmosphere. Many dipole molecules unknown to ordinary chemistry but recognized spectroscopically are apparently formed only under special conditions existing in the light sources.

No two substances yield the same spectrum, and, consequently, the chemical nature of substances can be determined spectroscopically. Thus, glowing atomic hydrogen is characterized by a bright line in the red, besides progressively weaker ones of shorter wavelength, and, since these are exhibited by nothing but hydrogen, they serve to disclose the presence of atomic hydrogen wherever it occurs in the luminous state. Each of the known chemical elements has its distinctive family of spectral lines, differing in distribution, relative intensity, and total number but always the same under the same conditions of excitation. The same element, however, under radically different excitation conditions will emit entirely different spectra. The most easily excited spectrum is that of neutral atoms in possession of all their valence electrons, but, if the excitation is violent enough to remove one electron from each atom, the resulting ions when excited radiate another characteristic spectrum. A third type of spectrum is emitted when the atoms are doubly ionized by removal of two electrons, and so on until only one outer electron remains. Thus, each element or atom has as many characteristic spectra as it has electrons. In most arc or spark sources generating atomic spectra of a pure element, two or more spectra appear simultaneously, with one usually predominating. When two or more elements with comparable excitation characteristics are abundant in a light source, each element emits its own spectra essentially the same as if all other elements were absent. The spectra of complex substances, such as ores, minerals, metal alloys, and chemical mixtures, thus reveal simultaneously all the types of atoms and ions that are excited in the light source. Such supernumerary line spectra often have superposed on them band spectra of molecules formed and excited in the same light



SPECTRA OF STRONTIUM: (ABOVE) SPARK, (BEL[illegible] FLAME LINES ARE REDUCED IN INTENSITY IN THE SPARK SPECTRUM, W[illegible] ENHANCED" OR "SPARK" LINES ARE INTENSIFIED

source, as well as lines and bands of atmospheric elements. Also seen is a more or less intense continuou[illegible]nd, which may be traced either to unquantized radiation [illegible]toms, to a continuous spectrum from incandescent oxid[illegible]ight source, or to stray reflected or scattered light in the spectrograph; sometimes to all three

In such complicated cases it is clear that only experienced spectroscopists can interpret the spectrogra[illegible]ly. Familiarity with the spectra of various elements ofte[illegible]n experienced observer to identify lines or bands at [illegible]he distribution and relative intensities in certain group[illegible]ized and recognized in the same way that constellations a[illegible]gnized among stars. In theory any unknown spectrun[illegible]ture of spectra can be identified by comparing it with al[illegible]ectra of known elements excited under similar condition[illegible]ractice the endless labour of repeatedly observing th[illegible]rison spectra is avoided by measuring wavelengths corres[illegible]o the unknown lines and comparing these with wavelength t[illegible]bles of identified lines compiled for this purpose

For purposes of chemical identification [illegible] wavelengths and the relative intensities of lines in spectra may be regarded as constants of nature, but closer examination shows that in any given spectrum certain variations are produced by different excitation conditions. Thus, spectral wavelengths are generally slightly increased as the total pressure of gas or vapour surrounding the radiating particles is increased. Such pressure shifts are further correlated with atomic diameters and with atomic energy levels.

Also the wavelengths are either decreased or increased, according to Doppler's principle (*see* DOPPLER EFFECT), when the radiating source approaches or recedes from the observer. In astrophysics this principle has disclosed the sun's rotation, the radial velocities of stars, and an apparently expanding universe; whereas in the laboratory it explains (when combined with the kinetic theory of gases) the intrinsic natural width of spectral lines as a function of $\sqrt{T/M}$, in which T represents absolute temperature and M molecular weight. When standard wavelengths in the solar spectrum are measured, it is necessary to correct the observations for Doppler effect due to the relative motion of the sun and the spectrograph.

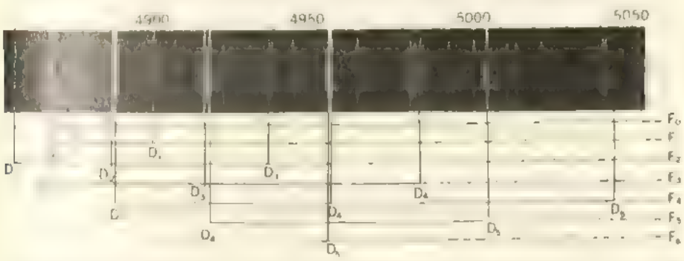
Again, with few exceptions, all atomic spectral lines are split into three or more components when the radiating particles are in strong magnetic fields (*see* ZEEMAN EFFECT), and, similarly, many lines either shift or widen or divide into resolved compo-

nents when the radiating particles are subjected to strong electric fields (*see* STARK EFFECT).

From the general theory of relativity (*q.v.*) A. Einstein in 1917 concluded that wavelengths characteristic of atoms should increase in strong gravitational fields. This effect is quite inappreciable in the laboratory, but it appears to be present in the solar spectrum, and especially in the spectra of very dense stars.

The most striking changes in relative intensities of lines belonging to a particular spectrum occur when excitation conditions favour the partial self-reversal of selected lines. These variations of spectral characteristics are highly objectionable in standard wavelengths and in spectrochemical analysis, but they are of primary importance in spectroscopic research, since they yield information about atomic structure and the mechanism of radiation. Consequently, when truly invariable spectra are desired, they may be produced by carefully specifying and reproducing the light source in such a manner that spectral variations are eliminated or minimized; but if atomic structure is being investigated, it is desirable to effect the largest practicable changes in spectra and relate these to variations in the light sources. In any case, it is evident that some means of accurately expressing the positions of lines in a spectrum is of primary importance.

Wavelengths.—Measurements of spectra become directly comparable with each other only when they have been converted to the scale of wavelengths (or frequencies) of the radiations that produce the spectral lines. The credit for introducing this scale into spectroscopy is due to Ångström, who described the spectrum of the sun in terms of wavelengths in 1868, and since



HIGH-DISPERSION ARC SPECTRUM OF IRON SHOWS MULTIPLET OF 15 LINES AND PLAN OF ITS STRUCTURE

that time the positions of practically all known spectral lines have been expressed in wavelengths. The unit of wavelength in spectroscopy is a ten-millionth part of a millimetre, or 10^{-10} m. It is called the angstrom and is symbolized by Å. In these units, the visible spectrum ranges from about 3800 Å (violet limit) to about 7700 Å (red limit). Extremely short waves, in particular those observed in Röntgen- or X-ray spectra, which range from about 0.1 Å to about 100 Å, are more conveniently expressed in still smaller units, and for this purpose the X unit, approximately 1/1000 Å, is generally used. Similarly, very long optical waves, in particular the far infrared heat waves, are more conveniently expressed in larger units, and for this purpose the micron μ = 10,000 Å is commonly used. Summarizing: 10^{-6} m. = 0.001 mm. = 1 μ = 10,000 Å = 10,000,000 X. Wavelengths ranging from 6 Å to 5,000,000 Å have been measured in spectra associated with valence electrons, but these limits will surely be extended in both directions.

Absolute wavelengths cannot be determined directly from prisms but have been measured with gratings and interferometers by methods outlined in the discussion of these instruments. Because such measurements are relatively difficult and tedious, they have been made only for the purpose of setting up standards of wavelength, and in all other spectroscopic work, whether with prisms or gratings, wavelengths are deduced by interpolation between adopted standards.

Wave Numbers.—In the interpretation of spectra as transitions between quantized energy states ($E_2 - E_1 = h\nu$), it is necessary to express the positions of the spectral lines on a frequency (ν) scale instead of a wavelength scale. Unfortunately, neither nature nor human ingenuity has disclosed a method of directly measuring the oscillation frequencies corresponding to light waves. These frequencies, defined as the number of waves that pass a given point in one second, must be derived from the relation $\nu = c/\lambda$, in which c is the velocity of light and λ is the observed wavelength. In order to make these frequencies independent of the medium through which the light passes, both c and λ must be evaluated *in vacuo*. Since the velocity of light is very great (about 29,979,300,000 cm. per second) and a light wave is very small (0.00000550 cm. for brightest visible light), the frequency is an extremely large and awkward number (545,000,000,000,000 in this example). In spectroscopy it has become common practice to replace the ungainly actual frequencies by wave numbers, representing the number of waves in one centimetre of vacuum and symbolized by cm.$^{-1}$. If the wavelength has been measured in air, it must be multiplied by the refractive index of air to change it to its value *in vacuo*. The National Bureau of Standards prepared a table of wave numbers for the convenient conversion of wavelengths in standard air to wave numbers *in vacuo*. For example, 5500.000 Å = 18,176.769 cm.$^{-1}$. Such wave numbers of observed spectral lines are strictly proportional to the natural frequencies and to the corresponding changes in atomic or molecular energy.

Standard Wavelengths.—In modern spectroscopy the primary standard of wavelengths is that of the red radiation from cadmium vapour first measured relative to the international metre bar by interferometer methods in a classical experiment by Michelson in 1892 (*see* PHYSICAL UNITS). Benoît, Fabry, and Pérot in 1905 found that 1 metre contained 1,553,164.13 of these waves, and the reciprocal of this number has been internationally adopted as the definition of the primary standard of wavelength: "the wavelength of the red ray of light from cadmium is 6,438.4696 angstroms in dry air at 15° C. on the hydrogen thermometer, at a pressure of 760 mm. of mercury, the value of g being 980.67 (45°)." The probable error of this absolute value is of the order of 1 part in 16,000,000; *i.e.*, somewhat less than the least error in the intercomparison of two metre bars.

The Eleventh General Conference of Weights and Measures, on Oct. 14, 1960, unanimously adopted a resolution defining the metre as the length equal to 1,650,763.73 wavelengths *in vacuo* of the radiation corresponding to the transitions between the $2p_{10}$ and $5d_5$ levels of the krypton-86 atom. The reciprocal of this number gives the wavelength 6057.80211 Å, and the angstrom is simply 10^{-10} m. That wavelength thus became the international standard of length, as well as the primary standard of wavelength.

A method of measuring, with high accuracy, the ratio of two different wavelengths was outlined in principle in the brief discussion of the Fabry-Pérot etalon interferometer. After 1907 this method was used in several laboratories to determine, relative to the primary standard, the wavelengths of selected lines that might serve as secondary standards. The wavelengths of several thousand atomic radiations characteristic of helium, neon, argon, krypton, xenon, iron, and titanium in laboratory spectra, and well over 1,000 lines in the sun's spectrum, have thus been measured, but only such mean values are accepted as secondary standards by the International Astronomical Union as have been independently and concordantly measured in at least three laboratories. For example, in the iron spectrum interferometric measurements have been made on thousands of lines with wavelengths ranging from 2101 Å to 10,216 Å, but only 306 secondary standards ranging from 2447 Å to 6677 Å have been adopted. These iron secondary standards are believed to be accurate within ±0.001 Å, but to insure their reproducibility the source is carefully specified as "the Pfund arc operated between 110 and 250 volts, with 5 amperes or less, at a length of 12–15 millimetres used over a central zone at right angles to the axis of the arc, not to exceed 1–1.5 millimetres in width, and with an iron rod 6–7 millimetres diameter as the upper pole and a bead of iron oxide as the lower pole."

It seemed certain in the 1960s that the iron standards would be superseded by superior thorium wavelengths emitted by electrodeless lamps excited by ultrahigh-frequency electromagnetic oscillations. In the absence of official standards in the extreme ultraviolet, spectroscopists have been forced to use preliminary

values obtained by the coincidence method of overlapping orders in grating spectra or by direct calculation from spectral series formulas or from spectral terms accurately determined from longer waves.

Interpolation of Wavelengths.—The wavelengths corresponding to lines in any spectrum whatsoever are readily determined by interpolation between standards, provided that the spectrum of the international iron arc is recorded with the same spectrograph. For this purpose the spectrum of the source under examination is photographed through one portion of the slit, and the arc spectrum of iron through one or both of the adjacent portions, so that two spectra in juxtaposition, or slightly overlapping,

SPECTRUM OF OXYGEN (MIDDLE) PLACED OVER THAT OF IRON TO COMPARE AND MEASURE WAVELENGTHS

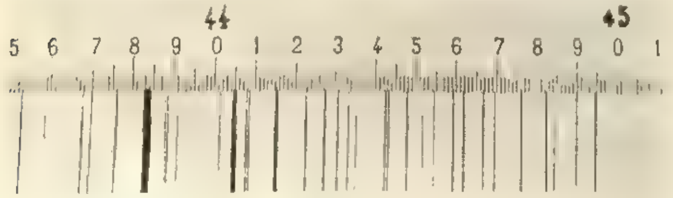
are obtained. For accurate determinations the relative positions of spectral lines and standards are measured with a micrometer or comparator consisting of an accurate screw designed to translate the spectrogram under a microscope so that the lines and standards are brought successively into coincidence with a reticle. The positions of the lines are then read from a scale that indicates complete turns of the screw, and a divided drumhead that registers fractions of a turn.

In prismatic spectra the dispersion increases rapidly as the wavelength diminishes, and linear interpolation will not give correct results. Approximate values of wavelengths may be obtained by interpolation on a graph of dispersion versus wavelength constructed by plotting scale readings of standard lines as abscissas and wavelengths as ordinates. The most accurate method of determining wavelengths in prismatic spectra is by mathematical interpolation with the Hartmann dispersion formula, which gives a simple relation between wavelength λ and linear scale reading r,

$$\lambda = \lambda_0 + \frac{c}{r - r_0}$$

where λ_0, c, and r_0 are constants that can be evaluated from the solution of three simultaneous equations containing three known wavelengths or standard lines and their corresponding scale readings.

Relative wavelengths are most readily obtained from diffraction spectra, since linear interpolation between standards is justified by the fact that the dispersion is constant (or nearly so) in a given grating order or setting. If r_i and r_f are millimetre scale readings for initial and final standards λ_i and λ_f on a spectrogram, the dispersion $d = (\lambda_f - \lambda_i)/(r_f - r_i)$ Å per mm. The product of this dispersion constant and r_i subtracted from λ_i gives a constant λ_o, which, on any calculating machine, can be automatically added to any dr_x, and the final result is $\lambda_x = \lambda_o + dr_x$. Both in prismatic and in grating spectrograms intermediate standards should be measured; in the former case to test the fit of the dispersion formula and in the latter to test the normality of the spectrum. If the calculated values deviate from the true standards, a plot of such deviations will reveal the proper corrections to apply to the wavelengths calculated for the other measured lines. Under the most favourable conditions such interpolated wavelengths are probably correct within a few thousandths angstrom. All wavelength measurements relative to iron standards are obviously facilitated by reference to a good photographic map of the iron-arc spectrum permitting instant identification of the standards.



PARTIAL PHOTOGRAPHIC MAP OF IRON-ARC SPECTRUM YIELDS APPROXIMATE WAVELENGTHS

Intensity Measurements.—The exact determination of relative intensities of the various parts or lines of a spectrum is in most cases as important as exact wavelength measurement, but this phase of spectroscopy had a tardy development, partly because of experimental difficulties and partly because the important applications were mostly unrecognized before 1925. Only heat-absorbing detectors show a linear and nonselective response to spectral intensity, and since they are typically inefficient heat engines, their use has been restricted mainly to infrared spectra. All other detectors, being nonlinear and selective, require calibration, but this is a complicated and difficult operation, especially in the extreme ultraviolet, where heat detectors are relatively insensitive. Consequently, even the most modern tables of identified spectral lines contain only rough estimates of line intensities on arbitrary scales, which differ from spectrum to spectrum and frequently from region to region in the same spectrum. These arbitrary intensity scales usually represent visual estimates of photographic blackening, and they range from 1 to 10 or 1 to 1,000 or 1 to 10,000. Such estimates have no definite meaning and no truly quantitative value.

After 1925 the need for accurate measurements of spectral-line intensities was stressed by the new quantum theory of atomic spectra and by a rebirth of quantitative spectrochemical analysis. Mathematicians predicted important relationships between line intensities and multiplet structures or quantum numbers, and experimental spectroscopists promptly devised methods of testing the theory. Simultaneously in applied spectroscopy quick methods of accurate chemical analysis based on the measurement of spectral-line intensity ratios were empirically developed. Since almost all spectroscopic observing was done photographically, the quantitative determination of intensities from spectrograms required the construction and use of suitable microphotometers or microdensitometers to measure the photographic densities of spectral-line images. By making a series of exposures of known ratio, either in intensity or in time, photographic density as a function of radiation intensity of a given wavelength can be determined for the particular type of photographic emulsion employed, and from density measurements at different wavelengths in a comparison spectrum of known energy distribution true intensity ratios can be deduced. Standard sources of known energy distribution are found in the blackbody or in metal-filament lamps operated at specified colour temperatures. Photographic sensitivity and contrast may be regarded as constant in narrow spectral regions, such as those encompassed by compact multiplets, hyperfine multiplets, Zeeman patterns, or close line pairs, and in these cases relative intensities of component lines are determinable without spectral energy calibration.

Absorption Spectra.—In 1859 Kirchhoff gave mathematical and experimental validity to the following law: "The ratio between the powers of emission and the powers of absorption for rays of the same wavelength is constant for all bodies at the same temperature." From this it follows that a transparent body cannot emit light, and one that radiates a continuous spectrum must be opaque. Further, a gas that radiates a line spectrum will absorb the lines that it radiates at the same temperature. For example, lines observed in laboratory arcs as emission spectra of 66 elements or compounds have been identified as absorption lines in the spectrum of the sun. As a consequence of Kirchhoff's law, practically everything that has been said about emission spectra applies also to absorption spectra. There are, however, some experimental conditions that permit molecular identification and quantitative chemical analysis by means of absorption spectra that cannot be duplicated in emission spectra.

Thus in the analysis of inorganic compounds (salts, minerals, etc.) by emission spectra, the light sources (flames, arcs, sparks) commonly used for exciting the spectra completely dissociate the compounds, revealing intense spectra of the metallic atoms and of unstable compounds formed in the light source, whereas the

spectra of the halogens or radicals are either unobserved or masked. Thus, essentially the same atomic emission spectrum is observed with various metallic compounds (halides, oxides, carbonates, sulfates, etc.). Similarly, organic molecules (tissues, dyes, etc.) are atomized by the heat and excitation of emission sources, revealing only spectra of metallic impurities that may be present, while the emission spectra of carbon, hydrogen, and oxygen are indistinguishable from those emitted by atmospheric constituents, dissociated carbon dioxide, and water vapour.

Often, however, inorganic compounds in solution and organic compounds in general exhibit characteristic absorption spectra that distinguish them from each other. Since absorption spectra of compounds are observed at relatively low temperatures, the molecules are stable and unimpaired, but since the substances are then usually in a liquid or solid state, rather than in a gaseous state, the absorption will depend on the energy content of the molecules and on the perturbing effects of surrounding molecules. These absorption spectra, therefore, usually consist of one or more broad bands in which the energy distribution is characteristic of the compound. If an unknown mixture contains two or more absorbing compounds, their identity and relative amounts can be determined if the absorption bands are sufficiently different in wavelength and accurately defined to permit an estimation of their relative intensities. Extraordinarily narrow absorption bands, almost linelike, are exhibited by solids, solutions, and glasses containing atoms that possess one or more *f*-type electrons (rare earths with atomic numbers 58 to 71 and 90 to 103).

ABSORPTION SPECTRA: NEODYMIUM GLASS (TOP); POTASSIUM VAPOUR (BOTTOM)

Narrow lines belonging to molecular spectra of oxygen (O_2) and water vapour (H_2O) are readily observed in absorption but not in emission. Red and infrared bands of these molecules invariably appear on spectrograms made in the laboratory with long-focus spectrographs provided that the light source emits a continuous background, and they are outstanding features of the solar spectrum, especially when the sun is near the horizon and shines through the thickest layer of terrestrial gas and vapour. Similarly, because of the great depth of absorbing vapours in the solar atmosphere, the atomic spectrum of iron is much more fully revealed in the Fraunhofer spectrum than has been observed in laboratory emission spectra. *See* also SPECTROPHOTOMETRY.

VI. APPLICATIONS

The principal applications of spectroscopy are to: (1) chemical and physical analysis; (2) research on atomic structure and the mechanism of radiation; and (3) investigations of some properties of atomic nuclei.

Spectrochemical Analysis.—That a system of qualitative chemical analysis could be based on spectroscopic observations was first convincingly demonstrated by Kirchhoff and Bunsen, who (1860) applied spectroscopy to the discovery of two alkali metals (Rb and Cs) and to the identification of chemical elements in the sun. Save for chemical analysis of meteorites that fall to earth, there was no way to learn anything concerning the chemical and physical conditions of celestial bodies except from an analysis and interpretation of the light those remote bodies emit. Consequently, the challenge and urge to apply spectroscopy to astronomy was irresistible, and an amazing accumulation of information on the radial motions, chemical compositions, and physical conditions of planets, comets, stars, and nebulae constituted the science of astrophysics in the 1960s.

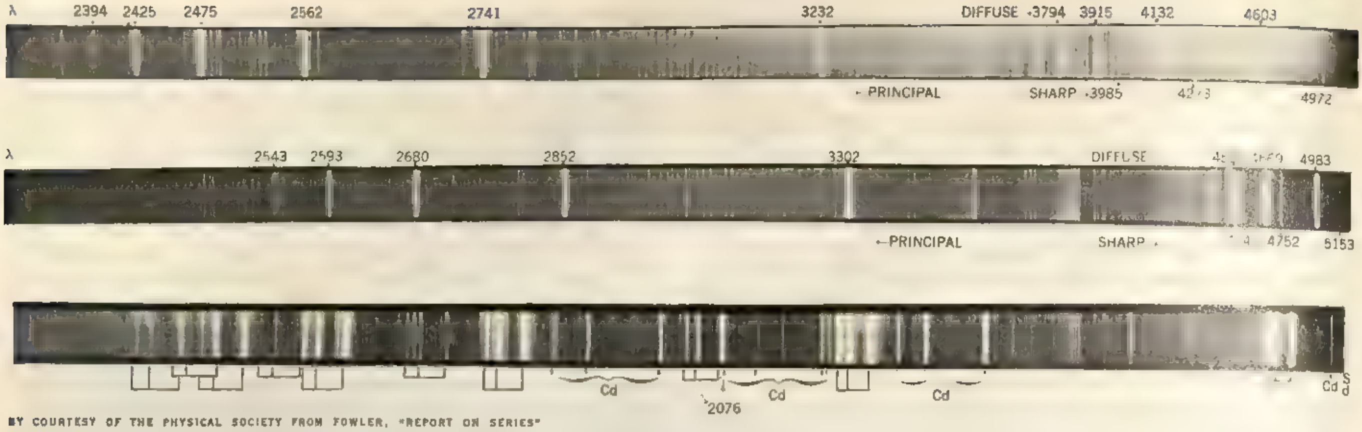
In terrestrial laboratories the application of spectroscopy to chemical analysis has always had to compete with conventional chemical methods; and progress was slow until precise photographic photometry was introduced about 1925, whereupon quantitative spectrochemical analysis quickly became common practice throughout the world. It is readily observed that when one element is progressively diluted in another, the spectra of the diluted element become progressively weaker and simpler until they finally vanish when the concentration is less than the limit of spectroscopic detection. Under specified conditions of observing, these partial or residual spectra may be correlated with the known composition of synthetic mixtures or standards and thus provide criteria for the quantitative estimation of comparable unknowns. This is essentially the principle of spectrochemical analysis as first applied to solutions by W. N. Hartley about 1880, to fused salts by A. de Gramont about 1900, and to metal alloys after about 1920. In 1925 W. Gerlach showed that unavoidable fluctuations in light sources and spectrum photography could be largely eliminated by referring the intensity of an impurity line to a neighbouring line of the major component as standard.

In the spectrum of any alloy or mixture of known composition, one or more pairs of lines can usually be found in which a strong (analysis) line of a minor constituent has the same intensity as a weak (internal standard) line of the major constituent. When such homologous pairs are found in spectrograms of samples containing different but known amounts of the analysis element, the concentration of this element in other samples is found by intercomparing the line pairs and selecting the one in which the two lines are of equal intensity. A further improvement of this internal-standard method resulted from the use of only one pair of lines the relative intensities of which in known samples are determined by photometric measurements and plotted as a function of concentration of the analysis element. If the two variables are plotted on logarithmic scales, the resulting analytical curve is usually a straight line with slope about one. After the preparation of such a calibration graph for each analysis element, any sample of comparable composition is easily analyzed by measuring line-intensity ratios and reading concentrations from the graphs. An excellent description of the preparation of samples and standards, spark excitation of spectra, and analytical procedure can be found in *Spark Spectrographic Analysis of Commercial Tin*, published (1942) as Research Paper 1,451 of the National Bureau of Standards.

Similar procedures have been perfected for the analysis of iron alloys, aluminum alloys, magnesium alloys, brasses, bronzes, type-metal, and practically all other types of metallic compounds. Careful chemical calibration and photographic photometry have reduced routine analytical errors to 1% or 2% of the amount present, and the time required for a complete analysis of a sample has been reduced to ten minutes or less. Since most of the time required for such an analysis is consumed in procuring and measuring a photographic spectrogram, elaborate spectrographs have been constructed with 55 photoelectron multiplier tubes to receive light from an internal standard line and from 54 different analysis lines. Radiant energies received by these phototubes are amplified to actuate indicators calibrated in percentage composition, thus permitting simultaneous and almost instantaneous determination of 20 to 35 elements. For composition control of metals and alloys in production, spectroscopy has surpassed chemistry in speed, sensitivity, and accuracy of analysis. It must be remembered, however, that no satisfactory absolute method of spectrochemical analysis had been devised; all the practical methods in the 1960s were relative to synthesized standards or chemically analyzed samples that served for spectrographic calibration.

Emission spectroscopy has also found extensive application to analytical problems in agricultural, biological, and pathological chemistry, in mineralogy and geochemistry, and in archaeology and criminology. In these problems the samples are often in liquid or powder form that can be vaporized and excited to emit spectra either in hot flames or in electric arcs or sparks between electrodes of pure carbon. Usually a definite amount of some suitable element is added to supply internal standards, and then the calibration and analytical procedures are essentially the same as for solid metal samples.

Empirically, it is found that, operating with light sources at atmospheric pressure and with spectral detectors sensitive between



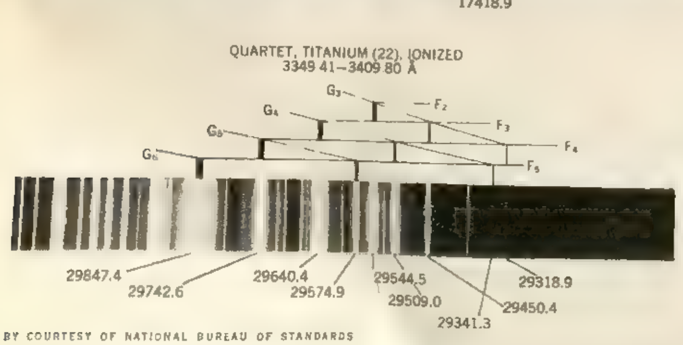
BY COURTESY OF THE PHYSICAL SOCIETY FROM FOWLER, "REPORT ON SERIES"

(ABOVE) ARC SPECTRUM OF LITHIUM SHOWS PRINCIPAL, SHARP, AND DIFFUSE SERIES; (CENTRE) ARC SPECTRUM OF SODIUM; (BELOW) ARC SPECTRUM OF ZINC AND CADMIUM SHOWS TRIPLET SYSTEMS

2000 and 9000 Å, at least 70 chemical elements can be readily determined by emission spectroscopy. This is related to the fact that these elements have ionization potentials less than ten volts and their strongest lines have wavelengths between the limits mentioned. Gaseous elements have higher ionization potentials; their strongest lines are in the far ultraviolet, and their spectra are fully developed only at reduced pressure in Geissler tubes. In principle, every one of the 103 chemical elements known in the 1960s can be detected and determined by means of emission spectra (*see* SPECTROCHEMICAL ANALYSIS).

Structure of Spectra and of Atoms.—Because atomic spectra appear, generally, to consist of a random distribution of lines of different intensities, a century elapsed between the discovery of spectral lines by Fraunhofer and a fruitful clue to their physical interpretation by Niels Bohr (1913). Then in little more than a decade an amazing development in analysis and theory of spectral structure made all spectra interpretable in terms of atomic properties and, conversely, permitted the deduction of atomic structure from observations on spectra.

Hydrogen is the lightest atom and possesses the simplest spectrum; in the visible it consists of a regular succession of lines with ever closer spacing and decrement of intensity in the direction of shorter wavelengths. In 1885 J. J. Balmer discovered an extremely simple mathematical relationship between these H lines. Expressed in wave numbers the Balmer formula gives $\nu = R/2^2 - R/n_1^2$, where the constant $R = 109{,}678.76$ cm.$^{-1}$, and n_1 represents the integers 3, 4, 5 . . . ∞. This Balmer series begins with 6562 Å and ends at 3646 Å. Similar H series were later found in the extreme ultraviolet, $\nu = R/1^2 - R/n_1^2$, by T. Lyman, and in the infrared, $\nu = R/3^2 - R/n_1^2$, by Paschen. All possible H series can be accurately represented by one formula, $\nu = R/n_2^2 - R/n_1^2$, where n_2 and $n_1 > n_2$ are integers and n_2 is constant for a given series. Thus, the wave number of any H line can be expressed as the difference of two quantities, the larger being a constant term called the limit of the series, whereas the other varies from line to line and is known as the running term.



BY COURTESY OF NATIONAL BUREAU OF STANDARDS

MULTIPLETS OF NEUTRAL AND IONIZED ATOMS WITH SIX-PLACE WAVE NUMBERS FOR SOME LINES: ALTERNATION LAW ILLUSTRATED VERTICALLY, DISPLACEMENT LAW HORIZONTALLY

In 1889 J. R. Rydberg reported finding series in the spectra of alkali and alkaline earth elements, and, notwithstanding the imperfect data then at his disposal, he discovered most of the important properties of spectral series. From certain characteristics of the lines, he distinguished three species of series (principal, sharp, and diffuse) superposed in the same spectrum, and he recognized that the members of each series were double for alkalis (Li, Na, K) and either single or triple for alkaline earths (Mg, Ca, Zn). Rydberg found that each series could be represented in wave numbers by a formula of the type $\nu = L - \frac{R}{(m+\mu)^2}$, in which L is the limit of the series, R the Rydberg constant (practically identical with R in the Balmer formula), and m has successive integral values to which a constant fractional part μ is added. A fourth type of series was discovered in 1907 by A. Bergmann, and because it was more hydrogenlike ($\mu \approx 0$) than the others, it came to be called fundamental. Rydberg also indicated that the difference between the limit of the P series and the common limit of the D and S series is equal to the wave number of the first line of the P series, and C. Runge in 1908 pointed out a similar relation between the D and F series, so that all the different series were interconnected. An abbreviated notation was then adopted for the four basic types of series:

Principal	$\nu = n\,\mathrm{S} - m\,\mathrm{P}$
Sharp	$\nu = n\,\mathrm{P} - m\,\mathrm{S}$
Diffuse	$\nu = n\,\mathrm{P} - m\,\mathrm{D}$
Fundamental	$\nu = n\,\mathrm{D} - m\,\mathrm{F}$

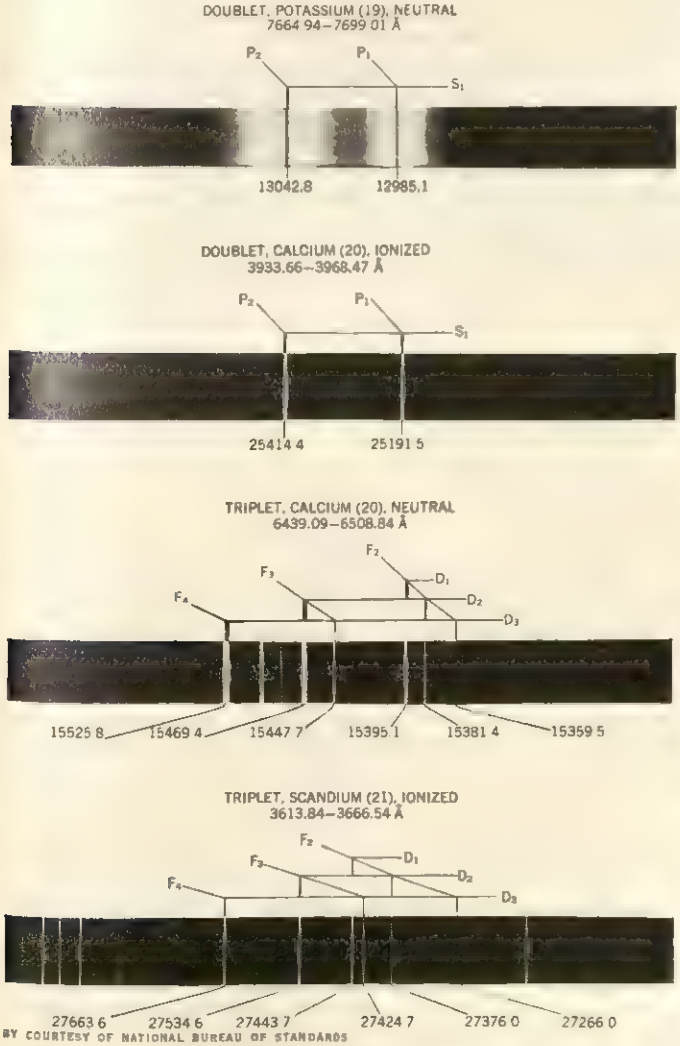
where n stands for a simple integer, and m for appropriate integers plus μ. In order to distinguish the components of doublets and triplets from each other, it was necessary to add subscripts to these term symbols. In addition to the regular series, others were detected when the limits were changed and still others between singlet and triplet terms (Ritz combination principle). Thus, it appeared that any given term might combine with many others, and although two terms were necessary for each observed line, the total number of terms required to represent any spectrum completely might be much smaller than the total number of lines.

In 1914 A. Fowler found ionized helium atoms (He^+) characterized by a hydrogenlike series except that the Rydberg constant R required multiplication by 4. He also discovered, in the spark spectra of Mg^+, Ca^+, and Sr^+, series of doublets resembling the P, S, and D doublets of alkalis but with 4 R in their formulas. By

1919 the regularities detected in various spectra sufficed to suggest two general laws: (1) the alternation law, which states that even and odd multiplicities of terms alternate in successive columns of the periodic system; and (2) the displacement law, which states that the spectrum of an ionized element resembles that of the neutral element preceding it in the periodic system. Term multiplicities of atoms or ions are thus determined solely by the number of electrons, whereas the atomic charge affects the position of the spectrum. These facts suggested a direct connection between spectral lines and the electrons and protons in atoms.

In 1913 Bohr proposed two fundamental postulates, viz., that atomic stationary states exist, and that the radiation frequency is equal to the energy difference of two states divided by Planck's constant ($h = 6.6256 \times 10^{-27}$ erg sec.). The empirically established spectral terms must, therefore, be interpreted as quantized stationary energy states characteristic of atoms and ions. The rapid development of the quantum theory of radiation constitutes one of the most thrilling chapters in the history of science (*see* QUANTUM MECHANICS).

Quantum Numbers and Atomic Energies.—Omitting further review of historical and experimental details, the following discussion is restricted to a vector model of the Rutherford-Bohr atom (*see* ATOM), which is assumed to consist of a tiny but massive nucleus (composed of protons and neutrons) around which circulate one or more electrons. The maximum number of electrons in any atom is expressed by its atomic number Z (identical with the number of protons in its nucleus), and this number ranges from 1 for H, the lightest, to 103 for Lw, the heaviest known in the 1960s. These electrons move in imaginary orbits about the nucleus at average distances represented by a principal quantum number n, which in turn numbers the successive shells of the periodic system and serves as a coefficient to the term symbol S, P, D, F. . . . A sequence of n term values subtracted from a limiting term gives the wave numbers of lines in a spectral series. In most cases an electron jump from one shell or value of n to any other shell or value of n constitutes the largest change in energy that can occur in a given atom. If an electron is moved from its lowest value of n to $n = \infty$, the atom is ionized, and the energy required to accomplish the removal of an electron is called the ionization potential. This energy is usually expressed either in wave numbers (cm.$^{-1}$) or in electron volts (ev); in the case of H it is 109,678.76 cm.$^{-1}$ (the Rydberg constant) or 13.595 ev. (*See* Table VII.) Hydrogenlike spectra are also emitted by ionized helium He^+, by doubly ionized lithium Li^{2+}, by trebly ionized beryllium Be^{3+}, etc., but the spectra are shifted successively toward higher frequencies because the energy of electron orbits as a function of n is proportional to Z^2/n^2. It follows that for He^+ the separation energy of the remaining electron from the lowest level (the ionization potential) will be 4 times that for the H atom; for Li^{2+} it will be 9 times as great and for Be^{3+} 16 times.



BY COURTESY OF NATIONAL BUREAU OF STANDARDS

MULTIPLETS

The next largest change in atomic energy (after that due to a change in n) is usually that associated with a constant orbital angular momentum of the optical electron. This momentum is pictured as a vector of magnitude 0, 1, 2, 3 . . . in $h/2\pi$ units; it is called the azimuthal quantum number and is symbolized by l. The numerical values of $l = 0, 1, 2, 3$ . . . correspond respectively to the empirical term symbols S, P, D, F . . . originally abbreviated from the different types of series first found in the spectra of alkali atoms. Electrons with $l = 0$ are called *s*-electrons; those with $l = 1$, *p*-electrons; those with $l = 2$, *d*-electrons; and those with $l = 3$, *f*-electrons. It may be interjected here that these four l values and the first seven n values suffice to describe the normal electron configurations of all possible atoms and ions. When two or more optical electrons are present, their individual orbital momenta, l_1, l_2, . . . are added vectorially to form a resultant L, but quantum theory dictates that such a resultant, as well as the individual l, can only exist as an integral multiple of $h/2\pi$. The individual l_i can, therefore, be oriented only in certain discrete directions to one another. For the case of two electrons with l_1 and l_2, the possible resultant L values are: $L = (l_1 + l_2)$, $(l_1 + l_2 - 1)$, $(l_1 + l_2 - 2)$ . . . $|l_1 - l_2|$. The vector addition for this case is shown graphically in fig. 13. If all but one of the l_i are zero, the resulting L will be identical with that one l_i. When there are three electrons for which $l \neq 0$, the vector addition is simply performed by combining the l values of two electrons and then combining each of the resulting L values with the l of the third electron. The term types given by various simple electron configurations are shown in Table I.

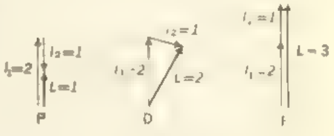


FIG. 13.—VECTOR ADDITION OF ORBITAL ANGULAR MOMENTA

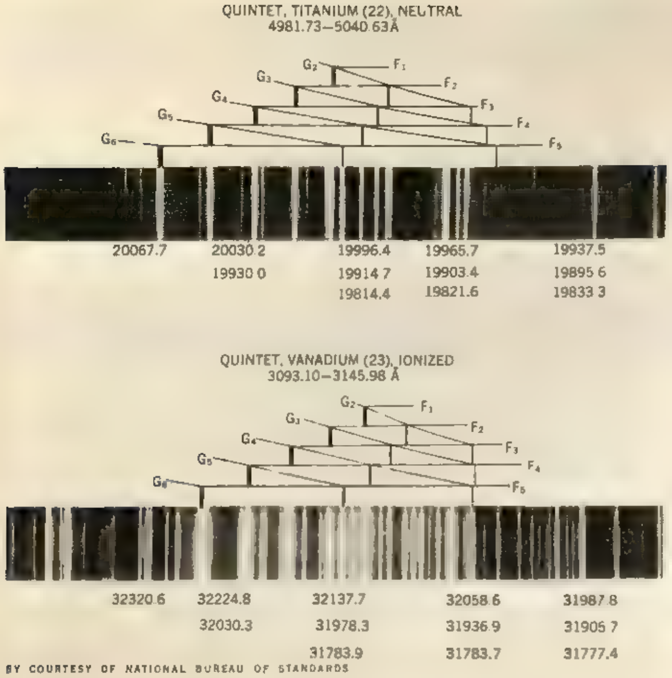
$l_1 = 2$ and $l_2 = 1$ to give resultants $L = 3, 2, 1$

Quantum theory and observation show that an important selection rule controls atomic energy changes resulting in radiation, viz., the greatest intensity of emission results when $\Delta L = \pm 1$. In addition, when the interaction of the electrons is small, only

TABLE I.—*L Values and Term Symbols Arising from Various Simple Electron Configurations*

Electron configurations	L	Term symbols
ss	0	S
sp	1	P
pp	0 1 2	S P D
pd	1 2 3	P D F
dd	0 1 2 3 4	S P D F G
df	1 2 3 4 5	P D F G H
ff	0 1 2 3 4 5 6	S P D F G H I
ppp	0 1 1 1 2 2 3	S P P P D D F

those quantum transitions take place for which only one of the electrons alters its l value in accordance with the selection rule $\Delta l = \pm 1$. Notwithstanding these simple rules, many spectra exhibit lines corresponding to $\Delta L = 0, \pm 2, \pm 3$, as well as to double-electron jumps.



BY COURTESY OF NATIONAL BUREAU OF STANDARDS

MULTIPLETS

A third contribution to the total energy of an electronic system comes from the rotation of each electron about its own axis. This axial angular momentum has the same magnitude for each electron, $s = \frac{1}{2}(h/2\pi)$; it is usually called the electron spin. When two or more electrons are present, the individual spin vectors s_i combine with each other to yield a resultant S (just as the l_i combine to form L). Analogous to L, the resultant spin S can take only certain discrete values, the maximum being obtained when all the s_i are parallel, and the minimum being one-half or zero according as the number of electrons is odd or even.

The spinning electrons and their resultant spin vectors S account for the splitting of most spectral terms into two or more components and provide a physical meaning for the subscripts (inner quantum numbers) originally attached to components of polyfold terms. The total angular momentum J of an atom is the vector sum of the resultant orbital angular momentum L and the resultant axial angular momentum S. According to quantum theory, L and S can be oriented to each other only in certain directions, and, therefore, only certain discrete values of the resultant J are possible. The largest and smallest values of J result from simple addition and subtraction of L and S, while all inter-

TABLE II.—*Resultant S Values and Possible Multiplicities*

Number of electrons	S	Possible multiplicities r
1	1/2	doublets
2	0, 1	singlets, triplets
3	1/2, 3/2	doublets, quartets
4	0, 1, 2	singlets, triplets, quintets
5	1/2, 3/2, 5/2	doublets, quartets, sextets
6	0, 1, 2, 3	singlets, triplets, quintets, septets
7	1/2, 3/2, 5/2, 7/2	doublets, quartets, sextets, octets
8	0, 1, 2, 3, 4	singlets, triplets, quintets, septets, nonets

mediate values that differ by integral amounts are allowed; *i.e.*,

$$J = (L+S),\ (L+S-1),\ (L+S-2),\ \ldots\ |L-S|$$

This rule is illustrated graphically for simple cases in fig. 14. When $L > S$, it is obvious that the number of permitted J values for a given value of L is $2S+1$. On the other hand, if $L < S$, the number of permitted J values for a given L is $2L+1$. Thus, for S terms ($L = 0$, $2L+1 = 1$) there is only one value of J; *i.e.*, $J = S$. Similarly, for P terms ($L = 1$, $2L+1 = 3$) the number of components can never exceed three. However, for a great majority of terms $L > S$, and for all these the possible number of components or term multiplicity r is equal to $2S+1$. The empirically indicated alternation law follows directly from this result, since the maximum multiplicity will be even or odd according as the number of electrons is odd or even.

In general, atomic energy states with different S are not expected to combine with one another because of a selection rule $\Delta S = 0$, but intersystem combinations are found to be fairly abundant and intense, especially as the atomic number increases. Table II displays the resultant S values and possible multiplicities for different numbers of electrons.

The total angular momentum of an atom can have integral or half integral $J(h/2\pi)$ values because $s = \frac{1}{2}(h/2\pi)$ for each electron. Furthermore, different sets of J values are associated with the same type of term (same L value) in systems of different multiplicity because S is different. And finally, different sets of J values apply to different terms in the same multiplicity system because $J = L + S$. Since J represents the total angular momentum, it is the most important quantum number for the fine structure or multiplet analysis of atomic and ionic spectra. It has a perfectly definite physical meaning and quantitative value under all circumstances. Atomic energy changes are rigorously regulated by the simple selection rule that $\Delta J = 0, \pm 1$, except that all 0 to 0 transitions are prohibited. Since the a priori probability that an atomic energy state exists is expressed by its statistical weight $2J+1$, it is obvious that J is important for the relative intensities of spectral lines. The J values that belong to a few different spectral terms in different multiplicity systems are shown in Table III.

For any given spectrum in which energy levels have been established, relative values of J attached to the levels are readily determined from their combining properties and the selection rule $\Delta J = 0, \pm 1$. In spectra resulting from terms of odd multiplicity the absolute value of J is fixed by the absence of the transition $J = 0$ to $J = 0$, which is forbidden. In other cases, the absolute value of J can be deduced from the multiplicity and type of term if these are unambiguous as, for example, in multiplets that are sufficiently law abiding so that the terms can be identified from the sum rule (the sum of the intensities of all the lines of a multiplet that belong to the same initial or final state is proportional to the statistical weight $2J+1$ of the initial or final state, respectively) or from the interval rule (the interval between two successive components, J and $J+1$, of a polyfold term is proportional to $J+1$). Analysis of the more complex spectra is usually hopeless without the aid of Zeeman effect; *i.e.*, the splitting of the

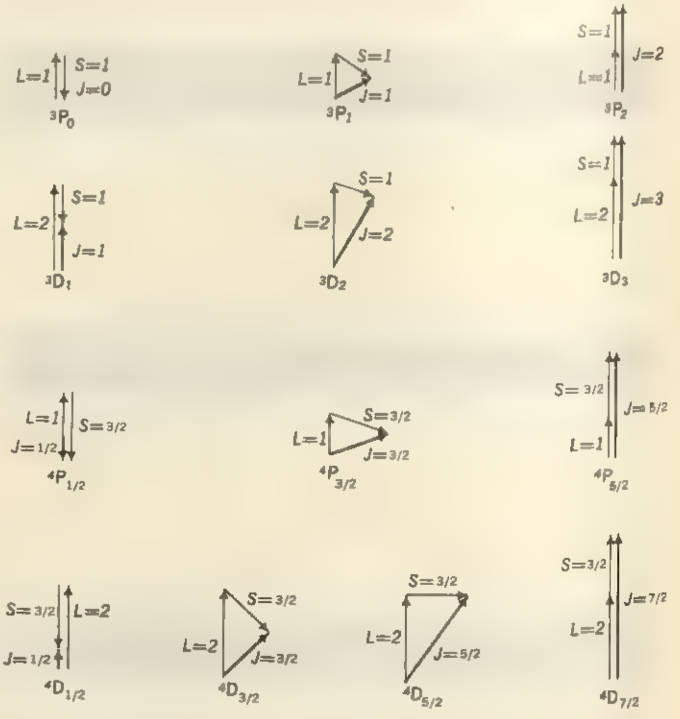


FIG. 14.—VECTOR ADDITION OF ORBITAL MOMENTUM L AND AXIAL MOMENTUM S FORMING RESULTANTS J FOR ^3P, ^3D, ^4P, AND ^4D TERMS

Table III.—*J Values for Terms of Odd and Even Multiplicity*

Terms	Singlet	Triplet	Quintet	Septet
	Odd multiplicities			
S	0	1	2	3
P	1	0 1 2	1 2 3	2 3 4
D	2	1 2 3	0 1 2 3 4	1 2 3 4 5
F	3	2 3 4	1 2 3 4 5	0 1 2 3 4 5 6
G	4	3 4 5	2 3 4 5 6	1 2 3 4 5 6 7
H	5	4 5 6	3 4 5 6 7	2 3 4 5 6 7 8

Terms	Doublet	Quartet	Sextet	Octet
	Even multiplicities			
S	1/2	3/2	5/2	7/2
P	1/2 3/2	1/2 3/2 5/2	3/2 5/2 7/2	5/2 7/2 9/2
D	3/2 5/2	1/2 3/2 5/2 7/2	1/2 3/2 5/2 7/2 9/2	3/2 5/2 7/2 9/2 11/2
F	5/2 7/2	3/2 5/2 7/2 9/2	1/2 3/2 5/2 7/2 9/2 11/2	1/2 3/2 5/2 7/2 9/2 11/2 13/2
G	7/2 9/2	5/2 7/2 9/2 11/2	3/2 5/2 7/2 9/2 11/2 13/2	1/2 3/2 5/2 7/2 9/2 11/2 13/2 15/2
H	9/2 11/2	7/2 9/2 11/2 13/2	5/2 7/2 9/2 11/2 13/2 15/2	3/2 5/2 7/2 9/2 11/2 13/2 15/2 17/2

Table IV.—*Possible States of an Electron in an Atom*

	K	L				M									N															
n	1	2				3									4															
	0	0	1			0	1			2					0	1			2					3						
	s	*s*	*p*			*s*	*p*			*d*					*s*	*p*			*d*					*f*						
	0	0	−1	0	+1	0	−1	0	+1	−2	−1	0	+1	+2	0	−1	0	+1	−2	−1	0	+1	+2	−3	−2	−1	0	+1	+2	+3
	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅	⇅

lines into polarized components when the source of light is placed in a strong magnetic field. The total angular momentum J is space quantized in a magnetic field and can assume $2J + 1$ discrete values, which are differentiated by magnetic quantum numbers M. The magnitude of the level splitting is the same for all singlet levels (normal Zeeman effect) but differs for all other terms according to the values of L and J (anomalous Zeeman effect). From any completely resolved Zeeman pattern the J values of the two combining levels are deduced at once from the number of components, and the L values are simultaneously derived from the separations of the components, since the splitting factors have been accurately given empirically and by the quantum theory (*see* Zeeman Effect).

Summarizing the above discussion of quantum numbers recalls that any atomic energy level (spectral term) is completely specified by four quantities: (1) its principal quantum number n, which indicates the electron shell or period that the optical electron occupies; (2) its type—S, P, D, F . . . where the capital letters stand for azimuthal quantum numbers or orbital angular momenta $L = 0, 1, 2, 3$ . . . respectively; (3) its inner quantum number or total angular momentum J (written as a suffix or subscript to the term type symbol); and (4) its multiplicity number $r = 2S + 1$, written as a superior prefix to the term type symbol. Thus 4^3D_2 represents the middle component ($J = 2$) of a D term ($L = 2$) in a triplet system ($r = 3$) for which the principal quantum number $n = 4$. Theoretically, the energy difference of two terms appears as intense radiation only when $\Delta L = \pm 1$, $\Delta J = 0, \pm 1$, $\Delta r = 0$, and the two terms are of opposite parity; *i.e.*, odd and even (Laporte rule). Typical energetic relationships of spectral terms and combining properties of the levels are shown graphically in fig. 15, which represents the spectrum of neutral mercury atoms.

Pauli Exclusion Principle.—Having found an electronic-quantum interpretation of spectral terms, it is next in order to seek the connection between spectral terms and atomic structure. In order to understand the development of the periodic system (*see* Periodic Law) and to account for striking periodicities in atomic properties, it is necessary to introduce a new assumption, known as the Pauli exclusion principle, which states that: "in one and the same atom, no two electrons can have the same set of values for the four quantum numbers, n, l, m_l, and m_s." Table IV shows the electron shells and subgroups that are possible for atoms as a consequence of this exclusion principle. The n cells are divided by l, and the l cells are further divided by m_l and by m_s, but, instead of inserting the last quantum numbers, the presence of an electron in a (n, l, m_l) cell is represented by arrows pointing up or down according as m_s is $+\frac{1}{2}$ or $-\frac{1}{2}$. Only two electrons can be in each such cell, and then only when they have antiparallel spin directions, since otherwise these two electrons would have the same four quantum numbers n, l, m_l, m_s. The maximum number of electrons that can have the same n and l is given by the number of arrows in the corresponding cells. This number is equal to $2(2l + 1)$, since $2l + 1$ is the number of possible m_l values

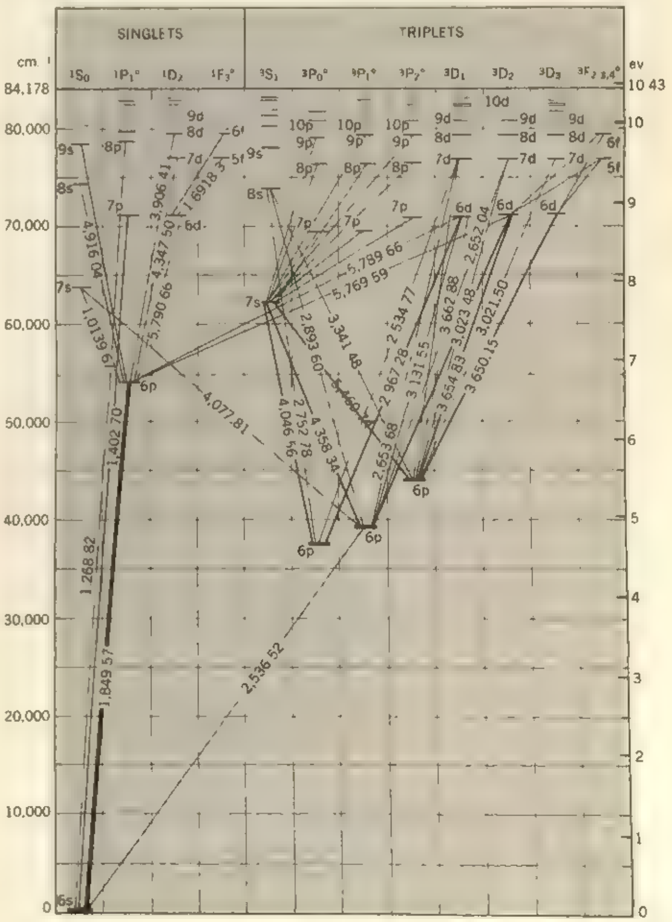


FIG. 15.—ENERGY-LEVEL DIAGRAM OF THE FIRST SPECTRUM OF MERCURY

TABLE V.—*Terms of Nonequivalent Electrons*

Electron configurations	Terms (omitting J values)
ss	^1S, ^3S
sp	^1P, ^3P
sd	^1D, ^3D
pp	^1S, ^1P, ^1D, ^3S, ^3P, ^3D
pd	^1P, ^1D, ^1F, ^3P, ^3D, ^3F
dd	^1S, ^1P, ^1D, ^1F, ^1G, ^3S, ^3P, ^3D, ^3F, ^3G
df	^1P, ^1D, ^1F, ^1G, ^1H, ^3P, ^3D, ^3F, ^3G, ^3H
sss	^2S, ^2S, ^4S
ssp	^2P, ^2P, ^4P
ssd	^2D, ^2D, ^4D
spp	^2S, ^2P, ^2D, ^2S, ^2P, ^2D, ^4S, ^4P, ^4D
spd	^2P, ^2D, ^2F, ^2P, ^2D, ^2F, ^4P, ^4D, ^4F
etc.	

for a given l. Thus, in any one shell no more than 2 s electrons (s^2), 6 p electrons (p^6), 10 d electrons (d^{10}), and 14 f electrons (f^{14}) can exist, and summing these numbers successively gives 2, 8, 18, and 32, which are precisely the totals of elements found in successive groups of atoms exhibiting periodic properties when arranged in atomic number sequence. Naturally the electron configurations of atoms in their lowest possible energy states, and the corresponding spectral terms, will show a periodicity, since, after a certain number of electrons of a given type have been added, the outermost electron will be once more, for example, an s-electron.

The spectral terms describing ground states of atoms may be deduced from assumed electron configurations by adding together the l and s moments of the individual electrons, but first it is necessary to make an assumption about the mutual coupling or interaction of the individual vectors. The coupling that seems to be most common is called LS coupling, expressed as

$$(l_1, l_2, \ldots) \ (s_1, s_2, \ldots) = (L, S) = J$$

This means that the individual l_i vectors are strongly coupled to produce resultant L values of different energies, and the individual s_i vectors are also strongly coupled to produce resultant S values of different energies. The resultants L and S are then less strongly coupled with one another, and their resultant is J. Other types of electron coupling, in particular jj coupling, occur relatively rarely, but coupling intermediate between LS and jj is found more frequently, especially among excited states of heavy elements. However, because LS coupling applies to all the lighter elements, predominates in many others, and is either accurately or very approximately valid for the ground states of all atoms and ions, it forms the basis for the standardized notation for spectral terms.

The spectral terms arising from nonequivalent electrons (belonging to different n, l groups of Table IV) may be obtained in simple cases by adding the L values of Table I to the S values of Table II. For example, two nonequivalent p electrons pp yield ($L = 0, 1, 2$) S, P, D terms, and since the electron spins can be either antiparallel or parallel ($S = 0, 1$), there will be singlet and triplet terms ($r = 2S + 1$); *i.e.*, pp always produces a family of terms ^1S, ^1P, ^1D, ^3S, ^3P, ^3D. Examples of the term families invariably associated with certain pairs or triads of nonequivalent electrons are given in Table V.

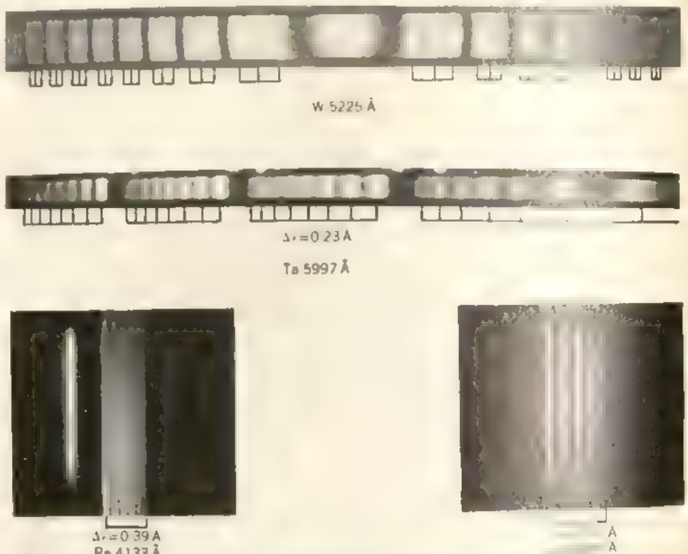
When two electrons are equivalent, *i.e.*, have the same n and the same l, they must at least differ in their values of m_l or m_s. For example, when two equivalent p-electrons (p^2) have l with the same direction (giving a D term), m_l is the same ($+1, 0, -1$) for both, and, therefore, both cannot have $m_s = +\frac{1}{2}$, or both $m_s = -\frac{1}{2}$. This means that their spins can only be antiparallel for a D term, thus producing ^1D but not ^3D. Altogether two equivalent p-electrons (p^2) can produce only the terms ^1S, ^1D, ^3P, whereas

TABLE VI.—*Terms of Equivalent Electrons*

Electron configurations	Terms (omitting J values)
s^2	^1S
p^2	^1S, ^1D, ^3P
p^3	^2P, ^2D, ^4S
d^2	^1S, ^1D, ^1G, ^3P, ^3F
d^3	^2P, $^2D(2)$, ^2F, ^2G, ^2H, ^4P, ^4F
d^4	$^1S(2)$, $^1D(2)$, ^1F, $^1G(2)$, ^1I, $^3P(2)$, ^3D, $^3F(2)$, ^3G, ^3H, ^5D
d^5	^2S, ^2P, $^2D(3)$, $^2G(2)$, ^2H, ^2I, ^4P, ^4D, ^4F, ^4G, ^6S
f^2	^1S, ^1D, ^1G, ^1I, ^3P, ^3F, ^3H
f^3	^2P, $^2D(2)$, $^2F(2)$, $^2G(2)$, $^2H(2)$, ^2I, ^2K, ^2L, ^4S, ^4D, ^4F, ^4G, ^4I
etc.	

two nonequivalent p-electrons (pp) yield ^1S, ^1P, ^1D, ^3S, ^3P, ^3D. Other examples of terms identified with equivalent electrons are displayed in Table VI.

An important consequence of the Pauli principle is that closed shells (in which the maximum number of equivalent electrons is present) have $L = 0$ and $S = 0$ and, therefore, always present the $^1S_0$ state. Closed shells can, therefore, be ignored when deriving the terms given by any electron configuration. Furthermore, $^1S_0$ for a closed shell must result when the shell is divided into two parts, the term types for each part derived, and the resulting angular momentum vectors added together. For example, adding the angular momenta of p^2 electrons to the corresponding quantities for p^4 electrons will give the resultant for p^6 electrons; *i.e.*, $^1S_0$. From this it follows L and S must be the same for these two electron configurations, and the terms arising from p^4 are the same as those from p^2. Similarly, d^6 gives the same terms as d^4, f^{12} the same as f^2, etc. In general, any subgroup lacking one or more electrons to fill the group behaves spectroscopically as if the lacking electrons alone were present, except that the terms are regular (smallest J level has least energy) when the group is less than half filled but inverted when more than half filled.



(TOP) REPRODUCED BY PERMISSION FROM "INTRODUCTION TO ATOMIC SPEC... E. WHITE, COPYRIGHTED, 1934, BY THE MCGRAW-HILL BOOK COMPANY, INC., (BOT... COURTESY OF NATIONAL BUREAU OF STANDARDS

(TOP) FABRY-PEROT INTERFEROMETER PATTERN OF TUNGSTEN LINE SHOWS COMPONENTS FOR THREE PRINCIPAL ISOTOPES, W[182], W[184], W[186]. PATTERN OF TANTALUM LINE SHOWS EIGHT COMPONENTS OF HYPERFINE STRUCTURE; (BOTTOM) PLANE-GRATING IMAGES OF RHENIUM LINES SHOW SIX COMPONENTS OF HYPERFINE STRUCTURE

Finally, of the terms given by equivalent electrons, those with greatest multiplicity generally lie lowest, and of these the lowest is that with the greatest L. In Table VI the last listed term is the lowest energy in each family.

Normal States and Electron Configurations of Atoms.—Quantum principles having thus specified the types of spectral terms arising from certain electrons, it became theoretically possible in 1925 to determine from identified spectral terms the electron configurations of all atoms and their ions. By 1964 the ground states of 96 atoms had been uniquely determined from spectral structure, and since these data are of great importance in spectroscopy, atomic physics, and chemistry, they are collected in Table VII. In addition, Table VII lists all the spectral multiplicities experimentally detected, all the reported spectroscopic ionization potentials, and the wavelengths of the strongest spectral lines characteristic of neutral atoms. In general, the strongest lines result from $s \leftrightarrow p$ electron transitions.

Hyperfine Structure of Spectral Lines and Properties of Atomic Nuclei.—When examined with apparatus of adequate resolving power, many spectral lines have been found to consist of two or more components lying extremely close together. This hyperfine structure arises from properties of the atomic nucleus

with which optical electrons interact more or less according to their orbits. The actual line splitting is exceedingly small (usually a fraction of a wave number) but nevertheless of very great importance because it provides a means of obtaining quantitative information about certain properties of isotopic nuclei. The spectroscopic influence of a nucleus may arise from its mass or from an intrinsic angular momentum or nuclear spin, analogous to electron spin. The former is called isotope effect and the latter nuclear-spin hyperfine structure, or *hfs* for short.

Isotope Effect.—Many natural elements consist of a mixture of two or more isotopes, each of which has an approximately whole number atomic weight. Different isotopes of an element have the same number and arrangement of outer electrons and, consequently, have the same multiplet structure for their spectra. They are, however, distinguished from each other by their mass because an electron necessarily revolves about the common centre of gravity of itself and the nucleus, and the Rydberg constant, therefore, depends on the nuclear mass. Thus, the Balmer lines of $1H^1$ atoms are accompanied on the short-wave side by very faint companions ascribed to $1H^2$, and the detection of these isotope-effect components in 1932 constituted the discovery of heavy hydrogen. Similar isotope effects have been observed in many other spectra, but as soon as several electrons are present, the simple mass effect observed in the hydrogen spectra is complicated by other factors that were not completely understood in the 1960s.

Nuclear-Spin Hyperfine Structure.—True *hfs* can be simply and quantitatively explained if it is assumed that the nucleus of the emitting atom possesses an angular momentum and an associated magnetic moment. According to quantum theory, the nuclear spin can be only an integral or half-integral multiple of $h/2\pi$. It is written $I(h/2\pi)$, where I is the spin vector or quantum number of the nuclear angular momentum. The spectrum of the H_2 molecule shows that for the simplest atomic nucleus (the proton) $I = \frac{1}{2}$. The proton thus has the same mechanical moment as the electron, but various values may be expected for heavier nuclei, since they contain more than one proton. As with electron spin, a magnetic moment is associated with the nuclear spin, since the nucleus is electrically charged and the rotation of electrically charged particles gives rise to a magnetic moment. Because a proton is the same size as an electron but 1,840 times heavier, the magnetic moment of an atomic nucleus is always 1,000 to 2,000 times smaller than that of an electron. For this reason the spectral term-splittings (energy differences) deduced from *hfs* are usually less than 1/1,000 those in multiplet structure (*see* NUCLEAR MOMENTS).

In the same way that (for multiplet structure) L and S were combined to obtain the total angular momentum J of the optical electrons, it is necessary (for *hfs*) to combine J and I in order to obtain the total angular momentum F of the whole atom, including nuclear spin. The hyperfine quantum number F can take the following values: $F = J + I, J + I - 1, J + I - 2, \ldots, |J - I|$. This gives, in all $2J + 1$ or $2I + 1$ different values, according as $J < I$ or $J > I$. The vector-addition diagrams for J and I will be identical with those for L and S where $J = L$ and $I = S$ (*see* fig. 14). In general, the number of *hfs* components of a spectral term is different for different terms of the same spectrum. Terms with $J = 0$ are always single, but all others terms are split into two components if $I = \frac{1}{2}$. If I is greater than $\frac{1}{2}$, terms with $J < I$ have $2J + 1$ components, whereas those with $J > I$ have $2I + 1$ components.

The magnitude of hyperfine splitting depends on the nuclear magnetic moment but is also dependent on the type of electron. Thus, spectral terms arising from electron configurations containing one s-type electron will be split wider than those from a p electron with the same principal quantum number, because the s electron approaches closer to the nucleus and undergoes a stronger interaction. The transitions between *hfs* states are regulated by the same selection rule for F that holds for J, *i.e.*, $\Delta F = 0, \pm 1$, but 0 to 0 transitions are forbidden. Furthermore, the same intensity rule and the same interval rule are valid for both *hfs* and ordinary multiplet structure, except F replaces J.

After deriving (from spectroscopic observations and quantum theory) a satisfactory explanation of *hfs*, it became possible conversely to deduce, from the results of spectroscopic experiments, several properties of atomic nuclei. In 30 years following 1927, the mechanical moments or spins of 226 atomic nuclei were determined, the magnetic moments for 194 and the quadrupole moments for 75. The results are compiled in *Progress in Nuclear*

TABLE VII.—*Atomic Properties From Spectroscopy*

Period	Atom	Ground level	Electron configuration	Spectral multiplicities	Ionization potential ev	Strongest wavelength A
1	1 H	$^2S_{0\frac{1}{2}}$	$1s^1$	2	13.595	1215.66
	2 He	$^1S_0$	$(1s^2)$	1, 3	24.581	584 33
2	3 Li	$^2S_{0\frac{1}{2}}$	$2s^1$	2	5.390	6707.85
	4 Be	$^1S_0$	$2s^2$	1, 3	9.320	2348.61
	5 B	$^2P^\circ_{0\frac{1}{2}}$	$2s^2\ 2p^1$	2	8.296	2497 73
	6 C	$^3P_0$	$2s^2\ 2p^2$	1, 3, 5	11.256	1657.01
	7 N	$^4S^\circ_{1\frac{1}{2}}$	$2s^2\ 2p^3$	2, 4	14.53	1134 98
	8 O	$^3P_2$	$2s^2\ 2p^4$	1, 3, 5	13.614	1302.19
	9 F	$^2P^\circ_{1\frac{1}{2}}$	$2s^2\ 2p^5$	2, 4	17.418	954.8
	10 Ne	$^1S_0$	$(2s^2\ 2p^6)$	1, 3	21.559	735 89
	11 Na	$^2S_{0\frac{1}{2}}$	$3s^1$	2	5.138	5889 95
3	12 Mg	$^1S_0$	$3s^2$	1, 3	7.644	2852.13
	13 Al	$^2P^\circ_{0\frac{1}{2}}$	$3s^2\ 3p^1$	2	5.984	3961.53
	14 Si	$^3P_0$	$3s^2\ 3p^2$	1, 3, 5	8.1512	2516.12
	15 P	$^4S^\circ_{1\frac{1}{2}}$	$3s^2\ 3p^3$	2, 4	10.484	1774.94
	16 S	$^3P_2$	$3s^2\ 3p^4$	1, 3, 5	10.357	1807.31
	17 Cl	$^2P^\circ_{1\frac{1}{2}}$	$3s^2\ 3p^5$	2, 4	13.01	1347.24
	18 Ar	$^1S_0$	$(3s^2\ 3p^6)$	1, 3	15.755	1048 22
4	19 K	$^2S_{0\frac{1}{2}}$	$4s^1$	2	4.339	7664.91
	20 Ca	$^1S_0$	$4s^2$	1, 3	6.111	4226.73
	21 Sc	$^2D_{1\frac{1}{2}}$	$3d^1\ 4s^2$	2, 4	6.54	3911.81
	22 Ti	$^3F_2$	$3d^2\ 4s^2$	1, 3, 5	6.82	3653.50
	23 V	$^4F_{1\frac{1}{2}}$	$3d^3\ 4s^2$	2, 4, 6	6.74	4379.24
	24 Cr	$^7S_3$	$3d^5\ 4s^1$	1, 3, 5, 7	6.764	3578.69
	25 Mn	$^6S_{2\frac{1}{2}}$	$3d^5\ 4s^2$	2, 4, 6, 8	7.432	4030.76
	26 Fe	$^5D_4$	$3d^6\ 4s^2$	1, 3, 5, 7	7.87	3719.94
	27 Co	$^4F_{4\frac{1}{2}}$	$3d^7\ 4s^2$	2, 4, 6	7.86	3453.50
	28 Ni	$^3F_4$	$3d^8\ 4s^2$	1, 3, 5	7.633	3414.76
	29 Cu	$^2S_{0\frac{1}{2}}$	$(3d^{10})\ 4s^1$	2, 4	7.724	3247 54
	30 Zn	$^1S_0$	$4s^2$	1, 3	9.391	2138.56
	31 Ga	$^2P^\circ_{0\frac{1}{2}}$	$4s^2\ 4p^1$	2, 4	6.00	4172.06
	32 Ge	$^3P_0$	$4s^2\ 4p^2$	1, 3, 5	7.90	2651.18
	33 As	$^4S^\circ_{1\frac{1}{2}}$	$4s^2\ 4p^3$	2, 4	9.81	1890.43
	34 Se	$^3P_2$	$4s^2\ 4p^4$	1, 3, 5	9.75	1960.91
	35 Br	$^2P^\circ_{1\frac{1}{2}}$	$4s^2\ 4p^5$	2, 4	11.813	1488 45
	36 Kr	$^1S_0$	$(4s^2\ 4p^6)$	1, 3	13.996	1164.87
5	37 Rb	$^2S_{0\frac{1}{2}}$	$5s^1$	2	4.176	7800 23
	38 Sr	$^1S_0$	$5s^2$	1, 3	5.693	4607.33
	39 Y	$^2D_{1\frac{1}{2}}$	$4d^1\ 5s^2$	2, 4	6.38	4102.38
	40 Zr	$^3F_2$	$4d^2\ 5s^2$	1, 3, 5	6.84	3601.19
	41 Nb	$^6D_{0\frac{1}{2}}$	$4d^4\ 5s^1$	2, 4, 6	6.88	4058.94
	42 Mo	$^7S_3$	$4d^5\ 5s^1$	1, 3, 5, 7	7.10	3798.25
	43 Tc	$^6S_{2\frac{1}{2}}$	$4d^5\ 5s^2$	2, 4, 6, 8	7.28	3636.10
	44 Ru	$^5F_5$	$4d^7\ 5s^1$	1, 3, 5, 7	7.364	3498.94
	45 Rh	$^4F_{4\frac{1}{2}}$	$4d^8\ 5s^1$	2, 4	7.46	3434.89
	46 Pd	$^1S_0$	$(4d^{10})$	1, 3	8.33	3404.58
	47 Ag	$^2S_{0\frac{1}{2}}$	$5s^1$	2, 4	7.574	3280.68
	48 Cd	$^1S_0$	$5s^2$	1, 3	8.991	2288.02
	49 In	$^2P^\circ_{0\frac{1}{2}}$	$5s^2\ 5p^1$	2, 4	5.785	4511.32
	50 Sn	$^3P_0$	$5s^2\ 5p^2$	1, 3, 5	7.342	2839.99
	51 Sb	$^4S^\circ_{1\frac{1}{2}}$	$5s^2\ 5p^3$	2, 4	8.639	2068.33
	52 Te	$^3P_2$	$5s^2\ 5p^4$	1, 3, 5	9.01	2142.75
	53 I	$^2P^\circ_{1\frac{1}{2}}$	$5s^2\ 5p^5$	2, 4	10.454	1782.76
	54 Xe	$^1S_0$	$(5s^2\ 5p^6)$	1, 3	12.127	1295.59
6	55 Cs	$^2S_{0\frac{1}{2}}$	$6s^1$	2	3.893	8521.10
	56 Ba	$^1S_0$	$6s^2$	1, 3	5.210	5535.55
	57 La	$^2D_{1\frac{1}{2}}$	$5d^1\ 6s^2$	2, 4	5.61	6249.93
	58 Ce	$^1G^\circ_4$	$4f^1\ 5d^1\ 6s^2$	1, 3, 5	(5.60)	5699.23
	59 Pr	$^4I^\circ_{4\frac{1}{2}}$	$4f^3\ 6s^2$	4	(5.48)	4951.36
	60 Nd	$^5I_4$	$4f^4\ 6s^2$	5	(5.51)	4924.53
	61 Pm	$^6H^\circ_{2\frac{1}{2}}$	$4f^5\ 6s^2$	6		4781.29
	62 Sm	$^7F_0$	$4f^6\ 6s^2$	7, 9	5.6	4296.74
	63 Eu	$^8S^\circ_{3\frac{1}{2}}$	$4f^7\ 6s^2$	6, 8, 10	5.67	4594 03
	64 Gd	$^9D^\circ_2$	$4f^7\ 5d^1\ 6s^2$	7, 9, 11	6.16	4225.85
	65 Tb	$^6H^\circ_{7\frac{1}{2}}$	$4f^9\ 6s^2$	6, 8	(5.98)	4326.47
	66 Dy	$^5I_8$	$4f^{10}\ 6s^2$	5		4211.72
	67 Ho	$^4I^\circ_{7\frac{1}{2}}$	$4f^{11}\ 6s^2$	4		4103.84
	68 Er	$^3H_6$	$4f^{12}\ 6s^2$	3, 5	(6.08)	4007.97
	69 Tm	$^2F^\circ_{3\frac{1}{2}}$	$4f^{13}\ 6s^2$	2	(5.81)	4094.19
	70 Yb	$^1S_0$	$(4f^{14})\ 6s^2$	1, 3	6.25	3987.99
	71 Lu	$^2D_{1\frac{1}{2}}$	$5d^1\ 6s^2$	2, 4	6.15	3359.56
	72 Hf	$^3F_2$	$5d^2\ 6s^2$	1, 3, 5	6.8	3072 88
	73 Ta	$^4F_{1\frac{1}{2}}$	$5d^3\ 6s^2$	2, 4, 6	7.88	2714 67
	74 W	$^5D_0$	$5d^4\ 6s^2$	3, 5, 7	7.98	4008 75
	75 Re	$^6S_{2\frac{1}{2}}$	$5d^5\ 6s^2$	2, 4, 6, 8	7.87	3460.47
	76 Os	$^5D_4$	$5d^6\ 6s^2$	3, 5, 7	8.5	2909.06
	77 Ir	$^4F_{4\frac{1}{2}}$	$5d^7\ 6s^2$	2, 4, 6	9.1	2543.97
	78 Pt	$^3D_3$	$5d^9\ 6s^1$	1, 3, 5	9.0	2659.45
	79 Au	$^2S_{0\frac{1}{2}}$	$(5d^{10})\ 6s^1$	2, 4	9.22	2427.95
	80 Hg	$^1S_0$	$6s^2$	1, 3	10.43	1849 68
	81 Tl	$^2P^\circ_{0\frac{1}{2}}$	$6s^2\ 6p^1$	2, 4	6.106	3519.24
	82 Pb	$^3P_0$	$6s^2\ 6p^2$	1, 3	7.415	4057.82
	83 Bi	$^4S^\circ_{1\frac{1}{2}}$	$6s^2\ 6p^3$	2, 4	7.287	3067.72
	84 Po	$^3P_2$	$6s^2\ 6p^4$	1, 3, 5	8.43	2449.99
	85 At	$^2P^\circ_{1\frac{1}{2}}$	$6s^2\ 6p^5$	2, 4		2244.01
	86 Rn	$^1S_0$	$6s^2\ 6p^6$	1, 3	10.746	1786 07
	87 Fr	$^2S_{0\frac{1}{2}}$	$7s^1$			
7	88 Ra	$^1S_0$	$7s^2$	1, 3	5.277	4825.91
	89 Ac	$^2D_{1\frac{1}{2}}$	$6d^1\ 7s^2$	2, 4	6.9	4179.98
	90 Th	$^3F_2$	$6d^2\ 7s^2$	1, 3, 5	(6 95)	3719.44
	91 Pa	$^4K_{5\frac{1}{2}}$	$5f^2\ 6d^1\ 7s^2$	4		
	92 U	$^5L^\circ_6$	$5f^3\ 6d^1\ 7s^2$	5, 7	(6.08)	3584.88
	93 Np	$^6L_{5\frac{1}{2}}$	$5f^4\ 6d^1\ 7s^2$	6		
	94 Pu	$^7F_0$	$5f^6\ 7s^2$	5, 7, 9	5.8	
	95 Am	$^8S^\circ_{3\frac{1}{2}}$	$5f^7\ 7s^2$	8, 10	6.0	6054 64
	96 Cm	$^9D^\circ_2$	$5f^7\ 6d^1\ 7s^2$	9		

TABLE VIII.—*Mechanical, Magnetic, and Quadrupole Moments of Atomic Nuclei*

Atomic nucleus	Mechanical moment $I(h/2\pi)$	Magnetic moment in nuclear magnetons $\mu\,(he/4\pi M_p c)$	Quadrupole moment $Q\ (e \times 10^{-24}\ \text{cm.}^2)$
n	1/2	−1.913139	
H^1	1/2	2.79281	
D^2	1	0.857407	0.002738
Li^6	1	0.822008	−0.0011
Li^7	3/2	3.256310	−0.040
Be^9	3/2	−1.17737	0.03
B^{10}	3	1.80081	0.074
B^{11}	3/2	2.68852	0.0355
N^{14}	1	0.40371	0.011
N^{15}	1/2	−0.28313	
F^{19}	5/2	3.50	
Na^{23}	3/2	2.21753	0.10
Al^{27}	5/2	3.641421	0.155
K^{39}	3/2	0.39146	0.09
K^{40}	4	−1.2981	−0.093
As^{75}	3/2	1.43896	0.3
Cs^{133}	7/2	2.57887	−0.003
La^{138}	5	3.7071	~2.7
La^{139}	7/2	2.77807	0.23
Bi^{209}	9/2	4.07970	−0.4

Physics, vol. 6 (1957). Typical data are shown in Table VIII in which the exponent on the chemical symbol represents the mass of the atomic nucleus. In general, nuclear mechanical moments are derived either from the number of observed *hfs* components, or from their relative intensities, or from the intervals between *hfs* levels. Magnetic moments are calculated from a theoretical formula containing the observed *hfs* intervals and other constants, but the most precise values have been obtained from measured deviations of atomic or molecular beams in inhomogeneous magnetic fields. A positive moment indicates that the *hfs* terms are regular; *i.e.*, the level with smallest F value has the lowest energy. Deviations of *hfs* levels from the interval rule are referred to a quadrupole moment arising from a nonspherical distribution of protons in the nuclei. There are two possibilities: if the positive charge is spread in the direction of the mechanical moment, the nucleus is elongated and has a positive quadrupole moment; otherwise the nucleus is flattened and has a negative quadrupole moment.

The only regularity apparent in Table VIII is that the nuclear mechanical moment is a half integer if the mass of the nucleus is odd, and a full integer if the mass is even. It is also seen that the algebraic sums of the mechanical moments and of the magnetic moments of the neutron (n) and proton (H^1) are exactly and nearly equal, respectively, to the same quantities for the deuteron (D^2).

The manner in which the moments of the elementary particles must be combined to yield the observed values for the other atomic nuclei has been partially explained by a nuclear shell model (*see* NUCLEUS: *Nuclear Structure and Nuclear Forces*). *See* also references under "Spectroscopy" in the Index.

BIBLIOGRAPHY.—H. Kayser, *Handbuch der Spektroskopie*, 8 vol. (1900–34); E. C. C. Baly, *Spectroscopy*, 3 vol. (1924–27); F. Hund, *Linienspektren und periodisches System der Elemente* (1927); A. Sommerfeld, *Atombau und Spektrallinien*, 5th ed. (1931); R. F. Bacher and S. A. Goudsmit, *Atomic Energy States* (1932); H. E. White, *Introduction to Atomic Spectra* (1934); E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (1935); H. Kayser and R. Ritschl, *Tabelle der Hauptlinien der Linienspektren aller Elemente* (1939); Massachusetts Institute of Technology, Spectroscopy Laboratory, *Wavelength Tables*, comp. under the direction of G. R. Harrison (1939); G. Herzberg, *Atomic Spectra and Atomic Structure* (1944); L. C. Pauling and S. A. Goudsmit, *Structure of Line Spectra* (1930); W. F. Meggers, "Spectroscopy, Past, Present, and Future," *J. Opt. Soc. Am.*, vol. 36 (1946); G. Harrison, *Practical Spectroscopy* (1948); C. E. Moore, *Atomic Energy Levels*, Circular 467, National Bureau of Standards, 3 vol. (1948–58); G. Herzberg, *Molecular Spectra and Molecular Structure*, 3 vol. (1950–); W. F. Meggers, C. H. Corliss, and B. F. Scribner, "Tables of Spectral-Line Intensities," *Natn. Bur. Stand.* Monogr. 32, 2 parts (1961); R. A. Sawyer, *Experimental Spectroscopy*, 3rd ed. (1963); N. B. Colthup *et al.*, *Introduction to Infrared and Raman Spectroscopy* (1964); A. D. Cross, *An Introduction to Practical Infrared Spectroscopy* (1964); L. Lang (ed.), *Absorption Spectra in the Ultraviolet and Visible Region*, 8 vol. (1961–1966); H. Griem, *Plasma Spectroscopy* (1964); Gerald W. King, *Spectroscopy and Molecular Structure* (1964); S. K. Freeman (ed.), *Interpretive Spectroscopy* (1965); D. W. Mathieson (ed.), *Interpretation of Organic Spectra* (1965); Lawrence Mertz, *Transformations in Optics* (1965); K. Siegbahn (ed.), *Alpha-, Beta- and Gamma-ray Spectroscopy*, 2 vol. (1965); D. H. Whiffen, *Spectroscopy* (1966); J. C. Henniker, *Infrared Spectrometry of Industrial Polymers* (1967). (W. F. M.)

SPECTROSCOPY, ASTRONOMICAL, is a branch of astrophysics based on information obtained by means of a prism, a diffraction grating, or some other light-dispersing device. By using such techniques data have become available that make it possible to identify the chemical elements present on the sun and other stars. In a widely celebrated series of investigations, for example, the element helium was detected on the sun about a generation before its presence on earth was demonstrated. Application of spectroscopic methods to the problems of astronomy permits indirect measurements to be made of the temperature of far-distant stars. Astronomical spectroscopy also revealed the phenomenon of the red shift. This is an application of the Doppler effect in regard to the wavelengths of electromagnetic radiation and manifests itself in spectroscopy by a shift of a light source's spectral lines toward the red end of the spectrum. The identification of this phenomenon provided a basis for the theory of the expanding universe (*see* COSMOLOGY).

This article is organized into the following sections and subsections:

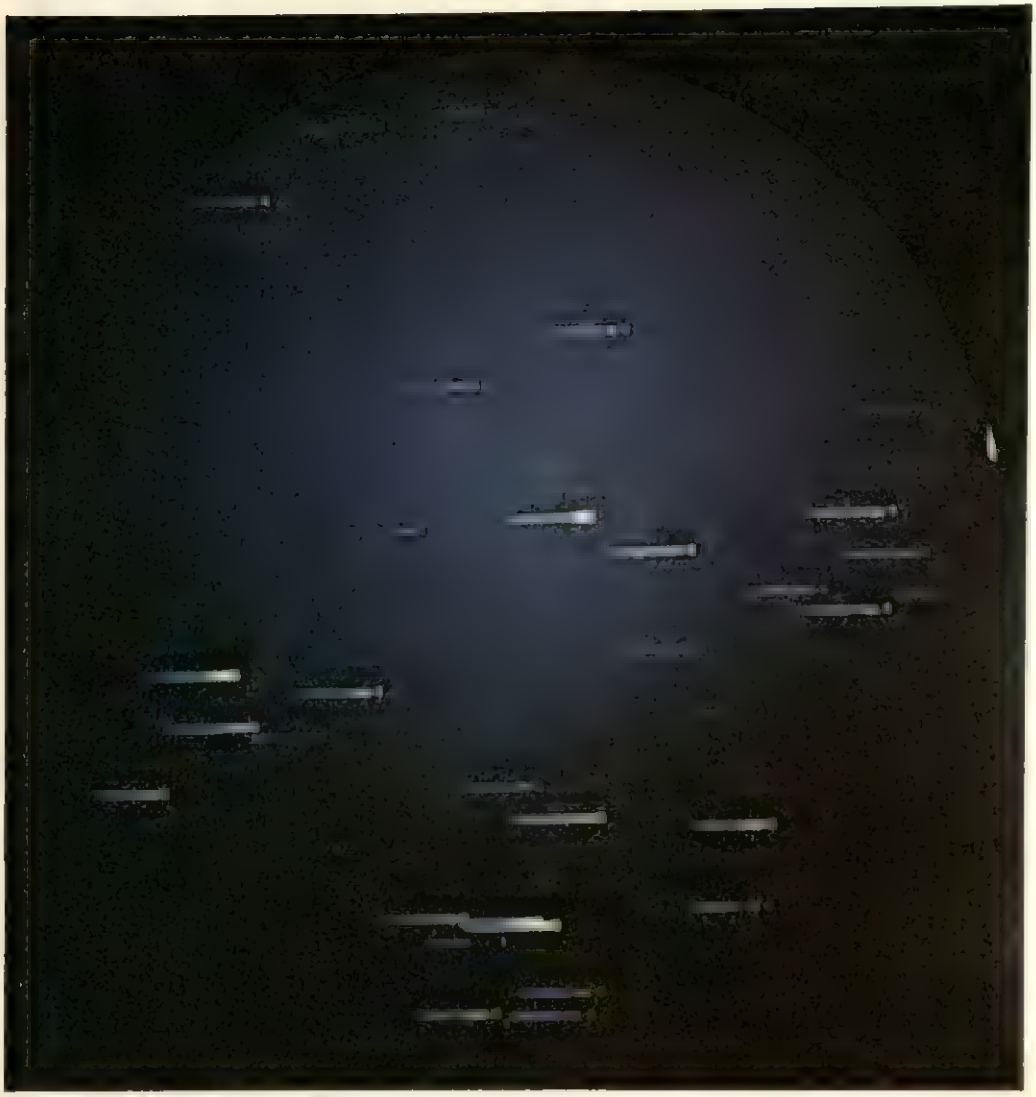
I. HISTORICAL DEVELOPMENT

Astronomical spectroscopy had its beginning in 1666 when Isaac Newton observed the separation of the colours of the rainbow in a prism. In 1802 William H. Wollaston placed a narrow slit in a window blind and with a prism of better quality than the one used by Newton, but without a collimator or a camera, noticed several dark lines in the spectrum of sunlight. Real progress in astronomical spectroscopy became possible when Joseph von Fraunhofer in 1814 made use of a telescope, in addition to the prism and a distant slit, and thereby constructed the first modern spectroscope. A collimator lens was first used by W. H. Simms in 1840. With his instrument Fraunhofer observed and mapped 754 dark lines in the spectrum of the sun. He also observed the spectra of several bright stars and found that, although they all had dark lines upon a coloured continuum, the lines were not the same as those of sunlight. Fraunhofer noticed that the yellow emission line from a flame fed with table salt, which had already been observed by John Herschel, coincides in position with a conspicuous black line in the solar spectrum, which he had designated by the letter D. But he did not follow up this result.

It was only after Gustav R. Kirchhoff had advanced his famous

three laws of spectroscopy that the connection between absorption and emission became obvious (*see* SPECTROSCOPY: *History*). Thereafter William Huggins (1824–1910) in England undertook a systematic survey of the spectra of all stars and nebulae within the reach of his instruments at Tulse Hill in London. This period was the richest in the number and importance of astrophysical discoveries. Many different types of stellar spectra were found; the Orion nebula was found to be gaseous, contrary to the opinion of other astronomers, while the Andromeda nebula showed a continuous spectrum similar to that of normal stars like the sun. Huggins concluded independently of Johann K. F. Zöllner that the solar prominences could be observed without an eclipse by means of a spectroscope, and that the planets shine by reflected sunlight. He was the first to try to measure the radial motions of stars from the Doppler displacements of their lines.

The work of Huggins was of a pioneering character, and it was he who laid the foundations for all further investigations. His tradition in England was carried on by J. N. Lockyer and F. E. Baxandall at the Solar Physics Observatory in South Kensington, which was transferred to Cambridge after Lockyer's retirement in [illegible].



BY COURTESY OF THE UNIVERSITY OF MICHIGAN

SPECTRAL IMAGE OF STARS SEEN THROUGH A PRISM PLACED IN FRONT OF THE TELESCOPE LENS REVEALS IN RAINBOW COLOURS BOTH THEIR BASIC COMPOSITION AND THEIR TEMPERATURES. THIS STAR GROUP IS CALLED THE HYADES

The problem of spectral classification was undertaken in the middle of the 19th century by Angelo Secchi in Italy and was later continued by H. C. Vogel in Germany, who was the first to introduce the photographic method into spectroscopy, and who succeeded in determining accurate radial velocities of many stars. Vogel's work, especially his measurements of radial velocities, stimulated several other observatories, near the turn of the century, to undertake systematic studies in this field. W. W. Campbell at the Lick Observatory undertook the determination of the motions of all stars brighter than magnitude 5.5 in both hemispheres of the sky. E. B. Frost at the Yerkes Observatory concentrated upon the motions of hot stars. J. S. Plaskett at Victoria, British Columbia, discovered many spectroscopic double stars and determined their orbits; he also extended to fainter stars the work of other observatories. W. S. Adams and his associates at Mount Wilson Observatory, California, determined the velocities of hundreds of faint stars and increased the precision to about ±0.01 km./sec. in the case of the bright star Arcturus. A. A. Belopolsky at Pulkovo, Russia, and, more recently, G. A. Shajn at Simeis, in the Crimea, also made contributions in this field. H. Spencer Jones, then at the Cape Observatory, published numerous measures of stellar motions in the southern sky. C. Fehrenbach in France has developed a reliable method of determining stellar radial velocities on objective-prism (slitless) plates.

The work on spectral classification received a new impetus when E. C. Pickering at Harvard, assisted by Miss A. J. Cannon, Miss A. C. Maury, and Mrs. W. P. Fleming, undertook the classification of several hundred thousand stars over the entire sky. This work, culminating in the *Henry Draper Catalogue* by Miss Cannon, published 1918–24, provides the basis for all modern astrophysical work.

The spectroscopic study of the sun was advanced by P. J. C. Janssen, of the Meudon Observatory in Paris, who followed Huggins' suggestion and succeeded, in 1868, in observing the prominences of the sun in full daylight. Further advances were made by N. C. Dunér in Sweden, who determined the rotation of the sun from the Doppler effect (*q.v.*) and found it to be different in different latitudes of the sun, and by H. A. Rowland at Johns Hopkins University, who prepared a catalogue of solar wavelengths. At the turn of the century great progress in this field was made by G. E. Hale, who successively founded and directed the Kenwood Observatory in Chicago, the Yerkes Observatory in Williams Bay, Wis., and the Mount Wilson Observatory at Pasadena, Calif. The spectroheliograph was invented independently by Hale, and by H. A. Deslandres of the Paris Observatory. In 1908 Hale discovered the large magnetic fields in sunspots by observing the Zeeman effect (*q.v.*), the shift in the positions and shapes of the sunspots' spectral lines due to the interaction of the atoms with the magnetic field.

Of all the results of astronomical spectroscopy none is more interesting than the evidence for the uniformity of distribution of the chemical elements throughout the universe. It was a revelation to the earlier spectroscopists when they discovered, first in the sun and later in the stars and nebulae, the same familiar substances—hydrogen, iron, calcium, and the rest—that they knew on the earth. The belief in the universality of the chemical ele-

ments was much strengthened when the lines of helium, originally found in the spectrum of the rim of the sun in 1868 by P. Janssen and by J. Norman Lockyer, were first produced in terrestrial laboratories about 25 years later. It was only in 1941 that the last great enigma of line identification was solved when B. Edlén, in Sweden, announced that the previously unidentified lines of the solar corona originate in the atoms of the common elements, iron, calcium, nickel, and argon, excited to a degree of ionization not even dreamed of previously. Henry Norris Russell had remarked earlier, when Ira S. Bowen in a similar manner disposed of the so-called "nebulium" lines in certain nebulae, that the mysterious substances of the astrophysicists one after another literally disappeared in "thin air." As a matter of fact, Bowen found that the nebulae consist of oxygen, nitrogen, hydrogen, and a few other gases; in other words, they are not very different in composition from air.

The principle of the uniformity of chemical elements has now been extended to include even the distant galaxies, whose light travels hundreds of millions of years at the rate of 186,000 mi. per second before it reaches the eye of the observer. The principle of uniformity of chemical elements means that the atomic building blocks of the universe are the same throughout space. It does not mean that the proportions of these elements are the same in all astronomical objects. We know, in fact, that certain stars differ strikingly from others in their composition, and among the most pressing spectroscopic problems of today is the measurement of these differences in the abundances of elements. Such measurements permit the nuclear processes responsible for the formation of the elements to be inferred.

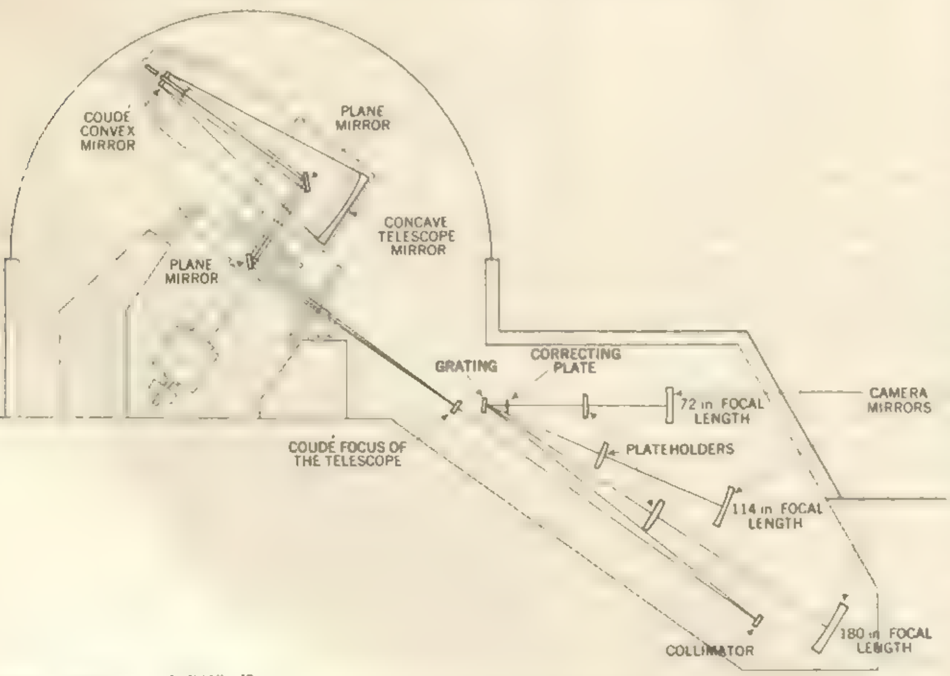
Many other applications of spectroscopy to astronomy, such as the measurement of the temperatures and the luminosities of stars, have developed so rapidly in recent years that their story can best be told separately in the sections that follow.

II. INSTRUMENTS

1. Slit Spectrograph.—The stellar spectrograph is an instrument constructed like an ordinary laboratory spectrograph (*see* SPECTROSCOPY: *Spectroscopes and Spectrographs*). But, since nearly all astronomical light sources are exceedingly faint, special precautions must be observed so that all the available light of the object be utilized in forming the spectrum. The instrument most commonly used is a slit spectrograph attached at the focus of an astronomical telescope. It is essential that the angular aperture of the collimator be the same as that of the telescope. Otherwise, light is spilled over the edges of the collimator (collimator of greater focal ratio than telescope), or the optical parts of the spectrograph are unnecessarily large and wasteful (collimator of smaller focal ratio than telescope). If the instrument is used in good seeing for the observation of stars, which appear nearly as points on the slit of the spectrograph, then the efficiency is proportional to the area of the telescope objective, while the purity is proportional to the length of the collimator. But, if the instrument is used for the observation of luminous surfaces, such as nebulae, the moon, the planets, etc., the efficiency is independent of the linear or angular apertures of the telescope and depends only upon the aperture of the collimator.

In practice the astronomer is usually concerned with intermediate conditions: the stars have tremor disks caused by irregularities in the air, and, while their images have a pronounced, sharp peak of light intensity in their contours, they cannot be regarded as mathematical points. Hence, it is advantageous to use a telescope of large aperture to gather in a large amount of light, and a long collimator to secure the maximum possible purity. In very "poor seeing," when the star images may be as large as ½ minute of arc in diameter (this would project as a disk of 4 mm. in diameter on the slit of the spectrograph of the 82-inch reflector of the McDonald Observatory), the advantages of the large aperture are largely lost, while in good seeing, when the diameter of the image is less than one second of arc, the efficiency is almost exactly proportional to the linear aperture.

For the greatest efficiency it is desirable to make the ratio of collimator to camera lengths as large as possible, so that the slit can be opened without increasing the width of the projected image of the slit on the plate to more than the size of the plate-grain clusters. Thus, short-focus cameras are favoured, which require dispersing units of great power in order to secure the desired scale of the spectrum. For ordinary work, with cameras having focal ratios in the order of $f/3$ to $f/5$, photographic lenses having three or four components give good results. For instruments of very short focus, microscope objectives were adapted successfully as camera lenses by W. B. Rayton and R. J. Bracey. The very fastest cameras, however, have usually been Schmidt cameras, those with twice-through corrector plates mounted directly in front of the gratings, as designed by I. S. Bowen at Mount Wilson and Palomar Observatory working down to about $f/1$. Even smaller focal ratios, extending to $f/0.35$, have been obtained with thick-mirror Schmidt cameras in which the photographic plate or film is pressed against the back of the corrector lens.



ADAPTED FROM A DRAWING BY T. DUNHAM, JR.

FIG. 1.—THE 70-IN. COUDÉ TELESCOPE SPECTROGRAPH

Slit spectrographs, mounted at either the prime focus or at the Newtonian or Cassegrain foci, move with the telescope and therefore must be made rigid enough to show no appreciable change in flexure during exposures lasting several hours. This difficulty is avoided in the fixed spectrographs mounted at the coudé focus of large reflectors, and in solar towers where coelostat mirrors send the light beam into underground shafts. Coudé spectrographs, similar to that sketched in fig. 1, are in use at the Mount Stromlo, Victoria, Lick, Mount Wilson, Palomar, McDonald, Okayama, and Haute Provence observatories.

For very bright objects long cameras and powerful dispersing units may be used. For example, the spectrum of the sun's surface has been photographed with the solar telescopes of the Mount

Wilson and Kitt Peak observatories with linear scales as great as 0.1 Å/mm. and resolutions up to 600,000. A few of the brightest stars also can be analyzed with such long cameras, but most stellar spectroscopy is carried out at much lower dispersions. While classification can be done with smaller scales, commonly from 40 to 250 Å/mm, radial velocities have usually been measured on plates having scales between 5 and 80 Å/mm. For very faint stars and intergalactic objects it has been necessary to work with scales as low as 500 to 1200 Å/mm.

The principal advantages of slit spectrographs consist of (1) greater purity than is possible with slitless instruments; (2) a comparison spectrum that can be photographed alongside the astronomical spectrum to serve in radial-velocity determinations; and (3) the elimination of the background of the night sky, which places a definite limit on the efficiency of slitless instruments.

The spectral region accessible to spectrographs on the surface of the earth is limited at the violet end by the absorption of the terrestrial atmosphere near λ (wavelength) 2900 Å. This limitation is dramatically removed by observations from rockets above the atmosphere. In 1946 the U.S. Naval Research Laboratory sent a spectrograph in a German V-2 rocket to a height of 33.5 mi. and photographed the sun's spectrum. Since that time the observations of the sun and a few of the brightest stars have been extended past the Lyman α line of hydrogen at λ 1216 Å and through the X ray region down to 1Å. On the long-wavelength end the practical limit set by the sensitivity of photographic emulsions or electron image tubes is about λ 12,000 Å (1.2 μ), but by using photoconductive cells of lead sulfide, mercury-germanium, etc., the observations in the far infrared have been extended out to 14 μ. Of course, there is atmospheric absorption in those regions also, due principally to bands of water vapour and carbon dioxide. Since molecules of the latter substances are most abundant in the lower part of the earth's atmosphere, it is possible to send telescopes and spectrographs suspended from balloons high enough to eliminate most of the terrestrial absorption. Spectra of planets and stars were taken at heights up to 16 mi. with the Stratoscope balloons of a group led by M. Schwarzschild at Princeton University, and planetary spectra of Venus and Mars were obtained by the balloon program of J. Strong at Johns Hopkins University.

The dispersing units of astronomical spectrographs are made in the form of prisms (glass, quartz, rock salt), or of diffraction gratings. The latter are particularly useful in the red and infrared regions, where the dispersion of glass is so low that a prohibitive thickness of the substance is required to obtain the same dispersion that a grating with 600 lines per millimetre gives in the second order.

The grating is also efficient in the ultraviolet region, where ordinary glass is opaque, and where quartz prisms are difficult to obtain in sufficiently large sizes for high-dispersion instruments. For rocket observations in the extreme ultraviolet below 800 Å, it is necessary to use the gratings at grazing incidence, which greatly enhances the reflectivity.

Although it has long been known that dispersion of light can be achieved by interference effects in a Fabry-Pérot etalon (a type of interferometer), the earlier astrophysical applications were directed mainly to the study of the auroral green line, which was photographed as early as 1922 by H. D. Babcock with a hand magnifier used as a camera lens. Because of the high resolution obtainable with a tunable etalon it is receiving increased attention, particularly for analysis of the solar spectrum in the infrared. Although the etalon can be used as a slitless device, it is generally combined with a conventional spectrometer.

In astronomical spectroscopy the photographic plate has reigned for many years as the most sensitive and convenient detector. Electronic enhancement of signal strength is now also used either to allow fainter spectra to be photographed, as in an image tube (electron camera), or to replace the plate by a photoelectric cell across which the spectrum is scanned. The increase in efficiency of an image tube is about a factor of ten as compared with direct photography.

Under ordinary observing conditions a large amount of light is lost on the slit jaws of the spectrograph (where it is used for

BY COURTESY OF (RIGHT) WARNER AND SWASEY OBSERVATORY, CASE INSTITUTE OF TECHNOLOGY; PHOTOGRAPH, (LEFT) GEORGE LEAVENS FROM PHOTO RESEARCHERS, INC

TWO TYPES OF SPECTRA

(Left) Spectrum of a rainbow, formed by refraction of sunlight in raindrops. (Right) A spectral photograph of the Ring nebula reveals the nebula as a series of coloured rings. The streaks in the picture are the spectra of nearby stars

guiding the telescope on the star image). It is not possible to filter this light into the spectrograph in such a way that it will reinforce the spectrum of the central portions of the star image. But an ingenious device has been constructed by Ira S. Bowen—an instrument called the image slicer—in which small plane mirrors are used to place the originally wasted light in the slit, alongside the main spectrum. In this manner a widened spectrum is obtained without letting the star trail on the slit, and this results in a substantial reduction of the exposure time.

2. Slitless Spectrograph.—The image of a star is approximately a point. Hence, if a prism is placed in front of the lens or mirror of a telescope, the spectrum is recorded as a narrow continuous band with interruptions in place of spectral lines. If during the exposure the star is permitted to trail slightly in a direction parallel to the refracting edge of the prism, a widened spectrum is obtained in which the star serves as its own slit. The great advantage of slitless spectrographs consists in the fact that they record a large field of stars in a single exposure instead of being limited to a single star on the slit of the instrument. The principal disadvantage lies in the fact that in a large telescope, such as would be required to record the spectra of very faint stars, the images of the stars are tremor disks of appreciable diameter, and spectra obtained in this way are not as pure as those of slit spectrographs.

Another difficulty lies in the absence of a comparison spectrum for determining the radial motions of stars. This can be circumvented by placing a cell filled with some absorbing solution in the path of the rays. Neodymium chloride has been used for this purpose, since it produces one sharp absorption band in the photographic region which can serve as a reference point for measuring the displacements of the stellar lines. A better standard of reference is furnished by the mean positions of the stellar lines themselves, if a direct and reversed image having the same dispersion can be produced side by side. This can be done by rotating the objective prism through 180° between exposures, but the difficult problem of constructing the rotating prism with sufficient mechanical and optical precision was not solved until 1947, when Charles Fehrenbach built his "normal-field prism" at Haute Provence. With improved versions of this instrument, Fehrenbach and his staff measured radial velocities of several hundred stars.

Most of the work on the spectral classification of faint stars is carried on with objective prisms. Although the earlier programs were done with photographic refractors, the wide field and critical definition of star images obtained with large Schmidt cameras has resulted in the general use of these telescopes for objective-prism work. Among the best-known Schmidt reflectors are the 24-in. of the Warner and Swasey Observatory at Cleveland, O., and the 32-in. of the Hamburg Observatory in West Germany.

3. Spectrophotometric Instruments.—For the measurement of the light intensities in continuous spectra and in spectral lines, microphotometers of various designs are used. Instruments for measuring spectrograms usually consist of a light source and a narrow slit; the latter—or its image—is placed in contact with the plate, parallel to the spectral lines, so that only a narrow band of light passes through the plate. This band falls upon a photoelectric

cell, which activates a galvanometer whose mirror produces a displaced light signal upon a sheet of photographic paper. As the plate is slowly moved across the slit, the drum carrying the paper is turned, and the signal produces a continuous record of the galvanometer deflection of the original beam, as modified by the continuous spectrum and by the lines on the plate. Tracings of spectra obtained in this manner must be calibrated by means of marks produced on the original plate by a step wedge, a tube sensitometer, or some similar device which changes intensities by known amounts. This is required in order to express the displacements of the signal on the tracings in terms of relative intensities. Several microphotometers have been designed to perform this wedge calibration automatically, and give tracings in which deflections are proportional to the intensities in the original spectrum.

The most accurate spectrophotometry is done directly at the telescope by mounting a second slit at the focus of the spectrograph and scanning the spectrum with a photoelectric cell. Scanners have been used extensively at both the coudé and Cassegrain foci of the Mount Wilson and Palomar reflectors.

4. Spectroheliograph.—Invented almost simultaneously in 1890 by G. E. Hale in Chicago and H. A. Deslandres in Paris, this is a large spectrograph into which a narrow slit has been built in front of the focal plane of the camera so that only the light of a narrow range of wavelengths falls upon the plate. This slit is so adjusted that the light of a particular absorption line in the sun's spectrum is transmitted. It must be remembered that the lines originate in the uppermost regions of the sun's atmosphere and are not completely dark but contain light of smaller intensity than that of the continuous spectrum. The continuous spectrum is formed by light which is produced, on the average, at a depth of several hundred kilometres in the atmosphere. The image of the sun is projected upon the first, the ordinary, slit of the spectrograph, and the entire telescope is slowly moved by means of an electric motor at right angles to the slit. Simultaneously the plateholder carrying the sensitive emulsion is moved past the second slit. An image of the sun thus is obtained on the plate which is produced by the residual light of a particular spectral line. Since not all regions on the surface of the sun are equally bright in such monochromatic light, it is possible to study features like calcium eruptions (or flocculi), hydrogen tornadoes (or vortexes), and other structures which cannot be seen in the brilliant light of the continuous spectrum, which alone is strong enough to register in the extremely short exposures required for obtaining a direct photograph.

A modification of the spectroheliograph, the spectrohelioscope, was invented by Hale for visual observation. Some phenomena, such as brilliant eruptions in hydrogen light (bombs), last only a few minutes and can best be detected and followed by means of prolonged visual observation.

5. Coronal Spectrograph.—In 1930 Bernard Lyot built on the Pic du Midi in France a coronagraph, an instrument with which he succeeded for the first time in photographing the corona when the sun was not in eclipse. The essential feature of its design is the elimination of scattered and diffracted light in the optical parts. The spectrum of the corona has since been observed regularly with various other coronagraphs.

III. CLASSIFICATION OF STELLAR SPECTRA

1. Methods.—The complex pattern of absorption features (and sometimes emission lines) presented in a stellar spectrum shows regularities which allow the stars to be arranged in sequences, which depend upon the physical and chemical conditions in their atmospheres. The principal groups, or divisions, in these sequences are the spectral types of the stars. These types can later be calibrated in terms of the physical variables (temperature, luminosity, etc.), if these variables can be measured independently for even a few of the stars. This whole procedure for rapid and direct assignment of physically significant types is known as spectral classification.

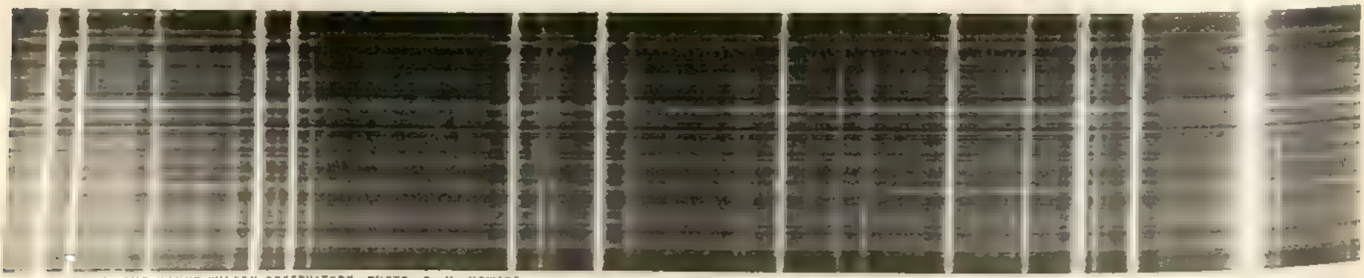
The first enduring system of classification resulted in the *Henry Draper Catalogue,* which gave spectral types of nearly 300,000 stars. The types: O, B, A, F, G, K, M, R, N, and S form a sequence according to effective temperature, which decreases from about 40,000° K for the earliest O-type stars to 2000° K or less for the latest stars of types M, N, or S. Since the calibration of types in terms of temperature is still subject to gradual improvement, the *Henry Draper Catalogue* types have remained useful as empirical indicators of temperature, particularly for the thousands of stars for which there are no later and more refined classifications.

It was noticed early by Miss A. Maury, and others, that stars having the same *Henry Draper Catalogue* type could differ in some features; and, after E. Hertzsprung and H. N. Russell showed that large differences in luminosity exist, W. S. Adams and A. Kohlschütter developed the spectroscopic criteria for classifying stars in two dimensions: luminosity and temperature. In a later system developed by W. W. Morgan, at the Yerkes Observatory, luminosity also is expressed by empirical units: the luminosity classes ranging from I for supergiants through II (bright giants), III (giants), IV (subgiants), to V for dwarfs and for main-sequence stars.

Stars of luminosity class I have absolute magnitudes of the order of -5, while the main sequence (class V) ranges in absolute magnitudes from -4 near type B down to the faintest M-type dwarfs at $+13$ or lower.

The hottest stars, class O, have spectra dominated by the lines of ionized helium; those of class B have hydrogen and neutral helium. In class A hydrogen is very strong, helium is weak or absent, and ionized metals, such as Fe II, Ti II, Mg II, are prominent. (The Roman numerals designate the state of ionization.) In types F, G, and K the lines of neutral metals, Fe I, Ti I, etc., become gradually stronger, and the lines of Ca II, Ca I, and Na I are conspicuous. Near type G we first observe bands of certain diatomic molecules, CH and CN, which also increase in strength as we pass toward the cooler stars. Types M, R, N, and S are characterized by very strong molecular bands of TiO in class M, C2 and CN in classes R and N, and ZrO in class S. Spectral types R and N have since been combined into a new class C with subdivisions which more accurately reflect temperature differences.

The luminosity criteria which are used within each spectral class differ greatly. In the hottest stars, classes O, B, and A, the hydrogen lines are much narrower in the more luminous stars (supergiants). In the intermediate and cool stars, lines of ionized



BY COURTESY OF THE MOUNT WILSON OBSERVATORY; PHOTO, R. H. HOWARD

SPECTRUM OF THE SUN

Spectrum of the sun photographed at Mount Wilson Observatory at scale of 0.1 Å/mm. and a resolution of about 400,000. The horizontal streaks are due to bright granules and the darker interstices between them, and their relative motions give rise to Doppler shifts which cause the lines to have a jagged appearance

elements, such as Sr II in classes G and K, are enhanced in the more luminous stars. The molecular bands of CN, and others, can also be used as luminosity criteria.

Types M, S, and C overlap in their temperature ranges, and illustrate the importance of considering the chemical composition as a third factor determining the spectrum. The S and C stars appear to contain relatively more of such heavy metals as zirconium, lanthanum, and strontium as compared to type M. The ratio of carbon to oxygen is also greater in these two groups, and the excess of carbon as revealed in strong bands of C_2 is the defining characteristic of type C.

Less conspicuous but very important effects of composition appear in the yellow stars of types F, G, and K. The stars divide into several populations which differ in the ratio of hydrogen to all the heavier elements, and the more extreme cases of metal-poor stars form the population II of stars in globular clusters. The metal-poor stars tend to have weaker lines of metals in types F and G and abnormally weak bands of CN in the range G5 to K3. These subtle differences cannot usually be detected in objective-prism spectrograms.

An accurate classification of stellar spectra is important for three widely different problems: (1) the study of the physical properties of stellar atmospheres; (2) the statistical study of stars according to their principal characteristics; and (3) the study of the spatial arrangement of the stars in our galaxy by means of their luminosities.

It is important to have a clear understanding of the purpose of an investigation before the methods and the criteria are chosen. Thus, certain spectral features may be useful in empirical determinations of luminosities, but may be almost useless in studies of physical characteristics because they consist of unresolved lines or bands of different sources.

In addition to the *Henry Draper Catalogue,* which covers the entire sky and includes practically all stars brighter than apparent magnitude 8.2, there are supplements extending to much fainter magnitudes in limited areas. Other more specialized catalogues include the *Mount Wilson Catalogue of Spectroscopic Absolute Magnitudes*, and the catalogues of spectral types of the Hamburg, Berlin Babelsberg, Crimean, Warner and Swasey, Leander McCormick, Stockholm, Uppsala, Haute Provence, and other observatories.

An atlas of MK spectral types was issued by the Yerkes Observatory, and a catalogue listing all stars classified in this system has been completed by C. and M. Jaschek of La Plata. Other spectral atlases based upon spectrograms of various scales have been prepared at the Kitt Peak and Córdoba observatories.

Stellar classification, based upon spectrophotometric measurement rather than visual estimation of line strengths in the cooler stars, has been developed by B. Strömgren and others at the Copenhagen and Kitt Peak observatories. For types B, A, and F a corresponding three-dimensional classification was carried out at the University of Paris under the leadership of D. Barbier and D. Chalonge. All these studies owe much to the pioneer measurements of the CN bands begun as early as 1922 by B. Lindblad.

2. Interpretation of Spectral Classification.—The theoretical interpretation of the two-dimensional classification of stellar spectra rests upon the theory of thermal ionization proposed by M. N. Saha in 1920, and developed in later years by H. N. Russell, E. A. Milne, R. H. Fowler, and others. Some of the ideas underlying this theory were already present in the pioneering work of Norman Lockyer, but the exact mathematical formulation was first given by Saha.

In a gas exposed to high-temperature radiation, neutral atoms are constantly being ionized by quanta of light, and free electrons are being captured by ions. An equilibrium is set up by which the number of ionization processes is exactly balanced by an equal number of recombinations. The relative numbers of ions (n') and neutral atoms (n) depend upon the temperature of the radiation (T), the ionization potential of the atoms (I), and the partial pressure (p_e) of the free electrons, many of which may result from the ionization of some other element than the one we are considering:

$$\log \frac{n'}{n} = -I\,\frac{5040}{T} + \frac{5}{2}\log\, T - 0.48 - \log\, p_e$$

The observed intensity of a neutral line is directly dependent upon n, that of an ionized line upon n'. As the temperature increases, n decreases while n' increases, by virtue of the first factor containing T, which is more important than the second factor. The ionization is smaller when I is large. But the ratio n'/n also depends upon p_e and increases as the pressure decreases.

Consider two stars of the same temperature, but of different pressures, and suppose that we observe in their spectra lines of two elements having different ionization potentials. In the dense star the ionization is less than in the more tenuous star. Moreover, the element of higher potential will be less ionized than the one of lower potential. In practice we do not choose the stars according to temperature but according to spectral class. The latter rests upon a compromise of several slightly conflicting criteria, corresponding to different ionization potentials; for example, two stars classified as *F* do not have exactly the same temperature if their pressures are not alike; the more tenuous star has the lower temperature.

It must also be remembered that many elements can be ionized more than once, so that as T increases or p_e decreases, n' at first increases at the expense of n, reaching a maximum at definite pairs of values of T and p_e, after which it again decreases, while n'', representing the next stage of ionization, begins to increase. Ionization temperatures are determined for a sequence of stars, for which it may be assumed that p_e is constant, by observing at which spectral class lines of different ionization potential reach their greatest intensity. The condition for maximum intensity is obtained by differentiating the ionization formula with respect to T.

The theory of ionization suffices to explain many of the observed absolute magnitude effects in stellar spectra. For a given temperature high luminosity is equivalent to large size (giants) and therefore to small surface gravity and, in turn, small surface gravity indicates low pressure. The tendency of ionized lines to become stronger in giants is explained by the fundamental formula, but for a more complete interpretation it has been necessary to develop a very detailed theory of the formation of the continuous spectrum and of the absorption lines.

IV. RADIAL VELOCITIES

The Doppler principle gives the displacement of a spectral line ($\Delta\lambda$) as a function of the wavelength (λ) and the relative radial velocity of the source and the observer (v):

$$\Delta\lambda = \frac{\lambda \cdot v}{c}$$

where c is the velocity of light. The displacement can be measured on plates furnished with a comparison spectrum and, accordingly, v can be determined. The quantity thus obtained must be corrected for the components of the Earth's motion around the sun and the Earth's rotation around its axis. The precision of the determination depends upon the dispersion of the spectrogram, the precision of the laboratory wavelengths used for star and comparison lines, the sharpness of the star lines, the quality of the optical parts used, the uniformity of the illumination of the slit and the collimator by the star image, and other factors.

The probable error of a radial velocity measured on a single spectrogram can be made as low as 0.1 km/sec for a star with sharp lines (type G or K) taken with a dispersion of about 3 km/sec. For B- or A-type stars the probable error is about twice as great. Since the greater number of velocities have been measured on plates of much smaller scale, and there are systematic differences of the order of 1 or 2 km/sec between the results of different observatories, the velocities of most of the stars brighter than the seventh magnitude are known only to within about 2 km/sec.

The *General Catalogue of Stellar Radial Velocities* prepared by R. E. Wilson in 1953 contained 15,106 stars. Since then several

thousand more measurements have been made, and the task of keeping a continuously revised card catalogue of all known velocities has been taken over by the Commission on Radial Velocities of the International Astronomical Union.

Among the results obtained with the aid of radial velocities are the following: (1) the determination of the velocity (about 20 km/sec), and the direction in space, of the motion of the solar system with respect to the nearer stars; (2) the estimation of the random motions of the stars, which for the massive B stars averages as low as 15 km/sec, increasing down the main sequence to nearly 60 km/sec for typical M-dwarfs; (3) the recognition that the entire galaxy rotates around a distant centre located in the direction of the constellation Sagittarius, with a velocity of between 200 and 300 km/sec in the region where the sun is located; (4) the comparison of radial velocities and proper motions to determine mean distances of groups of stars of any one kind; (5) the discovery of a group of high-velocity stars which are moving with respect to the sun with speeds of the order of 100 km/sec, and in a direction opposite to the normal direction of galactic rotation; (6) the measurement of red shifts in very dense stars (white dwarfs), which are probably the gravitational red shifts predicted by Einstein (*see* GRAVITATION); (7) the law of linear increase with distance in the velocity of recession of external galaxies; (8) the discovery of numerous stars showing periodic variations in radial velocity—due to orbital motions of close double stars or to pulsations in the radii of individual stars (such as Cepheid variables); (9) the study of eruptions in novae and in other expanding shells of gas which surround such stars as P Cygni; (10) the discovery of stationary lines, which do not show the Doppler shifts characteristic of the stars but are interpreted as absorption produced by clouds of interstellar gas.

V. CONTINUOUS SPECTRA

The continuous spectra of the stars show conspicuous differences in the distribution of energy as a function of wavelength. These differences define the colours of the stars. The O and B stars are bluish, while the M, C, and S stars are red. Observations of the absolute energy distributions have been made at many observatories, originally by photographic photometry but now photoelectrically. These observations provide information about the surface temperatures of the stars, and the way in which interstellar matter has modified the colours that would be observed if space were completely transparent.

From the observations of nearby stars the colour temperatures are derived by adjusting T (temperature) in the Planck formula (*see* HEAT: *Radiation: Distribution of Energy in the Spectrum*) to give the best fit with the observed distribution. Differences in the energy distribution of two stars are often given in the form of "relative gradients":

$$\Delta\phi = \frac{c_2}{T_1} - \frac{c_2}{T_2}$$

where c_2 is the constant of the Planck formula. This procedure rests upon the observed fact that the energy distributions of many stars resemble black-body curves fairly closely. Accurate observations, however, have shown a number of important departures:

(1) In spectra of stars of types F, G, K, M, C, and S, which contain thousands of absorption lines, these lines not only reduce the energy radiated in many spectral regions but are often so crowded together that they overlap and leave no undisturbed continuum between them. The observed colours must be corrected for the effect of this "line blanketing." (2) At the limit of the Balmer series of hydrogen the spectra of moderately hot stars (class A) show a discontinuity, usually described as the "Balmer jump"; the continuous spectrum is abnormally weak on the violet side of λ 3647 Å. (3) In some stars the hydrogen lines appear in emission, and the Balmer jump is replaced by an abnormal strengthening of the spectrum shortward of λ 3647 Å. (4) The continuous radiation itself, of most stars, does not follow exactly a black-body curve. In a few unusual stars, such as P Cygni, the discrepancies are very large and give quite different colour temperatures in various regions of the spectrum. (5) Molecular bands in the cooler stars depress large portions of the spectrum. Thus, the energy distribution in the visual region of M-stars gives a measure not of colour temperature but merely of the changing band absorption. (6) The continuous spectrum of the sun is weaker and redder at the edge than in the centre, and the combined light is therefore a mixture of different energy distributions. This phenomenon of "limb darkening" must be present in most stars and has been detected in many eclipsing variables. (7) A strange inconsistency between the spectral class inferred from the absorption lines and observed colour temperature is noticed in several close double stars. This is attributed to the "reflection effect," which causes lines of relatively high excitation to be produced in the atmosphere of a cool binary component by the radiation of a hot component.

All of these phenomena are reasonably well understood. The theory of the continuous spectrum, and especially of the effect of limb darkening, is closely related to the problem of the physical mechanism which causes the gases in the outer layers of stars to be more or less opaque to continuous radiation. Continuous absorption at the series limits of H and He, and to a lesser extent of other elements, scattering of light by free electrons, and, in the cooler stars, continuous absorption from negative ions of hydrogen, and probably of other elements, are known to contribute to the continuous absorption coefficient (as distinct from the absorption coefficient within a spectral line).

VI. ABSORPTION LINES

By means of a spectrophotometer it is possible to measure the intensity of a star's radiation at successive wavelengths, from one edge of a spectral line to the other. When the results of these measurements are expressed in units of the intensity of the continuous spectrum, in the immediate neighbourhood of the line, and are plotted as ordinates against the wavelengths as abscissae, the resulting curve is designated as the profile (or contour) of the line. The profiles of faint lines are strongly affected by the finite resolving power of the spectrograph. For these it is customary to integrate the profiles and to express them in units of one angstrom of complete absorption. These values are designated as the equivalent widths of spectral lines. But for strong lines, like those of Ca II in the sun, the hydrogen lines in the A stars, etc., it is possible to derive physically significant profiles whose interest consists in the information they yield concerning the variation of the line-absorption coefficient with wavelength, and the character of the propagation of light through the atmospheres of the stars. Important information can be obtained from a curve in which the equivalent widths of the absorption lines are plotted as ordinates against the corresponding numbers of atoms effective in producing each line, as abscissae. Such curves are designated as curves of growth because they indicate the manner in which the intensity of a line grows as the numbers of atoms are increased.

1. Radiation Damping.—Many lines, like those of Ca II in the sun, have broad profiles with very deep, narrow pointed centres, where the residual light is only from 1% to 10% of the continuous spectrum, and with wings which extend very far to both sides (fig. 2). These profiles are well represented by an absorption coefficient which is inversely proportional to the square of the distance from the centre of the line. Physical theory suggests that this is the common form of the absorption coefficient of a single, undisturbed atom, where the gradual damping of the radiation in consequence of the loss of energy is the sole broadening influence. Since the mathematical form of this expression is known for a single atom of Ca II, it is possible from the observed profile to derive the number of absorbing atoms of Ca II per cm² in the atmosphere of the sun. This was first done by Unsöld, who used earlier measurements by K. Schwarzschild. In some spectral lines, for example those of Mg I in the sun or some lines of He I in B stars, the profiles are similar in shape to those given by radiation damping but their widths are much greater—just as though the constant of the formula were several times as great as that predicted by atomic theory. This phenomenon is explained by the effect of collisional damping, which was

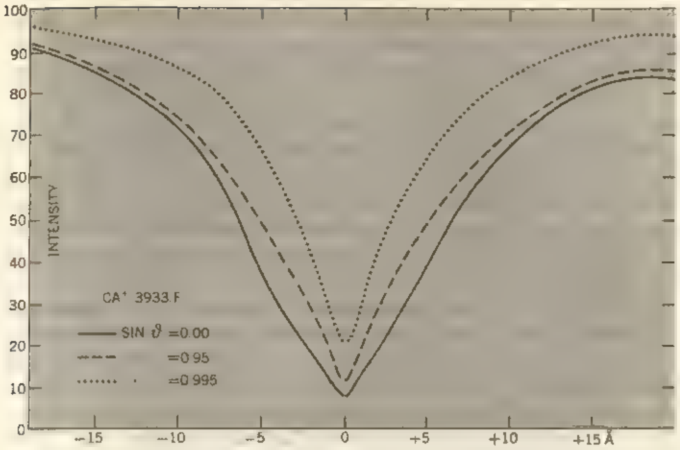


FIG. 2.—PROFILES OF THE STRONG LINE CA II K IN THE SPECTRUM OF THE SUN

Solid line: near the centre of the sun's disk; dashed line: at a point intermediate between centre and limb; dotted line: near the extreme edge of the sun. These profiles are broadened by the effects of radiation damping and collisional damping. The general weakening of the line from centre to limb is explained by the theory of transfer of radiation through a stellar atmosphere

first introduced into physical theory by H. Lorentz and later extended by Weisskopf and Wigner to include processes where atoms do not actually collide but experience close approaches to one another.

This form of broadening is important in dwarf stars of low luminosity and is unimportant in giants of high luminosity. In both types of broadening the equivalent width of a line increases as $\sqrt{N}$, where N is the number of absorbing atoms. For example, we know that the line K of Ca II is formed by twice as many atoms as the line H. In consequence the equivalent widths of the two lines in the sun are in ratio $\sqrt{2}$ to 1.

2. Thermal Doppler Effect.—In a stellar atmosphere the atoms are never quite undisturbed. They are moving about in all directions because of the high temperatures prevailing in these regions, and these motions are presumably in accordance with statistical theory.

Although every single atom radiates in accordance with radiation damping, some atoms are approaching while others are receding. The fraction which moves with a velocity v is, in accordance with Maxwell's law (*see* KINETIC THEORY OF MATTER), proportional to $e^{-(v/v_0)^2}$, where v_0 measures the average speed of the atoms. But if an atom moves with a velocity v in the line of sight it absorbs not in the normal position of the line but at a distance of $\Delta\lambda = \frac{\lambda \, . \, v}{c}$. If v_0 were very large, then radiation damping would be relatively unimportant. The profiles would be bell-shaped, in accordance with Maxwell's formula, and the equivalent widths would increase as N, and not as $\sqrt{N}$. For normal stellar temperatures radiation damping predominates for all but the very faintest lines. We therefore have a gradual transition in the relation between equivalent width and N: at first it varies as N, then as $\sqrt{N}$.

The corresponding bending of the curve of growth, as this relation is called, was observed in the sun by Minnaert, by Allen, and by others. (Fig. 3.)

3. Turbulence.—The curves of growth of many stars, such as α Persei or ϵ Aurigae, show a transition from the N to $\sqrt{N}$ relation at an equivalent width which is much too large to be explained by thermal motions. Moreover, these curves show a region of transition between the two major branches of the curve of growth, where the equivalent width changes little with N and remains almost horizontal. This phenomenon, first noticed by Struve and Elvey, is explained in terms of turbulent motions in the atmospheres of the stars. The sizes of the volumes of gas which are subject to these motions are not known, but they must be considerably smaller than the entire depth of the atmosphere, because otherwise the phenomenon would manifest itself differently in the curve of growth and would suggest convection, and not turbulence. The largest average turbulent velocity measured is about 67 km/sec, for the star 17 Leporis. The best analogue on the earth are the thermals which form on sunny days, but it must be remembered that the causes of the two phenomena are not the same.

4. Stark Effect.—In the presence of an electric field the spectral lines of many atoms are disturbed. In hydrogen the lines are split into several components; in helium some lines are little affected while some are displaced toward the red and others are displaced toward the violet, and completely new lines, never seen without electric fields, appear in the spectrum. It has never been possible to detect even the slightest evidence of a uniform electric field in the sun or stars. But the atmospheres of all stars contain vast numbers of charged particles, namely, free electrons and ions, each of which carries its electric charge and may in its thermal flight through the atmosphere come close enough to a radiating atom to produce an appreciable Stark effect (*q.v.*). We observe a combination of countless cases of this sort, with all possible distances between neighbouring particles and all orientations. The observed line should therefore have a blurred appearance, and its form can be predicted by a combination of statistical and physical theories.

The profiles which result from an application of the theory by Holtsmark differ greatly from those produced by damping or by Doppler effect. The wings are exceedingly broad and the centres of the lines are appreciably raised, so that the residual intensities may be as great as 50% of the continuous spectrum. Such profiles are observed in the hydrogen lines of dwarf stars, but, because of the symmetrical form of the Stark effect splitting in H, this was not sufficient to prove the existence of Stark effect in stars. The proof came when the lines of He I in B stars were found to show appreciable broadening in those lines which are sensitive to electric fields, and no broadening in those lines which are not sensitive to it. Stellar spectra also show several of the new lines forbidden by the quantum theory in the absence of fields; and the permitted lines show systematic shifts in accordance with the theory.

The Stark effect is appreciable in ordinary dwarfs and is very large in white dwarfs. It can be used to determine the luminosity of a star, because of the close relation between this quantity and the atmospheric density.

5. Rotation.—In many stars, especially of classes O, B, A, and F, all lines appear blurred, even those which are not sensitive to Stark effect, while at the same time the curve of growth shows no appreciable turbulence. The equivalent widths of the lines

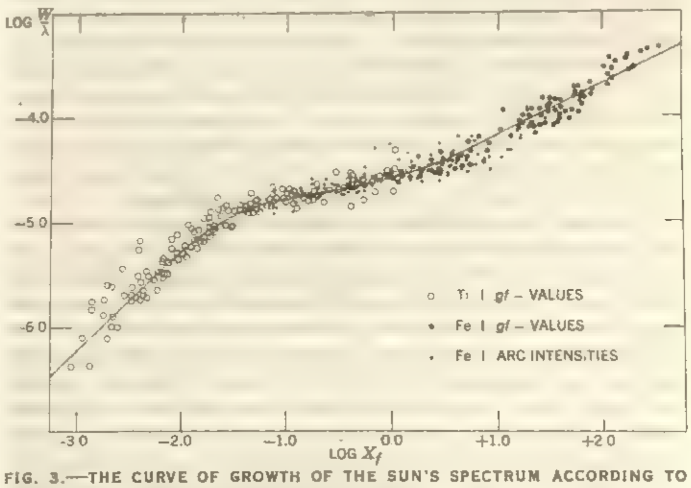


FIG. 3.—THE CURVE OF GROWTH OF THE SUN'S SPECTRUM ACCORDING TO K. O. WRIGHT

The abscissas are logarithms of the relative numbers of atoms (here designated as X_f) effective in producing the line and which can be inferred from multiplet and supermultiplet relations in spectroscopy, or from laboratory measurements. The ordinates are logarithms of the equivalent widths (W) of the lines divided by their wavelengths (λ). The curve is proportional to X_f in its lower left portion and is proportional to $X_f^{\frac{1}{2}}$ in its upper right portion. The bending over is the result of the thermal motions of the atoms in the sun's atmosphere

are not altered by this broadening influence, and the great widths are obtained at the expense of the central intensities. In some cases it has been shown that the widths of the lines are proportional to the wavelengths, if corrections are made to reduce them to the same equivalent width.

Since the Doppler effect is proportional to λ, and since no physical alteration of the absorption coefficient can fail to influence the curve of growth, or the equivalent widths of the lines, the conclusion is justified that we are dealing with some form of Doppler effect. Axial rotation, radial expansion, and convection suggest themselves. The effect of expansion is to produce an unsymmetrical profile which is deepest at the point of greatest negative displacement $\Delta\lambda_0$, and merges with the continuous spectrum at $\Delta\lambda = 0$. Since we can see only one hemisphere of an expanding star, the line does not extend into positive values of $\Delta\lambda$. Neither the asymmetry nor the preference for negative displacements agrees with the observations. We are left with rotation and convection.

To decide between them Struve and Shajn made numerous determinations of line profiles in close spectroscopic binaries. In these stars the components are almost in contact and it is reasonable to suppose that their rotational periods are equal to the period of orbital revolution. All have very broad and shallow lines, and this phenomenon is the more conspicuous the shorter the period. Since no such relation is expected in the case of convection, the "dish-shaped" line profiles can be attributed to axial rotation. In a few eclipsing binaries, for example in Algol, the line profiles become unsymmetrical during the partial phases of the eclipse and, as Rossiter and McLaughlin have shown, it is possible to derive from them the rotational velocity at the equator.

For single stars the line profiles furnish the values of $v \sin i$, where v is the equatorial velocity and i is the inclination of the star's axis of rotation to the line of sight. On the assumption of random orientation of the axes of rotation, mean values of the rotational velocity can be found for groups of stars. Following the early estimates of rotational velocities by Struve, Elvey, and Morgan at Yerkes, many more accurate velocities have been measured, and a catalogue by A. A. Boyarchuk and I. M. Kopylov of the Crimean Observatory (1964) listed $v \sin i$ for 2,558 stars.

The range of observed rotational speeds extends from 0 to more than 400 km/sec, with the longest values found in bright-line B stars. In ordinary absorption-line stars the mean $v \sin i$ reaches a maximum near B5 and declines steadily toward later types. A. E. Slettebak and others have shown that the giant stars of early type (B and A) rotate more slowly than dwarfs, while for types F and G the reverse holds and the giants are rotating faster. These results are consistent with current theories of the evolution of stars away from the main sequence. Among double stars the wide pairs rotate normally; as A. E. Slettebak expressed it, "Visual binary components apparently rotate as though each star were not aware that the other component is there." In contrast, close binary stars rotate very slowly and H. A. Abt has emphasized that there must be close interaction between the components, since their periods of rotation and of revolution are synchronized.

6. Stellar Line Profiles.—The formation of stellar absorption lines commonly involves the combination of two or more of the above mentioned widening agencies, and the resultant profile depends upon the manner in which the radiation is propagated through the star's atmosphere. It is necessary to take account of the fact that the atoms not only absorb but also reemit radiation, and that this emission occurs in all directions while the absorption takes place in radiation directed outward from the star. There are two extreme possibilities: the radiation may be reemitted without change of frequency but with change of direction only (pure scattering), or it may be redistributed in frequency according to the Kirchhoff-Planck function (pure absorption). The Balmer lines are believed to approach the case of pure absorption, while resonance lines involve a scattering process primarily.

The accurate calculation of line profiles for comparison with observations is now generally carried out by the method of model atmospheres. From preliminary "coarse analysis" by means of curves of growth a model structure giving the dependence of temperature, density, and continuous absorption on depth is adopted. Then the intensity at each point in the profile is computed by one of several possible methods of numerical integration of the equation for the transfer of radiation through the atmosphere. By using this technique of "fine analysis" of stellar atmospheres it has been possible to predict accurately the shapes of many lines in the spectra of stars of types B, A, and F. Extension to cooler stars has been hindered by incomplete knowledge of the sources of continuous opacity in their atmospheres, but rapid progress is being made.

7. Centre to Limb Differences.—The most powerful method for testing theories of line formation consists in the study of line profiles at different distances from the centre of the sun. It was shown originally by K. Schwarzschild that if a line is formed by pure absorption it should completely disappear at the limb of the sun. Actually, all strong solar lines become somewhat weaker at the limb, but they do not disappear (fig. 2). By combining these centre-limb differences in central intensities of lines with the limb darkening in continuous radiation, it has been possible to construct rather detailed models of the temperature distribution through the solar atmosphere.

8. Abundances of Elements.—Equivalent widths of atomic lines and depths of molecular bands are the data from which the chemical composition of stellar atmospheres is derived. Since the pioneer analysis of the solar spectrum by H. N. Russell in 1932, abundances in many stars have been estimated by means of curves of growth, and, in a few, by means of model-atmosphere calculations.

For the sun and most of the stars in its neighbourhood (stellar population I) abundances of the common elements are found. In their atmospheres the ratio of hydrogen to all other elements heavier than helium is about 5,000 to 1. There are, in addition, many stars whose spectra give evidence of quite different chemical composition. Of particular interest are the metal-poor stars (stellar population II) in which the content of heavy elements is reduced by additional factors of from 5 to 100 as compared with hydrogen. In the spectra of F- and early G-type stars, where the lines of common metals such as iron are just becoming conspicuous, this deficiency in abundance not surprisingly reveals itself by a weakening of the metal lines as compared to the Balmer lines of hydrogen. At lower temperatures, however, these metal lines are normally strong enough to lie on the flat part of the curve of growth, where even a considerable change in numbers of atoms will have little effect on the appearance of the lines. In these types the features, which are most responsive to the increased opacity corresponding to an excess of hydrogen (since continuous absorption in their atmospheres is proportional to the amount of H^- present), are molecular bands in which the absorption is spread over an appreciable range in wavelengths. Thus, the blue and violet bands of CN are weakened in metal-poor giants of types G5 to K2, while molecules containing hydrogen, such as CH, show their usual band strengths. The most extreme examples are found among the stars populating the nearby spherical halo surrounding the galactic disk—the members of globular clusters and isolated stars with very large velocities in space. It appears probable that such stars were formed in parts of space where the original hydrogen gas was not greatly contaminated by heavier elements, synthesized in the interiors of stars of earlier generations.

In contrast to the metal-poor stars are those in which the ratio of all the heavier elements to hydrogen is considerably greater than in the sun. Thus there are some strong-line G- and K-type stars in which there appears to be a moderate enrichment of nearly all the metals. More extreme differences are shown by barium stars, carbon stars, and S-type stars. In such objects the lines and bands due to heavy elements (Zr, Sr, La, etc.) are enhanced in relation to the metals of the iron peak (Fe, Ti, etc.). In the carbon stars there is also a great excess of carbon over oxygen, and there appears to be some increase in the proportion of carbon to oxygen in type S also. It appears that these heavy-metal stars are relatively young objects formed from the debris of earlier stars in which the heavy elements were formed in the interiors before

being blown out into space, but it is also likely that inside the S- and C-stars, which are giants, the burning of helium and carbon is probably going on at the present time. *See* STAR: *Formation of the Elements in Stars.*

VII. SPECTRA OF STARS OF UNUSUAL INTEREST

In addition to the main groups of stars, there are many individual stars and groups which possess spectral peculiarities that reveal significant astrophysical information.

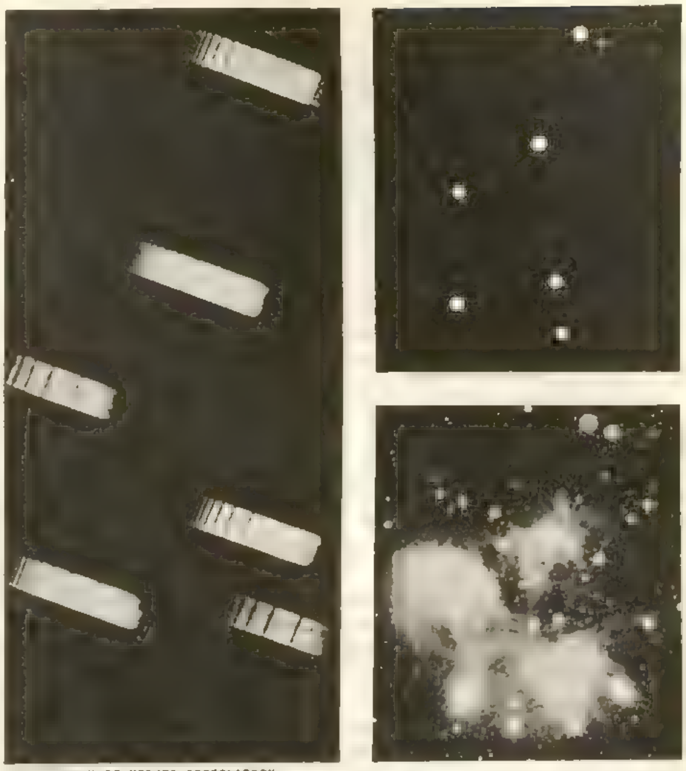
1. The Wolf-Rayet Stars.—Numbering fewer than 100 out of about 500,000 well-observed stars, these show continuous spectra upon which are superposed broad emission bands of atoms of a high degree of ionization, such as He II, C III, C IV, N III, N IV, N V, etc. These bands are often flanked on their violet sides by more or less diffuse absorption lines—a phenomenon which is believed to indicate that the lines and bands originate in shells expanding in all directions with velocities of several hundreds or even several thousands of km/sec. No forbidden lines have been found among the various features, which suggests that the shells cannot be very great in extent. A large fraction of the known Wolf-Rayet stars are close binaries, and in some the spectra of the companions turn out to be normal stars of class O or B. In these cases the Wolf-Rayet stars have much smaller masses.

C. S. Beals, at Victoria, discovered that there are two kinds of Wolf-Rayet stars, one group in which carbon lines are strong and another in which nitrogen predominates. Both groups have been arranged in sequences of decreasing ionization by Beals, and by W. A. Hiltner, but there are many peculiarities which need further explanation.

2. Stars with Bright Lines.—Some stars that are otherwise normal in spectrum have bright lines of hydrogen, and of other elements, in addition to the usual absorption lines. Sometimes these lines are broad, or even double; in other stars they are sharp and narrow. The broadening appears to be caused by Doppler effect, but the nature of the motions involved is not the same in all cases. In a large group of bright-line stars of classes B and A the motions are rotational in character, a tenuous shell or ring of gas revolving around a normal star which itself has such a rapid axial rotation that it approaches (or perhaps even reaches) the condition of rotational instability. Several stars of this kind happen to be members of close eclipsing double stars. The occulting star in its path eclipses first the approaching lobe of the rotating shell, then the star itself, and, finally, the receding lobe of the shell. In other stars the motions are those of expansion, as in Wolf-Rayet stars and in novae, and in still others they are of an irregular or turbulent character. In many stars the bright lines vary in the course of years.

A separate group is formed of various stars of the cooler classes which also show bright lines of H, Ca II, Fe I, Mg I, Si I, etc. Even the sun has weak central emissions in the lines designated as H and K of Ca II, and in other stars, like Arcturus, they are fairly strong. The origin of these emission lines is somewhat obscure. O. C. Wilson has shown that the widths of the Ca II emission lines are strongly correlated with the luminosities of the stars. F. Hoyle and Wilson have suggested that the turbulent motions which give rise to the line broadening is produced in the convective layer just below the photosphere.

Stronger emission lines are found in the spectra of Mira variable stars, but only at certain phases of their light curves. Just before maximum light the Balmer lines become conspicuous in emission. After this maximum they gradually weaken and disappear before minimum, while many metallic emissions, particularly Fe λ 4308 Å, Mg λ 4571 Å, and In λ 4511 Å reach their greatest strength on the descending branches of the light curves. This remarkable systematic behaviour has not been fully explained, although P. W. Merrill and A. D. Thackeray noticed that the unusual strength of Fe λ 4202 Å, and Fe λ 4308 Å, and In λ 4511 Å, involves fluorescence due to coincidence of their lower levels with the upper levels of strong ultraviolet lines of hydrogen and iron. Another significant fact, observed by Merrill and A. H. Joy, is that the Balmer emission lines are weakened by absorption of Ca II ions and TiO molecules, which must lie at greater heights than those



BY COURTESY OF YERKES OBSERVATORY

(LEFT) SLITLESS SPECTRA OF SOME BRIGHT STARS IN THE PLEIADES, TAKEN THROUGH A PRISM PLACED IN FRONT OF THE CAMERA LENS. THE SPECTRA APPEAR WIDENED BECAUSE OF THE MOVEMENT OF THE STARS, WHICH ARE ALSO SHOWN IN AN EXTRAFOCAL PHOTOGRAPH (TOP RIGHT). (BOTTOM RIGHT) A DIRECT PHOTOGRAPH OF THE CLUSTER OF THE PLEIADES CONTAINING HUNDREDS OF STARS ENMESHED IN A DIFFUSE NEBULOSITY

at which the emission lines are formed.

An interesting but small group of stars have spectra containing forbidden emission lines, *i.e.*, lines which are not produced in the laboratory because the corresponding transitions are not permitted by the rules of the quantum theory. Some of these stars have weak continuous spectra upon whose background are seen numerous bright lines of H, Fe II, [Fe II], Fe III, [Fe III], [Fe V], [Fe VI], [Fe VII], and probably even [Fe X] and other elements. The bracketed designations of elements stand for forbidden lines. Since the ionization potentials of these substances differ widely, by Saha's formula of ionization they could not all be observed in a gas of uniform temperature and pressure. The conclusion is that they come from different regions of a highly complex nebulous structure. P. W. Merrill found that most of these stars also show absorption bands of TiO. It is believed that they consist of three sources: a cool star, a hot companion, and a gaseous mass in which the radiation of the hot star causes ionization and excitation.

In some cases, like the star Antares, the binary can be observed directly. Its bright component is a cool giant, but it also has a faint, hot companion whose spectrum reveals forbidden [Fe II]. Swings and Struve found that in these stars the forbidden lines of [Fe II] and those permitted lines of Fe II, whose energy levels are low, are abnormally strong, so strong, in fact, that an impossibly low excitation temperature results when the intensities are interpreted by means of the Boltzmann exponential relation. The conclusion is that the mechanism of excitation consists in collisions between atoms and electrons, and not in the process of line absorption or ionization, and that the emission lines are produced in a small nebulosity, about 3 sec. of arc in diameter, which surrounds the hot star.

3. T Tauri Variables.—The spectra of these stars also show emission lines superposed upon fairly normal G or K type absorption spectra. According to A. H. Joy and G. Herbig, they are always found in or near large cosmic dust clouds. Herbig suggests that they are very young stars, of small mass, which are

still contracting and have not yet reached the main sequence. Their absorption lines are broadened by rapid axial rotation, as would be expected in young stars which have not experienced the braking action of magnetic fields or which have not had time to transfer their angular momenta to a system of planets. The spectra and brightness of the T Tauri variables undergo irregular variations.

4. Spectra of Shells.—In 1938 the spectrum of the star Pleione, one of the brighter members of the cluster of the Pleiades, underwent a sudden change. For about 30 years the spectrum had been that of a normal B star with very diffuse absorption lines of H and He. It now developed emission lines of H and a set of sharp absorption lines of H, Fe II, Ti II, Ca II, and other ionized elements. In the course of several years these sharp lines became very strong and indicated a tendency toward lowered ionization. A number of other stars which possess similar spectra are known, though they have not often been observed in the process of formation. The source of these spectra is designated as shells because it is inferred from the reduced intensities of a few characteristic lines that the gases, in which they are formed, are not at the surface of the star but are removed above it to several times its radius. The phenomenon producing these lowered intensities is designated as "dilution of radiation." To obtain normal intensities the populations of the different atomic levels must correspond to Boltzmann's relation. If the mechanism of excitation is the absorption of radiation this happens only when the radiation corresponds to conditions of thermodynamic equilibrium. When a gas is illuminated by a distant star, the radiation is "diluted" and those excited energy levels, which possess transitions to lower terms, are depopulated in the same ratio that the apparent area of the star as seen from the gas bears to the complete sphere. The phenomenon of dilution causes the lines of Mg II 4481 and Si II 4128, 4131 to appear abnormally weak. It also produces remarkable changes in the relative intensities of the He I lines. The formation of shells seems to be related to the process of axial rotation of the stars. In some shells (ζ Tauri and ϕ Persei) the shell is stratified and different layers of it revolve at different speeds around the main body.

5. Pulsating Stars.—Among the stars whose radial velocities undergo periodic changes are the variables of the Cepheid and the long-period groups. The former have velocity curves which may be described as the mirror images of their respective light curves. The variations in velocity are not caused by orbital motion, as in ordinary binaries, but are believed to be produced by periodic pulsations. The periods vary from a small fraction of a day to about 50 days. In some short-period variables two, three, or even more periods combine to produce beat effects in the radial velocity curves and the light curves. Some of these periods represent the higher modes of the fundamental period. The spectral classes also undergo periodic variations, being hottest near maximum light, or near minimum radial velocity. There is conflicting evidence from the luminosity criteria, but the hydrogen lines give maximum luminosity near maximum light; this correspondence is not physically obvious because the observed change in brightness (of the order of one stellar magnitude) is fully accounted for by the variation in temperature, and the elementary pulsation theory would predict maximum radius and, hence, minimum atmospheric pressure halfway from minimum toward maximum radial velocity. Evidently the hydrogen atmosphere is expanded at the time of maximum light, and the phenomenon is only indirectly related to the pulsation.

The Mira variables also are probably pulsating stars. Merrill and Joy showed that there is a relationship between their period and average spectral class, later spectral classes being generally associated with longer periods until type M7e is reached at periods of about 400 days. For the few stars of even longer period the type may be either early or late Me, which suggests that these long periods may not be the fundamental ones in some stars. There are also Mira variables of types Se or Ce.

Among the hotter stars there is a special group of pulsating variables, the β Canis Majoris stars, which have been studied most thoroughly by O. Struve. They have spectral types in the range B1 to B3, periods generally less than a quarter of a day, and very small amplitudes of variation. Their pulsations appear to be similar to those of Cepheids. At somewhat later types (F) there is another group of stars, the δ Scuti variables, which also have very short periods and small amplitudes.

6. Magnetic and Spectrum Variables.—In the *Henry Draper Catalogue* Miss Cannon noted that in the spectra of a few percent of the A-type stars the metallic lines are stronger than in normal stars. Those in which there is some enhancement of all the usual metals were defined as metallic-line (Am) stars by W. W. Morgan. More anomalous, however, are those in which the lines of certain ions only, such as Sr II, Eu II, Mn II, etc., stand out inconsistently. These are the peculiar A-stars (Ap). A. J. Deutsch pointed out that the hottest stars in the group tend to show enhancement of Mn II; as the temperature decreases down the sequence the stars tend to show, in turn, abnormally strong lines of Si II, Eu II, Cr II, and Sr II. The intensities of these lines in some Ap stars vary periodically, and in a few cases quite small variations in light occur also. The radial velocities tend to vary with the same periods as the line intensities.

The Ap stars that are spectrum variables have been found by H. W. Babcock to have large and variable magnetic fields. These fields may have strengths of several thousand gauss, and the changes in field strength are frequently great enough to reverse the polarity as determined from the Zeeman patterns of the spectral lines. Thus, the field of α^2 CVn changes from -1400 to $+1600$ gauss in a period of 5.7 days, with the Eu II lines strongest at -1400 gauss, and the Cr II lines at maximum when the field is positive. These great variations seem to rule out deep-seated abundance anomalies as the cause of the spectral peculiarities. The period of variation is apparently the period of axial rotation of the star, and it is generally agreed that some magneto-hydrodynamic process coupled to axial rotation must be involved. None of the detailed theories, however, has yet accounted for all the observations of the Ap stars.

Magnetic fields, extending up to several hundred gauss in strength, have been observed by Babcock in a few stars of later type also, including S-type stars, and in the cluster-type variable RR Lyrae.

7. Red Giants with Outflow of Matter.—As early as 1935 it was pointed out by W. S. Adams that on the Mount Wilson coudé spectrograms of α Orionis, and other cool stars, the strong absorption lines arising from the lowest atomic terms were double. The redward components were the normal, rather wide, lines of the reversing layer, but the sharp violet-displaced components were interpreted as being caused by upward-moving currents of gas. Later studies by A. J. Deutsch showed that these currents were expanding envelopes extending to several times the radii of the stars and moving at speeds of from 2 to 25 km/sec. This outflow of matter represents a considerable loss of mass, estimated by Deutsch to be at least 3×10^{-8} solar masses per year in the case of α Herculis. Similar circumstellar lines, but with smaller displacements, are found generally in ordinary giants of type later than MO.

This outflow is not surprising, since it had been pointed out earlier (G. A. Shajn) that the effective gravity may be nearly zero at the surfaces of cool luminous stars, so that only slight outward forces would be needed to allow atoms to exceed the velocity of escape. There are also arguments from evolutionary theory, and from the distribution of the several kinds of stars in clusters, that loss of mass should occur during some part of a star's lifetime.

Of course, loss of mass is involved in eruptive events such as novae, but the spectroscopic evidence now shows that gradual outflow of matter is also important, at least during the cool giant phase of evolution. That either sporadic or continual ejection of gases can occur in hot stars also is shown by spectroscopic observations of expanding shells in P Cygni stars and very luminous supergiants such as ρ Cassiopeiae.

8. White Dwarfs.—Since white dwarf stars have luminosities about 8 to 10 magnitudes below the main sequence, densities of

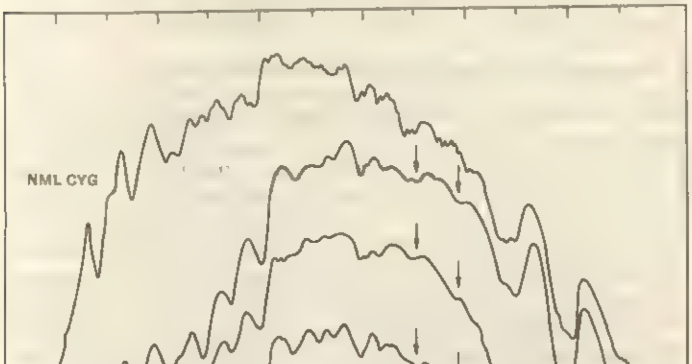
10,000 or more, and degenerate cores consisting largely of helium, with practically all their hydrogen confined in relatively thin surface layers, it might be expected that their spectra would be unusual. Actually they display a great variety of spectra, which have been divided by J. L. Greenstein into nine main types, each prefixed by the letter "D." Type DC has an almost continuous spectrum, but a wide range of colours. Others contain the usual Balmer lines, often greatly broadened by pressure effects. In the hotter stars helium lines are observed, with or without the hydrogen lines. In DF and DG spectra the H and K lines of Ca II dominate the spectrum. A few can be described as white-dwarf carbon stars, since they show two heads of the C_2 molecule near λ 4670 Å.

Perhaps the most puzzling is the spectrum of the star AC +70°8247, found by R. Minkowski to contain unidentified absorption bands near λ 4135 Å. For only a few of such white dwarfs has it been possible to construct model atmospheres that predict successfully their spectral characteristics.

9. Infrared Stars.—Until recently the coolest known stars were several Mira variables for which temperatures near 1600° K at minimum light had been measured radiometrically by E. Pettit and S. B. Nicholson as early as 1933. The marked improvement in the efficiency of cooled photoconducting cells as detectors of infrared radiation made it possible for the California Institute of Technology to organize in 1963 a systematic survey of the sky in two infrared ranges, 0.68–0.92 μ and 2.01–2.41 μ. Stars of known colour made up about 99% of those recorded, but the remaining 1% were so much redder that they are well described as infrared stars. Among them the only ones bright enough to appear even on the red plates of the Palomar Sky Survey were identified with very faint stars, below the 16th magnitude. Of these only the brightest, NML Cyg, could be seen visually with the 200-in. reflector.

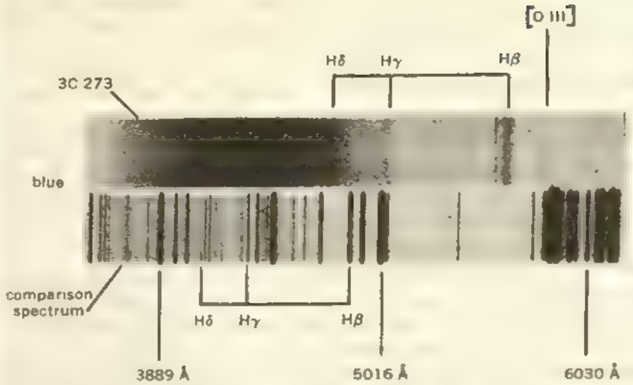
In general, the spectra of infrared stars in this region resemble those of carbon stars more closely than S- or M-type spectra, particularly in the deficiency of water vapour.

The Taurus infrared star has been shown to be a Mira variable, but the Cygnus object does not show large variations in light. The temperatures of these remarkable stars appear to be about 1000° K or even less.



BY COURTESY OF "ASTROPHYSICAL JOURNAL," © 1966. PHOTO, G. NEUGEBAUER

SPECTRAL SCANS OF INFRARED STARS NML CYG AND NML TAU. THE COOL CARBON STARS RS CYG AND WZ CAS ARE SHOWN FOR COMPARISON. THE SERIES OF DIPS BETWEEN 2.3 AND 2.45 μ ARE DUE TO STELLAR BANDS OF CARBON MONOXIDE



BY COURTESY OF THE CALIFORNIA INSTITUTE OF TECHNOLOGY, PHOTO, M. SCHMIDT

RED SHIFT OF HYDROGEN EMISSION LINES IN QUASI-STELLAR RADIO SOURCE 3C 273, PHOTOGRAPHED WITH 200-IN. HALE REFLECTOR; Hβ IS SHIFTED BY 771A, OR 16% ITS WAVELENGTH. FOR PURPOSES OF COMPARISON, THE SPECTRUM OF A DISCHARGE TUBE IS SHOWN BENEATH THAT OF 3C 273

VIII. QUASI-STELLAR RADIO SOURCES

Many radio sources have been identified with galaxies that can readily be observed optically. There were a few sources of very small angular diameter, however, that were found about 1960 to coincide with objects that appeared essentially stellar on direct photographs. These objects have dissimilar spectra with emission features for which no identification was found for several years. The clue to the problem was found by M. Schmidt in 1963 when he noted that some of the broad lines in the spectrum of the radio source 3C273 corresponded to Balmer lines of hydrogen shifted to the red by 16% of their wavelength, or 600 to 700 Å. The lines in other sources which lacked the Balmer lines were then identified by Schmidt, Greenstein, and T. A. Mathews with ultraviolet lines of other elements, including ionized magnesium and forbidden lines of ionized oxygen, etc., by assuming similar or even greater red shifts for those objects. Red shift is usually measured as the change in position of a line divided by its original wavelength, or $z = \Delta\lambda/_{\lambda}$. The largest values of z measured for any quasi-stellar sources are very nearly 2.0. With this shift the line of C IV normally observed in the vacuum ultraviolet at λ 1549 Å is actually found in the blue-green at about λ 4647 Å.

The quasi-stellar radio sources (or quasars) thus have greater red shifts than even the faintest and most distant of the ordinary galaxies that can be observed. This implies that the quasars are extremely distant and luminous objects, if their red shifts are considered as cosmological; *i.e.*, Doppler shifts obeying the same relation between shift and distance as those of other galaxies. One objection to this interpretation is that the light of several of the sources has been found to vary appreciably in intensity over intervals of even a few weeks, and it is difficult to account for such short-period variations in an object of galactic dimensions. If, on the other hand, they are objects lying within our own galaxy, another explanation must be found for the enormous red shifts. It seems certain, at least, that the quasi-stellar sources cannot be very dense dwarf stars close to the solar system, for those which are bright enough to have been found on old photographs show no significant angular motions in the sky.

See SUN; PLANETS; COMET; NEBULA; *see* also references under "Spectroscopy, Astronomical" in the Index.

BIBLIOGRAPHY.—One of the most comprehensive modern reference works is the series *Stars and Stellar Systems* (1966), under the general editorship of G. P. Kuiper. The topics discussed in the volumes of that series are as follows: Spectrographs—vol. 2, ch. 2 (I. S. Bowen), Spectral classification—vol. 3, ch. 8 (P. C. Keenan), Continuous spectra—vol. 6, ch. 8 (A. D. Code), Rotation and turbulence—vol. 6, ch. 8 (S. S. Huang and O. Struve), White dwarfs—vol. 6, ch. 19 (J. L. Greenstein). A more detailed account of spectral classification will be found in W. W. Morgan *et al.*, *An Atlas of Stellar Spectra* (1943). Magnetic Variables and Ap Stars are discussed in the chapter, "Magnetic Stars," by P. Ledoux and P. Renson in *Annual Review of Astronomy and Astrophysics*, vol. 4 (1966). Special types of stars brighter than magnitude 6.5 are given in *A Catalogue of Bright Stars*, edited by D. Hoffleit. The *Henry Draper Catalogue* was issued as *Harvard Observa-*

tory Annals, vol. 91–99 (1918–24). The theory of continuous spectra and absorption lines is covered by L. H. Aller, *Astrophysics*, vol. 1, *The Atmospheres of the Sun and Stars*, 2nd ed. (1963).

(O. St.; P. C. Ke.)

SPECTROSCOPY, X-RAY. A beam of short-wave electromagnetic radiation called X rays is scattered when it strikes a crystal, much in the same way a beam of light is diffracted by a ruled grating. The science that deals with the study of X-ray diffraction is X-ray spectroscopy. Its practical applications (including the chemical analysis of metal alloys, liquid solutions, and gases) are discussed in the concluding section of X Rays.

History.—Max von Laue discovered in 1912 that radiations in the X-ray wavelength range of 10 to 0.1 Å are diffracted by natural crystals whose layers of atoms are only a few angstrom units apart ($1 \text{ Å} = 10^{-8}$ cm.). Ruled gratings that work well in registering ordinary optical spectra with wavelengths down to about 2000 Å are of no value in diffracting X rays. To register the region between 2000 and 10 Å, other methods were subsequently developed.

In one of his first papers concerning X rays, W. C. Röntgen noted that X rays produced in an ordinary X-ray tube differ in quality when the vacuum of the tube and particularly the voltage on the tube are varied. It was shown early that the X radiations became more penetrating as the voltage on the tube was raised. The degree of penetration (usually measured in terms of the thickness of aluminum necessary to reduce their intensity by half) was used as a means of characterizing the quality of the radiation.

It was by this means also that C. G. Barkla was able to show that the different elements when excited so as to give off X rays all have their own characteristic radiations. For instance, for an element such as silver Barkla showed the existence of two characteristic radiations called the K and L radiation having very different penetrating powers.

These two characteristic radiations, the K series and the L series, were experimentally verified for a great number of elements. The measurements of their penetration in aluminum showed that their hardness (ability to penetrate) increases regularly for both series as the atomic weight of the emitting element increases.

Basis of X-ray Spectroscopy.—The diffraction patterns which W. Friedrich and P. Knipping obtained by allowing a fine beam of X rays to pass through a crystal as suggested by Laue induced W. H. and W. L. Bragg to perform the experiments which were the first step in the development of X-ray spectroscopy. As the Braggs showed, a monochromatic X-ray beam is reflected by a cleavage face (or any other atomic planes) of a crystal according to the ordinary laws of optical reflection; *i.e.*, the incident and the reflected beams are in the same plane, perpendicular to the reflecting face; further, the angles between these two beams and the reflecting face are equal. In addition to these laws the following condition must be fulfilled if reflection is to occur

$$n\lambda = 2d \sin \psi_n \tag{1}$$

where λ is the wavelength of the monochromatic radiation, d is the distance between two adjacent atomic layers parallel to the reflecting plane, ψ_n is the angle between the beam and the plane, and n indicates the order of the reflection (fig. 1).

This equation, generally known as Bragg's law, forms the basis for measuring the wavelengths of X rays. This law needs a small correction, due to the fact that the wavelength is slightly different in a vacuum (or air) than in a crystal.

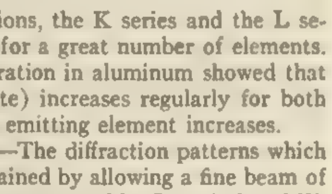


FIG. 1.—REFLECTION OF X RAYS ACCORDING TO BRAGG'S LAW (*see* TEXT)

Spectrometric Methods.—The wavelength region of ordinary X rays extends over the range 0.1 to 10 Å. The rays show different penetrating properties, and are termed hard (penetrating) or soft X rays. The region of hard X rays extends from 0.1 Å to about 1.5 Å, the wavelength where the absorption in air becomes important. In the soft X-ray region the great absorption in air makes it necessary to enclose the spectrometer in a vacuum container. The X-ray region from 10 Å and up to ultraviolet is called the ultrasoft region. Ultrasoft X rays can be studied by using ruled gratings.

FIG. 2.—REFLECTION OF X RAYS BY CRYSTALLOGRAPHIC PLANES: (A) REFLECTION MOUNTING FOR SOFT X RAYS, (B) MOUNTING FOR HARD X RAYS

In most spectrographs the X-ray beam incident on the crystal will be reflected by the crystallographic planes that are parallel to the crystal surface. This so-called reflection mounting of the crystal is best suited for the soft X-ray region (fig. 2A). In the hard X-ray region the angle ψ is small and the X-ray beam therefore sweeps almost parallel to the crystal. In this case the transmission mounting is the most suitable (fig. 2B). The short wavelength radiation can easily penetrate the crystal plate and the reflection occurs at crystal planes that are perpendicular to the crystal surface.

The first X-ray spectrographs used a plane crystal as is shown diagramatically in fig. 3. The different monochromatic rays which constitute the beam coming from the X-ray source and defined by a slit are reflected by the crystal at different angles according to Bragg's law (equation [1]). If the angles may vary from ψ_1 to ψ_2 only such rays are reflected whose wavelengths have values between λ_1 and λ_2 where

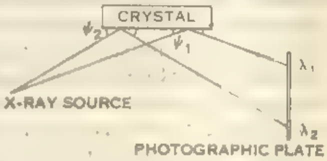


FIG. 3.—PRINCIPLE OF EARLY X-RAY SPECTROGRAPH USING PLANE CRYSTAL

$$n\lambda_1 = 2d \sin \psi_1$$
$$n\lambda_2 = 2d \sin \psi_2$$

The photographic plate therefore registers a spectrum ranging from λ_1 to λ_2. The region of wavelengths can be varied by turning the crystal. The wavelength of any spectral line on the photographic plate can be computed by equation (1) from its position on the plate.

FIG. 4.—PRINCIPLE OF THE TWO-CRYSTAL SPECTROMETER: (A) CRYSTAL AT PARALLEL POSITION, (B) CRYSTAL AT ANTIPARALLEL POSITION

The resolving power (and the sharpness of the lines) of a crystal spectrometer depends on the slit width and the distance between slit and crystal. In the two-crystal or double-crystal spectrometer the slit is replaced by another crystal (fig. 4). The first crystal selects a wavelength region which is analyzed by the second crystal. This crystal can be put in two different positions. At the parallel position the outgoing beam is parallel to the incoming beam throughout the wavelength region (fig. 4A). If the second crystal is turned to the antiparallel position the beam is deflected twice (fig. 4B). The crystal has thereby been turned through an angle of $180° - 2\psi$. The wavelength is then computed by equation (1). Each setting of the second crystal allows only a very narrow wavelength band to be reflected. The width of this wavelength band depends on the quality of the crystals and is usually narrower than the width of a spectral line. In using the instrument the crystal is gradually turned through the wavelength region to be investigated; the reflected intensity is measured in each setting of the crystal by means of some counting device; *e.g.*, a Geiger-Müller counter. Photographic registration is never used at this instrument.

The spectrometers mentioned above, especially the double-crystal spectrometer, have the disadvantage of low luminosity. A great improvement of luminosity was achieved by the principle of bent-crystal spectrometers, first formulated by J. W. M. DuMond and P. Kirkpatrick in 1930. A reflecting type of spectrograph with bent crystal was first constructed by H. H. Johann in 1931. A thin crystal plate, C in fig. 5A, is bent cylindrically to a radius r and a circle is drawn with r taken as its diameter. This circle, which corresponds to the Rowland circle in optical spectroscopy, causes rays coming from a point A on the circle to be reflected and focused on another point B on the circle. Because the centre of curvature of the crystal O also lies on the Rowland circle,

Bragg's law is met for the reflection. The focusing is not absolutely perfect, however, since the crystal does not follow the Rowland circle smoothly. The focusing defects are not severe if the crystal aperture α is small and the Bragg angles great (long wavelengths).

A transmission type of bent-crystal spectrograph was first built by Y. Cauchois in 1932. Rays from the focal spot directed toward a virtual object point A are focused at B (fig. 5B). The rays are reflected by the internal crystal planes and consequently this spectrometer is best suited for short wavelengths. As in the Johann spectrometer the radiation is detected by a photographic plate or by means of a counter with a slit.

Detectors for X rays.—In spite of its poor sensitivity, especially for hard radiation, the photographic plate is the most important X-ray detector and the only one which simultaneously registers all lines in the spectrum (*see* NUCLEAR INSTRUMENTS).

Modern precision spectrometers are often equipped with counting detectors, however, where the single quanta produce voltage pulses, counted and registered by electronic devices. Geiger-Müller counters or proportional counters are used for soft X rays and scintillation counters preferably for hard X rays. The proportional counter and the scintillation counter require more elaborate electronic devices than the Geiger-Müller counter but they offer the great advantage of giving output pulses that are proportional in voltage height to the energy of the radiation. By electronic circuits it is possible to discard pulses of wrong height

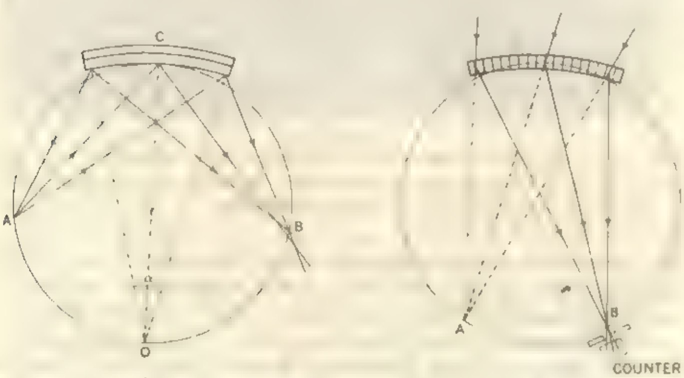


FIG. 5.—PRINCIPLE OF BENT-CRYSTAL SPECTROGRAPHS AS USED IN (A) A REFLECTING SPECTROGRAPH AND (B) A TRANSMISSION SPECTROGRAPH

arising from stray radiation, thus diminishing the background disturbances. Fig. 6 shows an X-ray spectrum detected by a curved-crystal spectrometer with a scintillation counter.

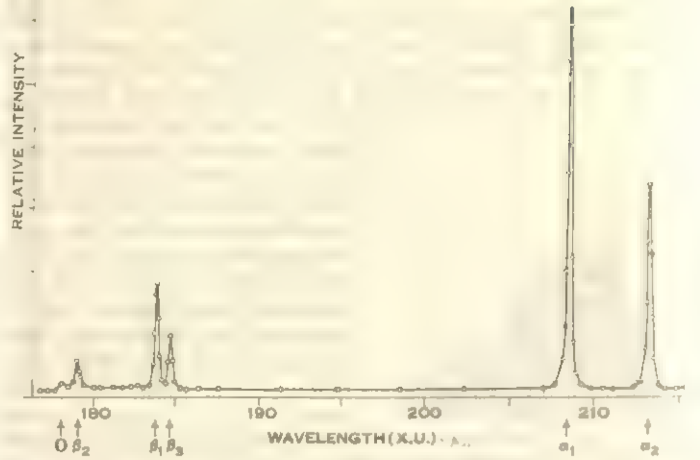


FIG. 6.—K-SPECTRUM OF TUNGSTEN REGISTERED BY A CURVED-CRYSTAL SPECTROMETER

Exciting X-ray Spectra.—To excite the X-ray spectrum of a substance, a small piece of it is placed as target material on the anticathode of an X-ray tube. After the tube has been exhausted to a suitable vacuum, a high potential (10 to 150 kv.) is applied to the electrodes of the tube, causing the anticathode to be bombarded by electrons from the cathode. The kinetic energy of the electrons imparted to the bombarded atoms is partially transformed into heat and light, and partially to X rays.

There are two main ways of producing a bombardment team of electrons with the X-ray tube. In the first method, the tube has a vacuum of 0.01 to 0.001 mm. Hg. The high voltage causes the gas remaining in the tube to ionize, with the positive ions being thrown against the aluminum cathode. The collision between the positive ions and the cathode sets free a number of electrons. The electrons travel in the opposite direction—from the cathode to the anticathode—and bombardment of the anticathode gives rise to X rays characteristic of the substance of the anticathode.

In the second method, the vacuum of the tube is much lower and there is no gas ionization of any importance. The bombarding electrons are supplied by the heated filament (usually of tungsten) in the cathode. This is the most commonly used type of X-ray tube because the electron beam can readily be controlled by adjusting the heating current in the cathode.

When analyzing materials, it is often necessary to change the substance on the anticathode; hence the tube must be built so the anticathode can be detached easily. This can be done by mounting the anticathode in a special joint. A tube which permits rapid change of both cathode and anticathode is shown in fig. 7. The tube is built of metal with water cooling to permit a high output of energy. A small window of aluminum or other low-absorbing material permits the study of radiation of longer wavelengths, which would be absorbed by the glass walls of an ordinary tube.

Another way of exciting X-ray spectra is to irradiate the substance with an intense beam of X rays. In this case the substance emits secondary rays which, with a few exceptions, are identical with the X radiation sent out by the substance when used as an anticathode in an X-ray tube. This method does not, however, give the same intensity as the former.

X-ray Wavelength Units.—In early X-ray work the wavelengths were given in angstrom (Å) units. Since this unit is too big for convenience, the X unit (X.U.; 10^{-11} cm.) was introduced.

The X-ray wavelengths are determined by crystal spectrometers and this requires an exact knowledge of the grating constant d of the crystal (*see* equation [1]). As experimental technique improved, the spectrometric measurements of X-ray lines became definitely more accurate than the measurements of the grating constants, which had to be determined from measurements of density and dimensions of cubic crystals. For these reasons the X unit has been defined by the lattice constant of the common plane in calcite set equal to 3,029.04 X.U. at 18° C. Later, E. Bäcklin, F. Tyrén, and other investigators succeeded in measuring some soft X-ray lines using ruled gratings. A direct connection between the angstrom and X units was thus established and by agreement 1,000 X.U. = 1.00202 Å.

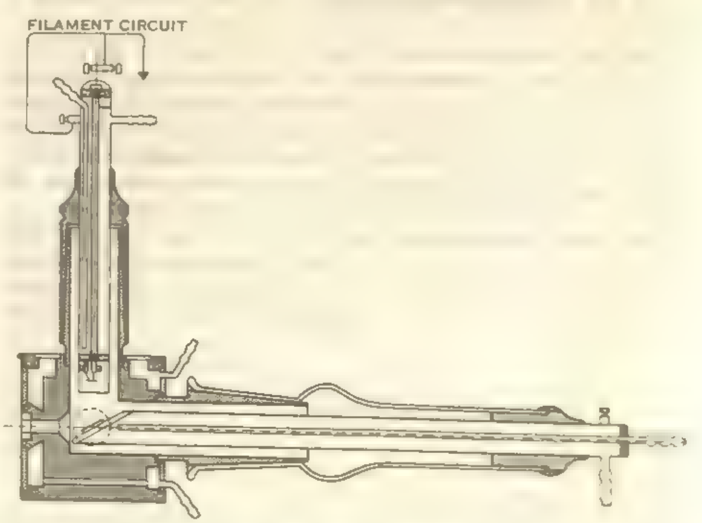


FIG. 7.—METAL X-RAY TUBE WITH HOT CATHODE FOR SPECTROSCOPIC USE

Different Kinds of X-ray Spectra.—The X-ray spectrum emitted from any substance is made up of two kinds of radiation, one of which shows a continuous distribution over a wide range of wavelengths, the other consisting of a few monochromatic rays overlapping the former (fig. 8). The first kind of radiation corresponds to the white light of optics, whereas the second is analogous to the line spectra. The continuous spectrum contains the greater part of the energy of radiation and is therefore the most important part for medical and other applications of X rays. This part of the X-ray spectrum shows qualitatively no dependence on the radiating substance, the intensity only being different from various anticathodes.

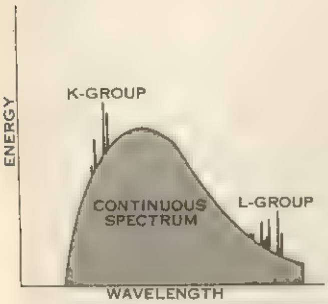


FIG. 8.—IDEALIZED X-RAY SPECTRUM SHOWING CONTINUOUS AND LINE-SPECTRA RADIATION

The line spectra on the other hand are characteristic of the emitting substance. They correspond to the K and L radiation of Barkla, but analysis has shown that each of these radiations consists of a group of monochromatic spectral lines.

The Continuous Spectrum.—The "white" or continuous radiation from an X-ray tube covers a rather wide region of wavelengths. The diagram of fig. 9 gives an idea of the distribution of the energy of the different wavelengths at X-ray tube voltages from 20 to 50 kv. The curves of distribution always start at a definite minimum wavelength (λ_{min}). The value of this wavelength decreases with increasing voltage. The wavelength (λ_{min}) or its corresponding frequency

$$\left(\nu_{max} = \frac{c}{\lambda_{min}}\right)$$

obeys the Einstein law

$$h\nu_{max} = eV \tag{2}$$

where e is the charge of the electron, V the voltage on the tube, and h Planck's constant. The amount eV is the energy received by the electron when accelerated by the voltage V. The most energetic radiation $h\nu_{max}$ in the spectrum occurs when all the electron energy is transferred to one X-ray quantum.

The energy eV of an electron or $h\nu$ of a radiation quantum is often expressed in the unit electron volt (ev). One electron volt is the energy of an electron (or another particle with the same charge) that has been accelerated by 1 v.

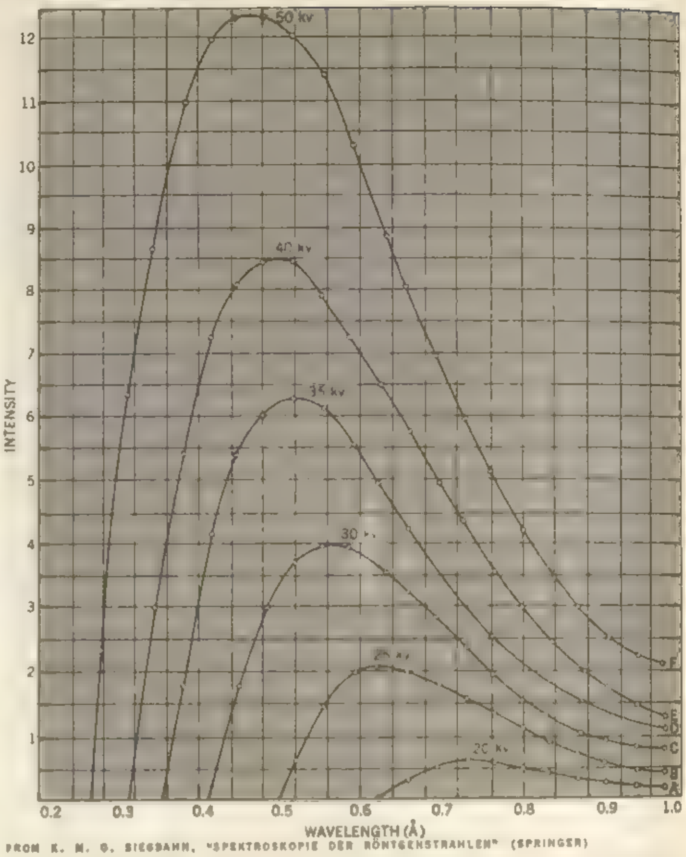
If the numerical values of e and h are introduced in equation (2) we obtain

$$\lambda\ \text{min} = \frac{12.3978}{V} \tag{3}$$

where the wavelength is expressed in angstrom units and V in kilovolts.

The total amount of energy (that is, the integral energy included by the distribution curves in fig. 9) is proportional to the square of the voltage on the tube, as has been verified by several investigators. Further, it has been found that for different elements used as anticathodes the total radiation is proportional to the atomic number of the element. For this reason, heavy elements have long been used as anticathodes in medical X-ray tubes.

The Line Spectra.—In 1913 and 1914 the young English physicist H. G.-J. Moseley published two remarkable papers in the *Philosophical Magazine* entitled "The High-Frequency Spectra of the Elements." Moseley showed that X-ray spectra include



FROM K. M. G. SIEGBAHN, "SPEKTROSKOPIE DER RÖNTGENSTRAHLEN" (SPRINGER)

FIG. 9.—INTENSITY DISTRIBUTION OF THE CONTINUOUS RADIATION FROM A TUNGSTEN ANTICATHODE AT VARIOUS VOLTAGES

line spectra of the same type as known in ordinary optics, and that these new spectra were built up in a more simple and regular way than is generally the case in spectra previously studied.

Two groups of spectral lines were found. One of these groups was identified with the K series of Barkla; the other, of longer wavelength and of more complex structure, corresponds to the L series. In 1916 M. Siegbahn discovered a new series of still longer wavelengths called the M series. The existence of series outside the K and L group had been suspected earlier by reasons of analogy. In the later development of X-ray spectroscopy, higher series, N, O, and so on, have been found.

A general scheme of the X-ray spectra is given in fig. 10, which contains the strongest lines of the three groups K, L, and M for every third element from sodium (11) to uranium (92). The diagram shows how all three series are regularly displaced to longer wavelengths as one proceeds from the heavier elements to the lighter. If the spectrum of some special element, say tungsten (74), a material commonly used for the anticathode, is considered, it will be seen that there are big gaps between the three groups where no lines are to be found. This fact is of predominant importance in interpreting the spectra in their relation to the structure of the atom.

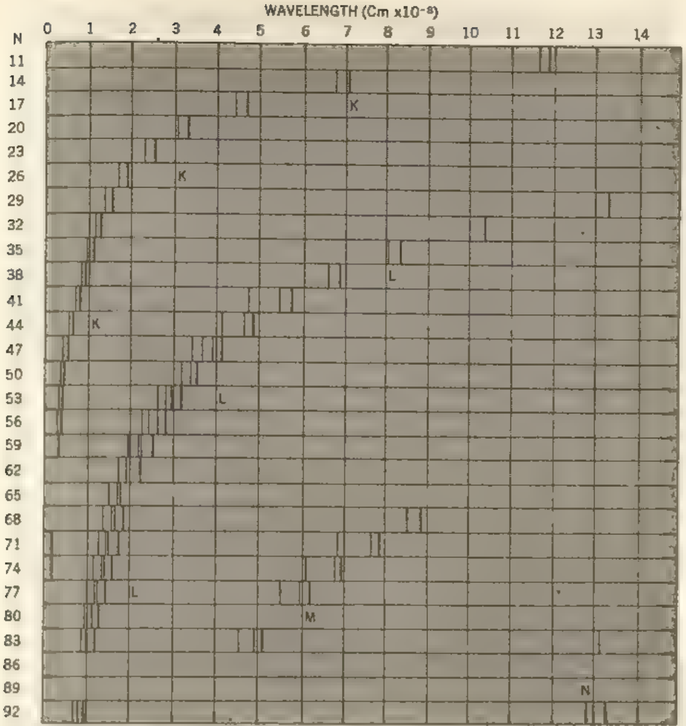
According to the Bohr-Rutherford hypothesis, atoms are built up of a positive nucleus of comparatively small dimensions which is surrounded by a number of electrons. The electrons are grouped in definite shells formed by given numbers of electrons. Closest to the nucleus and most strongly bound are two electrons forming the K shell. The K shell is surrounded by the more loosely bound L shell, consisting of 8 electrons, and farther out is the M shell with 18 electrons at the maximum.

In an X-ray tube an electron hits an atom in the anticathode and knocks out an electron from one of its shells.

An energetic electron is able to remove an electron from the

Electrons in Noble Gases

	K	L_I	L_{II}, L_{III}	M_I	M_{II}, M_{III}	M_{IV}, M_V	N_I	N_{II}, N_{III}	N_{IV}, N_V	N_{VI}, N_{VII}	O_I	O_{II}, O_{III}	O_{IV}, O_V	P_I	P_{II}, P_{III}
2 Helium	2	..	..	..	..	..	..	..	..	..	..	..	..		..
10 Neon	2	2	6	..	..	..	..	..	..	..	..	..	..	.	
18 Argon	2	2	6	2	6	..	..	..	..	..	..	..	..	.	..
36 Krypton	2	2	6	2	6	10	2	6	..	..	.	..			.
54 Xenon	2	2	6	2	6	10	2	6	10	..	2	6	..		.
86 Radon	2	2	6	2	6	10	2	6	10	14	2	6	10	2	6



FROM K. M. G. SIEGBAHN, "SPEKTROSKOPIE DER RÖNTGENSTRAHLEN" (SPRINGER)

FIG. 10.—THE STRONGEST LINES OF THE SEVERAL X-RAY SERIES ARE SHOWN FOR EVERY THIRD ELEMENT

K shell; in the case of tungsten the necessary energy is 69.51 kilo electron volts (Kev). The atom is now excited and contains the energy 69.51 Kev. The hole is filled immediately by an electron from an outer shell; *e.g.*, the L shell. The atom is now lacking one of its L electrons and contains the corresponding energy of 10.20 Kev. The energy difference 69.51 — 10.20 = 59.31 Kev is transmitted as an X-ray quantum $h\nu$ and the wavelength corresponding to 59.31 Kev can be calculated from equation (3) to be 0.209A (208.6 X.U.). If the hole in the K shell is filled by other electrons, other X-ray lines are obtained. They all have energies of the order of 60 to 70 Kev corresponding to 0.18 to 0.21 Å (178 to 213 X.U.). These lines form the K series of tungsten (fig. 6). A hole in the L shell involves an energy of about 10 Kev, and when this hole is filled up by transitions from the M, N, or other outer levels, the L series lines are emitted. The corresponding wavelengths are 1.2 to 1.6 Å (*see* fig. 10).

The transitions between the various energy states of the atom where different electrons are missing can be illustrated by an energy level scheme (fig. 11). The different levels are characterized by quantum numbers. The main quantum number n determines the various shells: K shell has $n = 1$, L shell has $n = 2$, etc. With one L electron missing, the atom is excited to the L level. An analysis shows, however, that the energy of the atomic excitation is dependent on which of the eight L electrons is ejected; in reality there are three different L levels, L_I, L_{II}, and L_{III}. Another quantum number l is now introduced. There are two L electrons with $l = 0$. If one of these electrons is missing, the atom is excited to the L_I level. The six other L electrons have $l = 1$, and the corresponding levels are L_{II} and L_{III}. The M shell with $n = 3$ gives rise to five levels: M_I ($l = 0$, 2 electrons); M_{II}, M_{III} ($l = 1$, 6 electrons); and M_{IV}, M_V ($l = 2$, 10 electrons). The building up of the electron configurations for some atoms (the noble gases) is given in the Table.

In fig. 11 some of the various transitions in the atom are represented by lines connecting the levels. The lines originating from the K level represent the K group, etc. (The four lines from the K level in fig. 11 correspond to the lines $\alpha_2\ \alpha_1\ \beta_3\ \beta_1$ in fig. 6.)

The electrons do not jump haphazardly from one level to another; the transitions are regulated by selection rules. The most easily occurring transitions, which give rise to the strongest lines, pass between levels where the quantum number l changes by one unit. The transitions noted in fig. 11 are some of the possible dipole lines. Other transitions are not strictly forbidden, but they give rise to very weak lines called quadrupole lines. Such a line is the KM_{IV}, KM_V line.

The Absorption Spectra.—In addition to the emission spectrum of X rays, there is also an absorption spectrum analogous to the absorption spectra of ordinary optics. Such spectra are obtained when a thin foil of some substance is placed between the X-ray source and the detector; *e.g.*, photographic plate. Instead of a continuous blackening from the white X radiation the photographic plate shows one or more sharp edges where there is rapid change in the blackening—the so-called absorption-edges.

These absorption discontinuities are explained as follows: An X-ray quantum $h\nu$ impinging on an atom may transfer its energy to a K or L electron, which will be ejected from the atom. To eject a K electron from a tungsten atom, an energy of 69.51 Kev is required, as mentioned before. This energy corresponds to a wavelength of 178 X.U. Obviously only quanta with an energy higher than 69.51 Kev (that is, radiation of wavelengths shorter than 178 X.U.) can eject K electrons. Radiation of wavelengths longer than 178 X.U. may pass relatively easily through a tungsten foil, but wavelengths shorter than the limit are strongly absorbed due to the process mentioned. When registered by a spectrometer, this results in a discontinuity in the intensity curve.

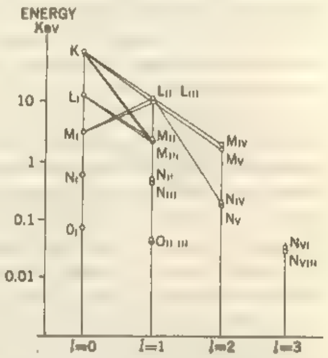


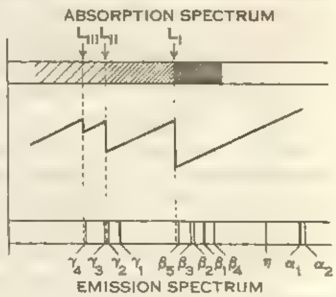
FIG. 11.—ENERGY LEVEL SCHEME FOR TUNGSTEN

Since the L shell has three energy levels, the L-absorption spectrum shows three discontinuities corresponding to the three levels L_I, L_{II}, and L_{III}. Fig. 12 shows an L-absorption spectrum together with the L-emission lines.

Ultrasoft X rays.—As already mentioned, the crystal grating method has its limitations in that wavelengths longer than double the distance between the atomic layers (say 10 or 20 Å) cannot be selectively reflected. This is clear from Bragg's law (equation [1]) which shows that λ cannot be larger than $2d$.

Fortunately it was found that ruled gratings when used in nearly grazing incidence gave nice spectra of the radiation with wavelengths down to the region where the crystals had their limits. This method was first used by R. A. Millikan for the optical spectra and then by A. H. Compton and R. L. Doan, M. G. Siegbahn, and others for the study of the ultrasoft X-ray spectra. An important result of this work was that it suggested a direct comparison between optical and X-ray wavelengths, resulting in a determination of atomic distances and the previously mentioned relation between the angstrom and X units.

Subsequently in this way a very exact determination of the value for the fundamental electric unit—the electronic charge—was worked out by Bäcklin. It was further possible to extend the different X-ray spectral series into the new wavelength region and to find lines belonging to the N and to the O series.



FROM K. M. G. SIEGBAHN, "SPEKTROSKOPIE DER RÖNTGENSTRAHLEN" (SPRINGER)

FIG. 12.—ABSORPTION AND EMISSION SPECTRUM SHOWING THE RELATIVE POSITIONS OF DISCONTINUITIES AND LINES

Miscellaneous Influences on X-ray Spectra.—The most striking difference between the ordinary optical spectra and the X-ray spectra is the fact that the former show a marked resemblance among elements with the same chemical character and change considerably from one group of chemically analogous elements to another, whereas in the X-ray spectra no indication of the chemical nature of the elements is to be found. This is readily explained by the Bohr-

Rutherford concept of the structure of the atom. From this interpretation, the optical spectra arise from the surface of the atom, which is also the seat of the chemical bonds, while the X-ray spectra are given off from the inner parts of the electronic atmosphere of the atom. It must, however, be expected that a small influence may be observed if an atom emitting an X ray is acted upon by neighbouring atoms. Determining the degree of this influence depends on the sensitivity and accuracy of methods used in studying X-ray spectra. By increasing the precision of study techniques, it has been possible to detect a number of such influences. It was first found by J. Bergengren that the K-absorption edge for phosphorus was displaced slightly when the element was modified. The line spectra are also influenced to a certain degree by the chemical bonds. *See* also QUANTUM MECHANICS; RADIOACTIVITY; RADIOLOGY; SPECTROSCOPY; X RAYS.

See B. D. Cullity, *Elements of X-ray Diffraction* (1956); A. Guinier, *X-ray Diffraction in Crystals. . . .* (1963). (M. SIE.)

SPEECH AND LANGUAGE, in psychology, present a set of ancient problems. Plato set one such question in his dialogue *Cratylus*: ". . . Cratylus has been arguing about names; he says that they are natural and not conventional; not a portion of the human voice which men agree to use; but that there is a truth or correctness in them, which is the same for Hellenes as for barbarians." Many centuries of argument added little to Plato's discussion of symbolism in language until a simple 20th-century experiment gave it new life.

People were asked to match words from their own language with those from another that was completely unfamiliar to them; speakers of English, for instance, were given the antonyms small-large and the unfamiliar equivalent pair from Chinese, *ch'ing-ch'ung*. Nearly 90% of those tested guessed correctly that large means *ch'ung* and small means *ch'ing*. If there were nothing in the sound of the foreign words to suggest the translation, by chance about 50% would be expected to guess correctly. Many such experiments have been conducted and, while improvement over chance performance is generally very small, the evidence suggests that parts of many languages do share a principle of correctness "which is the same for Hellenes as for barbarians."

New experiments and developments in research technology, structural linguistics, and information theory have shed light on many other aspects of psychology and language (*see* INFORMATION THEORY; LINGUISTICS). These include the recognition of speech, determinants of word association, the nature of meaning, the way children learn their native language, and the relation between language and thinking.

Recognition of Speech.—The sound spectrograph provides detailed visible records of the acoustic properties of speech sounds (*see* VOICE: *Methods of Investigation*). Such records (spectrograms) have been used in the search for acoustic cues that permit recognition of given vowels or consonants regardless of the speaker. The relative importance of each cue is then investigated by painting artificial spectrograms and converting these visual patterns into synthetic speech that listeners are asked to identify. Such studies show that cues for recognition are found primarily in the two or three most intense formants (frequency regions of the greatest acoustic energy), in the formant transitions between consonants and vowels, in the frequencies of the bursts in the release of stop consonants, and in the durational aspects of certain sounds (*see* PHONETICS: *Phonetic Description*). The work of R. Jakobson (*q.v.*), M. Halle, and their collaborators suggests that cues to recognition in any language may be very economically described in terms of a small number of acoustic features. Such progress has led to electronic equipment (*e.g.*, the "Voder") that produces recognizable speech (*see* ROBOT); by the 1960s experimental devices could recognize spoken sounds, heralding such applications as a voice typewriter and a telephone for the deaf.

In ordinary experience some sounds occur with relatively high frequency, and people approach the task of recognition with a set of weighted expectancies. A syllable like *srate*, which does not occur in English, is less likely to be correctly identified than is a word like crate. Visual presentation of a form like *vernalit* (which technically approximates English in terms of information theory) is more likely to elicit recognition among English speakers than is a form like *ozhgpmtj*.

Word Association.—In the word-association experiment a subject is presented with a stimulus word (*e.g.*, hot) and asked to provide a response word (*e.g.*, cold). The period of time between the two is called the response latency. After its invention by Sir Francis Galton, the experiment was applied to psychiatric diagnosis and to the detection of lies and guilty knowledge in criminology. A subject's immediate response may be revealing in an obvious way, or an unusually long latency may indicate that such a revealing response has been deliberately suppressed.

The usual determinants of an associative response are habits based on ordinary language usage (as are the weighted expectancies that affect recognition). The probability of a particular response word is related to how frequently the word occurs in speech and writing, and how often it follows the stimulus word in ordinary usage. The word counts of E. L. Thorndike (*q.v.*) and I. Lorge are commonly used compilations of such data from printed materials.

In 1910 G. H. Kent and A. J. Rosanoff published the responses of 1,000 adults to 100 English words; several sets of word-association norms have appeared since. The most popular (primary) response tends to be given by a large percentage of subjects (*e.g.*, 65% gave "light" in response to "lamp" in the Kent-Rosanoff study), the list trailing off with many responses given only once. J. J. Jenkins and W. A. Russell, who published the University of Minnesota word-association norms (1954), compared sets of data collected since 1910 and found an increasing proportion of popular responses. This may be a result of greater uniformity in usage arising from the growth of mass media communication.

Language associations also have been studied through verbal conditioning; in one type of experiment a word is presented along with the sound of a loud buzzer. The buzzer tends to elicit a change in the electrical resistance of the subject's body (*see* PSYCHOGALVANIC REFLEX). After several repetitions the reflex occurs in response to the word alone (*see* CONDITIONING). The varying intensities of the reflex in response to associated words have been used to infer the implicit verbal associations of the subject. Extensive use of the method in the Soviet Union and the U.S. has shown, for example, that young children tend to exhibit a generalized reflex to words that sound alike (*e.g.*, earn-urn); adults are more likely to generalize to words of similar meaning (*e.g.*, urn-vase).

Meaning.—One way to find out if someone knows the meaning of a word is to ask him for a definition or synonym as in verbal intelligence testing. Or, the person can be asked if the word applies to this or that referent object. J. Piaget, for example, in studying how children understand the word "alive" asked whether a dog is alive, a table is alive, and so on (*see* CONCEPT FORMATION, PSYCHOLOGY OF; MEANING; SEMANTICS IN LINGUISTICS). These methods are concerned with the central literal meaning of a word.

The semantic differential, developed by C. E. Osgood and others, is a technique designed to assess subtler aspects of meaning. This work began with efforts to identify basic dimensions of connotative meaning in English. In a typical study subjects were given a list of 20 assorted words (*e.g.*, lady, boulder, sin) plus 50 seven-point rating scales with extremes labeled as good-bad, light-heavy, red-green, sweet-sour, active-passive, and so on. The task was to place each of the 20 words on each scale. Although some of the required judgments were prosaic (*e.g.*, deciding whether boulder is light or heavy), the scales often lacked literal application (*e.g.*, is sin red or green?). Many scales were found to function as near-synonyms, with judgments on one being highly predictable from those on another. Statistical treatment of the data with a method called factor analysis revealed three major independent dimensions identified as evaluation, potency, and activity. Similar studies (including some with other languages) have yielded the same dimensions, permitting economical measurement of meaning through a few scales selected to represent each major dimension. Most commonly, four scales are used to represent evaluation (usually including good-bad); three to represent potency (usually including strong-weak); and three for activity

(usually including active-passive). Semantic differential measures of meaning have been widely applied, particularly to the study of social attitudes, experimental aesthetics, in consumer research, and in evaluating the effects of psychotherapy.

Children's Speech.—A very large part of language is learned in the typical child's first three years. Babbling in the first year and a half often includes sounds not used in the parental language (*e.g.*, umlaut vowels for an infant in an English-speaking household) and those the child later may have difficulty articulating (*e.g.*, [l] and [r]). Vowels predominate over consonants in babbling, and studies of early vocalization (O. C. Irwin and others) indicate that front vowels (*e.g.*, [I] as in hid) tend to appear before back vowels (*e.g.*, [o] as in blow); the first consonants are likely to be velar and glottal sounds formed in the back of the throat (*e.g.*, [g] as in good). Imitation of adult sounds tends first to appear about six months of age; the first intelligible word commonly at about the end of the first year; and simple sentences near the end of the second year. Although normal children may vary widely from this schedule, maturation does tend to set a timetable for human speech development.

Children's understanding of word and sentence meaning shows a long and complex development. L. S. Vygotsky and Piaget have shown that children at first conceive of the relation between words and referents as natural and unalterable rather than as local conventions. O. Jespersen, H. Werner, and many others have shown that a child often overgeneralizes in his first use of words; for example, calling all male adults "daddy" or all quadrupeds "doggie." Piaget has traced the development of understanding of the complex concepts associated with such words as "because," "although," and "more."

A terminal linguistic achievement is the ability of children to produce sentences they have never heard but that are meaningful and grammatical. The child learns to use a grammar or a set of rules (which he cannot explicitly state) for the combination of words. Studies of children learning Russian (K. Chukovsky), French (P. Guillaume and A. Grégoire), and English (J. Berko) show that children at first operate with a simply regular grammar. Children learning English, for instance, are likely to say "digged" and "ringed" rather than "dug" and "rang," or "gooses" rather than "geese." The rules of adult English involve many qualifications and exceptions, and young children treat the language as if it were a more simply consistent system than it is.

Thinking.—It is popularly believed that all normal people are aware of reality in much the same way and that languages are simply conventional vehicles for thinking and play no part in its formation. A very different view (E. Sapir, B. L. Whorf) holds that the same reality can be variously conceived and that each language embodies and perpetuates a particular conception.

There is very little evidence of a direct effect of language on thinking, but countless examples of linguistic differences strongly suggest a parallel difference in thought processes. Numerous languages have a single word for the colours green and blue. Eskimo have words for many more varieties of snow than do Europeans or Americans. The Hanunóo people of the Philippine Islands have 92 names for what an English-speaking layman would call simply rice. From many such examples the fineness of differentiation in vocabulary seems to increase with the importance of the given referent field to the speakers of a language. People whose language fails to make a given distinction are probably less likely to think and perceive in such terms than are those whose language does. It seems likely, however, that the distinction can be learned if it becomes important and that the influence of the native language is not irrevocable.

See also THINKING AND PROBLEM SOLVING.

BIBLIOGRAPHY.—R. W. Brown, *Words and Things* (1958); G. A. Miller, *Language and Communication* (1951); H. Rubenstein and M. Aborn, "Psycholinguistics," in *Annual Review of Psychology*, vol. 2 (1960); C. Cherry, *On Human Communication* (1957); C. E. Osgood, G. J. Suci, and P. H. Tannenbaum, *The Measurement of Meaning* (1957); S. Saporta (ed.), *Psycholinguistics: a Book of Readings* (1961); L. S. Vygotsky, *Thought and Language* (1962); C. E. Osgood, "Psycholinguistics," in S. Koch (ed.), *Psychology: a Study of a Science*, vol. 6 (1963); H. Lane, "The Motor Theory of Speech Perception," *Psychological Review*, vol. 72 (1965). (Ro. W. B.)

SPEECH DISORDERS. Speech may be said to be disordered, and so to present a problem for the speaker, or his listeners, or both, when listeners pay uneasy attention to the manner of speech, have difficulty making it out, or are distracted from what is said by the way in which it is said. The major personal effects of speech disorders depend on the basic role of speech in human relationships. Insofar as such disorders interfere with ability to relate to others through speech, and so to achieve social and personal adjustment, they affect personality. Some authorities regard certain kinds of speech disorders as symptoms of emotional disturbance or personality maladjustment. Research evidence, however, indicates that while most persons with speech disorders have educational, vocational, social, and personal problems because of impaired speech and because of their own and other people's reactions to it, as a group they do not significantly differ in basic personality from persons without speech disorders.

So few estimates of prevalence of speech disorders in different countries are available that meaningful comparison is impossible. Further, available estimates reflect cultural differences in the value attached to quality of speech, especially in small children. Substantial data indicate that among the Ute, Bannock, and Shoshone Indians of North America, as well as in certain cultures in other parts of the world, lack of fluency in children is uncritically accepted and the problem of stuttering does not occur. On the other hand, in modern Western cultures, the strong trend toward "standard speech," enhanced by radio and television, encourages conformity and so directs attention to unusual or disordered speech. It seems likely that, other things equal, the more value a society attaches to "correct" or "normal" or "good" speech, the larger will be the proportion of its members regarded as having disordered or unsatisfactory speech.

The United States Office of Education has variously estimated that 3% to 4%, or roughly 2,500,000, of school-age children in the United States have speech disorders. For the rest of the population, an extremely conservative figure is 3%, or close to 5,000,000. This gives a national total of roughly 7,500,000 speech-handicapped persons. Office of Education statistics indicate further that speech impairment is the commonest handicap in U.S. school-age children: of the 105 out of every 1,000 that are handicapped in some way, 35 have speech disorders, 23 are mentally retarded, 20 are emotionally disturbed, 10 are physically disabled, 10 have handicapping illnesses, 6 have hearing problems, and 1 is blind or partially seeing.

Speech pathology and audiology have developed as a theoretical and clinical service field in the United States. Clinical work with speech disorders (usually in closer relationship to medicine than it is in the United States) is fostered also in Great Britain, the Scandinavian countries, most of the rest of continental Europe including the U.S.S.R., Japan, South Africa, South America, and such parts of the British Commonwealth as Canada, Australia, and New Zealand. There is promising evidence that as teachers, physicians, and parents learn more about how, for example, stuttering begins, its prevalence is decreasing; and that new modes of treatment of stuttering in its early stages, stressing counseling of parents and teachers, are increasingly effective. There is compelling evidence that most children who misarticulate speech sounds can be taught to form the sounds normally; that most voice problems yield to treatment; that most patients who undergo laryngectomy (surgical removal of the "voice box") can learn to speak again; that patients with aphasia recover useful language and speech. Such organic impairments as cerebral palsy, severe hearing loss, and cleft palate, however, even with all possible physical restoration, in some cases set limits to speech improvement. Nevertheless, persons without tongues have learned to speak, and the large numbers of persons with cleft palates who speak normally or nearly so reflect recent advances in surgery, prosthetic and orthodontic dentistry, and speech pathology and audiology.

DESCRIPTION, CAUSE, AND TREATMENT OF SPEECH DISORDERS

Classification Schemes.—Disorders of speech are commonly classified as follows: (1) disorders of voice; (2) disorders of

articulation; (3) stuttering (stammering); (4) aphasia (dysphasia); (5) retarded development of speech or language; and (6) speech disorders associated with hearing loss or deafness, cleft lip and cleft palate, cerebral palsy and other neuromuscular impairments, laryngectomy, facial or oral injury or deficiency, mental retardation, and emotional illness.

There are other classification schemes. Some speech pathologists add a category called cluttering, usually said to be distinguished by rapid rate, indistinct articulation, and some disfluency, but without the tension, anxiety, and blocking reactions that characterize stuttering. The definition and clinical significance of "cluttering" are in process of clarification. A controversial diagnostic category is "congenital" aphasia, referring to a condition said to be due to brain deficiency or to brain damage before, during, or after birth but before development of speech. "Congenital" aphasia is not universally accepted because some insist that there can be no aphasia if there is no developed language function to be lost. The controversy is complicated by the difficulty in most cases of clearly demonstrating brain damage. "Retarded speech or language development" is a more widely accepted diagnostic term.

The Total Speech Problem.—In presenting, below, descriptions, causes, and treatments of the various speech disorders (*see* MOUTH; VOICE; and EAR, ANATOMY OF, for anatomical details mentioned), each disorder is considered to be ono part of a total speech problem; each such problem has as participants not only the speaker but also his listeners. Some listeners contribute to solving or coping with the problem, others make it worse. Some even help to cause it in the first place, as, for example, in the beginning of stuttering, when the child's hesitations and repetitions appear to his parents to be a problem before they do to the child.

The total speech problem has three aspects: first, the speech characteristic that is regarded as a disorder; second, the listeners' reactions to the speech; and third, the speaker's reactions to the listeners' reactions and to his own speaking behaviour. Listeners' revealed feelings—and in the case of children the most important listeners often are the parents—about speech affect the speech, making it more or less hesitant, disfluent, or tense. The speaker's reactions to the listeners' reactions may thus either complicate the disorder or help in the direction of more acceptable speech.

The three aspects of the speech problem may be affected by various factors: the speech characteristic itself may be complicated by such organic conditions as cleft palate, dental malformations, or brain damage, or by such nonorganic factors as poor adult speech models and lack of stimulation and motivation. Listener and speaker reactions may be affected by level of education, intelligence, experience with and understanding of speech disorders, and personal and social adjustment. Listener and speaker reactions of course vary. What one listener hears as severe disfluency, another may regard as moderate and a third as normal for age and circumstance; likewise, the speaker may be unconcerned or moderately or deeply concerned about listener reactions and about the speech characteristic to which they are related.

Diagnosis and Therapy.—For any type of speech disorder, diagnosis involves: (1) detailed description of what the speaker does that makes his speech clinically significant; (2) identification of the times, places, listeners, and types of speaking in relation to which he does or does not do these things; and (3) determination of whether physiological or psychological factors are fostering the disorder.

Basic aspects of therapy include: (1) identifying the persons other than the speaker who are most concerned in the problem, counseling them, and enlisting their support; (2) making use of family, school, and community resources; (3) counseling the speaker as his age and maturity warrant; (4) speech retraining by individual or group methods that enable the speaker to hear, see, and feel what he does that interferes with his speech; to listen to the differences between disordered and normal speech patterns; to practise the normal patterns, or his best approximation of them, in drill sessions and increasingly in social situations Tape recorders, mirrors, sound amplifiers, and other instruments and positive rewards and reinforcements are useful in facilitating the desired learning. In any case of impaired speech associated with physical or psychological factors—as, for example, misarticulations related to mouth injury, cleft palate, or cerebral palsy, or speech impairment related to emotional illness or loss of hearing—appropriate medical care or psychological counseling should precede or accompany speech therapy.

Disorders of Voice.—The voice may be pitched too high or too low; may be too loud or too soft; may be distractingly nasal, hoarse, harsh, or breathy; or may lack expressive variation. The voice may be affected by enlarged tonsils and adenoids, throat infections, the common cold, vocal abuse, growths on the vocal cords, and paralysis. Most voice disorders, however, are associated not with organic pathology but with negative emotional reactions (resentment, hostility, anger, or shame), faulty models for learning speech, shyness and insecurity, or adolescent "change of voice." In speech therapy, ear training may be emphasized for monitoring pitch, loudness, quality, and flexibility. A change of pitch level may be prescribed to reduce tension and increase expressiveness.

Disorders of Articulation.—About 80% of the children in public-school remedial speech programs in the United States have disorders of articulation. Often these result from faulty learning due to meagre speech stimulation or from inadvertent positive reward by parents through their acceptance of speech with such errors. It is an unresolved question whether many of these children, concentrated in the early grades, would learn to articulate correctly in the normal course of maturation. Practically all articulation errors are omissions of speech sounds (as in *pay* for *play*), distortions of sounds (slighting sounds or overarticulating them, as in the "whistling" *s* or in the "mushed" *s* that resembles *sh*), or substitutions of one sound for another (as *wun* for *run* or *thum* for *sum*). A speaker may misarticulate one or more sounds or sound blends (as *st*, *pl*) consistently or inconsistently.

Speech retraining is concerned chiefly with the defective sounds. If the speaker is a child still learning to speak and making the errors common for his age, he may need no speech instruction; the clinician works instead with the parents, teachers, and other listeners to encourage normal maturation of the child's speech.

Stuttering (Stammering).—Stuttering, like all disorders of speech, is not concerned with the speaker alone. The characteristic to which his listeners originally react is disfluency, as seen in repetitions, "uh uh," hesitations, and interruptions common in normal childhood speech and occurring also in normal adult speech. Half or more of all parents do not notice their young children's disfluencies. Most parents who do notice them regard them as normal. Roughly one parent in 100 reacts with serious concern. Sensing this reaction, the child may gradually come to doubt that he can speak well enough to please his parents, and to fear the consequences if he does not. As he tries harder to speak more fluently than he can, he becomes less spontaneous, more hesitant, tense, and uneasy. Straining to do better, he presses his lips together, holds his breath, and performs other actions that interfere with speech. He has more difficulty speaking; his parents become still more concerned. They try to be helpful by urging him to relax, go slowly, start over; he feels more discouraged and becomes more tense and distressed. And so stuttering develops.

The single question "What causes stuttering?" therefore breaks down into four questions: (1) Why does one parent in a hundred make a serious issue of the child's disfluencies? (2) What causes the disfluencies? (3) How are they complicated by speech-disorganizing reactions? (4) Why does the speaker react negatively to his disfluencies and to his listeners' negative reactions to them —and to him? With regard to (2), some speech pathologists hold that the disfluencies are caused by organic factors, including heredity, while others view them as symptoms of emotional disturbance, usually involving impaired parent-child relationships; but laboratory and clinical data do not firmly support these hypotheses. Another explanation is that simple disfluencies are part of the normal speech process, varying with such factors as age, language development, momentary excitement, and listener in-

terest; the child learns speech-disruptive behaviour as he tries to keep from stuttering and so to gain approval.

For a young child, treatment is directed mainly to the parents and other listeners important to the child, who are counseled to increase their understanding of the factors, including their own reactions, that affect the child's disfluencies; to help them understand how the problem is compounded by the child's misguided efforts to speak more clearly; to encourage them to make it easy for him to talk to them; and to provide conditions favourable for his spontaneous attempts at speech. Older children and adults are given much the same sort of counseling as key listeners, being treated clinically as both listeners and speakers. They are trained to think about what they do that interferes with speaking rather than about what happens or what it is that prevents the words from coming out. Once the stutterer recognizes what he does that disrupts his speech, he works to increase his awareness of it and to contrast it with his normal speech. Therapy proceeds then toward learning or relearning to talk without exerting the tensions and doing the other things that disrupt speaking.

Retarded Development of Speech or Language.—Development of speech or language may be retarded in respect to: (1) amount of vocalizing during infancy, (2) age at which first words and sentences are spoken, (3) articulation of speech sounds, (4) length and complexity of speech response, (5) amount of speaking, and (6) vocabulary. A child's speech development is to be compared not with any average (though speech development norms have been presented) but with that of the majority of children and with his own past performance. Generally speaking, a child may be a year behind the majority without being considered retarded. In the United States, for example, where most children have begun to say words at the age of 12 to 15 months, a child would not be judged retarded until he had passed 24 months without doing so. Factors commonly associated with retarded speech development include lack of speech stimulation in the home; inadequate or inconsistent rewards and punishments; isolation and inactivity, as in prolonged illness; impairment of muscles used in speech; damage to the brain; and loss of hearing.

Aphasia (Dysphasia).—Aphasia, or more appropriately dysphasia, is the impairment of language function due to brain damage. Loss of language function may be slight to total and is likely to be regained as, and in the measure that, the patient recovers from the condition that caused it. Recovery may be hastened or enhanced by speech therapy. Speech retraining is based on remaining functions and life experience with language. A patient relearns best the language concerning things with which he was once familiar and thus in all likelihood will make best progress in familiar surroundings. Objects and pictures are shown, the patient repeats the names, points to them, then uses them in phrases and sentences.

Speech Disorders Associated with Other Physical or Mental Disorder or Impairment.—*Impaired Hearing.*—Hearing loss may be slight, moderate, or severe and varies in character. (*See* DEAFNESS AND IMPAIRED HEARING.) For a person with simple conductive hearing impairment, all voices are muted except his own, which sounds louder to him than it does to his listeners. His main or only speech problem lies in speaking loudly enough. Other types of hearing impairment may result in a variety of distortions of hearing: the person may hear his own and others' voices as muted, hear low-pitched sounds better than high-pitched, hear some things and not others. When two kinds of hearing loss occur together the associated speech problem usually reflects the combination. The speech and language difficulties of hard-of-hearing persons therefore may be complex. The afflicted person may have a limited vocabulary or may misarticulate certain sounds; his voice may be unusually loud, monotonous in pitch, and muffled in quality as if the mouth were full. Therapy emphasizes monitoring speech, training in the use of a hearing aid if the patient can benefit from it, practice in the recognition of sounds, training in lip-reading (or speech-reading, as it is more appropriately called), and vocabulary building. Young children with profound hearing impairment tend not to learn speech unless taught by special methods in schools for the deaf (*see* DEAF AND HARD OF HEARING, TRAINING AND WELFARE OF).

Cleft Lip and Cleft Palate.—Clefts of the upper lip or the palate (roof of mouth), or both, are formed, for reasons not well understood, when normal growth is disrupted early in fetal life. A cleft may involve only the soft palate or both soft and hard; may extend through the front upper gum ridge; and may involve one or both sides of the upper lip. Cleft lip is usually repaired surgically shortly after birth, and if there is no other cleft no speech problem is likely to develop unless there remains a limitation of lip function that affects articulation of lip sounds. Cleft palate may be closed surgically or by means of an obturator (artificial palate). An obturator must be adapted periodically to the child's growth; surgery should wait until the age at which it will not interfere significantly with growth of the upper jaw and the face. After palatal closure, most children, at least in the United States, require speech training to counteract excessive nasality, misarticulations, inefficient regulation of air pressure within the mouth, and in some cases vocal monotony, low loudness level, indistinctness, and lack of spontaneity. Training to correct faulty jaw and tongue action may be started even before closure has been made.

Cerebral Palsy and Other Neuromuscular Impairments.—Paralysis (*q.v.*) impairs speech when it affects muscles used in speaking. Cerebral palsy (*q.v.*) affects these muscles in about three out of four cases, producing laborious and slow speech, of uneven rhythm, with breathing irregularities, faulty articulation, and monotonous or erratic patterns of voice quality, pitch, and loudness. The speech problem is often only one of several associated with this condition. Treatment of cerebral palsy requires the resources of medicine, special education, physical and occupational therapy, psychology, social work, and speech therapy.

Laryngectomy.—Laryngectomy, most commonly performed in later life as a means of controlling cancer, leaves the patient without vocal cords. Though he cannot vocalize, usually he can learn esophageal speech, drawing air into his esophagus, producing sound as he expels it, and articulating this sound into spoken words. An "artificial larynx" also may be used (*see* VOICE: *Substitutes for the Larynx*).

Facial or Oral Injury or Deficiency.—Problems of articulation and occasionally of vocal resonance and quality may be associated with high and narrow hard palate, large or sluggish tongue, surgery of the tongue, cancer, and missing or misaligned front teeth. Following dental and medical treatment, speech therapy is indicated.

Mental Retardation.—Learning difficulties of the mentally retarded extend to learning to speak, misarticulations being the main speech problem. Retraining methods are adapted to specific needs and abilities, with emphasis on simplicity, persistence, patience, and rewards for small gains.

Emotional Illness.—The chief effects of emotional illness on speaking behaviour are restrictions or disturbances of speech, impairment of voice, and deviant use of language. Often the voice is high pitched, nasal, harsh, and monotonous; or the patient may not speak at all. Treatment usually begins with a psychological evaluation. Speech therapy and psychotherapy may then proceed together, in a relationship that depends on whether the speech problem preceded, is independent of, or is basic to the personality problem.

DEVELOPMENT OF SPEECH CORRECTION

Persons with speech disorders, like other handicapped persons, have traditionally known the scorn, ridicule, and even revulsion of their society. Attitudes toward handicaps have changed only slowly, and the change is still going on. Hippocrates (fl. 400 B.C.) was among the first to look at speech disorders scientifically, writing a treatise on the cause of stuttering. Other early scientists who were interested in speech and speech problems included Albertus Magnus, who in the 13th century wrote about voice pitch and articulation; Paracelsus, who in the 16th century referred to speech disturbance after injuries to the head; Hieronymus Mercurialis (Geronimo Mercuriale), who published (1584) on children's speech; and J. C. Ammann, whose *Surdus loquens* (1692) set down the oral method of training deaf children. By the 19th century a large European medical literature on speech was de-

veloping, and concurrently in the New World the work went on through activities of such men as T. H. Gallaudet, who had been schooled in London and Paris. The Spanish singing teacher Manuel García at mid-century developed a method for examining the larynx; the French surgeon-anthropologist Paul Broca in 1861 published his pioneer work relating aphasia to brain lesions; H. Klencke, in Leipzig, was the first formally to suggest (1862) that effective speech correction must deal with the total speech problem; C. L. Merkel's *Anthropophonik* ("Human Phonetics") in 1876 and Adolf Kussmaul's *Die Störungen der Sprache* ("The Impediments of Speech") in 1877 stand as classics from which German logopedics grew.

By the early years of the 20th century, with the growth of knowledge about speech, with refinement of clinical techniques in speech correction, and with new appreciation of the human need to communicate, there emerged an interdisciplinary profession which in the United States went by the name of speech pathology and audiology and in Europe was called logopedics and phoniatrics. The contributing professions included medicine, psychology, dentistry, education, speech, psychiatry, engineering, and social welfare. All are represented in the membership of the International Association of Logopedics and Phoniatrics (IALP) and of the American Speech and Hearing Association (ASHA), two of the more inclusive organizations in the field. Both were founded in the 1920s; both are dedicated to support of high professional standards, encouragement of research, and improvement of clinical procedures.

Most countries of Western Europe have some kind of provision for speech correction, but programs and requirements for the training of speech correctionists vary. Much of the program is directed by medical doctors, often in hospital clinics or special schools. As compared with procedures in the United States, the initial evaluation of each case is considerably longer, use of speech stimulation techniques and motivational devices is considerably less.

Great Britain, with a long history of speech correction, officially recognized speech disorders as a handicap by the Education Act, 1944. The College of Speech Therapists, an organizing and examining body, was established in 1945; the first school for children with very severe speech handicaps was opened in 1947.

Japan, with schools for the blind since the 8th century and for the deaf since the 19th, is developing a school-connected program in speech correction with the assistance of therapists and clinicians who are largely self-taught and with the support of a research program.

In the U.S.S.R. speech correction is directed by the Department of Logopedics in the government Institute of Defectology of the Academy of Pedagogical Sciences. Children with speech disorders go to regular schools; each school has at least one logopedist, who works often with other specialists. Therapy may include repetition of sounds, long periods of silence, use of the logopedist's finger or a wire apparatus called a zond to place the tongue properly for a given sound. The child is encouraged to assume much responsibility in overcoming his handicap.

In the United States, 200 colleges and universities offer degrees in speech pathology and audiology, 25% at the doctoral level. The M.A. degree is required by the American Speech and Hearing Association for membership and for clinical certification of speech correctionists. Many states, as well, set certification requirements. There are about 7,000 active speech correctionists, most of them in public schools, some in community clinics or with special health agencies, some staffing the 320 speech and hearing clinics associated with hospitals, colleges, and universities. About one-fourth of the children who have some kind of speech disorder are receiving special speech training.

BIBLIOGRAPHY.—*Foundations of Speech Pathology*, series of 14 volumes (1964); Wendell Johnson, *People in Quandaries* (1946), *Stuttering and What You Can Do About It* (1961); Wendell Johnson and associates, *The Onset of Stuttering* (1959); Wendell Johnson *et al.*, *Speech Handicapped School Children*, rev. ed. (1956); Wendell Johnson, Frederic L. Darley, and D. C. Spriestersbach, *Diagnostic Methods in Speech Pathology* (1963); Romaine P. Mackie, Harold M. Williams, and Patricia P. Hunter, *Statistics of Special Education for Exceptional Children*, U.S. Office of Education (1963); Samuel D. Robbins, *A Dictionary of Speech Pathology and Therapy* (1963); Hildred Schuell *et al.*, *Aphasia in Adults* (1964); Speech Foundation of America Series: *Stuttering and Its Treatment* (1960), *Stuttering: Its Prevention* (1962), *Stuttering Words* (1963); Lee Travis (ed.), *Handbook of Speech Pathology* (1957); Charles Van Riper, *Speech Correction Principles and Methods*, 4th ed. (1963); *see* also recent issues of these journals: *Asha, dsh Abstracts, J. Speech Hearing Dis., J. Speech Hearing Res.* (U.S.); *Acta Otolaryng.* (Stockholm); *Folia Phoniat.* (Basel); *Speech Path. Ther.* (London). (W. A. L. J.)

SPEED, JOHN (1552?–1629), English antiquary and cartographer, whose *Theatre of the Empire of Great Britain* (1611–12; titlepage 1611) was the first printed atlas of the British Isles. The son of a tailor, he was born at Farringdon, Cheshire, and became a tailor in London, but in 1598, through Fulke Greville's patronage, received a royal sinecure. This gave him leisure for his topographical and historical researches, and, helped by Sir Robert Cotton, William Camden, Sir Henry Spelman, and other fellow members of the Society of Antiquaries, he published (1608–10) a series of county maps of the British Isles and gathered material for his *History of Great Britaine . . . from Julius Caesar to . . . King James* (1611). To this the *Theatre*, in which the maps were collected, formed a prologue.

In 1626 he published the first English general atlas, *A Prospect of the Most Famous Parts of the World . . . Asia, Africa, Europe, America*. Like the *Theatre*, this was very popular, and both were frequently reprinted, sometimes, as in 1676, combined in one volume. Another popular work was his *Genealogies Recorded in Sacred Scripture*. Despite blindness, Speed continued to revise his *History* until his death, in London, on July 28, 1629.

Although his *History* was admired and is of interest for including valuable records, *e.g.*, lists of the abbeys dissolved by Henry VIII, it is mainly based on earlier work, as its author acknowledged. Speed's maps, too, are based on those of earlier cartographers; their importance consists in their innovations. His additions to the maps of Christopher Saxton and John Norden (*see* MAP) resulted from his own journeys; he added inset town plans and showed his interest in history by recording antiquities—forts, castles, battle sites, earthworks, etc. His *Theatre* remained influential until the 19th century, and his maps are still of historical interest.

See R. A. Skelton, *Decorative Printed Maps of the 15th to 18th Centuries* (1952); E. Lynam, *The Mapmaker's Art* (1953).

SPEEDOMETER, an instrument that indicates the speed of a vehicle; usually, it is combined with an odometer, which records the distance traveled. The speed-indicating mechanism of the speedometer is actuated by a permanent magnet rotated at a speed of 1,000 revolutions per mile of vehicle travel by a flexible shaft driven by gears at the rear of the transmission. The magnet turns within a "speed cup" made of a light, nonmagnetic metal, such as aluminum, and the magnetic circuit is completed by a circular stationary field plate surrounding the speed cup.

The rotating magnetic field passing through the metal of the speed cup induces eddy currents that increase with the magnet speed and set up a proportional magnetic field attracting the speed cup to the rotating magnet, tending to rotate it against the restraint of a coiled hairspring. The torque acting on the speed cup is proportional to the eddy currents and thus to the rotational speed of the magnet. The speed cup turns until the restraining torque of the hairspring, increasing with the angle through which the cup turns, equals the applied torque. The angular position at which the speed cup comes to rest varies directly with the rotational speed of the magnet and hence with the speed of the vehicle. The speed cup may have a scale of miles (or kilometres) per hour on its periphery that is visible through a small window, or it may have a pointer attached to it that moves across a graduated scale.

The odometer, which registers the distance traveled by the vehicle, consists of a set of figure wheels, usually from four to six, arranged along a cross shaft and turned by a series of star pinion gears located inside the drum formed by the figure wheels. A train of intermediate gears connects the figure wheel at the right to the magnet shaft at a gear ratio of 1,000:1, causing it to complete one turn per mile. Its numerals thus indicate tenths of a mile.

The figure wheels are so geared together that, as one wheel completes a revolution, it turns the figure wheel next to the left one-tenth of a turn, thus summing up the miles traveled. A "trip odometer" is sometimes included that is provided with a reset mechanism to return it to zero. (O. C. C.)

SPEKE, JOHN HANNING (1827–1864), English explorer who solved the mystery of the Nile sources by his discovery of Lake Victoria, was the first European to visit Uganda. He was born May 3, 1827, at Orleigh Court, Bideford, Devon. Commissioned in the Indian Army in 1844, he served in the Punjab and traveled in the Himalayas and Tibet. In 1854 he joined an expedition to explore Somaliland under the command of Richard Burton (*q.v.*). This venture was wrecked at the outset in an attack by Somalis outside Berbera in April 1855 during which Speke was severely wounded. In 1856 Speke joined Burton on a Royal Geographical Society expedition to investigate the great lake said to lie at the heart of Africa and to be the source of the Nile. They arrived in Zanzibar in December 1856 and left for the interior from Bagamoyo on June 27. They reached Tabora (or Kazé) by the Arab trade route on Nov. 7 where they learned of the existence of three great lakes (Nyasa, Tanganyika, Victoria), and, although in ill health, they pressed on to Ujiji on Lake Tanganyika, where they arrived in February of 1858. Back at Tabora in June, Speke struck north alone to find the lake he guessed was the source of the Nile, and this he reached and named Victoria on July 30. Speke returned to England first and announced his discovery of the Nile source; his conclusions were disputed later, and Tanganyika suggested as the source. Sir Roderick Murchison, president of the Royal Geographical Society, supported Speke who received the Founder's Medal. He was invited to lead another expedition to confirm his findings and chose James Grant to accompany him. Leaving Zanzibar in September 1860, they reached the southwestern shore of Lake Victoria, via Tabora, a year later. Speke pressed on ahead up the west shore into Buganda, arriving at King Mtesa's capital (near Kampala) on Feb. 19, 1862. He marked the Nile's exit from the lake by naming the Ripon Falls on July 28. Speke and Grant then tried to follow the course of the Nile but were checked by civil war in Bunyoro. They reached Gondokoro on Feb. 15, 1863, and met (Sir) Samuel and Florence Baker, to whom Speke gave directions for finding the lake in Bunyoro through which he guessed the Nile must flow. He described his adventures in *Journal of the Discovery of the Source of the Nile* (1863) and *What Led to the Discovery of the Source of the Nile* (1864).

On his return to England, Speke's claims were again challenged. A discussion between him and Burton was arranged to take place at the British Association at Bath in September 1864, but, while the audience waited, news was brought that Speke had been found killed accidentally by his own gun while shooting at Neston Park, near Corsham, Wiltshire, on Sept. 15, 1864.

Bibliography.—R. F. Burton, *First Footsteps in East Africa* (1856), *The Lake Regions of Central Africa* (1860), (with J. MacQueen), *The Nile Basin* (1864); J. A. Grant, *A Walk Across Africa* (1865); S. W. Baker, *The Albert N'yanza, Great Basin of the Nile* (1866); H. H. Johnston, *The Nile Quest* (1903); D. Middleton, *Baker of the Nile* (1949); J. N. L. Baker, "Sir Richard Burton and the Nile Sources," *English Historical Review*, 59:48–61 (Jan. 1944). (D. Mn.)

END OF VOLUME TWENTY

PRINTED IN THE U. S. A. BY R. R. DONNELLEY & SONS CO.

Encyclopædia Britannica;

OR, A

DICTIONARY

OF

ARTS and SCIENCES,

COMPILED UPON A NEW PLAN.

IN WHICH

The different Sciences and Arts are digested into distinct Treatises or Systems;

AND

The various Technical Terms, &c. are explained as they occur in the order of the Alphabet.

ILLUSTRATED WITH ONE HUNDRED AND SIXTY COPPERPLATES.

By a Society of Gentlemen in Scotland.

IN THREE VOLUMES.

VOL. I.

EDINBURGH:
Printed for A. Bell and C. Macfarquhar;
And sold by Colin Macfarquhar, at his Printing-office, Nicolson-Street.
M.DCC.LXXI.


Encyclopædia Britannica;
OR, A
DICTIONARY
OF
ARTS and SCIENCES,
COMPILED UPON A NEW PLAN.
IN WHICH
The different SCIENCES and ARTS are digested into distinct Treatises or Systems;
AND
The various TECHNICAL TERMS, &c. are explained as they occur in the order of the Alphabet.
ILLUSTRATED WITH ONE HUNDRED AND SIXTY COPPERPLATES.
By a SOCIETY of GENTLEMEN in SCOTLAND.
IN THREE VOLUMES.
VOL. I.
EDINBURGH:
Printed for A. BELL and C. MACFARQUHAR;
MDCCLXXI.
GUNNERY
CONIC SECTIONS